

# *Clinical Trials Handbook*

**Edited by**

**Shayne Cox Gad, Ph.D., D.A.B.T.**

Gad Consulting Services  
Cary, North Carolina

 **WILEY**

*A John Wiley & Sons, Inc., Publication*



# *Clinical Trials Handbook*



# *Clinical Trials Handbook*

**Edited by**

**Shayne Cox Gad, Ph.D., D.A.B.T.**

Gad Consulting Services  
Cary, North Carolina

 **WILEY**

*A John Wiley & Sons, Inc., Publication*

Copyright © 2009 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey  
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data:***

Clinical trials handbook / [edited by] Shayne Cox Gad.

p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-471-21388-8 (cloth)

1. Drugs--Testing--Handbooks, manuals, etc. 2. Clinical trials--Handbooks, manuals, etc.

I. Gad, Shayne Cox, 1948-

[DNLM: 1. Clinical Trials as Topic--Handbooks. QV 39 C64175 2009]

RM301.27.C578 2009

615'.1--dc22

2009005648

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*To my mother and father (Norma and Leonard Gad),  
now both gone but always remembered for all they gave me.*





# Contents

|                                                                                  |             |
|----------------------------------------------------------------------------------|-------------|
| <b>Preface</b>                                                                   | <b>xi</b>   |
| <b>Contributors</b>                                                              | <b>xiii</b> |
| <br>                                                                             |             |
| <b>1 Introduction to Clinical Trials</b>                                         | <b>1</b>    |
| <i>John Goffin</i>                                                               |             |
| <br>                                                                             |             |
| <b>2 Regulatory Requirements for Investigational New Drug</b>                    | <b>23</b>   |
| <i>Venkat Rao</i>                                                                |             |
| <br>                                                                             |             |
| <b>3 Preclinical Assessment of Safety in Human Subjects</b>                      | <b>71</b>   |
| <i>Nancy Wintering and Andrew B. Newberg</i>                                     |             |
| <br>                                                                             |             |
| <b>4 Predicting Human Adverse Drug Reactions from Nonclinical Safety Studies</b> | <b>87</b>   |
| <i>Jean-Pierre Valentin, Marianne Keisu, and Tim G. Hammond</i>                  |             |
| <br>                                                                             |             |
| <b>5.1 History of Clinical Trial Development and the Pharmaceutical Industry</b> | <b>115</b>  |
| <i>Jeffrey Peppercorn, Thomas G. Roberts, Jr., and Tim G. Hammond</i>            |             |
| <br>                                                                             |             |
| <b>5.2 Adaptive Research</b>                                                     | <b>135</b>  |
| <i>Michael Rosenberg</i>                                                         |             |
| <br>                                                                             |             |
| <b>6 Organization and Planning</b>                                               | <b>161</b>  |
| <i>Sheila Sprague and Mohit Bhandari</i>                                         |             |
| <br>                                                                             |             |
| <b>7 Process of Data Management</b>                                              | <b>185</b>  |
| <i>Nina Trocky and Cynthia Brandt</i>                                            |             |
|                                                                                  | <b>vii</b>  |

|             |                                                                                         |            |
|-------------|-----------------------------------------------------------------------------------------|------------|
| <b>8</b>    | <b>Clinical Trials Data Management</b>                                                  | <b>203</b> |
|             | <i>Eugenio Santoro and Angelo Tinazzi</i>                                               |            |
| <b>9.1</b>  | <b>Clinical Trials and the Food and Drug Administration</b>                             | <b>227</b> |
|             | <i>Tarek M. Mahfouz and Janelle S. Crossgrove</i>                                       |            |
| <b>9.2</b>  | <b>Phase I Clinical Trials</b>                                                          | <b>245</b> |
|             | <i>Elizabeth Norfleet and Shayne Cox Gad</i>                                            |            |
| <b>9.3</b>  | <b>Phase II Clinical Trials</b>                                                         | <b>255</b> |
|             | <i>Say-Beng Tan and David Machin</i>                                                    |            |
| <b>9.4</b>  | <b>Designing and Conducting Phase III Studies</b>                                       | <b>279</b> |
|             | <i>Nabil Saba, John Kauh, and Dong M. Shin</i>                                          |            |
| <b>9.5</b>  | <b>Phase IV: Postmarketing Trials</b>                                                   | <b>303</b> |
|             | <i>Karl Wegscheider</i>                                                                 |            |
| <b>9.6</b>  | <b>Phase IV and Postmarketing Clinical Trials</b>                                       | <b>325</b> |
|             | <i>Ali Miraj Khan</i>                                                                   |            |
| <b>9.7</b>  | <b>Regulatory Approval</b>                                                              | <b>349</b> |
|             | <i>Fred Henry and Weichung J. Shih</i>                                                  |            |
| <b>9.8</b>  | <b>New Paradigm for Analyzing Adverse Drug Events</b>                                   | <b>373</b> |
|             | <i>Ana Szarfman, Jonathan G. Levine, and Joseph M. Tonning</i>                          |            |
| <b>10.1</b> | <b>Clinical Trials in Interventional Cardiology: Focus on XIENCE Drug-Eluting Stent</b> | <b>397</b> |
|             | <i>J. Doostzadeh, S. Bezenek, W.-F. Cheong, P. Sood, L. Schwartz, and K. Sudhir</i>     |            |
| <b>10.2</b> | <b>Clinical Trials Involving Oral Diseases</b>                                          | <b>435</b> |
|             | <i>Bruce L. Pihlstrom, Bryan Michalowicz, Jane Atkinson, and Albert Kingman</i>         |            |
| <b>10.3</b> | <b>Dermatology Clinical Trials</b>                                                      | <b>461</b> |
|             | <i>Maryanne Kazanis, Alicia Van Cott, and Alexa Boer Kimball</i>                        |            |
| <b>10.4</b> | <b>Emergency Clinical Trials</b>                                                        | <b>477</b> |
|             | <i>Joaquín Borrás-Blasco, Andrés Navarro-Ruiz, and Consuelo Borrás</i>                  |            |
| <b>10.5</b> | <b>Gastroenterology</b>                                                                 | <b>501</b> |
|             | <i>Lise Lotte Gluud and Jørgen Rask-Madsen</i>                                          |            |
| <b>10.6</b> | <b>Gynecology Randomized Control Trials</b>                                             | <b>519</b> |
|             | <i>Khalid S. Khan, Tara Selman, and Jane Daniels</i>                                    |            |

|                                                                                                                    |            |
|--------------------------------------------------------------------------------------------------------------------|------------|
| <b>10.7 Special Population Studies (Healthy Patient Studies)</b>                                                   | <b>531</b> |
| <i>Doris K. Weilert</i>                                                                                            |            |
| <b>10.8 Musculoskeletal Disorders</b>                                                                              | <b>563</b> |
| <i>Masami Akai</i>                                                                                                 |            |
| <b>10.9 Oncology</b>                                                                                               | <b>587</b> |
| <i>Matjaz Zwitter</i>                                                                                              |            |
| <b>10.10 Pharmacological Treatment Options for Nonexudative and Exudative Age-Related Macular Degeneration</b>     | <b>607</b> |
| <i>Alejandro Oliver, Thomas A. Ciulla, and Alon Harris</i>                                                         |            |
| <b>10.11 Paediatrics</b>                                                                                           | <b>627</b> |
| <i>Anne Cusick, Natasha Lannin, and Iona Novak</i>                                                                 |            |
| <b>10.12 Clinical Trials in Dementia</b>                                                                           | <b>661</b> |
| <i>Encarnita Raya-Ampil and Jeffrey L. Cummings</i>                                                                |            |
| <b>10.13 Clinical Trials in Urology</b>                                                                            | <b>695</b> |
| <i>Geoffrey R. Wignall, Carol Wernecke, Linda Nott, and Hassan Razvi</i>                                           |            |
| <b>10.14 Clinical Trials on Cognitive Drugs</b>                                                                    | <b>705</b> |
| <i>Elisabetta Farina and Francesca Baglio</i>                                                                      |            |
| <b>10.15 Bridging Studies in Pharmaceutical Safety Assessment</b>                                                  | <b>733</b> |
| <i>Jon Ruckle</i>                                                                                                  |            |
| <b>10.16 Brief History of Clinical Trials on Viral Vaccines</b>                                                    | <b>769</b> |
| <i>Megan J. Brooks, Joseph J. Sasadeusz, and Gregory A. Tannock</i>                                                |            |
| <b>11 Methods of Randomization</b>                                                                                 | <b>779</b> |
| <i>Gladys McPherson and Marion Campbell</i>                                                                        |            |
| <b>12 Randomized Controlled Trials</b>                                                                             | <b>807</b> |
| <i>Giuseppe Garcea and David P. Berry</i>                                                                          |            |
| <b>13 Cross-Over Designs</b>                                                                                       | <b>823</b> |
| <i>Raphaël Porcher and Sylvie Chevret</i>                                                                          |            |
| <b>14.1 Biomarkers</b>                                                                                             | <b>851</b> |
| <i>Michael R. Bleavins, Claudio Carini, Malle Jurima-Romet, and Ramin Rahbari</i>                                  |            |
| <b>14.2 Biomarkers in Clinical Drug Development: Parallel Analysis of Alzheimer Disease and Multiple Sclerosis</b> | <b>869</b> |
| <i>Christine Betard, Filippo Martinelli Boneschi, and Paulo Caramelli</i>                                          |            |

|             |                                                                                                                                                                                                                   |             |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>15</b>   | <b>Review Boards</b>                                                                                                                                                                                              | <b>895</b>  |
|             | <i>Maureen Hood, Jason F. Kaar, and Vincent B. Ho</i>                                                                                                                                                             |             |
| <b>16</b>   | <b>Size of Clinical Trials</b>                                                                                                                                                                                    | <b>913</b>  |
|             | <i>Jitendra Ganju</i>                                                                                                                                                                                             |             |
| <b>17</b>   | <b>Blinding and Placebo</b>                                                                                                                                                                                       | <b>933</b>  |
|             | <i>Artur Bauhofer</i>                                                                                                                                                                                             |             |
| <b>18</b>   | <b>Pharmacology</b>                                                                                                                                                                                               | <b>949</b>  |
|             | <i>Thierry Buclin</i>                                                                                                                                                                                             |             |
| <b>19</b>   | <b>Modeling and Simulation in Clinical Drug Development</b>                                                                                                                                                       | <b>989</b>  |
|             | <i>Jerry Nedelman, Frank Bretz, Roland Fisch, Anna Georgieva, Chyi-Hung Hsu, Joseph Kahn, Ryosei Kawai, Phil Lowe, Jeff Maca, José Pinheiro, Anthony Rossini, Heinz Schmidli, Jean-Louis Steimer, and Jing Yu</i> |             |
| <b>20</b>   | <b>Monitoring</b>                                                                                                                                                                                                 | <b>1019</b> |
|             | <i>Nigel Stallard and Susan Todd</i>                                                                                                                                                                              |             |
| <b>21</b>   | <b>Inference Following Sequential Clinical Trials</b>                                                                                                                                                             | <b>1043</b> |
|             | <i>Aiyi Liu and Kai F. Yu</i>                                                                                                                                                                                     |             |
| <b>22</b>   | <b>Statistical Methods for Analysis of Clinical Trials</b>                                                                                                                                                        | <b>1053</b> |
|             | <i>Duolao Wang, Ameet Bakhai, and Nicola Maffulli</i>                                                                                                                                                             |             |
| <b>23</b>   | <b>Explanatory and Pragmatic Clinical Trials</b>                                                                                                                                                                  | <b>1081</b> |
|             | <i>Rob Herbert</i>                                                                                                                                                                                                |             |
| <b>24.1</b> | <b>Ethics of Clinical Research in Durg Trials</b>                                                                                                                                                                 | <b>1099</b> |
|             | <i>Roy G. Beran</i>                                                                                                                                                                                               |             |
| <b>24.2</b> | <b>Ethical Issues in Clinical Research</b>                                                                                                                                                                        | <b>1111</b> |
|             | <i>Kelton Tremellen and David Belford</i>                                                                                                                                                                         |             |
| <b>25</b>   | <b>Regulations</b>                                                                                                                                                                                                | <b>1153</b> |
|             | <i>Ramzi Dagher, Rajeshwari Sridhara, Nallaperumal Chidambaram, and Brian P. Booth</i>                                                                                                                            |             |
| <b>26</b>   | <b>Future Challenges in Design and Ethics of Clinical Trials</b>                                                                                                                                                  | <b>1173</b> |
|             | <i>Carl-Fredrik Burman and Axel Carlberg</i>                                                                                                                                                                      |             |
| <b>27</b>   | <b>Proof-of-Principle/Proof-of-Concept Trials in Drug Development</b>                                                                                                                                             | <b>1201</b> |
|             | <i>Ayman Al-Shurbaji</i>                                                                                                                                                                                          |             |
|             | <b>Index</b>                                                                                                                                                                                                      | <b>1219</b> |

# *Preface*

The *Clinical Trials Handbook* represents a collective attempt to present the entire range of approaches to the clinical development process for potential new therapeutic moieties, assembled in the context of this Wiley series on the entire process of pharmaceutical discovery and development. This volume, in fact, is the seventh in this series, which is intended to be comprehensive in its coverage.

The volume is unique in that it seeks to cover the entire range of general topics in the field of clinical trials while also presenting chapters that focus on a specific therapeutic usage over a wide range of disease claims. The 52 chapters cover introductory, regulatory and logistical issues, data management, general study design issues, types of clinical trials, and ethical and oversight issues.

This book would not have occurred without the dedicated efforts of Wiley's managing editors, Zabrina Mok and Gladys Mok. Their persistence in the recruitment of contributors and ensuring follow through was essential.

While like all textbooks this one presents the state of the practice and field at a specific period in time, I hope that it will become a frequently consulted friend.

SHAYNE COX GAD



# Contributors

- MASAMI AKAI, Director, Rehabilitation Hospital, National Rehabilitation Center Japan, Saitama, Japan, *Musculoskeletal Disorders*
- AYMAN AL-SHURBAJI, Experimental Medicine, International PharmaScience Center, Ferring Pharmaceuticals A/S, Copenhagen S, Denmark, *Proof-of-Principle/ Proof-of-Concept Trials in Drug Development*
- JANE ATKINSON, National Institutes of Health/NIDCR, Bethesda, Maryland, *Clinical Trials Involving Oral Diseases*
- FRANCESCA BAGLIO, Neurorehabilitation Unit, Don Carlo Gnocchi Foundation, Scientific Institute and University, IRCCS, Milan, Italy, *Clinical Trials on Cognitive Drugs*
- AMEET BAKHAI, Barnet General & Royal Free Hospitals, London, United Kingdom, *Statistical Methods for Analysis of Clinical Trials*
- ARTUR BAUHOFFER, Institute of Theoretical Surgery, Philipps-University Marburg, Marburg, Germany; current address: CSL-Behring GmbH, Marburg, Germany, *Blinding and Placebo*
- DAVID BELFORD, GroPep Limited, Adelaide, South Australia, *Ethical Issues in Clinical Research*
- ROY G. BERAN, Strategic Health Evaluators, Chatswood NSW 2067, Australia, *Ethics of Clinical Research in Drug Trials*
- DAVID P. BERRY, Department of Hepatobiliary and Pancreatic Surgery, The Leicester General Hospital, United Kingdom, *Randomized Controlled Trials*
- CHRISTINE BETARD, Global Strategic Drug Development Unit, Quintiles, Levallois-Perret, Cedex, France, *Biomarkers in Clinical Drug Development: Parallel Analysis of Alzheimer Disease and Multiple Sclerosis*
- S. BEZENEK, Clinical Science Department, Abbott Vascular Inc., Santa Clara, California, *Clinical Trials in Interventional Cardiology: Focus on XIENCE Drug-Eluting Stent*
- MOHIT BHANDARI, Division of Orthopaedic Surgery, Department of Surgery, McMaster University, Hamilton, Ontario, *Organization and Planning*

- MICHAEL R. BLEAVINS, Michigan Technology and Research Institute, Ann Arbor, Michigan, *Biomarkers*
- BRIAN P. BOOTH, Office of Translational Science, Office of Clinical Pharmacology, Division of Clinical Pharmacology, Food and Drug Administration, Rockville, Maryland, *Regulations*
- CONSUELO BORRÁS, Department of Physiology, University of Valencia, Valencia, Spain, *Emergency Clinical Trials*
- JOAQUÍN BORRÁS-BLASCO, Pharmacy Service, Hospital de Sagunto, Sagunto, Spain, *Emergency Clinical Trials*
- CYNTHIA BRANDT, Center for Medical Informatics, Yale University School of Medicine, New Haven, Connecticut, *Process of Data Management*
- FRANK BRETZ, Clinical Information Sciences, Novartis Pharmaceuticals Corp., East Hanover, New Jersey, *Modeling and Simulation in Clinical Drug Development*
- MEGAN J. BROOKS, Victorian Infectious Diseases Service, Centre for Clinical Research Excellence in Infectious Diseases, The Royal Melbourne Hospital, Parkville, Victoria, Australia, *Brief History of Clinical Trials on Vaccines*
- THIERRY BUCLIN, Division of Clinical Pharmacology and Toxicology, University Hospital of Lausanne, Lausanne, Switzerland, *Pharmacology*
- CARL-FREDRIK BURMAN, Technical & Scientific Development, AstraZeneca, Mölndal, Sweden, *Future Challenges in Design and Ethics of Clinical Trials*
- MARION CAMPBELL, Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland, *Methods of Randomization*
- PAULO CARAMELLI, Cognitive Neurology Unit, Department of Internal Medicine, Faculty of Medicine, Federal University of Minas Gerais, Belo Horizonte, Brazil, *Biomarkers in Clinical Drug Development: Parallel Analysis of Alzheimer Disease and Multiple Sclerosis*
- CLAUDIO CARINI, Fresenius Biotech of North America, Waltham, Massachusetts, *Biomarkers*
- AXEL CARLBERG, Department of Cardiothoracic Surgery, Lund University Hospital, Lund, Sweden, *Future Challenges in Design and Ethics of Clinical Trials*
- W-F. CHEONG, Clinical Science Department, Abbott Vascular Inc., Santa Clara, California, *Clinical Trials in Interventional Cardiology: Focus on XIENCE Drug-Eluting Stent*
- SYLVIE CHEVRET, Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis, France, *Cross-Over Designs*
- NALLAPERUMAL CHIDAMBARAM, Office of New Drug Quality Assessment, Division of Post-Marketing Evaluation, Food and Drug Administration, Rockville, Maryland, *Regulations*
- THOMAS A. CIULLA, Department of Ophthalmology, Indiana University, Indianapolis, Indiana, *Pharmacological Treatment Options for Nonexudative and Exudative Age-Related Macular Degeneration*
- JANELLE S. CROSSGROVE, Raabe College of Pharmacy, Ohio Northern University, Ada, Ohio, *Clinical Trials and the Food and Drug Administration*
- JEFFREY L. CUMMINGS, Departments of Neurology and Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine at UCLA, Los Angeles, California, *Clinical Trials in Dementia*



- ANNE CUSICK, School of Biomedical and Health Sciences, University of Western Sydney, Sydney, Australia, *Paediatrics*
- RAMZI DAGHER, Pfizer, Inc., New London, Connecticut, *Regulations*
- JANE DANIELS, Clinical Trials Unit and Academic Department of Obstetrics and Gynaecology, University of Birmingham, Birmingham, United Kingdom, *Gynecology Randomized Control Trials*
- J. DOOSTZADEH, Clinical Science Department, Abbott Vascular Inc., Santa Clara, California, *Clinical Trials in Interventional Cardiology: Focus on XIENCE Drug-Eluting Stent*
- ELISABETTA FARINA, Neurorehabilitation Unit, Don Carlo Gnocchi Foundation, Scientific Institute and University, IRCCS, Milan, Italy, *Clinical Trials on Cognitive Drugs*
- ROLAND FISCH, Modeling and Simulation, Novartis Pharma AG, Basel, Switzerland, *Modeling and Simulation in Clinical Drug Development*
- SHAYNE COX GAD, Gad Consulting Services, Cary, North Carolina, *Phase I Clinical Trials*
- JITENDRA GANJU, Amgen, Inc., South San Francisco, California, *Size of Clinical Trials*
- GIUSEPPE GARCEA, Cancer Studies and Molecular Medicine, The Leicester Royal Infirmary, United Kingdom, *Randomized Controlled Trials*
- ANNA GEORGIEVA, Modeling and Simulation, Novartis Pharmaceuticals Corp., East Hanover, New Jersey, *Modeling and Simulation in Clinical Drug Development*
- LISE LOTTE GLUUD, Copenhagen Trial Unit, Cochrane Hepato-Biliary Group, Copenhagen, Denmark, *Gastroenterology*
- JOHN GOFFIN, Department of Oncology, Juravinski Cancer Center, McMaster University Hamilton, Ontario, Canada, *Introduction to Clinical Trials*
- TIM G. HAMMOND, Safety Assessment, AstraZeneca, Macclesfield, Cheshire, United Kingdom, *Predicting Human Adverse Drug Reactions from Nonclinical Safety Studies*
- ALON HARRIS, Department of Ophthalmology, Indiana University, Indianapolis, Indiana, *Pharmacological Treatment Options for Nonexudative and Exudative Age-Related Macular Degeneration*
- FRED HENRY, Drug Development and Regulatory Affairs, Taisho Pharmaceuticals R&D Inc., Morristown, New Jersey, *Regulatory Approval*
- ROB HERBERT, The George Institute for International Health, Sydney, Australia, *Explanatory and Pragmatic Clinical Trials*
- VINCENT B. HO, Department of Radiology and Radiological Sciences, Uniformed Services University of the Health Sciences, Bethesda, Maryland, *Review Boards*
- MAUREEN N. HOOD, Department of Radiology and Radiological Sciences, Uniformed Services University of the Health Sciences, Bethesda, Maryland, *Review Boards*
- CHYI-HUNG HSU, Clinical Information Sciences, Novartis Pharmaceuticals Corp., East Hanover, New Jersey, *Modeling and Simulation in Clinical Drug Development*
- MALLE JURIMA-ROMET, MDS Pharma Services, Montreal, Quebec, *Biomarkers*
- JASON F. KAAR, Office of General Counsel, Uniformed Services University of Health Sciences, Bethesda, Maryland, *Review Boards*
- JOSEPH KAHN, Modeling and Simulation, Novartis Pharmaceuticals Corp., East Hanover, New Jersey, *Modeling and Simulation in Clinical Drug Development*

- JOHN KAUH, Emory University School of Medicine, Winship Cancer Institute, Department of Hematology and Oncology, Atlanta, Georgia, *Designing and Conducting Phase III Studies*
- RYOSEI KAWAI, Modeling and Simulation, Novartis Institutes for BioMedical Research, Inc., Cambridge, Massachusetts, *Modeling and Simulation in Clinical Drug Development*
- MARYANNE KAZANIS, Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts, *Dermatology Clinical Trials*
- MARIANNE KEISU, Patient Safety, AstraZeneca, Södertälje, Sweden, *Predicting Human Adverse Drug Reactions from Nonclinical Safety Studies*
- ALI MIRAJ KHAN, *Phase IV and Postmarketing Clinical Trials*
- KHALID S. KHAN, Birmingham Women's Hospital, Birmingham, United Kingdom, *Gynecology Randomized Control Trials*
- ALEXA BOER KIMBALL, Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts, *Dermatology Clinical Trials*
- ALBERT KINGMAN, National Institutes of Health/NIDCR, Bethesda, Maryland, *Clinical Trials Involving Oral Diseases*
- NATASHA LANNIN, Rehabilitation Research Studies Unit, Faculty of Medicine, University of Sydney, Sydney, Australia, *Paediatrics*
- JONATHAN G. LEVINE, Food and Drug Administration, CDER, Silver Spring, Maryland, *New Paradigm for Analyzing Adverse Drug Events*
- AIYI LIU, Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, Maryland, *Inference Following Sequential Clinical Trials*
- PHIL LOWE, Modeling and Simulation, Novartis Pharma AG, Basel, Switzerland, *Modeling and Simulation in Clinical Drug Development*
- JEFF MACA, Clinical Information Sciences, Novartis Pharmaceuticals Corp., East Hanover, New Jersey, *Modeling and Simulation in Clinical Drug Development*
- DAVID MACHIN, Division of Clinical Trials and Epidemiological Sciences, National Cancer Centre, Singapore, *Phase II Clinical Trials*
- NICOLA MAFFULLI, Department of Trauma and Orthopaedic Surgery, Keele University School of Medicine, Keele, Staffordshire, United Kingdom, *Statistical Methods for Analysis of Clinical Trials*
- TAREK M. MAHFOUZ, Raabe College of Pharmacy, Ohio Northern University, Ada, Ohio, *Clinical Trials and the Food and Drug Administration*
- FILIPPO MARTINELLI BONESCHI, Neuro-Rehabilitation Unit, Department of Neurology, San Raffaele Scientific Milano, Italy, *Biomarkers in Clinical Drug Development: Parallel Analysis of Alzheimer Disease and Multiple Sclerosis*
- GLADYS MCPHERSON, Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland, *Methods of Randomization*
- BRYAN MICHALOWICZ, School of Dentistry, University of Minnesota, Minneapolis, Minnesota, *Clinical Trials Involving Oral Diseases*
- ANDRÉS NAVARRO-RUIZ, Pharmacy Service, Hospital General Universitario de Elche, Elche, Spain, *Emergency Clinical Trials*
- JERRY NEDELMAN, Modeling and Simulation, Novartis Pharmaceuticals Corp., East Hanover, New Jersey, *Modeling and Simulation in Clinical Drug Development*
- ANDREW B. NEWBERG, Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania, *Preclinical Assessment of Safety in Human Subjects*

- ELIZABETH NORFLEET, Gad Consulting Services, Cary, North Carolina, *Phase I Clinical Trials*
- LINDA NOTT, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada, *Clinical Trials in Urology*
- IONA NOVAK, Cerebral Palsy Institute, Sydney, Australia, *Paediatrics*
- ALEJANDRO OLIVER, Department of Ophthalmology, Indiana University, Indianapolis, Indiana, *Pharmacological Treatment Options for Nonexudative and Exudative Age-Related Macular Degeneration*
- JEFFREY PEPPERCORN, Division of Medical Oncology, Duke University, Durham, North Carolina, *History of Clinical Trial Development and the Pharmaceutical Industry*
- BRUCE L. PIHLSTROM, School of Dentistry, University of Minnesota, Minneapolis, Minnesota, *Clinical Trials Involving Oral Diseases*
- JOSÉ PINHEIRO, Clinical Information Sciences, Novartis Pharmaceuticals Corp., East Hanover, New Jersey, *Modeling and Simulation in Clinical Drug Development*
- RAPHAËL PORCHER, Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis, France, *Cross-Over Designs*
- RAMIN RAHBARI, Innovative Scientific Management, New York, New York, *Biomarkers*
- VENKAT RAO, National and Defense Programs, Defense Division, Alexandria, Virginia, *Regulatory Requirements for Investigational New Drug*
- JØRGEN RASK-MADSEN, Department of Medical Gastroenterology, Herlev Hospital, University of Copenhagen, Herlev, Denmark, *Gastroenterology*
- ENCARNITA RAYA-AMPIL, Department of Neurology and Psychiatry, University of Santo Tomas, Manila, Philippines, *Clinical Trials in Dementia*
- HASSAN RAZVI, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada, *Clinical Trials in Urology*
- THOMAS G. ROBERTS, Jr., Noonday Asset Management, L.P., Charlotte, North Carolina, *History of Clinical Trial Development and the Pharmaceutical Industry*
- MICHAEL ROSENBERG, Health Decisions, Inc., Durham, North Carolina, *Adaptive Research*
- ANTHONY ROSSINI, Modeling and Simulation, Novartis Pharma AG, Basel, Switzerland, *Modeling and Simulation in Clinical Drug Development*
- JON RUCKLE, Covance Clinical Research Unit, Honolulu, Hawaii, *Bridging Studies in Pharmaceutical Safety Assessment*
- NABIL SABA, Emory University School of Medicine, Winship Cancer Institute, Department of Hematology and Oncology, Atlanta, Georgia, *Designing and Conducting Phase III Studies*
- EUGENIO SANTORO, Laboratory of Medical Informatics, Department of Epidemiology, “Mario Negri” Institute for Pharmacological Research, Milan, Italy, *Clinical Trials Data Management*
- JOSEPH J. SASADEUSZ, Victorian Infectious Diseases Service, Centre for Clinical Research Excellence in Infectious Diseases, The Royal Melbourne Hospital, Parkville, Victoria, Australia, *Brief History of Clinical Trials on Vaccines*
- HEINZ SCHMIDLI, Clinical Information Sciences, Novartis Pharma AG, Basel, Switzerland, *Modeling and Simulation in Clinical Drug Development*
- L. SCHWARTZ, Clinical Science Department, Abbott Vascular Inc., Santa Clara, California, *Clinical Trials in Interventional Cardiology: Focus on XIENCE Drug-Eluting Stent*

- TARA SELMAN, Birmingham Women's Hospital, Birmingham, United Kingdom, *Gynecology Randomized Control Trials*
- WEICHUNG J. SHIH, Department of Biostatistics, School of Public Health, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey, *Regulatory Approval*
- DONG M. SHIN, Emory University School of Medicine, Winship Cancer Institute, Department of Hematology and Oncology, Atlanta, Georgia, *Designing and Conducting Phase III Studies*
- P. SOOD, Clinical Science Department, Abbott Vascular Inc., Santa Clara, California, *Clinical Trials in Interventional Cardiology: Focus on XIENCE Drug-Eluting Stent*
- SHELIA SPRAGUE, Department of Clinical Epidemiology & Biostatistics, Department of Surgery, McMaster University, Hamilton, Ontario, *Organization and Planning*
- RAJESHWARI SRIDHARA, Office of Translational Science, Office of Biostatistics, Division of Biometrics, Food and Drug Administration, Rockville, Maryland, *Regulations*
- NIGEL STALLARD, Warwick Medical School, University of Warwick, Warwick, United Kingdom, *Monitoring*
- JEAN-LOUIS STEIMER, Modeling and Simulation, Novartis Pharma AG, Basel, Switzerland, *Modeling and Simulation in Clinical Drug Development*
- K. SUDHIR, Clinical Science Department, Abbott Vascular Inc., Santa Clara, California, *Clinical Trials in Interventional Cardiology: Focus on XIENCE Drug-Eluting Stent*
- ANA SZAREFMAN, Food and Drug Administration, CDER, Silver Spring, Maryland, *New Paradigm for Analyzing Adverse Drug Events*
- SAY-BENG TAN, Singapore Clinical Research Institute, Singapore, *Phase II Clinical Trials*
- GREGORY A. TANNOCK, Department of Biotechnology and Environmental Biology, RMIT University, Bundoora, Victoria, Australia, *Brief History of Clinical Trials on Vaccines*
- ANGELO TINAZZI, Merck Serono, Global Biostatistics, Geneva, Switzerland, *Clinical Trials Data Management*
- SUSAN TODD, Applied Statistics, University of Reading, Reading, United Kingdom, *Monitoring*
- JOSEPH M. TONNING, Food and Drug Administration, CDER, Silver Spring, Maryland, *New Paradigm for Analyzing Adverse Drug Events*
- KELTON TREMLLEN, Repromed, Dulwich, South Australia, *Ethical Issues in Clinical Research*
- NINA TROCKY, The University of Maryland Baltimore School of Nursing, *Process and Data Management*
- JEAN-PIERRE VALENTIN, Safety Assessment, AstraZeneca, Macclesfield, Cheshire, United Kingdom, *Predicting Human Adverse Drug Reactions from Nonclinical Safety Studies*
- ALICIA VAN COTT, Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts, *Dermatology Clinical Trials*
- DUOLAO WANG, Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, United Kingdom, *Statistical Methods for Analysis of Clinical Trials*

- KARL WEGSCHEIDER, Department of Medical Biometry and Epidemiology, University Hospital Eppendorf, Hamburg, Germany, *Phase IV: Postmarketing Trials*
- DORIS K. WEILERT, Clinical Pharmacology, Quintiles, Inc., Kansas City, Missouri, *Special Population Studies (Healthy Patient Studies)*
- CAROL WERNECKE, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada, *Clinical Trials in Urology*
- GEOFFREY R. WIGNALL, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada, *Clinical Trials in Urology*
- NANCY WINTERING, Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania, *Preclinical Assessment of Safety in Human Subjects*
- JING YU, Modeling and Simulation, Novartis Institutes for BioMedical Research, Inc., Cambridge, Massachusetts, *Modeling and Simulation in Clinical Drug Development*
- KAI F. YU, Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, Maryland, *Inference Following Sequential Clinical Trials*
- MATJAZ ZWITTER, Institute of Oncology, Ljubljana, Slovenia, and Department of Medical Ethics, Medical School, University of Maribor, Slovenia, *Oncology*



# 1

---

## *Introduction to Clinical Trials*

**John Goffin**

*Department of Oncology, Juravinski Cancer Center, McMaster University Hamilton,  
Ontario, Canada*

### **Contents**

|                                                   |    |
|---------------------------------------------------|----|
| 1.1 Goals of Chapter                              | 1  |
| 1.2 Goals of Clinical Trials and What Is at Stake | 2  |
| 1.3 Introduction to Phase I–IV Clinical Trials    | 2  |
| 1.3.1 Introduction to Phase I Trials              | 2  |
| 1.3.2 Introduction to Phase II Trials             | 4  |
| 1.3.3 Introduction to Phase III Trials            | 5  |
| 1.3.4 Introduction to Phase IV Trials             | 6  |
| 1.4 Principles of Trials Development              | 7  |
| 1.4.1 Big Picture, Small Picture                  | 7  |
| 1.4.2 Human Element                               | 8  |
| 1.4.3 Multidisciplinary Nature of Clinical Trials | 10 |
| 1.4.4 Know Your Audience, Know Your Market        | 12 |
| 1.5 Example in Drug Development                   | 14 |
| References                                        | 16 |

### **1.1 GOALS OF CHAPTER**

The purpose of this chapter is to consider the overall goals and requirements of conducting clinical trials. It is an opportunity to avoid pitfalls by viewing the larger picture. This chapter seeks to provoke consideration of key issues without duplicating the more detailed work of later chapters.

---

*Clinical Trials Handbook*, Edited by Shayne Cox Gad  
Copyright © 2009 John Wiley & Sons, Inc.

## 1.2 GOALS OF CLINICAL TRIALS AND WHAT IS AT STAKE

The ultimate goal of drug development is the creation of new, safe, and effective compounds for treating human disease. Clinical trials comprise the portion of this endeavor involving human subjects. While the basic tenants of scientific inquiry do not differ from preclinical research, the stakes are higher and the regulations more stringent.

The cost of conducting clinical trials can be measured in two ways: the human cost and the resource cost. The human cost is the cost from the patient's perspective. The patient suffers from a condition dire enough that experimental therapy is a consideration. He or she holds out hope for this therapy and trusts to the scientific skill and integrity of those conducting the trial. The patients expose themselves to an incompletely understood therapy and usually suffer some degree of toxicity in order to gain uncertain benefit. Prior to a drug being declared useful or not, hundreds or thousands of patients may be involved in trials related to the drug.

On another balance sheet, there is the impressive economic burden of drug development. The cost of successfully bringing a new drug to market is now in the range of \$800 million [1]. The interval between the start of clinical testing and the submission of an application for regulatory approval of a new drug is estimated at 6 years [1]. Even so, fields such as oncology are seeing an increase in drugs under study [2]. Yet there are limits to the number of clinical centers able to conduct trials. More importantly, there is a limit to the number of patients that are eligible to participate in a given trial, either by reason of demographic factors, comorbidity, incompatible disease parameters, or willingness. These limitations suggest that investigators must be selective about which drugs they study in clinical trials.

While drug discovery still involves an element of happenstance, contemporary drug development is ever more focused on mechanisms specific to a given disease. Frequently, therefore, a disease population will have been targeted during preclinical development. It is up to the clinical trials process to assess whether the new agent is both safe and effective in this or other populations. Generally, the first concern is assessing drug toxicity and the related dosing and pharmacokinetics. Following this, some evidence of efficacy is sought. If it is found, efficacy must be confirmed in larger, randomized trials. Finally, postmarketing surveillance studies may be performed. These successive clinical trials are usually categorized by phase, and these phases will be introduced below.

## 1.3 INTRODUCTION TO PHASE I-IV CLINICAL TRIALS

### 1.3.1 Introduction to Phase I Trials

**Purpose** New drugs are first introduced into human subjects in phase I trials. The primary goal of these first studies is to assess the safety of the agent and to determine an acceptable dose for further study. Related goals include the assessment of pharmacokinetics as well as pharmacodynamics. To study pharmacokinetics is to study how the body affects the drug: How is the drug absorbed? How is the drug distributed between body compartments? How is the drug metabolized and excreted? Pharmacodynamics is the relationship between drug exposure and drug effect. Here



we ask what normal physiological or disease processes are altered when a drug is administered at varying doses.

**Methods** The method used is to some extent dictated by the drug and disease under consideration. In fields other than oncology, phase I trials are typically undertaken in healthy volunteers. Typically, increasing doses of a drug are employed in small successive cohorts of patients. Each cohort is assessed, and subsequent dose levels are only used if excessive toxicity (often termed dose-limiting toxicity) is not encountered. At each dose level, blood or other body fluid is taken for pharmacokinetic studies.

In oncology studies, the first and lowest dose level may be based upon animal toxicities (e.g., 10% of the dose that is lethal in 10% of mice ( $LD_{10}$ )) and dose increments are often based upon a modified Fibonacci sequence (1, 2, 3, 5, 8, 13, ...), a scheme that decreases the dose increment with each subsequent level. The notion is to limit patient exposure to dose-limiting toxicity through more cautious later stage dose increases. Alternative dosing schemes employ one patient per dose level or a continuously modified dosing increment based upon observed toxicities; the goal of such alternative methods is to increase phase I efficiency and limit the number of patients who receive too little or too much drug [3]. At some point, toxicity is deemed to be excessive, and the appropriate dose level is then established, typically at the dose just below this point of excessive toxicity.

Pharmacokinetics is the study of the drug absorption, transport, distribution, metabolism, and elimination; the goal is to improve drug delivery and efficacy. An understanding of the molecular target may have implications for drug exposure. For example, antimetabolites used against cancer are considered to be most effective in the DNA (deoxyribonucleic acid) synthesis phase (S-phase) of the cell cycle. To best inhibit tumor growth, it is considered optimal to maintain a constant or prolonged exposure of the cancer to drug such that most cells are caught as they transit through S-phase. Pharmacokinetic analysis can tell the investigator if such an exposure is occurring and may prompt alternative dose schedules in subsequent studies.

Pharmacodynamic assays—assays that assess the effect of the drug on normal physiology or disease—may be useful in assessing whether a drug is likely to have a clinical effect. In cardiology, for example, the effects of a new agent on subjects' blood pressure or electrocardiogram may be relevant [4]. In studies of new antibodies or other targeted therapies, a therapeutic effect may be seen without the dose-dependent toxicities expected with other agents (e.g., the antimetabolite methotrexate used in rheumatoid arthritis or cancer). Conducting assays that demonstrate molecular changes in the relevant target could serve as a proof of concept for the agent; this, in turn, could prevent the need for higher dose levels, levels that could induce toxicity and would increase the duration of the study.

**Results** At the end of a phase I study, acute toxicities should be understood. Toxicities related to more long-term exposure may not be apparent until future studies are undertaken. In conjunction with the pharmacokinetic assays and any pharmacodynamic work, an assessment must be made as to whether further studies should be conducted, and, if so, at what dose. Pharmacokinetic analysis may suggest that changes in dose or dosing frequency are required. In instances where toxicity may be excessive at doses not expected or observed to have a useful biological

effect, further phase I studies may be designed to circumvent the toxicity. While preliminary activity against disease may be observed in phase I studies, the initial assessment of positive clinical outcomes is primarily the arena of phase II studies.

### 1.3.2 Introduction to Phase II Trials

**Purpose** Phase II studies are conducted to assess the initial activity of an agent against disease. Further information is gathered about an agent's adverse effects, and additional pharmacokinetic or pharmacodynamic studies may be conducted.

**Methods** Unlike phase I studies, which may employ many different doses of an agent, phase II trials typically employ one or occasionally a few dose levels. Larger cohorts of patients are exposed to the drug in order to observe one or more clinical endpoints. The measured endpoints will vary depending upon the drug and field of study. In trials of heart failure, for example, physiological parameters (e.g., ventricular remodeling) may be assessed in addition to clinical measures such as exercise tolerance [5, 6]. Vaccine studies typically assess safety and immune responses and may involve both treatment and control groups [7]. In oncology, tumor response (shrinkage) rates have traditionally been used as a measure of response, but newer targeted drugs have led to greater reliance upon endpoints such as stable disease rates. Prior to conducting the study, investigators should specify what minimal level of drug activity will be accepted as evidence to warrant subsequent investigation. Phase II studies should be designed as precursors to phase III studies.

Phase II studies may be single-arm assessments of drug activity; such studies have an implied comparator of prior trials or clinical experience. Alternatively, randomized studies may be conducted, comparing the experimental arm with either a placebo, a standard therapy control arm, another experimental arm, or different doses of the experimental arm itself. The randomized study, while of limited power, may improve drug development by increasing the likelihood of selecting the best drug or dose for further development [8]. When a standard treatment arm is used as a comparator, that arm may serve as a barometer for the severity or nature of the disease in the overall study cohort. Excellent or poor results in the experimental arm are interpreted in light of the control arm.

A more recent study type, the randomized discontinuation study, begins with a lead-in period in which all subjects receive the experimental arm. After a predetermined period, subjects are randomized between continuing the study drug and receiving a placebo or no therapy. The lead-in period eliminates noncompliant subjects and unresponsive disease, increasing the likelihood of differences being observed in the randomized portion of the study. The cost is in the greater number of patients required for the study due to drop-out in the initial nonrandomized period [9].

**Results** As noted, the clinical endpoints vary widely based upon disease and agent type. If a drug effect was seen, it must be considered whether the effect was sufficiently interesting in light of existing therapies or other study arms. If a clinical effect was not seen, one must assess whether this could be explained by any biological surrogates or pharmacokinetic studies also undertaken. The clinical efficacy must

be assessed in the face of observed toxicities. More severe toxicities might be acceptable for lifesaving therapies but not for agents directed at minor ailments. At the end of the phase II study, the investigator should have an initial assessment of a new agent's impact on a disease as well as a better understanding of the toxicity profile.

Two important and frequently used statistical concepts should be introduced here. The first is power. In clinical terms, power is the probability that a study will find that a drug is effective when the drug truly *is* effective. Statistically, it may be described as  $\text{Power} = 1 - \beta$ , where  $\beta$  is the probability of a study finding a drug ineffective despite the truth being that the drug *is* effective— $\beta$  is therefore also called the  $\beta$  error. A related term, the  $\alpha$  error, represents the opposite mistake; it is the chance that a study will find a drug effective when in truth the drug is ineffective. By general agreement, the value of  $\alpha$  is usually set at 0.05. Power increases with larger studies (i.e., more patients) and when more prespecified clinical events occur. Phases I and II trials typically employ small numbers of patients, which tends to increase error rates and limit statistical options. Nevertheless, statistics can inform us of the limitations of our knowledge. For example, if we observed 3 of 25 patients with cancer to have tumor responses, we could determine that—with 95% likelihood—the true response rate was from about 3–30% [10]. If we had hoped for better, we would need to carefully consider any next trial. Phases III and IV studies, described below, rely heavily on thoughtful consideration of  $\alpha$  and  $\beta$  errors.

### 1.3.3 Introduction to Phase III Trials

**Purpose** Phase III studies are typically large randomized studies designed to demonstrate useful clinical activity in a specific disease setting. The process of randomizing patients between different treatment arms is fundamental to avoiding biased interpretations of outcomes.

**Methods** The design of phase III studies is critical both in addressing a specific hypothesis and in the pragmatic sense of making a drug useful in clinical practice. Fundamentally, this means that an appropriate patient population must be selected, all treatments must be clinically relevant, and the expected improvement in outcome must be both clinically meaningful and statistically measurable. Eligibility criteria—those criteria that determine which patients may join the study—must define a population that is both adequately generalizable to include patients representative of the diseased cohort but also homogeneous enough to retain statistical power and to be applicable to a usefully recognizable disease group. For example, studies may be difficult to interpret when they include both early- and late-stage patients. If a study is positive, to which population is it best applied? If negative, might it be positive in one of the disease subpopulations if a study were done only in that group.

Treatment arms cannot ignore previously existing therapies. With respect to heart failure, a new drug must take into account that many patients will also be on ACE (angiotensin-converting enzyme) inhibitors,  $\beta$ -blockers, diuretics, antiplatelet agents, and possibly other medications. Excluding these medications may make the study uninterpretable in the real-world clinical context and, more importantly, it may be unethical.

The endpoint of a phase III study should be an accepted and clinically relevant one that is specified before the trial is conducted. For example, in many cancers, an improvement in response rate is not considered an adequate phase III endpoint, whereas improvements in survival or disease-free survival may be accepted. Secondary endpoints—quality of life, for example—may be employed but must be recognized as such at study completion.

A common difficulty with phase III studies is inadequate power. This is often due to an overly optimistic estimate of improvement in a clinical outcome, an estimate that may be a product of resource limitations. A lesser and potentially meaningful improvement may be missed if too few patients are accrued to the study or follow-up is too short.

**Results** The primary and any secondary clinical outcomes must be assessed and interpreted as planned. In circumstances where the primary outcome is of borderline significance or where the primary and secondary clinical outcomes are disparate, explanations may be considered and used as hypotheses for future study. Post hoc analyses are frequently conducted but can only be hypothesis generating.

### 1.3.4 Introduction to Phase IV Trials

**Purpose** Phase IV studies, sometimes called pharmacoepidemiologic studies, are those that are conducted after a drug has been approved for marketing. Such studies, often large, may assess a drug for uncommon toxicities that may be undetectable in smaller phases I–III studies, or they may establish the activity or tolerability of a drug in a particular population or practice setting.

Studies conducted to assess new methods of drug administration, combinations with other agents, or activity in other diseases—that is, studies seeking a new marketing indication—are better described and conducted as the phases I–III studies they represent. Similarly, a distinction can be made between trials seeking to answer a specific postmarketing question and those conducted solely to increase market share, so-called seeding trials. In the latter, there may be an incentive for the involved physicians to prescribe the drug in question and there may be no intent to publish the results [11, 12].

**Methods** Phase IV studies may be conducted in several ways.

1. Descriptive studies, sometimes collections of drug toxicities captured over time, may identify new problems. These may range from case studies to series of patients collected by companies or regulatory bodies. Although resource intensive, large prospective cohort studies may also be conducted to capture infrequent adverse events.
2. Randomized studies may be used to compare an agent to other similar agents or to confirm earlier results.
3. Case-control studies or retrospective cohort studies can be conducted after data on a drug has accumulated. This would typically be done to assess for unusual side effects or associations of a drug with the development of a subsequent disease, such as malignancies or autoimmune sequelae.

4. Cross-sectional studies, although perhaps less useful, assess drug exposure and outcomes in a population at a specific time. Causality may be more difficult to assess if a sequential temporal relationship cannot be determined [12].

**Results** The results of phase IV studies may be required to fulfill regulatory requirements after accelerated approval of a new drug. The additional numbers and prolonged follow-up provided by postmarketing studies may also be crucial in revealing important but infrequent toxicities. On occasion, these findings may lead to the withdrawal of a drug from the market, as, for example, after cardiovascular complications were associated with the anti-inflammatory drug rofecoxib [13, 14].

## 1.4 PRINCIPLES OF TRIALS DEVELOPMENT

### 1.4.1 Big Picture, Small Picture

**Overall Goal: Improved Patient Care** The details involved in protocol design and regulatory requirements can be overwhelming. Remembering the fundamental goal of clinical research—improved patient care—can be an aid; study design and decision making should be influenced by the consideration of what is best for patients.

Patients seek relief from suffering. The investigator should therefore choose the most relevant endpoint for a given trial. Studies of rhinitis may reasonably examine patient reporting of nasal discharge and congestion [15], while studies of pancreatic cancer must consider an agent's impact on survival or more relevant measures of symptoms or quality of life. Research protocols must be designed with these parameters in mind. The outcome of interest must be described in sufficient detail that it may be easily replicated, a matter as important in assessing a study's value in support of regulatory approval as it is to an understanding of what benefit a drug may be to future patients. Any clinical trial must assess the toxicities associated with treatment. Known adverse effects must be clearly described and provisions made for the adjustment of treatment to mitigate such toxicities should they occur. Of course, for sufficiently severe toxicities, a warning system must be in place to inform patients, investigators, and the companies and agencies overseeing the study. The details of such reporting requirement may vary, but the act of sharing such information is sensible.

**Quality** After careful protocol development comes the messy process of administering a protocol. Invariably, aspects of the protocol appear to be open to interpretation, and at some point there will be lapses in study conduct or paperwork. The maintenance of quality in a study means always trying to adhere to the letter and spirit of the protocol. It means that the responsible investigator must be available to arbitrate whether patients are actually eligible and whether protocol violations have occurred. It means that study coordinators must vigorously pursue the complete assessment of patients and the related documentation. Every effort must be made to follow patients to the completion of study. A poorly followed or documented study may be difficult to interpret and may not be acceptable to regulatory agencies or other entities overseeing the trial.

**Nothing in Isolation—The Bench and the Bedside** The present era is one of exciting new agents, many directed at specific targets in the disease process. Even while such agents must undertake the staged clinical trials process, they may evoke interesting biological questions with implications for ongoing or future studies. The prospective collection, banking, and analysis of biological specimens may reveal subsets of patients for whom a new agent may have particular benefit.

For example, small-molecule tyrosine kinase inhibitors directed at the endothelial growth factor receptor (EGFR) have been investigated in patients with non-small cell lung cancer. Despite good preclinical data [16], clinical studies demonstrated more limited benefit, ultimately resulting in limitations of access to one such drug, gefitinib, previously approved by the Food and Drug Administration (FDA) under accelerated approval [17]. The investigation of tumor samples, however, revealed that some tumors had mutations in the tyrosine kinase domain of the *EGFR* gene, with corresponding protein changes and apparent improvements in clinical responses [18, 19]. Unfortunately, this finding was made posthumously for gefitinib, but the implications for future development of this class of drug are clear. When feasible, biological investigations and specimen preservation should continue during the clinical period of study.

#### 1.4.2 Human Element

**Differences between Mice and Humans** Despite the fact that 99% of mouse genes have human counterparts [20], several important issues separate the species. First, important differences in biology can mean significantly different drug metabolism and elimination, such that pharmacokinetics can only be generally predicted [21]. Second, human xenografts planted in mice may respond to drug therapy, but such responses are not consistently predictive of response phase II clinical studies [22]. This supports the necessity of clinical studies. Third, ethics dictates that both the goals and conduct of preclinical and clinical studies must differ. In animals, while the suffering and distress of animals is to be minimized [23], it is accepted that toxicities must be observed in other species to understand new agents and protect the humans that are subsequently exposed. By contrast, the very structure of trials in humans is one of careful staging to avoid excessive toxicity or any death. Earlier studies establish safety while later studies assess for useful clinical activity of a drug.

**Relevance of Ethics** There are more and less obvious aspects of ethics involved in clinical drug development. We have fortunately recognized and codified the obvious, so, for example, it is universally recognized that withholding effective treatment for the sole purpose of observing natural disease history is unethical [24]. But there are less flagrant examples that affect study design.

The phase I study by its nature poses ethical conundrums. It is a study designed to assess toxicity and an acceptable dose for a drug, with clinical benefit being a secondary consideration. Thus, subjects put themselves at risk for uncertain benefit, and healthy volunteers stand no chance of clinical benefit. But the phase I trial is accepted for several reasons. First and foremost, if one accepts that our society wishes to continue to make progress against disease, it becomes an unavoidable necessity. A new drug must at some point be introduced into the human population.



This must be done in a careful and systematic fashion, but risk can only be minimized, not eliminated.

Second, patients who face the option of a phase I study are often those who have a disease without further standard therapeutic options. Although the chance of benefit for a given patient is likely to be very low, a chance for therapeutic success may be motivation enough [25], and altruism may play a smaller role in patient decision making than frequently thought [26]. Yet even when informed consent may be forthcoming, phase I studies are at greater risk than later phase studies for violating the principle of beneficence (i.e., offering insufficient benefit to justify risk) and for abusing the desperation of a vulnerable patient population at the expense of the ethical principle of justice [27].

Another challenging aspect of phase I studies is drug dosing. In oncology, it has been observed that benefit derived from new cytotoxic drugs occurs more frequently when doses are near the limit of acceptable toleration of side effects [28]. This means that patients who receive lower drug doses earlier in the study are less likely to have benefit, although they may also have less toxicity. Phase I dosing is therefore a balance between minimizing toxicity and maximizing any possible benefit for the greatest number of patients [25]. It is thus incumbent on investigators to carefully plan dosing increments during protocol development and assess side effects as the trial progresses.

Phase III studies, though more likely to confer benefit than phase I studies, still pose ethical challenges. One such difficulty is the decision about whether to stop a trial during interim analysis. A trial of hormone therapy (letrozole) after curative surgery for breast cancer was stopped at an interim analysis when the treated patients demonstrated lower rates of disease recurrence [29]. It may reasonably be asked whether such a study might better be continued blinded until longer follow-up was available or a survival difference was or was not found. While unquestionably it is better to avoid recurrence of breast cancer, the cost of adopting such therapy must be balanced against an incomplete study, other potentially better therapies, or trials that might be aborted by early adoption of the considered drug [30]. We are also accepting the financial cost of a new drug by its adoption. A society may reasonably consider for any therapy whether the gains so achieved are incurred at a reasonable cost in terms of other societal concerns. Such issues make it apparent that ethics is not a matter of nebulous constructs but an integral consideration for clinical trials.

**Quality of Life** Another aspect of research that separates the clinical from the preclinical phase is the human interpretation of ailments. From pain to dyspnea, humans demonstrate a range of subjective degrees of discomfort from the insults of disease [31, 32].

Although less concrete and more difficult to assess than endpoints such as survival or hospital admissions, quality of life or symptom control data can be meaningful to patients and clinicians. In circumstances where endpoints such as survival are not readily demonstrated, such as in rheumatoid arthritis, measurements of quality of life, symptoms, and function are useful to assess drug efficacy [33]. Investigators should endeavor to use validated scales so that the results are less open to question. Still, quality of life measures have provided challenges. How often does one conduct measurements? How does one account for the inevitably missing data points [34]?

In the field of oncology, quality of life scales alone have yet to prove sufficient for drug approval by the FDA. In contrast, other simple and easily comprehensible measures of pain or composite endpoints that include pain have been accepted as a basis for drug marketing [35].

1.4.3 Multidisciplinary Nature of Clinical Trials

**Actors** The manifold tasks and varied expertise required to conduct contemporary clinic trials necessitate the input and assistance of several groups. Prior to initiating a clinical trial, it must be assured that all the players are properly cued. Table 1 lists the persons and groups that typically must be available to conduct a trial, listed roughly in order of appearance but not importance.

Due to the diverse resources required to conduct clinical trials, it is not always practical for an organization to maintain capacity for every aspect of study conduct.

TABLE 1 Entities Involved in Clinical Trials

| Entity                                                | Role                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Principle investigator                                | While not all trials are conceived by the principle investigator, the principle investigator is responsible for the overall conduct of the trial.                                                                                                                    |
| Funding agency/<br>company                            | This may be a corporate, government, or charitable agency. In addition to funding, companies may supply drug. These bodies are frequently involved in receiving and disseminating reports of adverse events.                                                         |
| Statistician                                          | Statisticians are involved in study design, interim analyses, and the final analysis.                                                                                                                                                                                |
| Study coordinators                                    | Study coordinators are involved in all aspects of trials: protocol and form creation, submission of the protocol to various review boards and government regulatory agencies, patient consent and registration, as well as data collection, cleaning, and summation. |
| Contract and financial<br>administrators              | These persons negotiate agreements between funding agencies and centers conducting the trial, aid in the creation of budgets, and distribute funds necessary to conduct the trial.                                                                                   |
| Scientific review<br>committee                        | This body reviews the scientific merit of a clinical trial and may suggest improvements.                                                                                                                                                                             |
| Health/safety<br>committee                            | Although not involved in all studies, this group is responsible for ensuring that investigators adhere to regulations regarding infectious and hazardous substances.                                                                                                 |
| Institutional review<br>board/ethics<br>committee     | This body assesses whether the study meets the standards of respect for persons, beneficence, and justice and will prohibit substandard studies.                                                                                                                     |
| Data safety<br>monitoring board                       | Created before the initiation of the trial, this body provides objective oversight of the study and may recommend early closure of a study for reasons of either significant early benefit or excessive toxicity.                                                    |
| Pharmacists                                           | Pharmacists are responsible for research drug control and accounting.                                                                                                                                                                                                |
| Nursing staff                                         | Drug administration and sample collection requires both nursing staff and physical space, sometimes including facilities for overnight visits.                                                                                                                       |
| Pharmacokinetics<br>specialists                       | Pharmacokineticists are usually involved in phase I drug design and sample collection and analysis but may also be involved in later phase studies.                                                                                                                  |
| Outcomes<br>assessments staff<br>(e.g., radiologists) | Depending upon the outcomes being assessed, radiologists or other specialists may be required to interpret study data. In some instances, independent and blinded individuals or groups may be used to assess study data in a more objective fashion.                |



For this reason, an industry of contract research organizations has arisen to provide research services not available from in-house sources. These organizations can provide services such as research ethics review, protocol preparation, study administration, regulatory consultation, and radiologic imaging support. They can offer the advantages of expertise and efficiency in trial conduct, with offsetting disadvantages of decreased control over details, the need to rely on the contract agency for quality, and the need for careful communication with respect to the hired agency's responsibilities and goals [36].

**Statisticians** The early inclusion of an experienced statistician is advisable for most studies. In order to obtain a useful study result, a hypothesis must be generated and a statistical test must be chosen prior to study conduct. *Post hoc* statistical analyses can lead to new hypotheses for future research but cannot generate definitive answers [37].

A statistician can help to clarify the question under consideration. For example, when conducting a phase II study in heart failure, one may wish to assess the difference in exercise duration between two treatment arms [6]. Using the expected or minimally acceptable difference and the desired error rates, a statistician can advise on the number of patients that need to be recruited to the trial. Failure to determine this need may result in a futile, underpowered study or one which unnecessarily exposes excess patients to an experimental therapy.

In larger, phase III studies, the patient exposure and resource stakes are typically greater. As with our phase II example, realistic expected differences between endpoints must be considered. It must be decided whether the new therapy is likely to be superior, or whether the investigator wishes only to demonstrate that it is noninferior (although either less toxic, more convenient, or substantially cheaper), as the sample size will be larger in the latter case and the hypothesis test different. The ethical challenges of the interim analyses were previously mentioned, but the statistical challenges can also be substantial. One must estimate how many events are required in a population to sufficiently conduct the analysis, then employ a test that will assess the difference while accounting for repeated statistical testing. The goal is to avoid both false-positive studies and prolongation of a futile trial [38].

**Setting** During study development, investigators must decide where the trial will be conducted: primarily among academic centers and cooperative organizations or in community centers, usually under the auspices of a pharmaceutical company and frequently organized by contract research organizations. In addition, a study will be domestic or international.

Traditionally, academic centers and organizations have conducted clinical trials, although this has been changing [39]. While the clinical trials infrastructure is more commonly in place in academic centers, community centers have demonstrated the ability to conduct clinical trials as effectively as academic centers [40–43], and organizations have formed that may efficiently recruit patients within such centers [39]. Community-center-based trials may have the advantage of a more generalizable patient population than that seen in academic centers [44, 45]. Limitations of community trials may include limited recruitment despite declared interest, a need for financial incentives, a need for easy documentation, and a lack

of perceived benefit to the physician [46] or managed care organization [47], although these characteristics are by no means exclusive to community trials [48]. The corporate control of data, use of for-hire ethics boards, and the greater dependence on financial incentives can leave some community trials more open to question [39]. Indeed, possibly as a result of publication bias and data control, publications of industry-sponsored work tend more often to report in favor of the experimental therapy [49]. For this reason, studies conducted by academic centers may offer superior credibility.

While the logistical and regulatory convenience of domestically conducted clinical trials is undisputed, there may be advantages to studies conducted on an international basis. Most evidently, the recruitment pool may be vastly increased, particularly when countries are included where nonexperimental options are relatively limited—a source also of some ethical debate [50]. Dollar costs may also be reduced when developing countries are involved [51]. The result of international studies may be more generalizable and more readily accepted by clinicians debating the applicability of a trial to their setting. While the international adoption of standards such as the Guideline for Good Clinical Practice aims to facilitate drug development by improving the acceptance of trial results by the regulatory bodies of differing countries [52], the actual conduct of such studies can still be challenging. Differing bureaucracies and approval methods for experimental studies can mean expensive or prolonged approval processes. In developing countries, the conduct of trials may require increased support for centers with little experience conducting clinical trials, and simplified and minimized information collection. Despite the sometimes difficult logistics, it is recommended that randomization remain centralized [53].

#### 1.4.4 Know Your Audience, Know Your Market

**Who Is the Audience?** When developing a clinical trial, one must take into consideration the interested parties. First and foremost, there is the patient, who must deem the trial safe and attractive. There is the clinical investigator (and institutional review board), who must find the trial to be of sufficient scientific and ethical merit to allow accrual. There are the regulators, who may need to approve the trial for it to proceed and who will eventually need to approve an agent for nonexperimental use. And finally, there is the market, really an amalgam of the wills of patients and clinicians as influenced by competing therapies. While the term market connotes a mercenary purpose, the consideration of a drug's market is both worthy of time and compatible with the goal of optimal patient care.

**Considering Current and Evolving Practice** Clinical trials are not conducted in isolation. Rather, they become available to patients as an option alongside existing standard therapies. This imposes limitations on the experimental and control arms for a trial. For example, in many instances a patient commencing treatment for more than very mild rheumatoid arthritis (RA) would be a candidate for methotrexate [54]. Starting a patient purely on an experimental therapy could thus be deemed inadequate, and the experimental arm may need to employ both methotrexate and the experimental agent in combination. Similarly, it is considered unethical to unnecessarily delay treatment through the use of a placebo in the control arm of

a patient with RA [55]. Obviously, a trial that fails to consider these points is unlikely to be allowed to proceed, and even if approved may be unable to accrue.

Recognizing variations in clinical practice, a flexible treatment scheme has sometimes been adopted by trialists. In lieu of defining a specific control arm in a clinical trial, investigators may be allowed to choose the particular control or treatment arm that will be employed at their center [56, 57]. Such a trial is more likely to be attractive to a wider range of clinicians, as they may adopt local practices to the trial in question. This has obvious benefits for accrual and may enhance the generalizability of the study. On the other hand, such an open model may make it less clear what is being compared. For example, if the experimental arm contains several forms of therapy, a local investigator may not know if his or her standard regimen has been adequately compared to the experimental arm based on the primary analysis. While subset analyses may be performed, they are typically exploratory.

In addition to present practice, other concurrently operating clinical trials may impact on future practice and the ability to conduct a trial under development. First, a trial in the same population may compete for the finite pool of potential participants. Second, if a competing study is finished and found to be positive before the developing or ongoing study is complete, study completion may become impossible. The competing study may change the standard treatment landscape, alter investigator equipoise over the developing or ongoing study, and inhibit patient accrual. Patients will need to be informed of the evolving standard, and they may choose to avoid or withdraw from the trial.

Just as standards exist in clinical practice, methods of conducting clinical trials are largely standardized. While trial methodology is evolving, investigators and review boards may be uncomfortable with new methods. For example, in phase I studies in oncology, the common method of accrual is to admit cohorts of three to six patients at successive drug doses. Alternative methods, such as the accrual of one patient per dose level, or the continuous reassessment of the maximum tolerated dose using Bayesian methods have been advocated as potentially more efficient [3]. However, there is evidence that the implementation of new study methods is delayed, suggesting the discomfort of physicians or reviewing committees [58].

**Considering Endpoints** The choice of endpoint depends upon both the disease under consideration and the phase of clinical development of the drug. In congestive heart failure (CHF), for example, past successes in improving clinical outcomes have made it difficult to further improve results and to detect such improvements in phase III trials [59]. For drug development, this means that having an early, phase II assessment of activity is important to determine whether a drug should go on to phase III study. Given that phase II trials are intended to be shorter and smaller than phase III trials, using longer term endpoints such as hospitalization or mortality is unlikely to be practical. Surrogate endpoints are therefore considered for these phase II trials. While clinical endpoints represent measures of disease important to patient well-being or survival, surrogate endpoints are alternative endpoints that represent disease biology or a secondary clinical outcome and are intended to shorten the investigative timeline. To be valid, surrogates must correlate well with improvements in important clinical endpoints. One example of such a surrogate is brain natriuretic peptide, a neurohormone that predicts left ventricular function and prognosis and that has also become a diagnostic test [60]. While there is disagree-

ment about which surrogates are useful in CHF [61], the patient exposure to experimental therapy and the cost required by phase III studies dictate that an effort be made to use phase II studies, and surrogate markers can serve a useful role. Phase III endpoints must be more clinically relevant, in part because surrogate endpoints are not entirely reliable. In CHF, therefore, mortality is still a preferred measure of efficacy, although hospitalization rates and other secondary measures may be considered [62].

Endpoints once deemed of limited clinical value may gain importance through greater experience. Improvements in disease-free survival, an endpoint less concrete than overall survival, have historically not been regarded as sufficient to merit a change in clinical practice in many areas of oncology. More recently, analysis of accumulated studies has suggested that 3-year disease-free survival is an accurate surrogate of 5-year survival when administering adjuvant chemotherapy to patients who have had curative surgery for colon cancer [63]. The use of oxaliplatin in the adjuvant colon cancer setting was approved by the FDA on the basis of a disease-free survival benefit, and there is the potential to use such surrogates to shorten drug development time [64].

## 1.5 EXAMPLE IN DRUG DEVELOPMENT

To further understand the clinical trial process, it is useful to consider an example. The field of oncology has seen an increase in the number of experimental agents directed at specific disease mechanisms. These targeted drugs are sometimes considered to have the ability to prevent tumor growth while not actually causing tumor shrinkage (tumor response), and may be termed cytostatic agents. Typically, new drugs are first studied in patients with advanced, metastatic disease, and tumor response has been employed as a surrogate for clinically important endpoints such as survival. The challenge in studying cytostatic drugs is that they may not induce tumor response and may be less effective in patients with greater burdens of disease. Hence, useful drugs may be missed if tumor response is relied upon to demonstrate activity [65].

Such were the considerations during the development of marimastat, a matrix metalloproteinase inhibitor. Matrix metalloproteinases are a family of proteins that degrade extracellular matrix and thus facilitate the migration and metastasis of tumor cells and facilitate vascular growth. Preclinical work suggested marimastat inhibited this process [66]. Except for the first study, performed in healthy volunteers [67], phase I studies suggested a dose-limiting arthritis [68, 69]. These studies indicated doses for further work and suggested that achievable plasma levels were likely sufficient to achieve target inhibition.

Few single-agent phase II studies were performed, and tumor responses were rare [70–72]. With the understanding that marimastat might not show typical responses in tumors, a large study was performed with various tumor types to assess a surrogate endpoint, a change in tumor markers [73]. With the exception of prostate-specific antigen, the tumor markers that were used are not sufficiently associated with clinical endpoints that they are usually accepted as surrogates [65]. While an impact on tumor markers was suggested by this and another study [74], there was no clear evidence of improvement in any clinical endpoint.

Acknowledging the difficulty in detecting activity in metastatic disease, Miller et al. conducted a randomized phase II study in the adjuvant breast cancer setting [75]. This trial encountered musculoskeletal toxicity that prevented drug administration from being sufficiently sustained to warrant further adjuvant study.

Phase II data could thus be regarded as tenuous, but optimism was such that phase III drug development proceeded. In fact, for both the lung cancer and gastric cancer trials, there was no phase II data to support phase III efforts [76, 77]; the study in gastric cancer was based in part on pathological changes noted in a phase I trial [78]. The results of phase III studies were almost universally disappointing [77, 79–81], although minimal activity was seen in gastric cancer [76]. Development of the drug ceased.

It is unfair to be overly critical of the participants in such a story, but certain issues may be usefully considered. First, phase I studies may demonstrate some aspects of a drug's toxicity, but only with more patients and longer term follow-up will toxicity become clear. This became more evident in the phase II study in the adjuvant breast cancer setting, and flushing out the toxicity profile is another argument for phase II studies beyond looking for initial clinical activity. A resource-intensive phase III study would likely have been aborted in the same adjuvant situation. Second, surrogate markers can be misleading [60, 82, 83]. To be considered true surrogate markers, they must be biologically relevant, show a consistent and proportional relationship between a change in the marker and a clinically meaningful endpoint, and this relationship should be demonstrable in repeated studies [60]. Most tumor markers do not satisfy these requirements, and thus their use was probably not justified. That said, even markers directly in the biological pathway of a drug are not a guarantee of adequate surrogacy, as redundant and alternative molecular pathways may dilute or eliminate the relationship of the surrogate to a clinical endpoint. Unfortunately, an adequate biological surrogate test had not been established for *marimastat*. Proceeding to phase III studies based on uncertain surrogate markers was thus a gamble.

How does one decide when to carry out phase III studies in oncology for cytostatic drugs? This is still an evolving field. In terms of using clinical outcomes, the use of stable disease is being used by default, although there is modest evidence of a relationship between this and the more concrete endpoint of survival [84–88]. As response and even stable disease may be difficult to demonstrate in advanced malignancy, biomarkers are likely to remain relevant. Measuring direct effects on tumor is likely ideal, but many tumors are not readily accessible for repeat biopsy after treatment. In this instance, one might pursue changes in biomarkers in accessible tissue such as skin. There is still the hazard, however, that skin changes may not be representative of tumor changes. In either case, unless a similar drug has established a true surrogate relationship for the biomarker in question, investigators are left to establish the relationship, a very difficult task during the limited number of trials undertaken with a developing drug. In the absence of a validated surrogate or true clinical evidence of activity, the preclinical or clinical biological data must be compelling to proceed with large randomized studies. If it is, investigators might consider whether it is better to study the drug in the setting of earlier disease, perhaps in the adjuvant setting. While the benefit of a cytostatic agent may be more evident in this setting, larger treatment groups and longer follow-up are typically required to detect the small improvements in outcome often seen in early disease.

## REFERENCES

1. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22, 151–185.
2. Roberts, T. G., Jr., Lynch, T. J., Jr., and Chabner, B. A. (2003), The phase III trial in the era of targeted therapy: Unraveling the “go or no go” decision, *J. Clin. Oncol.*, 21, 3683–3695.
3. Eisenhauer, E. A., O'Dwyer, P. J., Christian, M., and Humphrey, J. S. (2000), Phase I clinical trial design in cancer drug development, *J. Clin. Oncol.*, 18, 684–692.
4. Kuhlmann, J. (1997), Drug research: From the idea to the product, *Int. J. Clin. Pharmacol. Ther.*, 35, 541–552.
5. Konstam, M. A. (2005), Reliability of ventricular remodeling as a surrogate for use in conjunction with clinical outcomes in heart failure, *Am. J. Cardiol.*, 96, 867–871.
6. Narang, R., Swedberg, K., and Cleland, J. G. (1996), What is the ideal study design for evaluation of treatment for heart failure? Insights from trials assessing the effect of ACE inhibitors on exercise capacity, *Eur. Heart J.*, 17, 120–134.
7. Farrington, P., and Miller, E. (2003), Clinical trials, *Methods Mol. Med.*, 87, 335–352.
8. Simon, R., Wittes, R. E., and Ellenberg, S. S. (1985), Randomized phase II clinical trials, *Cancer Treat. Rep.*, 69, 1375–1381.
9. Freidlin, B., and Simon, R. (2005), Evaluation of randomized discontinuation design, *J. Clin. Oncol.*, 23, 5094–5098.
10. Simon, R. (1987), How large should a phase II trial of a new drug be? *Cancer Treat. Rep.*, 71, 1079–1085.
11. Spilker, B. (1991), Marketing-oriented clinical studies, in *Guide to Clinical Trials*, Raven Press, New York, pp. 367–369.
12. Spilker, B. (1991), Classification and description of phase IV postmarketing study designs, in *Guide to Clinical Trials*, Raven Press, New York, pp. 44–58.
13. Bombardier, C., Laine, L., Reicin, A., Shapiro, D., Burgos-Vargas, R., Davis, B., Day, R., Ferraz, M. B., Hawkey, C. J., Hochberg, M. C., Kvien, T. K., and Schnitzer, T. J. (2000), Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group, *N. Engl. J. Med.*, 343, 1520–1528.
14. Bresalier, R. S., Sandler, R. S., Quan, H., Bolognese, J. A., Oxenius, B., Horgan, K., Lines, C., Riddell, R., Morton, D., Lanus, A., Konstam, M. A., and Baron, J. A. (2005), Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial, *N. Engl. J. Med.*, 352, 1092–1102.
15. U.S. Department of Health and Humans Services F.a.D.A.C.f.D.E.a.R. Guidance for Industry—Allergic Rhinitis: Clinical Development Programs for Drug Products, <http://www.fda.gov/cder/guidance/2718dft.pdf>, 2000, accessed Nov. 10, 2005.
16. Bunn, P. A., Jr., and Franklin, W. (2002), Epidermal growth factor receptor expression, signal pathway, and inhibitors in non-small cell lung cancer, *Semin. Oncol.*, 29, 38–44.
17. U.S. Food and Drug Administration, FDA Public Health Advisory—New Labelling and Distribution Program for Gefinitib (Iressa), <http://www.fda.gov/cder/drug/advisory/iressa.htm>, 2005, accessed Nov. 10, 2005.
18. Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J., and Haber, D. A. (2004), Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib, *N. Engl. J. Med.*, 350, 2129–2139.



19. Paez, J. G., Janne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M. J., Sellers, W. R., Johnson, B. E., and Meyerson, M. (2004), EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy, *Science*, 304, 1497–1500.
20. Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., et al. (2002), Initial sequencing and comparative analysis of the mouse genome, *Nature*, 420, 520–562.
21. Collins, J. M., Zaharko, D. S., Dedrick, R. L., and Chabner, B. A. (1986), Potential roles for preclinical pharmacology in phase I clinical trials, *Cancer Treat. Rep.*, 70, 73–80.
22. Voskoglou-Nomikos, T., Pater, J. L., and Seymour, L. (2003), Clinical predictive value of the in vitro cell line, human xenograft, and mouse allograft preclinical cancer models, *Clin. Cancer Res.*, 9, 4227–4239.
23. U.S. Department of Agriculture A.a.P.I.S, The Animal Welfare Act, <http://www.aphis.usda.gov/lpa/pubs/awact.html>, 2002.
24. World Medical Association, World Medical Association Declaration of Helsinki, <http://www.wma.net/e/policy/b3.htm#note1>, 2004.
25. Agrawal, M., and Emanuel, E. J. (2003), Ethics of phase 1 oncology studies: Reexamining the arguments and data, *JAMA*, 290, 1075–1082.
26. Joffe, S., Cook, E., Clark, J., and Weeks, J. (2003), Altruism among participants in cancer clinical trials, *Proc. Am. Soc. Clin. Oncol.*, 22, 523.
27. Kong, W. M. (2005), Legitimate requests and indecent proposals: Matters of justice in the ethical assessment of phase I trials involving competent patients, *J. Med. Ethics*, 31, 205–208.
28. Von Hoff, D. D., and Turner, J. (1991), Response rates, duration of response, and dose response effects in phase I studies of antineoplastics, *Invest. New Drugs*, 9, 115–122.
29. Goss, P. E., Ingle, J. N., Martino, S., Robert, N. J., Muss, H. B., Piccart, M. J., Castiglione, M., Tu, D., Shepherd, L. E., Pritchard, K. I., Livingston, R. B., Davidson, N. E., Norton, L., Perez, E. A., Abrams, J. S., Therasse, P., Palmer, M. J., and Pater, J. L. (2003), A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer, *N. Engl. J. Med.*, 349, 1793–1802.
30. Bryant, J., and Wolmark, N. (2003), Letrozole after tamoxifen for breast cancer—what is the price of success? *N. Engl. J. Med.*, 349, 1855–1857.
31. Carter, L. E., McNeil, D. W., Vowles, K. E., Sorrell, J. T., Turk, C. L., Ries, B. J., and Hopko, D. R. (2002), Effects of emotion on pain reports, tolerance and physiology. *Pain Res. Manag.*, 7, 21–30.
32. Rietveld, S. (1998), Symptom perception in asthma: A multidisciplinary review, *J. Asthma*, 35, 137–146.
33. Kreutz, G. (1999), European regulatory aspects on new medicines targeted at treatment of rheumatoid arthritis, *Ann. Rheum. Dis.*, 58(Suppl 1), 192–195.
34. Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., Mosconi, P., Osoba, D., Simes, J., and Hurny, C. (1998), Missing quality of life data in cancer clinical trials: Serious problems and challenges, *Stat. Med.*, 17, 517–532.
35. Johnson, J. R., Williams, G., and Pazdur, R. (2003), End points and United States Food and Drug Administration approval of oncology drugs, *J. Clin. Oncol.*, 21, 1404–1411.
36. Welling, P., Lasagna, L., and Banakar, U. (1996), *The Drug Development Process—Increasing Efficiency and Cost-Effectiveness*, Marcel Dekker, New York, pp. 317–351.
37. Moye, L. A., and Deswal, A. (2002), The fragility of cardiovascular clinical trial results, *J. Card. Fail.*, 8, 247–253.

38. Eisenhauer, E., Bonetti, M., and Gelber, R. (2005), Principles of clinical trials, in Cavalli, F., Hansen, H., and Kaye, S., Eds., *Textbook of Medical Oncology*, Martin Dunitz, London, pp. 99–136.
39. Silversides, A. (2004), The tribulations of community-based trials, *CMAJ*, 170, 33.
40. Bressler, N. M., Hawkins, B. S., Bressler, S. B., Miskala, P. H., and Marsh, M. J. (2004), Clinical trial performance of community- vs university-based practices in the submacular surgery trials (SST): SST report no. 2, *Arch. Ophthalmol.*, 122, 857–863.
41. Hjorth, M., Holmberg, E., Rodjer, S., Taube, A., and Westin, J. (1995), Patient accrual and quality of participation in a multicentre study on myeloma: A comparison between major and minor participating centres, *Br. J. Haematol.*, 91, 109–115.
42. Koretz, M. M., Jackson, P. M., Torti, F. M., and Carter, S. K. (1983), A comparison of the quality of participation of community affiliates and that of universities in the Northern California Oncology Group, *J. Clin. Oncol.*, 1, 640–644.
43. Begg, C. B., Carbone, P. P., Elson, P. J., and Zelen, M. (1982), Participation of community hospitals in clinical trials: Analysis of five years of experience in the Eastern Cooperative Oncology Group, *N. Engl. J. Med.*, 306, 1076–1080.
44. Layde, P. M., Broste, S. K., Desbiens, N., Follen, M., Lynn, J., Reding, D., and Vidaillet, H. (1996), Generalizability of clinical studies conducted at tertiary care medical centers: A population-based analysis, *J. Clin. Epidemiol.*, 49, 835–841.
45. Sharpe, N. (2002), Clinical trials and the real world: Selection bias and generalisability of trial results, *Cardiovasc. Drugs Ther.*, 16, 75–77.
46. Pearl, A., Wright, S., Gamble, G., Doughty, R., and Sharpe, N. (2003), Randomised trials in general practice—a New Zealand experience in recruitment, *N. Z. Med. J.*, 116, U681.
47. Donahue, D. C., Lewis, B. E., Ockene, I. S., and Saperia, G. (1996), Research collaboration between an HMO and an academic medical center: Lessons learned. *Acad. Med.*, 71, 126–132.
48. Keinonen, T., Keranen, T., Klaukka, T., Saano, V., Ylitalo, P., and Enlund, H. (2003), Investigator barriers and preferences to conduct clinical drug trials in Finland: A qualitative study, *Pharm. World Sci.*, 25, 251–259.
49. Montaner, J. S., O'Shaughnessy, M. V., and Schechter, M. T. (2001), Industry-sponsored clinical research: A double-edged sword, *Lancet*, 358, 1893–1895.
50. Emanuel, E. J., Currie, X. E., and Herman, A. (2005), Undue inducement in clinical research in developing countries: Is it a worry? *Lancet*, 366, 336–340.
51. Hayasaka, E. (2005), Approaches vary for clinical trials in developing countries, *J. Natl. Cancer Inst.*, 97, 1401–1403.
52. International Conference on Harmonisation Steering Committee, Guideline for Good Clinical Practice, [http://www.ich.org/MediaServer.jserv?@\\_ID=482&@\\_MODE=GLB](http://www.ich.org/MediaServer.jserv?@_ID=482&@_MODE=GLB), 1996, accessed 2005.
53. Yusuf, S., Mehta, S. R., Diaz, R., Paolasso, E., Pais, P., Xavier, D., Xie, C., Ahmed, R. J., Khazmi, K., Zhu, J., and Liu, L. (2004), Challenges in the conduct of large simple trials of important generic questions in resource-poor settings: The CREATE and ECLA trial program evaluating GIK (glucose, insulin and potassium) and low-molecular-weight heparin in acute myocardial infarction, *Am. Heart J.*, 148, 1068–1078.
54. American College of Rheumatology Subcommittee on Rheumatoid Arthritis. Guidelines for the management of rheumatoid arthritis: 2002 Update (2002), *Arthritis Rheum.*, 46, 328–346.
55. Strand, V. (2004), Counterpoint from the trenches: A pragmatic approach to therapeutic trials in rheumatoid arthritis, *Arthritis Rheum.*, 50, 1344–1347.



56. Arriagada, R., Bergman, B., Dunant, A., Le Chevalier, T., Pignon, J. P., and Vansteenkiste, J. (2004), Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer, *N. Engl. J. Med.*, 350, 351–360.
57. Neoptolemos, J. P., Dunn, J. A., Stocken, D. D., Almond, J., Link, K., Beger, H., Bassi, C., Falconi, M., Pederzoli, P., Dervenis, C., Fernandez-Cruz, L., Lacaine, F., Pap, A., Spooner, D., Kerr, D. J., Friess, H., and Buchler, M. W. (2001), Adjuvant chemoradiotherapy and chemotherapy in resectable pancreatic cancer: A randomised controlled trial, *Lancet*, 358, 1576–1585.
58. Dent, S. F., and Eisenhauer, E. A. (1996), Phase I trial design: Are new methodologies being put into practice? *Ann. Oncol.*, 7, 561–566.
59. Massie, B. M. (2003), The dilemma of drug development for heart failure: When is the time to initiate large clinical trials? *J. Card. Fail.*, 9, 347–349.
60. Anand, I. S., Florea, V. G., and Fisher, L. (2002), Surrogate end points in heart failure, *J. Am. Coll. Cardiol.*, 39, 1414–1421.
61. DeMets, D. L. (2000), Design of phase II trials in congestive heart failure, *Am. Heart J.*, 139, S207–S210.
62. Committee for Medicinal Products for Human Use, Note for Guidance on Clinical Investigation of Medicinal Products for the Treatment of Cardiac Failure, Addendum on Acute Heart Failure, <http://www.emea.eu.int/pdfs/human/ewp/298603en.pdf>, 2004.
63. Sargent, D. J., Wieand, H. S., Haller, D. G., Gray, R., Benedetti, J. K., Buyse, M., Labianca, R., Seitz, J. F., O'Callaghan, C. J., Francini, G., Grothey, A., O'Connell, M., Catalano, P. J., Blanke, C. D., Kerr, D., Green, E., Wolmark, N., Andre, T., Goldberg, R. M., and De Gramont, A. (2005), Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials, *J. Clin. Oncol.*, 23, 8664–8670.
64. Andre, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., Topham, C., Zaninelli, M., Clingan, P., Bridgewater, J., Tabah-Fisch, I., and De Gramont, A. (2004), Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer, *N. Engl. J. Med.*, 350, 2343–2351.
65. Gelmon, K. A., Eisenhauer, E. A., Harris, A. L., Ratain, M. J., and Workman, P. (1999), Anticancer agents targeting signaling molecules and cancer cell environment: Challenges for drug development? *J. Natl. Cancer Inst.*, 91, 1281–1287.
66. Hidalgo, M., and Eckhardt, S. G. (2001), Development of matrix metalloproteinase inhibitors in cancer therapy, *J. Natl. Cancer Inst.*, 93, 178–193.
67. Millar, A. W., Brown, P. D., Moore, J., Galloway, W. A., Cornish, A. G., Lenehan, T. J., and Lynch, K. P. (1998), Results of single and repeat dose studies of the oral matrix metalloproteinase inhibitor marimastat in healthy male volunteers, *Br. J. Clin. Pharmacol.*, 45, 21–26.
68. Rosemurgy, A., Harris, J., Langleben, A., Casper, E., Goode, S., and Rasmussen, H. (1999), Marimastat in patients with advanced pancreatic cancer: A dose-finding study, *Am. J. Clin. Oncol.*, 22, 247–252.
69. Wojtowicz-Praga, S., Torri, J., Johnson, M., Steen, V., Marshall, J., Ness, E., Dickson, R., Sale, M., Rasmussen, H. S., Chiodo, T. A., and Hawkins, M. J. (1998), Phase I trial of Marimastat, a novel matrix metalloproteinase inhibitor, administered orally to patients with advanced lung cancer, *J. Clin. Oncol.*, 16, 2150–2156.
70. Evans, J. D., Stark, A., Johnson, C. D., Daniel, F., Carmichael, J., Buckels, J., Imrie, C. W., Brown, P., and Neoptolemos, J. P. (2001), A phase II trial of marimastat in advanced pancreatic cancer, *Br. J. Cancer*, 85, 1865–1870.

71. Quirt, I., Bodurth, A., Lohmann, R., Rusthoven, J., Belanger, K., Young, V., Wainman, N., Stewar, W., and Eisenhauer, E. (2002), Phase II study of marimastat (BB-2516) in malignant melanoma: A clinical and tumor biopsy study of the National Cancer Institute of Canada Clinical Trials Group, *Invest. New Drugs*, 20, 431–437.
72. Rosenbaum, E., Zahurak, M., Sinibaldi, V., Carducci, M. A., Pili, R., Laufer, M., DeWeese, T. L., and Eisenberger, M. A. (2005), Marimastat in the treatment of patients with biochemically relapsed prostate cancer: A prospective randomized, double-blind, phase I/II trial, *Clin. Cancer Res.*, 11, 4437–4443.
73. Nemunaitis, J., Poole, C., Primrose, J., Rosemurgy, A., Malfetano, J., Brown, P., Berrington, A., Cornish, A., Lynch, K., Rasmussen, H., Kerr, D., Cox, D., and Millar, A. (1998), Combined analysis of studies of the effects of the matrix metalloproteinase inhibitor marimastat on serum tumor markers in advanced cancer: Selection of a biologically active and tolerable dose for longer-term studies, *Clin. Cancer Res.*, 4, 1101–1109.
74. Primrose, J. N., Bleiberg, H., Daniel, F., Van, Belle, S., Mansi, J. L., Seymour, M., Johnson, P. W., Neoptolemos, J. P., Baillet, M., Barker, K., Berrington, A., Brown, P. D., Millar, A. W., and Lynch, K. P. (1999), Marimastat in recurrent colorectal cancer: Exploratory evaluation of biological activity by measurement of carcinoembryonic antigen, *Br. J. Cancer*, 79, 509–514.
75. Miller, K. D., Gradishar, W., Schuchter, L., Sparano, J. A., Cobleigh, M., Robert, N., Rasmussen, H., and Sledge, G. W. (2002), A randomized phase II pilot trial of adjuvant marimastat in patients with early-stage breast cancer, *Ann. Oncol.*, 13, 1220–1224.
76. Bramhall, S. R., Hallissey, M. T., Whiting, J., Scholefield, J., Tierney, G., Stuart, R. C., Hawkins, R. E., McCulloch, P., Maughan, T., Brown, P. D., Baillet, M., and Fielding, J. W. (2002), Marimastat as maintenance therapy for patients with advanced gastric cancer: A randomised trial, *Br. J. Cancer*, 86, 1864–1870.
77. Shepherd, F. A., Giaccone, G., Seymour, L., Debruyne, C., Bezjak, A., Hirsh, V., Smylie, M., Rubin, S., Martins, H., Lamont, A., Krzakowski, M., Sadura, A., and Zee, B. (2002), Prospective, randomized, double-blind, placebo-controlled trial of marimastat after response to first-line chemotherapy in patients with small-cell lung cancer: A trial of the National Cancer Institute of Canada—Clinical Trials Group and the European Organization for Research and Treatment of Cancer, *J. Clin. Oncol.*, 20, 4434–4439.
78. Tierney, G. M., Griffin, N. R., Stuart, R. C., Kasem, H., Lynch, K. P., Lury, J. T., Brown, P. D., Millar, A. W., Steele, R. J., and Parsons, S. L. (1999), A pilot study of the safety and effects of the matrix metalloproteinase inhibitor marimastat in gastric cancer, *Eur. J. Cancer*, 35, 563–568.
79. Bramhall, S. R., Rosemurgy, A., Brown, P. D., Bowry, C., and Buckels, J. A. (2001), Marimastat as first-line therapy for patients with unresectable pancreatic cancer: A randomized trial, *J. Clin. Oncol.*, 19, 3447–3455.
80. Bramhall, S. R., Schulz, J., Nemunaitis, J., Brown, P. D., Baillet, M., and Buckels, J. A. (2002), A double-blind placebo-controlled, randomised study comparing gemcitabine and marimastat with gemcitabine and placebo as first line therapy in patients with advanced pancreatic cancer, *Br. J. Cancer*, 87, 161–167.
81. King, J., Zhao, J., Clingan, P., and Morris, D. (2003), Randomised double blind placebo control study of adjuvant treatment with the metalloproteinase inhibitor, Marimastat in patients with inoperable colorectal hepatic metastases: Significant survival advantage in patients with musculoskeletal side-effects, *Anticancer Res.*, 23, 639–645.
82. Thompson, D. F. (2002), Surrogate end points, skepticism, and the CAST study, *Ann. Pharmacother.*, 36, 170–171.
83. Stadler, W. M., and Ratain, M. J. (2000), Development of target-based antineoplastic agents, *Invest. New Drugs*, 18, 7–16.

84. Cesano, A., Lane, S. R., Poulin, R., Ross, G., and Fields, S. Z. (1999), Stabilization of disease as a useful predictor of survival following second-line chemotherapy in small cell lung cancer and ovarian cancer patients, *Int. J. Oncol.*, 15, 1233–1238.
85. Howell, A., Mackintosh, J., Jones, M., Redford, J., Wagstaff, J., and Sellwood, R. A. (1988), The definition of the “no change” category in patients treated with endocrine therapy and chemotherapy for advanced carcinoma of the breast, *Eur. J. Cancer Clin. Oncol.*, 24, 1567–1572.
86. Murray, N., Coppin, C., Coldman, A., Pater, J., and Rapp, E. (1994), Drug delivery analysis of the Canadian multicenter trial in non-small-cell lung cancer, *J. Clin. Oncol.*, 12, 2333–2339.
87. Rapp, E., Pater, J. L., Willan, A., Cormier, Y., Murray, N., Evans, W. K., Hodson, D. I., Clark, D. A., Feld, R., and Arnold, A. M. (1988), Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer—report of a Canadian multicenter randomized trial, *J. Clin. Oncol.*, 6, 633–641.
88. Sargent, D. J., Wieand, H. S., Haller, D. G., Gray, R., Benedetti, J. K., Buyse, M., Labianca, R., Seitz, J. F., O’Callaghan, C. J., Francini, G., Grothey, A., O’Connell, M., Catalano, P. J., Blanke, C. D., Kerr, D., Green, E., Wolmark, N., Andre, T., Goldberg, R. M., and De Gramont, A. (2005), Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: Individual patient data from 20,898 patients on 18 randomized trials, *J. Clin. Oncol.*, 23, 8664–8670.



# 2

---

## *Regulatory Requirements for Investigational New Drug*

**Venkat Rao**

*National and Defense Programs, Defense Division, Alexandria, Virginia*

### **Contents**

|       |                                                       |    |
|-------|-------------------------------------------------------|----|
| 2.1   | Introduction                                          | 24 |
| 2.2   | Investigational New Drug Application Process          | 26 |
| 2.2.1 | Roadmap for Future IND Product Development            | 34 |
| 2.3   | GLP Regulations in Nonclinical Investigations         | 36 |
| 2.4   | Investigational New Drug cGMP Compliance Requirements | 38 |
| 2.4.1 | cGMP for IND Phase I Clinical Trials                  | 39 |
| 2.4.2 | cGMP for IND                                          | 41 |
| 2.4.3 | From cGMP to Quality Systems                          | 41 |
| 2.5   | Role of Orphan Drug Act in Investigational New Drug   | 43 |
| 2.6   | Regulatory Requirements to Protect Human Subjects     | 46 |
| 2.7   | Requirements for Oversight: IRB                       | 49 |
| 2.7.1 | Composition of IRB                                    | 50 |
| 2.8   | Requirements of Financial Disclosure                  | 52 |
| 2.8.1 | Covered Clinical Studies                              | 52 |
| 2.8.2 | Certification and Disclosure of Requirements          | 53 |
| 2.8.3 | Disclosure Statement Evaluation                       | 53 |
| 2.9   | Requirements for Good Tissue Practice Compliance      | 54 |
| 2.10  | Requirements for IND Labeling                         | 55 |

|        |                                                            |    |
|--------|------------------------------------------------------------|----|
| 2.11   | Monitoring of Investigational New Drug Research            | 57 |
| 2.11.1 | Clinical Risk Assessment                                   | 58 |
| 2.11.2 | Computerized Systems in Clinical Trials                    | 60 |
| 2.11.3 | Quality Assurance                                          | 62 |
| 2.12   | Emerging Biosafety and Biosecurity Requirements            | 63 |
| 2.13   | Conclusions                                                | 66 |
|        | Appendix: Applicable and Relevant Regulations Covering IND | 67 |
|        | References                                                 | 68 |
|        | Bibliography                                               | 69 |

## 2.1 INTRODUCTION

Investigational new drug (IND) is a key phase within the drug development life cycle that requires considerable interactions between pharmaceutical companies and the Food and Drug Administration (FDA), the federal agency with principal responsibility to license new medical products for human use. As the IND approval authority it is the FDA's responsibility to protect public health by ensuring the safety, efficacy, and security of human and veterinary drugs and biological products. At the same time, the FDA's mission is to enable rapid advancement of pharmacological therapeutics development through technological innovations that make medicines more effective, safer, and affordable.

In the context of the IND process, these dual missions may seem contradictory and require a carefully balanced consideration of the risk–benefit of new medical products. For on the one hand, IND is the mechanism for new innovations in pharmacological therapeutics to meet the public health challenges; it is also the venue for introducing new risks with potentially devastating impact on public health and general environment. With the increasing number of new product development projects involving genetic engineering and recombinant DNA (deoxyribonucleic acid) technology in biological product development, the scope and volume of cell-based new biologics product development have expanded dramatically in the past two decades. For example, a new class of cell-based recombinant technology products generally known as the human cells, tissues, and cellular and tissue-based products (HCT/Ps) have created an entire array of investigational biologics. Similarly, new biodefense-oriented medical countermeasures such as vaccines, immunoglobulins, and monoclonal antibodies to protect and counter bioterrorism-related threats have potentials to introduce long-term human health and ecological risks.

Therefore, IND review and approval process will have to take into consideration not only the inherent benefits of introducing novel technological solutions for medical countermeasures during clinical trial, but consider the potential adverse impact to clinical trial subjects and the long-term public health and environmental consequences.

Nevertheless, medical product development is never complete without safety and efficacy data collected directly from studies on human subjects, which is why every drug or biologics product developer is bound to submit an IND application. This application process requires considerable forethought and preparation to ensure that the product under development is suitable for studies with human subjects.

Clinical investigators participating in the clinical studies and the study sponsors must document every facet of the study planning, data generation, and data management as long as the IND is in effect.

Regulatory affairs covering new medical product development are influenced by external factors driven by industry interests and new legislations promulgated by Congress aimed at improving the quality and safety of pharmaceuticals to protect and promote public health. Proactive efforts of the FDA in international forums and multilateral engagements with entities such as the International Conference on Harmonisation (ICH) and the Organization for Economic Cooperation and Development (OECD) brings scientific and technical discussions and harmonization of quality systems and safety guidelines for medical products registration.

In the past three decades a couple of major external events posed unique challenges to the drug development policies in general and IND-related regulatory affairs in particular.

First, the advent of AIDS (acquired immunodeficiency syndrome) in the 1980s drew unprecedented attention to the entire drug development and approval process. Changes in the new drug approval process was aimed at broadening involvement of patient community during the early phase of drug development activities and an overall pressure on the government agency to expedite the new drug approval process. Enormous efforts were made by those affected by the epidemic to gain “expanded access” to investigational drugs even before FDA granted a formal approval.

There simply was no mechanism in place within the regulatory affairs environment to circumvent the key milestones required in the new drug approval process and hesitation in the scientific circles to provide expanded access to highly toxic investigational drugs with potentially dangerous adverse effects. Combined efforts from the regulatory affairs and scientific community created the new National Institutes for Allergy and Infectious Diseases (NIAID) and funded the AIDS treatment research initiative known as Community Programs for Clinical Research on AIDS (CPCRA). A network of clinical study centers were created under the CPCRA to provide HIV-infected community access to IND products through participation in clinical trials. Expanded access to AIDS-related investigational drugs were made possible through IND treatment and parallel track protocols.

The complex nature of the HIV epidemic and highly toxic properties of some IND hindered broader access to possible pharmacological interventions for the affected community. However, the FDA finally issued regulations that broadly interpret use of IND for therapeutic purposes, which provided access to thousands of critically ill patients drugs not yet formally licensed for commercial use. NIAID created a “parallel track” program where selected INDs were made available to HIV-infected patients who could not participate in the clinical trial and have no other clinical alternatives. The parallel track policy was invoked to access IND for broader therapeutic uses, when available clinical evidence is less than adequate to apply the treatment IND clause to broaden access to investigational drugs for therapeutic uses.

Second, post-9/11 challenges to develop medical countermeasures to manage the threat of bioterrorism introduced unexpected demands on the new drug development regulations. The global threat environment called for expedited development of medical countermeasures such as vaccines, antidotes, and diagnostics to prevent

and protect the public against biological warfare agents. With very limited production capacity for vaccine products in the biotechnology industry community, and an assortment of experimental candidate products at very early stages in development, challenges were daunting at the starting point of this major national undertaking. Next, federal regulators and the biodefense product development community were faced with a major technological hurdle by way of potency testing for protective effects of experimental vaccines requiring live-agent tests, which completely eliminated the option of clinical trials for product efficacy determination. With clinical data only on the safety of the product, for most part regulatory decision to grant new biological licensure application would be based on efficacy data from experimental animal studies.

In 2002, the FDA published, a landmark “animal efficacy rule” amending the drug and biologics product regulations to allow appropriate studies in animals in cases where excessive toxicity would make human efficacy studies not ethical or feasible [1]. For example, the FDA approved ciprofloxacin for postexposure management of inhalational anthrax based on serum concentrations of the drug in human studies as a surrogate measure to serum concentrations associated with survival of animals exposed to aerosolized *Bacillus anthracis* spores. This data together with related data on the efficacy of ciprofloxacin in humans on other infections was the basis for the approval.

Use of the anthrax vaccine as an investigational new drug by the U.S. Department of Defense is yet another area where regulatory affairs relating to IND came head on with the national priority to rapidly develop and deploy medical countermeasures to protect military personnel and civilians from biological warfare agents. Controversies surrounding the military management of the unprecedented waiver to the FDA requirements of informed consent process and the role of institutional review board (IRB), the institutional level oversight authority over the conduct of the anthrax clinical trial, are yet to be resolved. As there are no published clinical studies on the efficacy or long-term safety of the anthrax vaccine, the protective value of the vaccine in humans is unknown.

With the best of intentions to protect the military and the public from the real threat of biological weapons, the wrenching decision by the military authorities to vaccinate uniformed personnel with an IND vaccine product through an unprecedented waiver of informed consent put the government in uncharted territory on the regulatory policy relating to IND development.

The following sections in this chapter will examine in detail the investigational new drug application process; the prevailing regulations relating to various aspects of IND investigation; latest requirements under good tissue practice for highly sophisticated biotechnology-derived products; various mechanisms, tools, and performance matrix to monitor IND clinical investigations; and emerging biosafety and biosecurity considerations in IND clinical investigations.

## 2.2 INVESTIGATIONAL NEW DRUG APPLICATION PROCESS

The IND is an embedded phase within the long drug development life cycle that could run anywhere from 10 to 13 years depending on the candidate product and study requirements. During the preclinical experimental investigations, a candidate



drug product with promising pharmacological potentials against a target disease or a physiological condition, and did not cause unacceptable damage to healthy tissue, is eligible to move into the IND phase. Through the IND process, the study sponsor requests permission from the FDA to begin clinical trials.

The IND is also a mechanism through which the pharmaceutical industry sponsor will obtain exemption to transport the IND product across state lines. Under the current federal law only licensed drugs are permitted for broader distribution and transportation across state lines. Thus, IND allows transport of investigational drug products for expanded clinical trials involving geographically dispersed multiple study locations.

The FDA grants three types of INDs: (a) investigator-initiated IND, (b) emergency use of IND, and (c) treatment IND. These INDs are granted either for commercial uses or for research use categories. A commercial uses category basically refers to an application from a pharmaceutical industry sponsor involved in medical product development for commerce. Whereas the research use category covers clinical investigations performed with the academic objective to better understand the clinical pathologies and effectiveness of medical countermeasure strategies. Research-oriented clinical investigations may involve either experimental products or an approved drug for new medical interventions.

*Investigator-Initiated IND* The application is submitted by a clinical investigator who is the study sponsor as well as the principal clinical investigator. An investigator-initiated IND may not have a commercial intent and primarily to investigate an unapproved medical countermeasure, or an approved product for a new indication not covered under the approved label. Investigators are required to file an IND application if the planned study with an approved drug involves a new patient population, which was not the basis in the clinical studies used in the earlier new drug licensure application.

*Emergency Use of IND* During a public health emergency the need for an experimental product may arise, but without sufficient time to complete the formal IND application process. Under such extraordinary circumstances the regulatory agency may authorize shipment of an investigational product for a specified use in advance of an IND. A public health official should make a request to the FDA under emergency situations to obtain the authorization for emergency uses.

*Treatment IND* The FDA issues a treatment IND for interim broader clinical uses if the experimental product shows promise in clinical trials for serious public health problems or immediately life-threatening conditions while the final clinical work is still under progress and the FDA review underway. For example, the 1980s AIDS epidemic introduced treatment INDs as part of a long-term effort to incorporate the concept of expanded access into the IND regulations. With very limited therapeutic options available to control and manage an expanding epidemic, use of investigational products for therapeutic purposes was made possible through this regulatory mechanism. In 2006, the FDA proposed new rules allowing seriously ill patients, and those with HIV/AIDS with limited or no treatment options, easier access to unapproved investigational products.

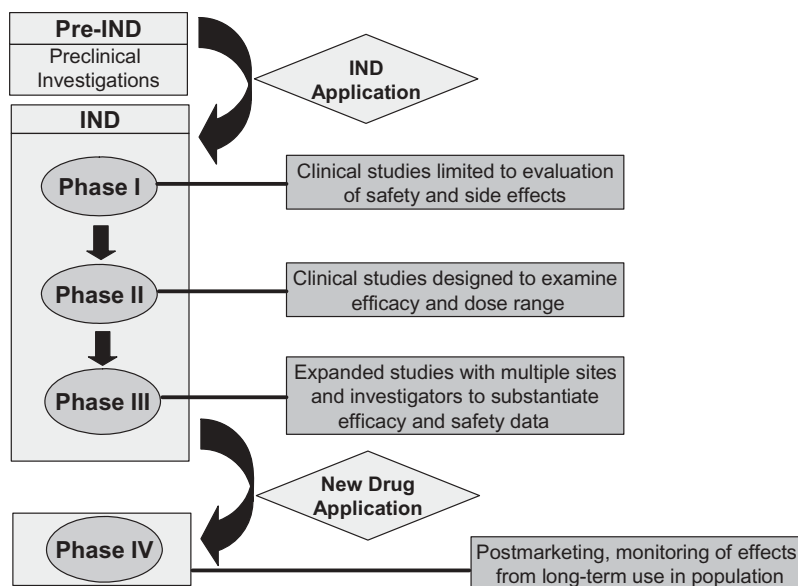


**EXHIBIT 1** Key development milestones in the investigational new drug development life cycle. Phase IV generally refers to a postmarketing surveillance of a newly approved medical product.

Exhibit 1 illustrates the key milestones in the IND phase of the drug development life cycle. At the conclusion of pharmacological and toxicological investigations on candidate experimental products during pre-IND phase, the drug sponsor makes a key decision to proceed with the IND application for those candidates with promising pharmacological data on the efficacy and acceptable toxicological data on safety. Hence, the IND application must have a sound rationale and supportive experimental data covering the following three broad areas:

**Pharmacology and Toxicological Data** Data from studies on experimental animals or similar model systems allow an assessment on the safety of the product to begin testing in human subjects and promising pharmacological effectiveness together with the underlying biochemical and molecular mechanism of action. This set of experimental data is collectively known as the pharmacokinetics and pharmacodynamics (PK/PD) and serves as a key benchmark on the bioavailability and mode of action.

**Clinical Protocol and Investigator Information** IND application must have a detailed protocol of the proposed clinical studies and risks to human subjects, if any, during the course of the clinical investigation. This section should also include sufficient information on the educational background and technical qualifications of the clinical investigators, who are generally practicing clinicians, and other medical and scientific professional with oversight responsibility. Also, included under this section is a certified undertaking by the principal investigator to obtain informed consent from all human subjects participating in the clinical trial and approved by the institutional review board (IRB) specially created to oversee the conduct of a proposed clinical trial.

**Product Manufacture Information** This section will provide a summary of the product composition, chemistry, formulation information, product stability

data, and manufacturer-related information, controls used for manufacture, and product quality assurance.

Phase I clinical studies focus on clinical pharmacology and short-term tolerance tests involving only a small number (not more than 50) healthy volunteers. The primary goal of phase I is the safety of the IND and a determination of dose ranges and appropriate routes of administration. Depending on the study design, phase I may involve limited pharmacokinetic analysis. Although in most cases phase I studies are performed on healthy individuals, there are instances patients are enrolled as human subjects during this phase when clinical investigation involves highly toxic products or targeted against life-threatening diseases such as AIDS or cancer.

Phase II clinical trials are designed to examine efficacy and refine the dose range from previous investigation [2]. These trials are longer in duration, taking 2 or more years, compared to phase I and involve a larger pool of human subjects and may involve more than one study location. Most phase II study designs are randomized, case-controlled investigations, where a group of patients receiving the IND drug, the “treatment group,” is compared with a matched group of patients with comparable clinical profiles, case history, and factors such as age, sex, and other demographic background, the “control group,” receives a placebo or standard therapy. Most phase II studies are double blinded by design in that both the patients and clinical investigators administering the study do not know the composition of the treatment and control groups and who is getting the IND product under study. The randomized, case-controlled study design greatly improves the validity of the clinical outcome data, and the double-blind design reduces errors in the study interpretation or other forms of bias.

Phase III trials are expanded investigations involving an even larger pool of human subjects running into thousands and involving multiple study locations and clinical investigators to address differences in responses due to demographic and geographic factors. Phase III trials are designed to further substantiate observations on safety and efficacy from previous clinical investigations, and a larger pool of human subjects allow the appearance of potentially less frequent adverse events not captured in smaller study populations.

At the conclusion of the phase III trial phase, the study sponsor would file applications with the FDA to obtain commercial licensure for the IND. A new drug application (NDA) is filed for a drug candidate, and for a biologics product this application is called biological licensure application (BLA). A product with an NDA/BLA is no longer an IND and is now commercially available for the public. According to industry reports, about 20% of the IND candidates make it through all phases of the research and are finally approved.

Phase IV could be either a postmarketing study to obtain additional information on the new drug in terms of rare adverse events, additional benefits, and optimal usage not captured during the clinical trials. Given a much broader population now exposed to the new medical product, early phase monitoring in the postmarketing period is generally considered phase IV. This phase is unique in that clinical investigation up until this point is a controlled study with a “treatment” group compared with a matched “control” group. Phase IV has no control group but merely tracks reports for adverse events or other product-related performance data. A phase IV data collection requirement is not mandatory for every approved medical product.

Depending on the potentials for adverse events from controlled clinical studies, the FDA may direct a sponsor to perform postmarketing surveillance for known and unexpected adverse events. Alternatively, a sponsor may proactively implement a phase IV plan to collect data on the potential benefits from the new drug to capture or increase market share for the new medical product.

Phase IV is sometimes differentiated from postmarketing surveillance when this phase actually involves a more structured design and clinical intervention within the terms of the product license. This differs from a postmarketing surveillance, which is mostly noninterventional and observational of the general population using a newly approved medical product. An interventional phase IV is carried out at hospitals or clinics to further consolidate the efficacy database of the newly approved product. This could be part of a risk management plan by the sponsor to obtain additional efficacy and safety data before general release of a new product.

The totality of findings from a well-designed, multiple clinical studies using targeted patient population with a demographic makeup similar to the general population forms the scientific rationale and data analytics to support submission of an NDA or a BLA to the FDA. It could take as many as 3 years for the FDA to review the nonclinical and clinical trials of data submitted in support of an NDA/BLA package before making a final decision on approval.

As part of this protracted engagement between the study sponsors and the FDA during the IND phase, several meetings are held as milestones for key decision points.

Exhibit 2 illustrates a work flow diagram for the industry meetings with the FDA during the IND phase and description of the meeting objectives and outcomes.

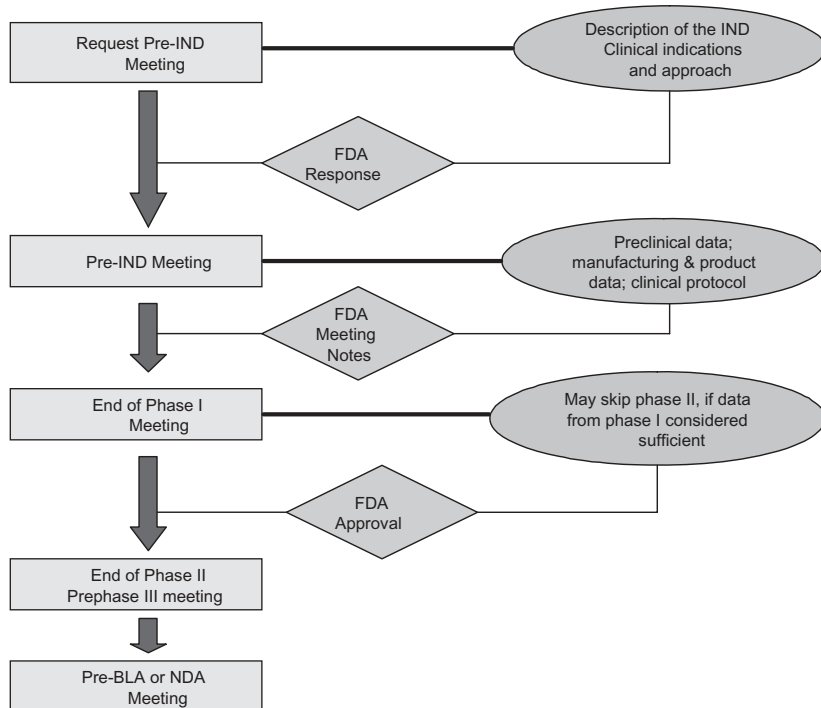


**EXHIBIT 2** Flow diagram of industry meeting with the Food and Drug Administration during various phases of the IND process.

Product sponsor with plans to begin the IND phase must first submit a request for a pre-IND meeting with the FDA and include in that request a brief description of the experimental product, description of the clinical indication, and clinical study approach. Once the meeting date is established, the product sponsor must submit a pre-IND package (as required under 21 CFR 312.82) providing more detailed information on the preclinical experimental data, product manufacturing information, preliminary information on physiochemical characterization of the product and manufacturing specifications, and the outline of the proposed clinical protocol(s).

Exhibit 3 summarizes the general focus of the key IND meetings and the nature of information requirements for discussions. For example, during the pre-IND meeting, the general focus of the information requirements are chemistry and formulation-centric data, whereas the end of phase II (EOP-II) meeting covers study progress review, manufacturing, product stability, safety issues, process validation, and quality systems related information unique to drugs, biologics, and rDNA protein biotechnology drugs. The EOP-II meeting will review the efficacy study results, data gaps and deficiencies, and phase III study plans. Sponsors may update the FDA on potential problems identified and resolved, and information necessary in support of the marketing application.

The IND submissions are a complicated undertaking that should not only meet the technical content requirements, but also strictly comply with the content and format specifications (21 CFR 312.23).

Exhibit 4 is an example of the content requirements in a typical IND submission compliant with the regulatory requirements.

An investigator brochure must be included in the IND submission if the candidate product is supplied to clinical investigators who are not part of the study sponsor's organization. An investigator brochure together with the clinical trial protocol are the fundamental documents required in the IND submission. The investigator brochure must provide a description of the candidate product, a summary of the pharmacology and toxicology, summary information on safety and effectiveness, and a description of the risks of adverse events and recommended precautions or special monitoring.

Once the IND package is received from the sponsor-investigator, the FDA will assign an IND number. A regulatory project manager (RPM) for each submission handles all administrative matters related to processing the IND and serves as the regulatory point of contact for the study sponsor and/or investigator.

Exhibit 5 is a work flow diagram of the IND review and approval process. IND submissions could be made either as hardcopy document or electronic submission. An alternate mechanism to submitting IND is through the master file submission (21 CFR 314.420). A master file will contain product and manufacturing information but does not include clinical protocol. Permit holder for the master file could at a later period add the clinical protocol information when filing the IND. The master file submission format protects proprietary product and manufacturing information from other outside organizations participating in the clinical trial. Using a cross-reference filing format, authorized persons would then file relevant clinical information without access to other proprietary information. The FDA will access both IND and master file information to begin the review. The review process begins only when the IND files are populated in the master file.

Essentially, the IND review team for a drug or biologics product candidate consists of the (a) RPM, (b) product reviewer, (c) pharmacology/toxicology reviewer,

| IND Meetings                            | General Objectives                                                                                                                                                                                                                                                                                                                      | Drugs/Biotechnology Products/Conventional Biologics                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                     |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                         |                                                                                                                                                                                                                                                                                                                                         | Safety                                                                                                                                                                                                                                                                                                                                  | Dose/Formulation                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                     |
| <b>Pre-IND Meeting</b>                  | <ul style="list-style-type: none"> <li>Physical, chemical, and biological characteristics</li> <li>Manufacturers</li> <li>Sources and method of preparation</li> <li>Removal of toxic agents</li> <li>Quality controls</li> <li>Formulations</li> <li>Sterility</li> <li>Pharmacology</li> <li>Toxicology</li> <li>Stability</li> </ul> | <ul style="list-style-type: none"> <li>Drugs from human sources</li> <li>Drugs from animal sources</li> <li>Biotechnology drugs</li> <li>Botanical drugs</li> <li>Reagents from animal or cell line sources</li> </ul>                                                                                                                  | <ul style="list-style-type: none"> <li>Novel excipients</li> <li>Novel dosage forms</li> <li>Drug-device delivery systems</li> </ul>                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                     |
| EOP-II                                  | General Objectives                                                                                                                                                                                                                                                                                                                      | Drugs                                                                                                                                                                                                                                                                                                                                   | rDNA Protein Biotechnology Drugs                                                                                                                                                                                                                                                                                                                     | Conventional Biologics <sup>a</sup>                                                                                                                                                                                                                                 |
| <b>End-of-Phase II Meeting (EOP-II)</b> | <ul style="list-style-type: none"> <li>Physical, chemical, and biological characterization</li> <li>Process validation</li> <li>Environmental consideration</li> <li>Manufacturing considerations</li> <li>Facility-related issues</li> <li>Novel policy issues or concerns</li> </ul>                                                  | Unique physicochemical and biological properties<br><br>Physicochemical characterization<br><br>Starting material designation<br><br>Qualification of impurities<br><br>Removal of adventitious agents<br><br>Approach to specifications<br><br>Sterilization process validation<br><br>Stability protocols<br><br>Environmental impact | Removal of adventitious agents<br><br>Approach to specifications<br><br>Sterilization process validation<br><br>Stability protocols<br><br>Environmental impact<br><br>Bioassay<br><br>Adequacy of cell bank characterization <sup>b</sup><br><br>Removal of product and product-related impurities<br><br>Bioactivity of product-related substances | Removal of adventitious agents<br><br>Approach to specifications<br><br>Sterilization process validation<br><br>Stability protocols<br><br>Environmental impact<br><br>Coordination of facility design<br><br>Process validation consideration<br><br>Potency assay |

<sup>a</sup>Nonrecombinant vaccines and blood products.

<sup>b</sup>Would include, but not limited to, biochemical characterizations such as peptide map, amino acid sequence, disulfide linkages, higher order structure, glycosylation sites and structures, other posttranslational modifications and plans for completion, if still incomplete.

**EXHIBIT 3** Summary of the objectives of IND meetings and information requirements for drugs/biologics candidates.

(d) clinical reviewer, and (e) statistical reviewer. During the first 30 days of the review period, the RPM will communicate with the study sponsor to obtain clarification on review comments from the IND review team and resolve any pending issues from the review team.

| General IND Submission Format and Content              |
|--------------------------------------------------------|
| Cover Sheet                                            |
| Table of Contents                                      |
| Introductory Statements and General Investigation Plan |
| Investigator's Brochure                                |
| Protocol for Each Planned Study                        |
| Chemistry, Manufacturing, and Control Information      |
| Toxicology and Pharmacology Information                |
| IRB Approved Consent Form                              |
| Previous Human Experience                              |
| Additional Information                                 |

**EXHIBIT 4** Illustrative example of the information content requirements in a typical IND submission by sponsor of the medical product development.

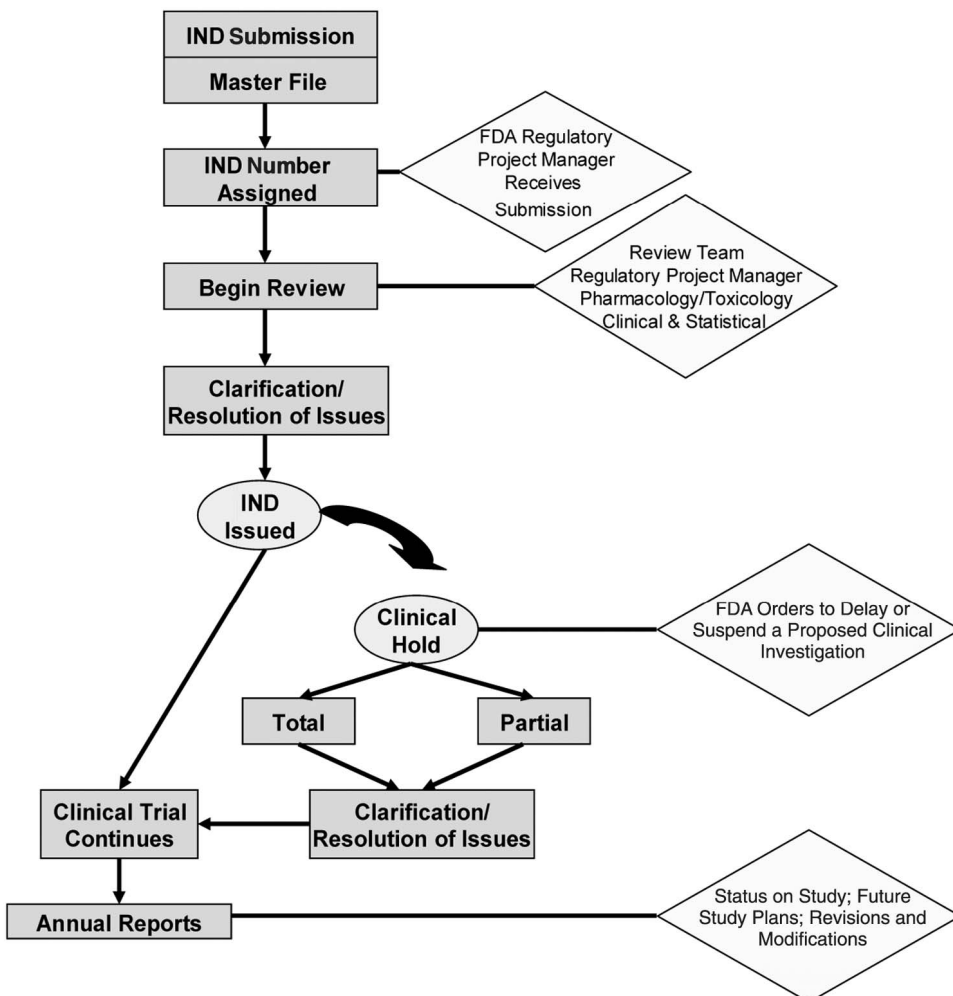


**EXHIBIT 5** Flow diagram of the IND review and approval process.



Usually within 30 days after submission, the IND goes into effect and the decision is communicated via a letter as long as there are no holds placed on the IND.

During the IND review, regulators may decide to place a clinical hold. A clinical hold is an order issued by the FDA to delay a proposed clinical investigation or to suspend an ongoing investigation. When a complete clinical hold is placed on an IND, all clinical work comes to a halt. The FDA may place a partial clinical hold to delay or suspend part of the proposed clinical investigation, meaning that holds may be placed on some clinical protocols within the study while others are allowed to proceed without delay. The regulatory criteria guiding the FDA's decision to place a clinical hold may be one or a combination of risk factors such as: (a) human subjects would be exposed to unreasonable and significant risk; (b) the investigator brochure was misleading, or erroneous, or incomplete; (c) the clinical investigators are not qualified; or (d) IND does not contain sufficient information to assess the risks to subjects of the proposed study (21 CFR 312.42). A hold is placed on more advanced phase II/III trials if a proposed study design is deficient to meet the study objectives.

It is up to the study sponsor to provide the deficient information and clarifications to release the clinical hold placed on the trial. If the sponsor response addresses all the issues detailed in the clinical hold letter, the response is considered complete and the FDA is required to respond within 30 days of receipt of the submission. However, the 30-day clock does not apply to partial or incomplete responses to the clinical hold.

Clinical hold is a risk management mechanism available to the regulators to address potential deficiencies and safety-related issues during the IND phase of the product development. For example, the new generation of tissue-derived biologics products and vaccines pose considerable challenge in the areas of safety, quality, and product potency determination. To a large measure, clinical trials are designed to collecting data to addressing these sorts of issues during the product licensure application. As stated above, one of the more common reasons a clinical hold may be placed is when the FDA cannot determine from an IND clinical protocol the risks to human subjects during clinical investigations.

One of the alternative options to circumvent clinical trial requirements and facilitate product development in areas related to national security such as bioterrorism threat, or a major public health emergency such as pandemic influenza, is to approve medical treatment based on efficacy data from experimental animals. The regulation, known as "animal rule," provides a mechanism for the FDA to expedite the approval process as long as the efficacy data from preclinical investigations on experimental animals are well-designed and data from the studies considered sufficient and adequate by the expert review committee. Only safety data from a phase I clinical trial is required as supportive evidence in making the final determination.

### **2.2.1 Roadmap for Future IND Product Development**

There is a growing concern in the pharmaceutical industry, scientific community, and among regulators on the decrease in the number of IND applications submitted over the past several years. Regulatory burden is cited as a major concern by the industry, although other factors related to extremely expensive up-front investments



required in new drug/biological development projects and the unpredictability of the product development activities. The product development community is beginning to employ the power of computational tools through bioinformatics to develop predicative tools on safety, effectiveness, and scale-up and manufacture of candidate therapeutics products.

Recognizing the importance of technology and regulatory compliance alignment in the development of novel therapies, the FDA unveiled a strategic document in 2004, the Critical Path Initiatives [3], which was aimed at identifying existing challenges in pharmaceuticals-oriented technology domains having a direct impact on the discovery and development of novel medicinal products. As part of this effort the Critical Path Opportunities List was compiled grouping program areas of emphasis under six categories covering mostly new product development strategies and technologies, development of new product development, testing and evaluation methods, and application of best business practices in product development. A brief summary of the six grouped critical path opportunities aimed at IND development activities are:

1. *Better evaluation tools such as biomarkers* for pregnancy, cardiovascular diseases, infectious diseases, cancer, neuropsychiatric diseases, and the like.
2. *Streamline clinical trials* in terms of innovative trial designs, application of best business practices in clinical data management, trial protocol development, and data analytics.
3. *Harnessing the power of bioinformatics* to identify new drugs and biologics candidates for development, identify safety biomarkers, adverse effects data mining, modeling and analysis, and failure analysis.
4. *Moving manufacturing toward twenty-first century* covering biotechnology-derived products, new generation vaccines, detection of contaminants in biologics during product development, scale-up and manufacturing, tissue engineering, product potency evaluation, and application of nanotechnology for therapeutics development.
5. *Development of products to meet public health needs* such as antimicrobial testing strategies, safety of blood and blood-derived products, and novel animal models in biodefense product development.
6. *Product development for at-risk population* with a specific focus on pediatrics. This would include better extrapolation of dose–response regimen in pediatric clinical trials, drug metabolism and therapeutic response, and new therapies for juvenile diabetes.

The Critical Path Initiative is unclear on market-driven factors critical to successful new drug development. First, biotechnology companies are involved in highly sophisticated genomic and proteomics-based product development and clinical gene therapy work in a domain with unclear regulatory implications and product liability issues due to potentials for long-term risks to public health. Commercially oriented research organizations are afraid of long-term liability lawsuits and have fewer incentives to take an active role in government-funded product development projects. This has to a large extent adversely reduced the available capacity to undertake a broad range of new product development projects.

Second, insurance coverage reform requires an equally forward leaning regulatory policy environment to encourage present shortage of providers of insurance to cover clinical trials. Whereas funded by the National Institutes of Health product development programs provide a modest coverage of the insurance cost incurred as part of the clinical trials, most of the cost is borne by the industry, hospitals, and academic institutions involved in medical product development related projects. At present, there are no defined mechanisms available within the existing regulations to the industry to directly address these issues with the regulatory agencies.

Finally, a roadmap for the future is incomplete if channeling of the R&D investment fails to reflect the requirements identified in the Critical Path Initiative. These industry investments in the R&D may not be consistent with the goals identified in the Critical Path Opportunities List. Industry investment in R&D priorities are defined and driven by market consideration. For example, considerable R&D resources are invested in protecting existing market shares, or to develop what are called “me too” drugs, with focus mostly in the formulary redesign and testing and not in breakthrough research in novel therapeutics development. Likewise, considerable product R&D and development is invested in derivative drugs taking the resources away from targets identified in the Critical Path Opportunities List.

Regulatory analysts are of the opinion that the national policy toward advanced therapeutics development requires a broader engagement of the R&D community, academia, and the industry, together with increased funding for processes development and assessments. Similarly, additional policy-level initiatives are required to simplify legal and financial barriers to clinical trials and additional resources to emerging genomics and proteomics-based product development.

## 2.3 GLP REGULATIONS IN NONCLINICAL INVESTIGATIONS

In the drug development process, preclinical investigation is a critical step during which drug molecules discovered from early phases of toxicological screening are subjected to a comprehensive animal testing before an IND can be filed. The regulatory requirements governing this process have imposed strict rules as to how a newly discovered chemical or biochemical molecule will be tested and evaluated prior to approval for testing in humans. The existing good laboratory practice (GLP) guidelines clearly lay out how preclinical experimentations is done in order to ensure the safety of the drug molecules, which then forms the basis for filing an IND application for approval to begin clinical trials (21 CFR 58).

Fundamentally, GLP is a quality system concerned with the organizational process and conditions under which nonclinical health and environmental safety studies are conducted. The organizational, personnel, and facility related GLP compliance requirements are written to ensure quality of data produced by nonclinical and clinical laboratories meet the best business practices and provide international acceptance to the quality of data generated in support of a new medical product licensure application.

The GLP regulations are part of the broad good laboratory practices for conducting preclinical experimental investigations in pharmaceutical product development for research or marketing permits. The scope of GLP regulations includes food and color additives, animal food additives, human and animal drugs, medical devices for

human use, biological products, and electronic products (biomedical devices). Compliance with GLP regulations is part of the requirements under the IND application for drugs and biologics (21 CFR312).

Existing GLP guidelines clearly define the parameters of preclinical experimental investigation to include *in vivo* and *in vitro* experiments in which test articles are studied prospectively in test systems under laboratory conditions to determine their safety. However, these tests do not include studies utilizing human subjects or clinical studies or field trials in animals. Also, preclinical experiments covered under this guideline does not include basic exploratory studies carried out to determine whether a test article has any potential utility or to determine physical or chemical characteristics of a test article (21 CFR58.3).

When product development companies outsource nonclinical studies to contract research facilities or academic research and development establishments, the GLP regulations require every entity participating in the nonclinical study included as part of the IND application and must comply with the provisions set forth under this regulation. As part of the enforcement mechanism the FDA may conduct facility inspection or authorize a designated third party with credentials to perform facility visit to ensure all laboratory records and study specimens related to the study are maintained and remain within the scope of the IND application under investigation.

Exhibit 6 schematically represents the key GLP requirements as part of pre-clinical investigations of an IND product, comprised of facility infrastructure components, business process, and personnel experience.

The GLP guidelines related to infrastructure cover experimental animal facility performing clinical research sites. Facility floor plans and material and process flows to attain operational isolation and prevent cross-contamination is one of the key criteria in the GLP compliance. Facility floor design and process flow plans are assisted by standard operating procedures to all study methods, operational-related

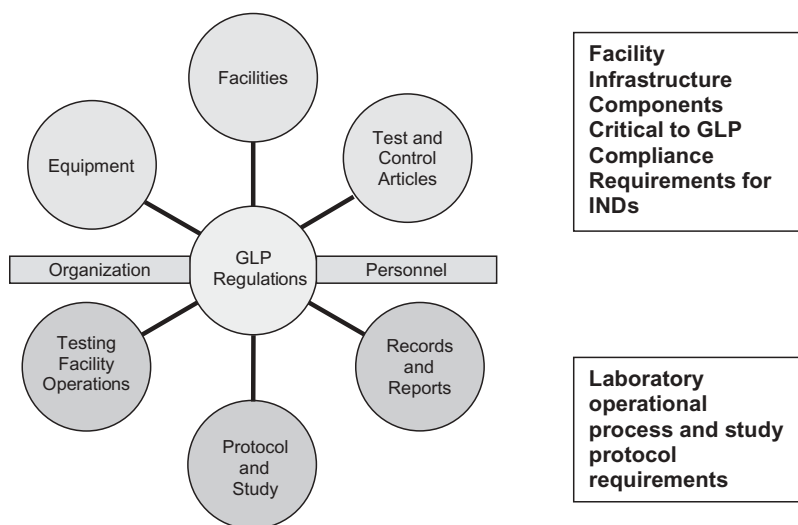


**EXHIBIT 6** Key Good Laboratory Practices requirements during the IND product development phase.

issues such as cleaning and maintenance, equipment calibration and certifications, and routine inspections.

The GLP guidelines require that experimental product-related nonclinical studies and clinical investigations on the IND should have approved written protocols. These protocols should clearly outline the objectives of the study, proposed study methodology and details related to the test materials, procedures, analytical plans, and maintain supportive documentation.

The organization and personnel aspects relating to GLP cover strict reporting requirements for laboratory personnel and testing facility management in nonclinical laboratory studies. For example, individuals found at any time to have an illness that might adversely affect the quality of work and pose risk of infection to others in the facility must be reported immediately. The quality assurance unit within the facility has the responsibility to monitor studies and assure management that the facilities, equipment, personnel, methods, practice, records, and controls are in compliance with the GLP regulations. During an FDA inspection, the written procedures and internal quality assurance (QA) logs and records must be available for inspection.

A key international initiative aimed at instituting and harmonizing GLP is the Organization for Economic Cooperation and Development (OECD) guidelines, a multilateral organization with 30 member states and 70 other participating nations. The goal of OECD GLP guidelines is to ensure “data generated in the testing of chemicals in an OECD member country in accordance with OECD Test Guidelines and OECD principles of GLP are accepted in other member countries for purposes of assessment and other uses relating to the protection of man and the environment” [4].

The OECD GLP guidelines provide a framework for laboratories to plan, perform, monitor, record, report, and archive experimental laboratory studies. These guidelines would assist regulatory agencies within the member state, where the GLP-certified laboratory facility exists, inspect, and ensure compliance with the national GLP guidelines set in compliance with OECD. At the same time, data generated from a GLP-certified facility assures other international regulatory agencies within the OECD that the study results on pharmaceutical compounds could be relied upon as to the hazard and potential risks to users, consumers, and the general environment.

## **2.4 INVESTIGATIONAL NEW DRUG cGMP COMPLIANCE REQUIREMENTS**

Although the U.S. federal government efforts to mandate the safety and purity of drugs goes as far back as 1902, when Congress decided to have biological products manufacturing facility licensed individually to protect the public from dangerously contaminated sera and vaccines, it was only in 1962 that the concept of “manufacturing controls” was introduced in the legislative statute, which was promulgated as “current good manufacturing practices” or simply cGMP.

cGMPs are essentially a family of systems consisting of policy procedures and written analytical documentation to guide a facility at the process levels on medical product manufacturing related activities. The goal of the cGMP is to ensure reli-

ability of a product manufactured at the facility through an established set of standards and processes for quality, purity, potency, composition, and identity claimed by the product sponsor. As a result, cGMP covers the entire gamut of the production systems, which includes plant and grounds, equipment and utensils, sanitation of building and facilities, quality assurance and quality control, production and process controls, warehousing, distribution and postdistribution process, and records access and archival system.

Compliance with cGMP is a fundamental requirement for a medical product development company whether involved in IND or routine manufacturing of licensed products. During the early years after cGMP promulgation, pharmaceutical industries experienced problems relating to potency, cross-contamination, sterility, and labeling related issues. As a result, the FDA initiated the Intensified Drug Inspection Program (IDIP) as an inspection mechanism to regulate the industry. If violations are found during an inspection, product(s) in the entire production line cannot be distributed until the industry demonstrates full compliance. For example, if during a routine GMP audit, some unknown particles were discovered in a production process, the plant will be ordered to temporarily shut down until the contaminants can be identified and removed from the production system. The facility management will perform a full-blown analysis involving sampling from the entire production lines, complete a battery of tests to identify the contaminants, and take measures to eliminate the problem. Only after a clear demonstration of these efforts would a resumption of normal production activity be allowed at the facility.

#### **2.4.1 cGMP for IND During Phase I Clinical Trials**

The 2006 FDA Guideline on the Preparation of IND Products (for human and animal uses) primarily addresses the regulatory compliance to cGMP regulations required under the Federal Food, Drug and Cosmetic Act (FD&C Act). These guidelines have no legally enforceable authority but are viewed as recommendations to address cGMP requirements in the production of INDs for phase I studies.

The earlier 1991 guideline addressed primarily the large-scale industrial manufacturing environment and not others such as small- or laboratory-level production of investigational new drugs. Also, the 1991 guideline did not clarify fully the FDA's programmatic expectation to adopt an incremental approach to institute manufacturing controls for the INDs. While addressing these issues, the 2006 guidelines represent the FDA's efforts to formally establish an approach guiding implementation of manufacturing controls in relation to IND products for phase I clinical trials. IND production settings covered in these guidelines include small-scale manufacturing at laboratory, batch productions for exploratory studies, and multiproduct and multibatch testing of IND products manufactured for phase I clinical investigations.

The 2006 FDA guidelines apply to all IND drug and biological products (including finished dosage forms used as placebos) for phase I clinical studies, which includes investigational recombinant and nonrecombinant therapeutic products, vaccines, allergenic products, *in vivo* diagnostics, plasma derivatives, blood and blood components, gene therapy products, and somatic cellular therapy products that are subject to cGMP requirements. However, these FDA guidelines do not apply

to (a) human cell or tissue products, (b) clinical trial of products subject to the device approval or clearance provisions within the existing regulations, (c) INDs manufactured for phase II/III clinical studies, and (d) approved products that are being used during phase I studies for other study endpoints such as route of exposure and new indications not covered in the approved label.

Exhibit 7 illustrates the key elements of a cGMP program for a facility involved in the manufacture of IND for clinical studies. The system components to establish

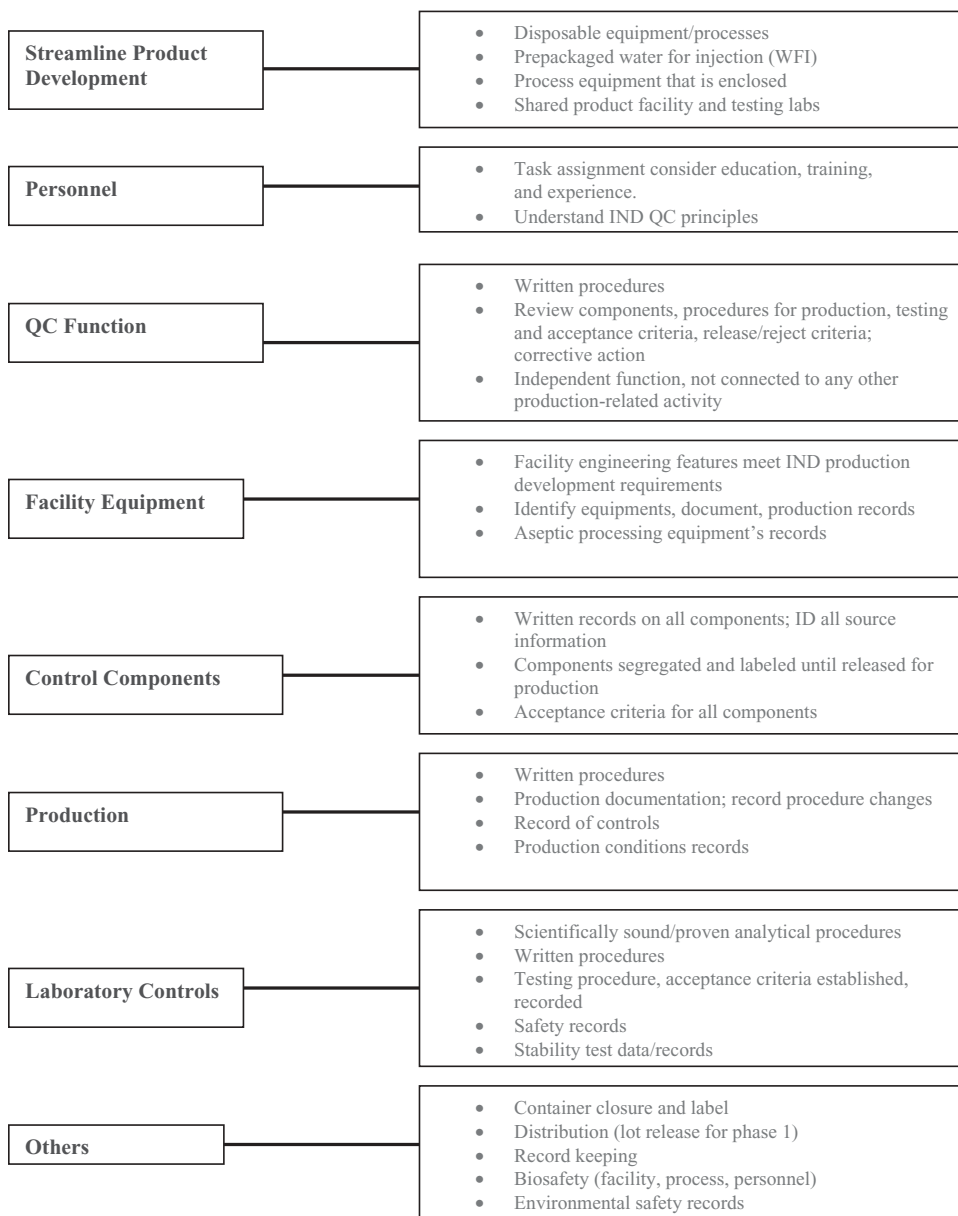


**EXHIBIT 7** FDA recommended approaches to complying with current good manufacturing practices during IND phase I investigations. (Based on FDA [5].)

manufacturing controls for an IND are similar to routine cGMP programs at pharmaceutical facilities covering process elements, facility and personnel, production and laboratory controls, quality assurance, and control. In the case of investigational biological products requiring special precautions, biosafety and biosecurity-related issues are part of the overall facility and process-level compliance requirements.

With safety and quality of product as the primary focus, the guidelines for the production of IND for use in phase I clinical studies are centered on the establishment of cGMP-driven quality control process. The nature and extent of manufacturing controls needed to achieve the desired quality criteria differ not only between the IND products and commercial manufacture but also among the IND products manufactured for various phases of clinical studies. However, regulatory guidelines are yet to delineate cGMP requirements to these product development phases.

### 2.4.2 cGMP for IND

In its efforts intended to streamline the IND process, while at the same time ensure the safety and quality of drugs at the earliest stages in the development pipeline, the FDA excluded most of the phase I INDs from the cGMP regulations for human drugs, including biological products [5]. The FDA maintains regulatory oversight on the production of INDs under the general statutory cGMP authority and the requirements set forth under the IND application authority. The amendment of the cGMP regulation was considered as part of the FDA's overall efforts to guide the industry with a consistent framework to manage and establish controls on the manufacturing at the early product development stages. However, the FDA withdrew the final rule published in the 2006 *Federal Register* to evaluate comments received from the industry and published a modified rule in 2008. The final rule specifies that 21 CFR Part 211 no longer applies for IND, including those under the exploratory products category manufactured for use in phase I clinical trials.

The regulatory implications invoking the general statutory cGMP authority for IND means that overarching goals to ensure quality, purity, potency, and composition of the investigational product under clinical trial meet the general standards set forth under the cGMP. Hence, facilities manufacturing the IND for clinical studies must comply with all requirements that include the facility, equipment and utensils, sanitation of building and facilities, quality assurance and quality control, production and process controls, warehousing, distribution and postdistribution process, and records access and archival system.

### 2.4.3 From cGMP to Quality Systems

The concept of quality systems and its relevance to cGMP requirements were first identified under the FDA guidelines for finished medical devices intended for human use (21 CFR 820).

Exhibit 8 is a comparative summary of the quality systems applicable to cGMP requirements under the FDA guidelines and the International Organization for Standardization (ISO) 9001 requirements for medical devices intended for human use. The quality systems components under these guidelines cover organizational structure and control, management systems, quality audits, and personnel and training. The requirements for subcontractors involved in the product manufacturing



| cGMP Quality Systems Requirements <sup>a</sup>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | ISO 9001 Requirements <sup>b</sup>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Design plans to be reviewed, updated, and approved as the project evolves</li> <li>• Design transfer control</li> <li>• The effective dates of documents must be identified</li> <li>• Label controls must be implemented</li> <li>• Limitations or allowable tolerances must be available to personnel performing equipment adjustments</li> <li>• Manufacturer maintenance of specific distribution records</li> <li>• The manufacturer must follow specific evaluation, reporting, and recall procedures for defective products</li> <li>• The manufacturer must mark confidential quality records</li> <li>• Manufacturer maintenance of quality records for no less than 2 years from the date of release of the product for commercial distribution</li> <li>• Manufacturer recording of specific data on records supporting the manufacture of product</li> <li>• The development of a quality system record</li> <li>• Manufacturer recording of specific data on servicing records</li> </ul> | <ul style="list-style-type: none"> <li>• Activities for consideration during quality planning</li> <li>• Provision for contract review activities</li> <li>• Design projects to be assigned to qualified personnel equipped with adequate resources</li> <li>• Organizational and technical interfaces to be defined, documented, transmitted, and reviewed during the design process</li> <li>• The verification of subcontracted products</li> <li>• The control of customer-supplied products</li> <li>• Control process for inspection, measuring, and test equipment</li> </ul> |

<sup>a</sup>Based on 21CFR, Part 820; <sup>b</sup>ISO 9001 (2000).

**EXHIBIT 8** Comparative summary of the quality systems applicable to current good manufacturing practices within the FDA and ISO 9001 guidelines.

activities are required to adhere to the quality assurance and quality control standards established by the sponsoring entity.

In 2004, the FDA introduced a transformational paradigm to cGMP through a quality systems-based approach for drugs and biologics products. The quality systems approach is based on the policy to encourage the pharmaceutical and biotechnology industries to adopt risk management principles. Regulatory shift to quality-systems-driven cGMP is in response to consistent problems with quality assurance and quality assurance operations in the medical products manufacturing sector. The FDA has declared quality systems approach as the mechanism to improve the predictability, consistency, integration, and overall effectiveness of its regulatory operation [3].

A risk-based approach to management controls, inspection, and oversight is one of the key pillars of the quality systems. Need for a risk-based regulatory approach was considered critical in the context of rapid advances in the process development and complex manufacturing processes for biotechnology-derived products. Process elements in these complex operations are at best ill defined, with limited manage-



ment controls and potentials for failures to meet quality assurance/quality control (QA/QC) targets. Other product-related quality systems considerations relate to safety issues such as contamination in the production systems, potentials for adverse human health, and environmental impact due to the biological agents and systems involved in the manufacture of specialized medical countermeasures such as vaccines against bioterrorism.

The risk-based approach essentially involves a scheme to prioritize the pharmaceutical facility inspection decision-making process. Through this prioritization scheme, the FDA will determine inspection of those facilities posing the greatest public health impact. The nature and extent of inspections at these facilities remains flexible and change with risk reduction strategies implemented at these facilities. A similar risk-based approach is used to assess product safety review as well. This includes product quality of INDs, preapproval chemistry, manufacturing controls, and postapproval supplement processes. The performance matrix to measure effectiveness of risk-based quality system implementation are (a) continuous improvement in product manufacturing, (b) increased product quality and process efficiency, and (c) availability of new medical products.

As part of the risk-based management initiative, the FDA established the Office of New Drug Chemistry (ONDC) within CDER, to establish a risk-based pharmaceutical quality assessment system with a focus on critical product quality attributes as it relates to safety and efficacy. Quality attributes such as product chemistry, formulation, manufacturing processes, and performance are targeted for process optimization and continued improvements.

Management controls are critical to addressing GMP issues such as the establishment of institutional-level policy and governance structures that communicate management intentions and priorities. Such a framework includes establishment of a systematic review of quality data trends on a regularly scheduled basis, resourcing plans and prioritization, and setting performance matrix and incentive systems. A risk-based decision framework is ideally suited to develop up front a decision scheme to prioritize these activities, set an internal monitoring to track status, and adjust the oversight requirement based on quality-systems-based performance matrix and internal audits. In particular, internal audits are a proactive mechanism to discover deviations from quality systems and correct deficiencies before a minor error escalates to a crisis level. Thus, management systems consistent with the risk-based framework for quality systems offers a win-win solution in terms of protecting the core business interests of the industry and the regulatory goals for cGMP compliance.

## **2.5 ROLE OF ORPHAN DRUG ACT IN INVESTIGATIONAL NEW DRUG**

Orphan drugs belong to a FDA category designation for medical countermeasures intended for use in a rare disease or condition defined under Section 526 of the Food and Drug Act. Designation of an IND under the exclusive approval provisions of the orphan drug provides the manufacturer with treatment use of investigational orphan drugs. Orphan drug status also guarantees a 7-year period of exclusive marketing of the licensed drug. As of December 2008, the FDA has listed a total of 1951 pharmaceutical products under the orphan drug designation [6]. Orphan

drug products are expected to be used in treatment of over 6 million people annually in the United States.

In order for an IND product sponsor to avail the orphan drug status for the candidate product under development, sufficient documentary evidence should be prepared for submission demonstrating that:

1. The IND candidate is developed to treat a disease or condition for which the drug is intended to affect fewer than 200,000 people in the United States or, if the drug is a vaccine, diagnostic drug, or preventive drug, fewer than 200,000 dose-units are administered annually.
2. There is no reasonable expectation that the costs in research and development for the IND candidate will be recovered from sales of the IND drug even to more than 200,000 people, or dose-units, within the United States.

IND applicants seeking orphan drug provisions at the early investigational phase of the product development must write and obtain written recommendations from the FDA clarifying the nonclinical and clinical investigational data requirements to satisfy the regulatory provisions. In particular, the IND application must provide considerable details with authoritative references to description of the disease or condition for which the drug is proposed to be investigated and the proposed indication or indications for use for such disease conditions, and the basis for conclusion that the drug is for the disease or condition that is rare in the United States (21CFR 316.10).

As part of this application requirements, IND applicants must provide sufficient data backed up with analysis of available nonclinical and clinical data pertinent to the drug and the disease to be studied with supportive publications. This requirement is considered particularly advantageous for an IND product intended to treat a life-threatening or severely debilitating illness, especially when there are no other satisfactory alternative therapy available. IND orphan drug applicants meeting these requirements are expected to go through an expedited FDA review.

On the basis of background information submitted by the IND applicant seeking orphan drug status as part of the preclinical and clinical investigations, the FDA will determine (a) whether the disease or condition for which the drug is intended is rare, or not so, in the United States, and (b) whether there exists sufficient evidence and supportive rationale for permitting investigational use of the drug for the rare disease condition.

A product manufacturer could avail the orphan drug designation for a previously unapproved drug, as long as the supportive documents specify a rare disease or condition. Alternatively, a new orphan drug designation could be requested for a drug already in the market, as long as the new indications based on new research and development suggests its use in the treatment of a rare disease or condition. The current regulation allows even an approved drug, which does not have an orphan drug designation, to apply and receive such a status as long as the drug has indicated use like the orphan drug for the same rare disease or condition, but is shown to have superior clinical response or demonstrated safety by way of adverse drug reactions. More than one sponsor may receive orphan drug designation of the same drug for the same disease or condition, as long as each applicant files an independent request with the FDA seeking such a designation.

| IND Orphan “Same Drug” Category                                                                                                                                                                                                                                                                                                                                                                                                                | “Same Drug” Regulatory Definition                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Small Drug Molecules</b>                                                                                                                                                                                                                                                                                                                                                                                                                    | <p>An IND candidate that contains the same active moiety (small drug molecule) as a previously approved drug and is intended for the same use as the previously approved drug, even if the particular ester or salt (including a salt with hydrogen or coordination bonds) or other noncovalent derivative such as a complex, chelate, or clathrate has not been previously approved, except that if the subsequent drug can be shown to be clinically superior to the first drug, it will not be considered to be the same drug.</p> |
| <b>Large Drug Molecule (Macromolecule)</b><br>An IND candidate that contains the same principal molecular structural features (but not necessarily all of the same structural features) and is intended for the same use as a previously approved drug, except that, if the subsequent drug can be shown to be clinically superior, it will not be considered to be the same drug. Criterion applied for various categories of macromolecules. | <p>Two protein drugs would be considered the same if the only differences in structure between them were due to posttranslational events or infidelity of translation or transcription or were minor differences in amino acid sequence; other potentially important differences, such as different glycosylation patterns or different tertiary structures, would not cause the drugs to be considered different unless the differences were shown to be clinically superior.</p>                                                    |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                | <p>Two polysaccharide drugs would be considered the same if they had identical saccharide repeating units, even if the number of units were to vary and even if there were postpolymerization modifications, unless the subsequent drug could be shown to be clinically superior.</p>                                                                                                                                                                                                                                                 |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                | <p>Two polynucleotide drugs consisting of two or more distinct nucleotides would be considered the same if they had an identical sequence of purine and pyrimidine bases (or their derivatives) bound to an identical sugar backbone (ribose, deoxyribose, or modifications of these sugars), unless the subsequent drug were shown to be clinically superior.</p>                                                                                                                                                                    |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                | <p>Closely related, complex partly definable drugs with similar therapeutic intent, such as two live viral vaccines for the same indication, would be considered the same unless the subsequent drug was shown to be clinically superior.</p>                                                                                                                                                                                                                                                                                         |

**EXHIBIT 9** Summary of the definitions for IND orphan “same drug” categories for small and large drug molecules.

Exhibit 9 summarizes the definitions for IND orphan “same drug” categories for small and large drug molecules. The decision criteria as to when an IND can be considered an orphan drug are driven by the inherent differences in the chemical composition of the moiety. For example, covalent or noncovalently modified derivatives of previously approved small drug molecules, which are mostly synthetic organic compounds, could be classified as an orphan drug as long as it is indicated for a rare disease or condition, or is shown to have a superior clinical response compared to the original compound.

The definition for what constitutes “same drug” is more complex for biologics, which are macromolecules such as proteins, complex polysaccharide drugs, and polynucleotide drugs. Here the differences are based not on the primary polymer sequence but on differences in the higher structure due to posttranslational modifications as in the case of proteins, or postpolymerization modification as in the case of polysaccharide- or polynucleotide-based drugs.

The provisions within the orphan drug approval process provide a market protection of 7 years during which time the FDA will not approve another sponsor’s marketing application for the same drug. Once approved, the updated list of products with orphan drug status is published periodically in a report titled “Approved Drug Products with Therapeutic Equivalence Evaluations.” The FDA also publishes a cumulative list of pharmaceutical products that have received orphan drug designation and a separate list of those that have received marketing approval. As of December 2008, there are a total of 1951 pharmaceutical products with orphan drug designation and 325 products with marketing approval.

## 2.6 REGULATORY REQUIREMENTS TO PROTECT HUMAN SUBJECTS

Protection of human subjects in a clinical trial is a critical requirement under existing regulations with the aim that participants in an investigational research are familiar with the study and are provided with a minimum amount of information to provide informed consent. Therefore, the scope of regulations covering protection of human subjects (21 CFR, Part 50) pertains mostly to compliance with the overarching regulatory goals protecting the rights and safety of subjects involved in clinical research investigations. The roles and responsibilities of the IRB (discussed in the following sections) covers additional specific obligations and commitments at the institutional levels to the standards of conduct by the investigators, sponsoring agencies, and the institutional authority to the overall safety and protection of clinical research subjects.

Exhibit 10 is a flowchart illustrating a general guideline for a clinical investigator to determine if a waiver or alteration of an informed consent requirement is practical or feasible to their proposed IND clinical investigations. The principal investigator leading the clinical study must first establish whether the proposed clinical study poses greater than minimal risk to the human subjects. If it is determined that the study poses more than minimal risk, then a waiver/alteration of the informed consent requirement is not allowed. However, if it is determined that the proposed study poses only a minimal risk, the principal investigator must establish the rationale that (a) the proposed waiver will not affect the rights and welfare of the human subjects, and (b) it is appropriate to provide patient information to the study subject later. The IRB should review these request waivers as displayed in the flowchart and communicate to the principal investigator the decision on waiver request.

Hence, informed consent constitutes a fundamental requirement in clinical research planning and management. Under the prevailing regulations, no clinical investigation involving human subjects could proceed without the investigator having first obtained the legally effective informed consent of the subject or the subject’s legally authorized representative. In order for the human subject or his or

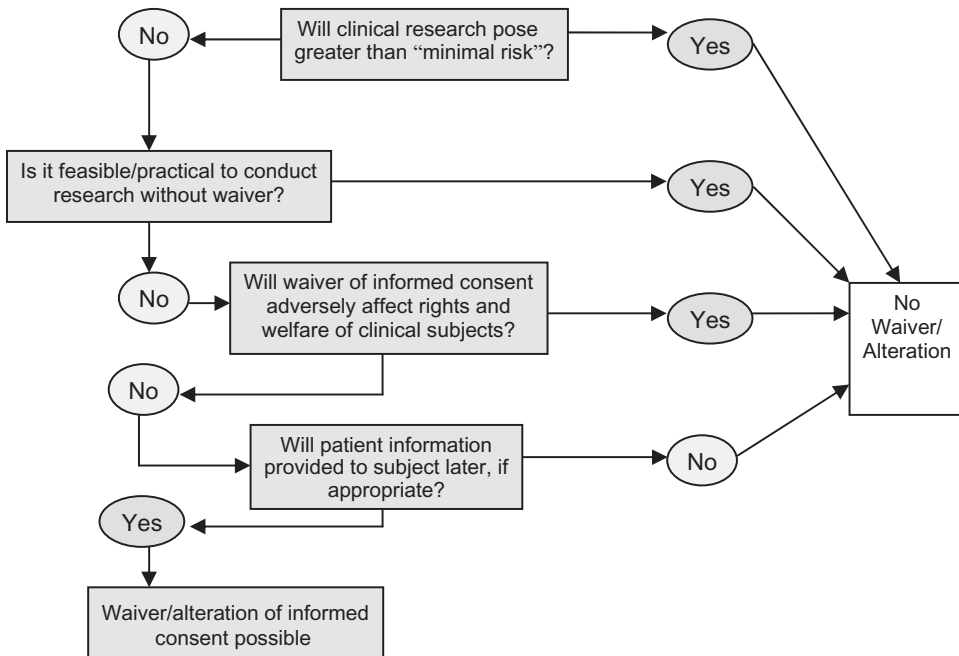


**EXHIBIT 10** Flowchart illustrating a general guideline for clinical investigator to determine if a waiver or alteration of an informed consent requirement is practical or feasible to their proposed IND clinical investigations.

her authorized representative to make an informed decision, it is contingent upon the clinical investigating team to provide the information in a format easy to understand and disclose all relevant scientific, technical, and legally binding issues pertaining to the proposed study. The informed consent paperwork should not have any legal language that might waive any of the subjects' legal rights and release the investigator, sponsoring agency, and the performing institution from liability due to negligence.

Although informed consent is a fundamental prerequisite, there may be instances when a investigator working on an IND may not be able to obtain consent, such as (a) when the human subject is in a life-threatening situation and requires immediate administration of the investigational product, or (b) when the human subject is unable to communicate with, or could not obtain a legally effective consent form, and insufficient time to obtain consent from subject's legal representative, or (c) when no alternative methods or approved, or generally recognized therapy is available that has an equal or greater likelihood of saving the subject's life. Under a combination of circumstances listed here, the investigator and the physician conducting the clinical trial must certify in writing citing the reason(s) and substantiate with additional records as needed and submit to the IRB within 5 working days after the use of the IND test article.

The president of the United States may waive the informed consent requirement for the administration of an IND (including an antibiotic or a biological product) to the members of the U.S. armed forces in connection with a particular military

operation, in support of a specific protocol under an IND application by a Department of Defense sponsored clinical investigation (10 U.S.C. 1107[f]). Under this statute the president must first establish, at the request of the secretary of defense, that obtaining consent is not feasible, despite the best interests of the military member, and apply the standards and criteria set forth under the FDA regulations. A key consideration in the presidential request for waiver is the determination that the military operation presents substantial risk of exposure to chemical, biological, nuclear, or other form of exposure from the deployment that are likely to produce serious life-threatening injury or illness, or even death, and that there are no known or acceptable biodefense-related medical countermeasures currently available that could meet the potential health threat posed by the military operation.

As required under this statute, the Department of Defense will constitute an IRB with at least three members not affiliated to the Defense Department or the federal government to review and approve the proposed IND protocol without informed consent.

The flowchart in Exhibit 10 illustrates a general guideline for a clinical investigator to determine if a waiver or alteration of an informed consent requirement is practical or feasible to their proposed IND research. It is incumbent upon the investigator and the study sponsor to determine response to each of these questions in the decision tree and provide supportive documentation to the IRB as required under the regulations.

Clinical investigators must provide a clear statement of objective and explain the purpose of the research, expected duration of subject participation, description of the study protocol and actual procedures, and clearly identify procedures that are experimental in nature. The informed consent disclosures must provide sufficient information on the potential risks to health and well-being of the subject, while describing the anticipated benefits from participation in the study. The subjects must be aware of their rights and confidentiality of health information such as disease condition, medical treatments, and so forth.

The patient information and consent forms (PICF) must be approved by the IRB prior to use. A fully executed informed consent should be dated and signed by the subject or the subjects legally authorized representative at the time of the consent. All signed consent forms should be retained by the clinical investigator throughout the study period and a copy of the consent form provided to the subject.

A recent study by Beardsley et al. [7] investigated the patient knowledge and satisfaction regarding the informed consent process for cancer clinical trials and found that the lengths of PICFs, submitted to the IRB over the past few years, have increased with time.

Exhibit 11 illustrates the number of pages in the PICFs of 102 patients participating in 27 therapeutics clinical trials across 4 hospitals over the past 6 years on approved cancer clinical trials. A notable observation from this study was that although the number of pages in the PICF has grown dramatically from a mean of 7 pages (range 3–9) in 2000 compared to 11 pages (range 7–21) in 2005, important information for the patient was missing in several cases and that the patient understanding was inversely proportional to the page count of the PICFs. Evidently, the number of pages in the PICF does not correspond to the effectiveness in communicating the complex details of a clinical trial protocol.

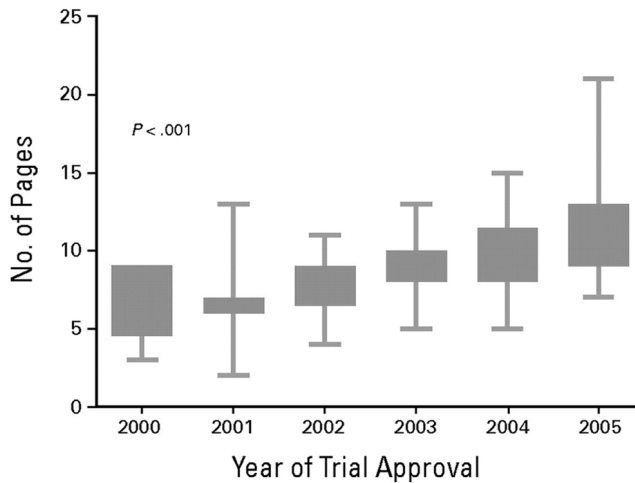


**EXHIBIT 11** Number of pages in the participant information and consent form in a sample of clinical studies performed during the past 5 years. (Source: Beardsley et al. [7].)

## 2.7 REQUIREMENTS FOR OVERSIGHT: IRB

At the institutional level clinical trials require a thoroughly objective and unbiased oversight and review of proposed investigations by a duly appointed IRB. It is the responsibility of the IRB to conduct a technical review upfront of a proposed clinical investigation that supports applications for research involving INDs or marketing permits for products regulated by the FDA. The goal of an IRB review process is to ensure protection of the rights and welfare of human subjects proposed under a clinical investigation (21 CFR, Part 56).

No clinical investigation could begin without proper review and approval of an IRB, unless the FDA provides a formal waiver of any of the IRB requirements, including the requirement for review for specific research activities otherwise covered under these regulations. All clinical investigations involving INDs must meet the IRB requirements and obtain proper approvals prior to submission to the FDA for a formal review of the application. IND product applications with data generated from a clinical research activity conducted without proper review and documentation of an initial and continuing IRB review process may be rejected by the FDA from further consideration.

Under the existing regulations, some categories of clinical investigations may be exempt from a formal IRB review process. A couple of these exemptions apply to IND candidates; for example, investigational new products of test articles for emergency use, provided that such emergency use is reported to the IRB within 5 working days. However, subsequent use of the test article must proceed only after completion of a proper IRB review. There are other likely scenarios under which an investigational new product may be used during an emergency such as biological countermeasures products such as vaccines and therapeutics against bioterrorism-related biological agents. With much of the product development activity in the preclinical experimental research for new generation biodefense products, an emergency use of promising candidates may become necessary.



As another example, a waiver to current requirements under the formal IRB review process may be obtained if a clinical investigation for an investigational new drug for research purposes has commenced before July 27, 1981, and was subject to the requirements of IRB review under the then prevailing regulations, and provided that the investigation remained under the review of an IRB, and compliant with requirements in effect before July 27, 1981.

The IRB represents an institutional-level mechanism to approve, monitor, and report on regulatory aspects relating to clinical research set forth under FDA regulations and guidelines. As part of this effort IRB has:

1. Authority to review and approve or reject all clinical research activities.
2. Authority to seek additional information given to subject as part of informed consent to ensure that sufficient information is provided to protect the rights and welfare of subjects in the clinical research.
3. Verify all documentation relating to informed consent as required under the FDA regulations. Furthermore, IRB will determine whether a waiver is required to allow a legally authorized representative to sign a written consent form under certain conditions as long as it does not present more than minimal risk of harm to subjects.
4. Authority to ask and obtain from the principal investigator of an IND clinical research relevant information to clarify, supplement, or modify the research plan to meet the institutional and regulatory requirements.
5. Responsibility to notify clinical investigator and the institution in writing on the decision to approve or reject a proposed research activity or modification required to obtain the IRB approval.
6. Responsibility to perform reviews of research at appropriate intervals to ensure the clinical investigations are in accordance with the IRB approved protocols and plans.
7. Responsibility to inform study sponsor any exceptions to informed consent and duly document that the disclosure has occurred.
8. Responsibility to ensure that if children are some or all the subjects for a clinical study, the study plans are in compliance with appropriate regulations (under 21 CFR, Part 50).

Exhibit 12 illustrates key criteria for approval of a clinical research by the IRB. As the core of the approval process, IRB must establish that the clinical research proposal poses potentially minimal risk to human subjects, and the anticipated benefits associated outweigh potential risks associated with the clinical study.

### **2.7.1 Composition of IRB**

The IRB consists of at least five members drawn from a multidisciplinary background in order to facilitate a proper review of the clinical research activities undertaken by the institution. From the standpoint of a review of a proposed clinical research activity involving an IND product, it is crucial that the IRB members be thoroughly familiar with the technical requirements, potential vulnerability of the



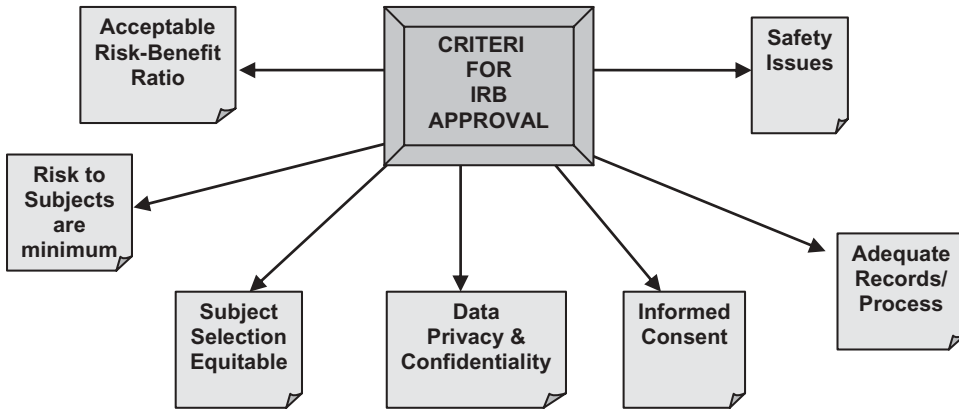


**EXHIBIT 12** Key criteria considered by the institutional review board as part of review and approval of a clinical research proposal.

clinical subjects under the proposed investigation, the nature and extent of institutional commitments, other considerations including cultural background, sensitivity to such issues as community attitudes toward race and gender issues, and the safeguard of rights and welfare of human subjects. Vulnerability of clinical subjects to a clinical trial involving an IND would have to consider cases when children, pregnant women, prisoners, elderly subjects, or mentally and/or physically disabled persons are part of a clinical investigation.

Under existing regulations an IRB should be comprised of (a) qualified men and women drawn not entirely from one profession to minimize bias; (b) members representing both scientific and nonscientific disciplines relevant to bringing a balanced consideration of the overall institutional review process; (c) members from external community drawn either from professional societies and community organizations not affiliated with the institution; and (d) members without a conflict of interest, either as participants in the proposed study or as study sponsors.

The IRB is required to follow a systematic review process, follow written procedures, and document the review process (a) prior to the commencement of a clinical research activity, (b) continuing review of the research and reporting its findings and actions, and (c) document IRB review findings, approvals, and reporting to regulatory agencies as required.

Upon submission of all relevant documents of a proposed clinical study, an IRB is established to approve and oversee on a continuing basis to determine the nature, frequency, and extent of reviews required and the need for verifications from sources other than investigators that no material changes have occurred since previous IRB review. The IRB will establish written procedures for reporting changes in research protocols and ensure that changes in approved protocols were reviewed and approved prior to actual changes made to the study. A formal approval by the IRB requires that the review include at least one IRB member whose primary area of focus is in the nonscientific area, and that a majority on the IRB committee approves the proposed clinical research.

It is the responsibility of the IRB to provide formal reports to the appropriate institutional officials and the FDA of any untoward findings in the course of ongoing clinical investigations such as unanticipated problems involving risks to human

subjects, breaches in compliance to approved study protocols, noncompliance with regulations or requirements, or determinations of the IRB. If a study is suspended or terminated for any of these reasons, the decision must be communicated by the IRB immediately to the appropriate institutional authorities and the FDA.

## **2.8 REQUIREMENTS OF FINANCIAL DISCLOSURE**

As part of the overall review of all clinical studies submitted in marketing applications, existing regulations require the study outcomes not tainted by potential sources of bias in clinical studies due to financial interests of clinical investigators. This review must ensure disclosure, if any, of the potential financial interest for clinical investigators participating in the clinical studies in the form of payment arrangements such as royalties, or proprietary interest in the product such as a patent, or equity interest with the study sponsor submitting the new product application.

The FDA is required to review financial disclosure information for all new human drugs and biological products and marketing applications and reclassification petitions for medical devices. IND applications may have certain exemptions to these requirements, depending on the nature of the clinical investigation, study design, and the type of data collected during these investigations.

This requirement applies to all sponsors submitting marketing applications to the FDA for approval of a drug, device, or biological product. It is the responsibility of the sponsor to ensure all financial disclosures in the form of disclosure statements and certifications are submitted for each clinical investigator involved in the clinical studies. Financial disclosure statements are required for even submissions of clinical study results from other investigations directly relevant to the application but obtained from investigators who are not part of the sponsor-funded clinical studies. As required under the law, regulators will use the clinical study technical information submitted by the sponsor together with disclosure statements and information collected during on-site inspections to determine the reliability of data and ensure no biases are introduced in the interpretation of the study results from clinical investigations toward a new product development, including INDs.

### **2.8.1 Covered Clinical Studies**

As part of the financial disclosure requirements, the FDA defines what constitutes covered clinical studies (20 CFR, Part 54.2). A covered clinical study could be the complete study or parts of the study protocol dealing with the efficacy of a drug or device in humans submitted in a marketing application or reclassification petition. The key to this requirement is the determination that the study, either in total or parts, deals with the efficacy of a product and/or a study outcome making a singular but significant demonstration of safety. This would not include the phase I clinical investigation of INDs, which would generally involve dose-thresholds determination, clinical tolerance, pharmacokinetics, and general pharmacological studies except those involved in the determination of product efficacy.

Financial disclosures include clinical investigators directly involved in the study and those under subcontract to the principal clinical investigator working on a large

study at multiple locations. These disclosures also include spouses and children of clinical investigators so as to eliminate potential sources of bias in clinical studies.

## **2.8.2 Certification and Disclosure Requirements**

Regulations require the sponsors of clinical studies to submit to the FDA a list of all clinical investigators who were involved in the clinical studies evaluating the efficacy and safety of the product, including the exemptions afforded to certain INDs. These submissions must provide complete and accurate details of financial emoluments in any form accrued by the investigators either directly or indirectly (through a subcontract) from the study sponsors. Clinical investigators with IND exemptions must provide the sponsors with sufficient accurate information to allow subsequent disclosure or certification.

In order to certify that a clinical investigator has no financial interest, the study sponsor should obtain from each clinical investigator directly or indirectly (through a subcontract) a copy of FDA Form 3454 declaring absence of any financial interests or arrangements, duly signed by the chief financial officer or other responsible corporate official of the sponsor. If the certification covers only parts of the clinical data in the application, those parts should be clearly identified as such and appended with a list of studies covered under the certification. It is the responsibility of the clinical investigator to update relevant changes to the financial relationship with the sponsor or its affiliated contractor during the course of the study.

## **2.8.3 Disclosure Statement Evaluation**

Existing regulations provide a combination of evaluation strategies on the potential impact of any disclosed financial interest on the reliability of the study through (a) financial disclosure information and certifications furnished by the sponsor, (b) information collected from on-site inspection, and (c) effect of study design.

As part of the financial disclosure evaluation, regulations will consider potential for impact on study reliability based on the nature and extent of a disclosed financial interest, extent of downside financial benefits from an approved product, and steps taken by the sponsors to mitigate the potential biases on the study outcome. A key to this evaluation is the assessment of the overall study design, particularly large studies with multiple investigators scattered at different geographic locations and investigating on different study endpoints. Single and double-blind clinical trial designs, quantitatively verifiable endpoints at multiple locations involving different investigators, study design development, and actual study administration and data collection by unconnected investigating team working at different locations are some of the mitigating measures to minimize, or possibly eliminate, potentials for bias, should there be a financial interest among some investigators participating in the clinical study.

A key goal of the regulatory evaluation is to ensure reliability of clinical data submitted as part of a marketing application of human drug, biological product or device, and applicable INDs. If during the evaluation process it was revealed that financial interests on the part of a clinical investigator may have influenced the study outcome, regulators could respond by (a) initiating audits of data derived from the clinical investigation in question, (b) request submission of additional analysis of

data to reassess the interpretation of study results, (c) request the sponsor to conduct additional investigation through an independent third party, and (d) refuse to consider the study results as part of overall consideration for an agency action (21 CFR 54.5).

Study sponsors must retain all relevant financial records of clinical investigators involved in the study. These records must have all the changes in financial status submitted by the clinical investigator to the sponsor during the course of the investigation.

## 2.9 REQUIREMENTS FOR GOOD TISSUE PRACTICE COMPLIANCE

With the advent of recombinant technologies in biological product development, the scope and volume of cell-based new biologics product development have expanded dramatically in the past decade. This group of therapeutic products, generally known as the human cells, tissues, and cellular and tissue-based products (HCT/Ps) has opened a vast range of therapeutic products development using cell-based recombinant technologies. Although the first-generation cell-based therapeutic products were mostly blood and tissue transplantation, the scope of cell-based therapeutics has now expanded into other areas such as tissue repair and regeneration, modification of immune functions, facilitate tissue/organ regeneration, and gene replacement therapies. The potential for adverse effects and health and environmental risks associated with HCT/Ps presents a unique regulatory challenge.

Recognizing the need for an entirely new regulatory framework, the FDA proposed current good tissue practices (cGTP) rule together with the cGMP to prevent introduction, transmission, and spread of communicable diseases and mitigate potential environmental contamination from HCT/Ps infectious agents. This together with the requirements under cGMP for production of safe, pure, and efficacious products addresses both process controls at the manufacturing phases as well as the potential impact of releases into the general environment. As a result, the cGTP guidelines outline the methods used in the manufacture of HCT/Ps as well as recordkeeping and establishment of a quality program (21 CFR, Parts 16, 1270, and 1271).

All HCT/Ps are required to comply with the cGTP guidelines, and those candidate products, considered from initial assessment as having potentials for inducing adverse environmental impact, are required to comply with both the cGTP and cGMP requirements and are required to obtain premarket approval through the IND application process for biologics products.

The cGTP guidelines introduce additional regulatory compliance requirements in the IND clinical research investigations. Both sponsors and clinical study investigators are expected to comply with additional cGTP-driven requirements in terms of facility-level process controls, recordkeeping, and establishment of quality programs. Additional requirements of labeling, reporting, inspection, and enforcement apply under the cGTP to all HCT/Ps IND product development related activities. It is relevant to note that the cGTP requirements for HCT/Ps are regulated under the authority of the Public Health Services Act (PHS Act) and not as drugs, devices, and/or biological products. The regulatory community acknowledges the overall intent of cGTP is to improve protection of public health through good clinical care

and application of sound scientific principles to improve quality and reliability of IND product related data [8].

Facilities working on HCT/Ps are required to investigate any adverse reaction with the potential for communicable disease related to a HCT/P under development and report immediately for broader distribution. Under the cGTP guidelines, adverse reaction are defined as a noxious and unintended response to any HCT/P for which there is a reasonable possibility that the response may have been caused by the product or a causal relationship cannot be ruled out [3]. As set forth in the cGTP guidelines, it is the responsibility of the facility handling the HCT/Ps and sponsors to investigate any noxious or unintended adverse reaction and report. Regulatory agencies receiving these reports from multiple sources will look for a general pattern in the adverse events reporting to determine the nature of the emerging trend and seriousness in terms of public health impact. For example, several hospitals reporting an outbreak of methicillin-resistant *Staphylococcus aureus* (MRSA) following a procedure involving a therapeutic product from a human tissue may trace its origin to a single establishment with contamination with MRSA.

Facilities handling HCT/Ps are required to report to the FDA of any serious adverse reactions involving a communicable disease. The cGTP regulatory guidelines define reporting requirements for adverse reactions involving a communicable disease as those that are (a) fatal, (b) life-threatening, (c) result in permanent impairment of body function or permanent damage to body structure, or (d) necessitate medical or surgical intervention including hospitalization. This report must be filed with the FDA within 15 days of initial receipt of information. According to published records, this will be the first federal requirements for reporting of adverse reactions from transplanted HCT/Ps.

Perhaps, adverse event surveillance linked to facilities and operations involving HCT/P is the most critical element in the cGTP guidelines for it requires a mechanism to detect and identify adverse events by physicians and infection control practitioners in a clinical setting, and a reporting mechanism to risk managers at the facility level and onward to the regulatory agencies. The FDA has established the MedWatch program, the safety information and adverse event reporting program, to encourage and enable voluntary reporting, which is crucial for the overall program success. Although MedWatch handles issues involving commercially available pharmaceutical products, facilities performing HCT/P-related IND clinical investigational activities are required to use the same systems for reporting adverse events.

## 2.10 REQUIREMENTS FOR IND LABELING

The drug label, also known as the package insert, is intended to provide an accurate and concise summary of all relevant information necessary for the user to determine the safety and efficacy for approved indications. Current regulations outline the overall requirements on content and format of labeling for human prescription drugs (21 CFR, Part 201.56). As required under this regulation, the label must provide information that is accurate and should not be written in a promotional tone, and most of all misleading in terms of the effectiveness, approved indications, and contraindications. As far as possible the label claims must be based on data

XXX Research Center  
 YYY Address St, City, State 00000  
 Phone: 000-00-0000

Pt Name (or study ID number):      Date (dispensed)  
 Dr. ZZZ (must be MD)                      Visit # (or way to track):

**Drug name and strength or study acronym (include Manufacturer)**  
 Take as directed. Bring bottle to each clinic visit. Do not discard when empty.

**Caution:** New Drug.

(Limited by Federal Law to Investigational Use.)

**EXHIBIT 13** Sample investigational drug label for exclusive use during the clinical trial.

derived from clinical studies and not based on experimental data from nonclinical studies.

The labeling requirements set forth under the current regulations for all drugs and biological products apply to the IND as well.

Exhibit 13 illustrates a sample investigational drug label for exclusive use during the clinical trials. The label format and content should be in strict compliance with the regulatory requirements for package inserts with the exception that the label for the IND must clearly mention that under federal law the product is meant only for investigational uses and a note of caution that the content is a new drug. The name of the physician on the label must be a participating clinician in the IND clinical trial.

The IND labeling issue is particularly relevant in the “off-label” and investigational use of marketed drugs, biologics, and medical devices. Off-label use of an approved drug refers to a use, by a practicing physician, of a drug that is not included in the approved label for human prescription drugs.

A general topic of discussion in the clinical research and regulatory affairs community is whether an off-label use of an approved drug constitutes experimental investigation that normally requires formal review and approval by an IRB and strictly follows the informed consent process. Likewise, there is some confusion as to whether an off-label use of an approved drug requires formal submission of an IND application to FDA.

The requirements under good medical practice and the references within the existing drug approval process address these issues.

Good medical practices establish the principles and values that guide medical professionals in clinical care and service delivery. Although the good medical practices are essentially addressed to clinical practitioners, its broader goal is to let the public know what they can expect from medical services rendered at a clinical setting. This is particularly relevant in the context of off-label use of an approved drug for an indication not included in the approved label.

If a physician uses a product for an indication not on the approved label based on his or her clinical knowledge and patient history, it is extremely important that the use be based on sound scientific rationale and a thorough understanding of the product indications and potentials for adverse effects and maintain complete records

of all treatment procedures and therapeutic regime for off-label uses. However, under the current regulations there is no requirement of off-label uses for a formal submission of an IND application, investigational device exemption, or an IRB review. Instead, the clinical knowledge and expert judgment of the physician based on a detailed knowledge of the product and the patient clinical condition guide the off-label use of an approved drug or medical devices. Off-label uses within an institutional medical practice may require an IRB review or other existing institutional oversight and governance process.

It is important to note that the off-label use of an approved drug is not the same as the investigational use of an approved pharmaceutical product since investigational use indicates the use of an approved product in a clinical study protocol must comply with the FDA requirement under the IND application procedures (21 CFR 312). This compliance requirement is essential when the objective of the investigation is to develop additional information relating to product safety or efficacy, in which case an IND application and due process is required. However, this requirement is not considered essential if the proposed investigation meets all the following six conditions: (a) Proposed investigation will not lead to a new NDA application in support of a new indication for use or support any significant change in the approved labeling for the pharmaceutical product; (b) proposed investigation is not intended to support a significant change in advertising of the product; (c) proposed investigation does not involve a route of administration, or dosage level, or use in a subject population, or any other factor that could significantly increase the risks associated with the use of the pharmaceutical product; (d) proposed investigation requires a formal IRB review or any other related IND application filing requirements; (e) proposed investigation is an attempt to comply with a requirement concerning promotion and sale of a pharmaceutical product; and (f) proposed investigation does not invoke the exemption from informed consent requirements for emergency research (21 CFR 50.24).

Hence, the off-label use of an approved drug is guided by the good medical practice and is the practitioner's responsibility for prescribing a drug for uses not indicated in the approved label. The off-label uses in a clinical setting should be based on a sound scientific rationale and a thorough understanding of the pharmacological data determining the label content. Although the IND application requirements do not apply under these conditions, legal actions are more important in the off-label use of an approved drug. Physicians are at the risk of facing a malpractice lawsuit for negligent use of any drug whether or not the FDA has approved the use of that drug (both label and off-label). Therefore, labeling does not preclude a physician from applying his or her accumulated clinical knowledge and expert judgment in the determination of off-label uses.

In contrast to off-label use, most investigational use of an approved pharmaceutical product in a clinical study protocol must comply with the existing requirement under the IND application procedures.

## **2.11 MONITORING OF INVESTIGATIONAL NEW DRUG RESEARCH**

From a regulatory compliance standpoint, monitoring IND research revolves around nonclinical and clinical research facilities and data generated from these



investigations. Apart from the core requirements of quality and integrity of investigational data submitted in support of an IND application, monitoring ought to protect the rights and welfare of the human subjects involved in clinical research. Recognizing the need for a comprehensive monitoring program, the FDA has instituted the Bioresearch Monitoring (BIMO) program to fundamentally ensure the quality and integrity of the data submitted to support IND applications, and simultaneously ensure protection of the rights and welfare of human subjects during this process.

The BIMO program covers both on-site inspections and data audits to monitor all aspects of the experimental and clinical research conducted in support of the FDA regulatory approval process. According to published reports, the BIMO program on average conducts over 1000 inspections annually and covering GLP audits of nonclinical testing labs, clinical investigators, sponsors/monitors, and the IRBs [6]. Implemented through four multicenter compliance programs, BIMO covers facilities and clinical research activities at both domestic and international locations.

Large medical institutions, both private clinical research facilities and academic teaching institutions with significant clinical research related activities, have established a detailed institutional-level monitoring of the IND-related clinical research studies, including standard operating procedures and GCP both at the institutional and study levels.

Exhibit 14 illustrates institutional-level GCP compliance monitoring elements during an IND clinical research activity. The goal of institutional-level monitoring is to oversee clinical research activities and ensure that these activities are conducted, recorded, and reported in compliance with the established protocol and standard operating procedures and good clinical procedures.

Monitoring is a continuous activity requiring reporting and regularly scheduled reviews by the IRB, which has the oversight responsibility. Depending on the clinical study design, a medical review committee may be established to specially monitor a clinical protocol. On the contrary, audit is a systematic and independent investigation conducted by an external team of all aspects of a clinical study and its overall compliance with the standard operating procedures, GCP, and applicable regulations.

Most often the goal of monitoring review is to improve performance, establish data integrity, protect human subjects, and establish compliance with internal procedures and regulatory requirements. Exhibit 14 lists typical monitoring elements related to GCP compliance in IND clinical research relating to process improvements, data integrity, protection of human subject, and compliance with GCP and applicable regulations. The responsibility for most of the GCP compliance rests with the principal investigator and the monitoring responsibility with the IRB.

### **2.11.1 Clinical Risk Assessment**

A key element of the institutional-level monitoring of the IND research activity is the establishment of a risk assessment and risk management program. As part of the risk assessment process, it is important to distinguish between hazard and risk. Whereas hazard may be defined as any factor—internal or external to the clinical investigation—that could cause harm, risk is the measure of the probability that



| Compliance Element                          | Description                                                                                                                                              | Lead Responsibility                      | Monitoring Responsibility                |
|---------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|------------------------------------------|
| Regulatory documentation                    | Documentation and resources to track communications with FDA, work flow, other regulatory compliance documentation (EPA, OSHA, biosafety, etc).          | Principal investigator                   | IRB/<br>regulatory compliance department |
| Documentation of roles and responsibilities | Filed by the principal investigator on the roles and responsibility of clinical research team members; educational qualifications; credentials; etc.     | Principal investigator                   | IRB/<br>chief medical officer            |
| Clinical research management system (CRMIS) | CIO establishes a customized CRMIS to track work flow and link database.                                                                                 | CIO/principal investigator               | IRB                                      |
| Patient outreach program                    | Oversee clinical subject recruitment process; resources for outreach and enrollment and retention; cost-effectiveness analysis.                          | Principal investigator                   | IRB/study operations support office      |
| Informed consent                            | Electronic documentation to track all paperwork; quality assurance; educational resources to assist patient education on the study and informed consent. | Principal investigator                   | IRB                                      |
| Inclusion/exclusion criteria documentation  | Case report forms, detailed checklists; supportive documentation; links to database.                                                                     | Principal investigator                   | IRB/medical review committee             |
| Adverse effects reporting system            | Serious adverse event (SAE), unexpected adverse event (UAE) reports in real time; and periodic; links to database.                                       | Principal investigator                   | IRB/medical review committee/<br>FDA     |
| Drug/biologics accountability               | IND/IDE application templates; SOPs; sponsor IND documentation.                                                                                          | Principal investigator/<br>study sponsor | IRB/pharmacy department                  |

**EXHIBIT 14** Institutional level monitoring of IND of good clinical practice compliance during IND clinical investigations.

harm will be caused by the hazard. Therefore, presence of hazard alone does not constitute establishment of risk to a research study. What is needed is a structured risk assessment process that takes into account all technical, operational, and information systems and processes related to an IND clinical investigation and a clearly worked out risk mitigation strategy.

Exhibit 15 is a notional illustration of various elements of a clinical risk assessment and risk mitigation process. It is important to note that the clinical team should identify potential hazards inherent in a trial and their associated risks, potential consequences, and a reasonable approach mitigating the risks.

Hazard identification and assessment should identify and sort hazard by its origin and potential impact such as:

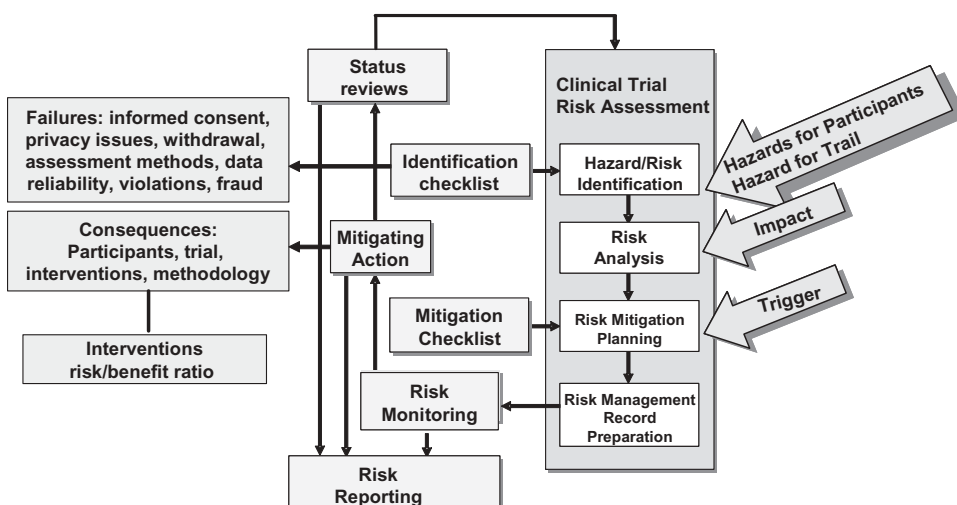


**EXHIBIT 15** Notional illustration of various clinical trial risk assessment and risk mitigation process as part of an institutional monitoring of IND research activities.

1. Hazard for the participants arising due to failure to properly complete the informed consent process or information systems failure to protect the privacy of participants.
2. Hazard for participants due to the nature of intervention mechanism set up as part of the study design such as unexpected and/or expected adverse effects, assessment methods such as biopsy, radiation, and pharmaceutical adjuvants used in the IND formulation.
3. Hazard for the trial such as potentials for an incomplete study due to failures in human subjects recruitment and follow-up, violation of the inclusion/exclusion criteria, reliability of the study results, procedural errors, improper information assurance, quality systems failure, failure in adherence to study protocol, general fraud, and misrepresentation.

Risk management approaches must carefully weigh the consequences of the hazard, either to the study participant or the clinical trial, or both, before developing options for mitigating the risks. For example, establishment of a training program for informed consent process, privacy protection, and information assurance and quality systems would address several participant- and study-specific hazards. Similarly, establishment of systems to monitor and report adverse effects and systems to maintain awareness and ability to report adverse events, systematic review of the study design and clinical trial protocol to assess the statistical power and reliability, and a well-designed monitoring and reporting program to track and report study violations are some of the more commonly employed options to mitigate risks associated with IND clinical investigations.

### 2.11.2 Computerized Systems in Clinical Trials

Clinical research management systems are the indispensable platform to seamlessly integrate clinical research, IND activities and regulatory management. Clinical

research systems employ clinical-rules-based decision systems to help guide clinical practices that are key to clinical trials and establish a collaborative environment for information exchange, storage, retrieval, and analysis. Recognizing the importance of clinical information systems to IND, the BIMO program inspections and audits are centered on the data resident in the clinical research management systems to ensure highest standards of quality, reliability, and conformity with the regulations.

Information gathered during clinical studies should meet the established criteria of quality to remain compliant with requirements for electronic data and a mechanism to audit the system for data attribution, accuracy, and originality. Guidance is available to the industry on how these data quality requirements might be satisfied where computerized clinical research management systems are employed to generate, analyze, modify, archive, and transmit clinical data [2]. This guidance also addresses requirements of the Electronic Records/Electronic Signatures rule (21 CFR Part 11). The guidance may be applicable to source documents created in hardcopy and later entered into a computerized clinical research management system, or directly entered into a computerized system, or automatically captured by a computerized system.

The guidance to industry identifies standard operating procedures relevant to use of computerized systems such as:

1. *Data Entry*: To ensure data attributability through identification of individuals entering the data, passwords protection to limit and track access control, electronic signatures, audit trails, and date and time stamps.
2. *System Features*: Features that will facilitate collection of quality data such as consistent use of terminologies, data tags to facilitate inspection and review, ability to retrieve data, maintain collateral information relevant to data integrity, and system capability to reconstruct a study to backtrack how data were obtained and managed in support of an audit.
3. *Security*: To include physical and logical security. Physical security refers to internal safeguards built into the system and external safeguards to restrict access to authorized users. The system must have a robust feature to prevent unauthorized access. Logical security refers more to maintaining data integrity and to ensure that the information resident in the system is not altered, browsed, or transferred using external software applications.
4. *System Dependability*: To ensure that the system is in conformity with the study sponsors' established requirements for completeness, accuracy, and reliability. System documentation should be readily available for inspection during site visits. Sufficient documentation on software systems validation such as written design specifications, test plan, and test results and demonstration validating the system design specifications.
5. *System Controls*: To include software version controls, contingency plans in the event of a system failure, and a backup and recovery plan to retrieve electronic records.
6. *Training Records*: To include documentation on qualification of individuals managing the database systems and data entry activities, and training records for verification that suitable training was provided to individuals performing these functions.

7. *Records Inspection:* During a facility inspection all records submitted to the Agency may be audited for track changes, no matter how they were created or maintained.
8. *Electronic Signatures:* Electronic signatures are intended to be the legally binding equivalent of conventional handwritten signatures.

### 2.11.3 Quality Assurance

As part of an IND product development activity, it is the responsibility of the sponsor to ensure compliance with all pertinent regulations and quality standards. For example, product development methodologies will be compliant with section 505(b) of the Food, Drug, and Cosmetic Act, 21 U.S.C. § 355(b), methodology used for preparation of the drug substance, and the control testing used to monitor its identity, strength, quality, and purity, as required for IND submission.

Exhibit 16 illustrates the dramatic shift in the regulatory landscape of QA before and now, propelled due to an increasing emphasis on quality systems and the establishment of a robust QA/QC as part of the entire product development life cycle. In the last decade, the regulatory affairs related to QA during IND have shifted to a more activist role with explicit requirements for audits and greater contacts between the industry and the regulators on QA-related matters. The reporting requirements have increased transparency of the IND information collection processes and improved access to QA data.

From the industry standpoint, the IND product development and clinical trials landscape has changed significantly in the past two decades, propelling the need for a reform in regulatory compliance requirements in quality control and quality assurance. More and more clinical investigations now involve extended studies, involving multiple sites and having a large volume of clinical trial subjects at each site. Added

| Regulatory Landscape Before                                                                                                       | Regulatory Landscape Now                                                                                                                                                                                                                                                                      |
|-----------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| No explicit requirement for QA audits                                                                                             | Audit required under the FDA International Conference on Harmonisation (ICH) GCP Consolidated Guidelines. Section 5.19 (Audit).                                                                                                                                                               |
| Mostly left to the sponsors to implement QA as an in-house program and considered a cornerstone of program management success.    | Explicit requirement under the FDA International Conference on Harmonisation (ICH) GCP Consolidated Guidelines. Section 6.11 (Quality Control and Quality Assurance—Documents). Other QA guidelines are: European Union (EU) The Engage Guidelines; World Health Organization GCP Guidelines. |
| No explicit QA related reporting requirements                                                                                     | Greater contact with the industry on QA.                                                                                                                                                                                                                                                      |
| QA was a “black box” as far as FDA was considered and interpreted as a tool for success for sponsor’s product development program | Increased transparency and reporting requirements on misconduct and activities leading to compromise on the safety and security of human subjects.                                                                                                                                            |

**EXHIBIT 16** Comparison of the regulatory affairs role in the quality assurance and quality control during the IND phase of product development.

to this is the new and expanded role now played by clinical research organizations and a vastly expanded pool of clinical investigators now taking part in IND-related clinical studies. Large pharmaceutical companies have reached out to global destinations for IND-related clinical trials bring in countries and medical research institutions that were traditionally not part of the product development pipelines. Access to highly sophisticated clinical research information systems and electronic connectivity have made possible information collection, sharing, and analysis in a widely distributed global environment. Finally, an increasing number of IND clinical studies are now designed that allow participation of sensitive human subjects based on age group and pre-existing clinical conditions. These are compelling enough reasons for the regulators to recognize the need for a compliance framework addressing quality control and assurance as an overall requirement throughout the product development process.

The quality assurance and product technical support departments of the sponsors are responsible for assuring that IND study participants meet the quality objectives and comply with the established guidelines. Participating institutions (hospitals, academic medical institutions, and clinical research organizations) are required to implement and follow written approved procedures to ensure all operations are performed in accordance with quality systems guidelines under the GCP and cGMP regulations and in-house policies.

All personnel supporting or engaged in the IND-related manufacture, testing, or quality assurance are required to comply with the established written approved procedures. In addition, these institutions must have a comprehensive quality program that controls all manufacturing including a complete document control system, documentation review and approval, auditing, and training.

As part of the overall document control process related to QA, study sponsors and participating institutions must maintain an approved document entry and tracking system that includes specifications, standard operating procedures, testing protocols, batch records, and test reports. Study sponsors may establish management oversight of the entire process with real-time oversight of critical production and testing activities, as well as audits.

Exhibit 17 illustrates a typical list of areas covered during QA audits of a biologics facility. The audit area covers floor areas for proposed activities, equipment identified for the projects, documentation, employee training, and facility audit/biosafety inspection records. Facilities used for the IND product development may be inspected to verify the availability of appropriate space, compliance with biosafety requirements, equipment and personnel to operate the following functional areas: (a) fermentation suite, including downstream processing, (b) purification suite separated into upstream and downstream purification rooms, (c) production support areas, (d) process and facility utility area, (e) waste treatment area, (f) warehousing, (g) QA/QC, (h) cell banking, (i) finished and in-process material storage, and (j) offices.

## 2.12 EMERGING BIOSAFETY AND BIOSECURITY REQUIREMENTS

Biotechnology and pharmaceutical companies with an array of biologics product development portfolios and clinical research organizations are beginning to place

| Facilities and Equipment                                                                                                                                                                                                                                                 | Qualification and Calibration                                                                                                                                                                                                                                                                     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>Production and control laboratory equipment</li> <li>Normal maintenance</li> <li>Preventative maintenance</li> <li>Maintenance and servicing of equipment</li> </ul>                                                              | <ul style="list-style-type: none"> <li>Equipment installation and operational qualification</li> <li>Equipment performance validation</li> <li>Cleaning validation</li> <li>Equipment calibration</li> <li>Computer software validation</li> </ul>                                                |
| Sanitization                                                                                                                                                                                                                                                             | Production                                                                                                                                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>Analytical methods validation in support of cleaning</li> <li>Facility cleaning</li> <li>Equipment cleaning</li> </ul>                                                                                                            | <ul style="list-style-type: none"> <li>Raw materials/supplies</li> <li>Sampling</li> <li>Quarantine</li> <li>Bulk manufacture</li> <li>Packing</li> <li>Rejected materials/supplies</li> </ul>                                                                                                    |
| Storage and Distribution                                                                                                                                                                                                                                                 | Process Validation Quality Control                                                                                                                                                                                                                                                                |
| <ul style="list-style-type: none"> <li>Complaints</li> <li>Product recalls</li> <li>Self-inspection</li> </ul>                                                                                                                                                           | <ul style="list-style-type: none"> <li>Microbiological/environmental support</li> <li>Analytical support</li> <li>Raw material support</li> </ul>                                                                                                                                                 |
| Documentation (Technical)                                                                                                                                                                                                                                                | Documentation (Information/Systems)                                                                                                                                                                                                                                                               |
| <ul style="list-style-type: none"> <li>Production records</li> <li>Standard operating procedures</li> <li>Standard manufacturing procedures</li> <li>Protocols</li> <li>Change control</li> <li>Equipment specifications</li> <li>Raw material specifications</li> </ul> | <ul style="list-style-type: none"> <li>Training procedures</li> <li>Computer program specifications</li> <li>Documentation control of process deviations</li> <li>Calibration and test documents</li> <li>Validation documents</li> <li>Purchase orders</li> <li>Authorization to ship</li> </ul> |

**EXHIBIT 17** Typical list of areas of compliance investigated during audits of IND product development programs.

considerable importance on laboratory biosecurity with a focus on improving security at microbiological research facilities, clinical laboratories, and ancillary laboratory services such as biological material storage and distribution facilities. A key element of this growing awareness requires a clear delineation of the concepts of biosafety and biosecurity in IND product development activities in the context of new regulations. Whereas biosafety refers to institutional-level measures to prevent and mitigate the accidental release of biologic agents and toxins, biosecurity refers to instructional measures that guard against the deliberate release of pathogens for malicious purposes (including bioterrorism). Thus far, existing U.S. and international regulations and guidelines have focused on biosafety rather than biosecurity.

In the aftermath of the 9/11 terrorist attacks, followed in the same year by a string of anthrax attacks in the United States, the Congress passed two significant pieces of legislation. First, the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act of

2001 established criminal penalties for possession, shipping, and receiving of certain biological agents, known as select agents (SA) and toxins, if used as a weapon or for any reason not plausibly justified for prophylactic, protective, bona fide research, or other peaceful purposes.

Second, the Public Health Security and Bioterrorism Preparedness and Response Act (PHSBPRA) of 2002 greatly expanded controls over dangerous pathogens and toxins stored, used, and transferred between laboratory and ancillary facilities within the United States. These legislations establish the regulatory premise to introduce biosecurity practices at research laboratories handling dangerous etiologic agents and toxins as part of an overall national security program.

For the most part, biosafety and biosecurity practices are guided by institutional policy and governance set up for this purpose. The environmental health and safety (EH&S) division is most often charged with addressing all biosafety and environmental regulatory affairs as they apply to development, production, and testing of IND products. It is up to the management leadership in an organization to recognize and prioritize the importance of biosafety and biosecurity as fundamental requirements at all phases of product development activities. This requires a careful integration of EH&S-related guidelines and regulations throughout the product life-cycle development. For an effective EH&S program it is important that staff be highly qualified in the areas biosafety, environmental health, and risk assessment and having working knowledge of the regulations covering product development.

Exhibit 17 is a summary of typical areas of compliance investigated during an audit of an IND product development facility. Evidently, the audit areas are cGMP and quality systems related areas but with a focus on the safety and security of the product and processes with a goal of protecting the workers in the occupational setting and the general environment. As part of the safety and security-related audits, the EH&S team may review floor plans as they relate to potential for cross-contamination and release of potentially infectious materials in the working environment. Production process and laboratory equipment, air and water distribution systems, and sanitary systems inspected to ensure built-in engineering protection segregate and contain hazardous chemicals and biological materials from drinking water systems, heating, ventilating, and air-conditioning (HVAC) systems. Warehouse and receiving facility for nonclinical experimental facilities are where animal housing is maintained and bulk raw chemicals are stored. Finally, biocontainment facilities involving biosafety levels 2 or 3 (BSL-2 or BSL-3) are required for handling potentially dangerous etiologic agents and infectious materials for experimental purposes.

As required, the EH&S will perform a biosafety-related risk analysis, known as the maximum credible event (MCE) analysis to determine the potential for accidental release of dangerous biological agents into the occupational environment within the facility and general environment outside, and assess the facility-level crisis and consequence management capability and resources for effective containment of an accidental release during manufacturing, processing, storage, or animal testing operations.

EH&S staff perform preaward audits of subcontractor facilities to ensure compliance with all applicable and current federal, state, and local laws, codes, ordinances, and regulations, as well as all public health services safety and health provisions. These activities would include:



- Review of subcontractor's facility safety plan and complete site visit to verify floor plan details on the ground and review safety records, biosafety oversight committee meeting minutes, and standard operating procedures (SOPs).
- Review the type of hazardous chemicals regulated under the Resource Conservation and Recovery Act (RCRA, 1976) and perform hazard assessment as required.
- Conduct detailed site visits and proposed testing of production areas and associated support areas. All areas are assessed for the appropriateness of the biological safety level employed and compliance with the biosafety standard including general housekeeping, security, and general safety (fire, chemical, radiation, and electrical).
- Review all documentation required by federal, state, and local regulations and related in-house documentation. Documents that are reviewed may include, but not limited to, chemical hygiene plans, safety SOPs, biosafety/safety committee meeting minutes, HAZCOM plan, biological safety plan, training records, engineering control equipment maintenance and certification records, written respirator program, and medical surveillance program.
- Written recommendations are provided as part of the facility biosafety visit and a corrective action plan is requested.
- Monitor facility compliance with biosafety and environmental practices through annual inspections and frequent communications with the facility EH&S staff.

The EH&S staff from the sponsoring organization may be required to provide support and additional oversight over the entire spectrum of facility biosafety, environmental regulatory affairs, workers protection requirements for hazard analysis, and risk assessment for IND product-related development activities.

Requirements for biosafety practices and compliance with environmental and worker protection are unique for each project and could change as the nature of activities change. For example, the volume of a potentially infectious material stored at a facility could determine the biosafety requirement. Thus, a lower biosafety level for a storage operation (with no volume changes) may not apply for a production/testing-related operation where volume changes are likely. Likewise, a higher volume of a hazardous chemical stored at the facility could warrant application of RCRA regulations, whereas a lower volume would be excluded from the regulation.

Both sponsoring organization and participants in the product development activities can make no compromise on biosafety and environmental regulations compliance, designed to be used as screens upfront in facility selection process and monitored by designated staff from both organizations to ensure that the facility remains fully compliant with regulatory requirements throughout the IND activities.

## **2.13 CONCLUSIONS**

The IND is a key phase within the drug development life cycle. The regulatory affairs relating to IND requires a balancing consideration of the inherent benefits



of introducing novel medical products for human use, while at the same time protect clinical trial subjects from potentially harmful test candidates during the preclinical product development phase.

The FDA regulations are the principal drivers of the regulatory affairs related to IND development. Existing FDA regulations relating to IND clinical trials such as cGMP, GLP, and GCP guide IND development for drugs and biologics, and new initiatives such as the GTP attempt to incorporate the expanded scope and volume of recombinant technology-based HCT/Ps and genomic therapies. The FDA has reinvigorated the scope of regulatory affairs related to clinical risk assessment, quality assurance, quality control, and use of computerized systems in clinical trials. Electronic IND submissions have eliminated the need for voluminous hardcopy submission, which require the entire study planning, preparation, execution, and management within the information systems domain.

Biotechnology companies are involved in highly sophisticated biologics products development based on genomics and functional proteomics with unclear regulatory implications, risks to clinical subjects, product liability, and potentials for long-term public health risks. Recognizing the importance of technology and regulatory compliance in the development of novel therapies, the FDA launched the Critical Path Initiatives covering product development strategies and technologies in these IND development areas.

Recently, the FDA initiated the BIMO to proactively ensure the quality and integrity of data submitted in support of IND applications and to protect the rights and welfare of clinical trial subjects. Through a combination of on-site inspections and data audits BIMO attempts to monitor all aspects of the experimental and clinical research conducted in support of the regulatory approval process.

In the aftermath of 9/11, research laboratories are beginning to place considerable importance to improving biosecurity at microbiological research facilities, clinical laboratories, and ancillary laboratory facilities to guard against misuse of pathogenic materials and select agents used in IND research and medical countermeasure products development activities. At the institutional level, integration of a robust EH&S program for biosafety and biosecurity as part of the product development life cycle is an essential regulatory requirement.

## **APPENDIX: APPLICABLE AND RELEVANT REGULATIONS COVERING IND**

### **Key IND-Related Regulations**

|                    |                                                                                      |
|--------------------|--------------------------------------------------------------------------------------|
| 21CFR Part 312     | Investigational New Drug Application                                                 |
| 21CFR Part 312.82  | Early Consultation for Pre-Investigational New Drug Meeting                          |
| 21CFR Part 314     | INDA and NDA Applications for FDA Approvals to Market New Drug                       |
| 21CFR Part 314.42  | Revisions to Agency Requirements on IND Applications                                 |
| 21CFR Part 314.420 | INDA Application, Master File Submission                                             |
| 21CFR Part 316     | Orphan Drugs                                                                         |
| 21CFR Part 316.10  | Content and Format of a Request for Written Recommendation to get Orphan Drug Status |

|                   |                                                                                     |
|-------------------|-------------------------------------------------------------------------------------|
| 21CFR Part 58     | Good Laboratory Practice for Non-clinical Laboratory Studies                        |
| 21CFR Part 58.3   | Physical and Chemical Characteristics of Test Article in GLP guidelines             |
| 21CFR Part 50     | Protection of Human Subjects                                                        |
| 21CFR Part 50.24  | Informed Consent Requirements for Emergency Research                                |
| 21CFR Part 56     | Institutional Review Boards                                                         |
| 21CFR Part 201.56 | Content and Format for Drug Labeling                                                |
| 21CFR Part 54     | Financial Disclosure by Clinical Investigators                                      |
| 21CFR Part 54.2   | Covered Clinical Studies for Financial Disclosures by Clinical Investigators        |
| 21CFR Part 54.5   | Agency Evaluation of Financial Interests for Clinical Investigators                 |
| 21CFR Part 820    | cGMP Quality Systems Guidelines for Finished Medical Devices Intended for Human Use |
| 10 U.S.C. 1107    | United States Code—Notice of Use of an IND or a Drug Unapproved for its Applied Use |
| 21CFR Part 16     | Protection of Human Research Subjects                                               |
| 21CFR Part 1270   | Current Good Tissue Practice Compliance                                             |
| 21CFR Part 1271   | Current Good Tissue Practice Compliance                                             |
| 21CVR Part 11     | Electronic Records/Electronic Signatures Rule                                       |

### **Other Relevant Regulations Applicable to IND Development**

|             |                                                                                               |
|-------------|-----------------------------------------------------------------------------------------------|
| FD&C        | Federal Food, Drug and Cosmetic Act, 1938                                                     |
| USA PATRIOT | Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act, 2001            |
| PHSBPRA     | Public Health Security and Bioterrorism Preparedness and Response Act, 2002                   |
| RCRA        | Resource Conservation and Recovery Act, 1976                                                  |
| BMBL        | Biosafety in Microbiological and Biomedical Laboratories Guidelines (4 <sup>th</sup> Edition) |
| OSHA        | Occupational Safety and Health Act, 1970                                                      |
| ISO9001     | International Organization for Standardization 9000 Series for Quality Management System      |

### **ACKNOWLEDGMENT**

Author would like to acknowledge the interest and commitment of the National Defense Program of the Computer Sciences Corporation to support technical excellence, and Mr. Alvin Keith, the Business Unit Executive, for the support in the preparation of this chapter.

### **REFERENCES**

1. FDA (2002), New Drug and Biological Drug Products; Evidence Needed to Demonstrate Effectiveness of New Drugs When Human Efficacy Studies Are Not Ethical or Feasible,

- Final Rule, 21 CFR Parts 314 and 601, Docket No. 98N-0237, Health and Human Services, Washington DC.
2. FDA (1999). Computerized Systems Used in Clinical Trials—Guidance to Industry. Bio-research Monitoring Program, Food and Drug Administration, Rockville, MD.
  3. FDA (2004), *Pharmaceutical cGMPs for the 21st Century—A Risk-Based Approach*, Department of Health and Human Services, U.S. Food and Drug Administration, Final Report. Washington, DC.
  4. Organization for Economic Cooperation and Development (OECD) (1998), OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, ENV/MC/CHEM (98)17, 34.
  5. FDA (2008). CGMP for Phase I Investigational Drugs—Guidance to Industry. Division of Drug Information, HFD-240, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, MD.
  6. FDA (2008). Cumulative List of Products with Orphan Designation. The Office of Orphan Products Developments, Health and Human Services, Washington, DC.
  7. Beardsley, E., Jefford, M., and Mileschkin, L. (2007), Longer consent forms for clinical trials compromise patient understanding: So why are they lengthening? *J. Clin. Oncol.*, 25(9), 13–14.
  8. Burger, S. R. (2003), Current regulatory issues in cell and tissue therapy, *Cytotherapy*, 5(4), 289–298.

## BIBLIOGRAPHY

- Bhattacharyya, S. (2006), in *Product Development, Preclinical Testing and Toxicity Studies*, CBER Presentation at the Drug Information Association Meeting, June 18–22, Philadelphia, PA.
- Hirschfeld, S. (2006), in IND “202” Clinical Perspective. CBER Presentation at ISCT 6th Annual Somatic Cell Therapy Symposium, September 26–28, Bethesda, MD.



# 3

---

## *Preclinical Assessment of Safety in Human Subjects*

**Nancy Wintering and Andrew B. Newberg**

*Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania*

### **Contents**

|      |                                                                 |    |
|------|-----------------------------------------------------------------|----|
| 3.1  | Preparing Human Studies in Light of Preclinical Studies         | 71 |
| 3.2  | Estimated Biodistribution and Pharmacokinetics in Humans        | 73 |
| 3.3  | Initial IND Process                                             | 74 |
| 3.4  | Overview of Sections of IND                                     | 75 |
| 3.5  | IND Exemption Status                                            | 78 |
| 3.6  | Institutional Review Board Issues                               | 79 |
| 3.7  | Study Design                                                    | 80 |
| 3.8  | Subject Selection                                               | 81 |
| 3.9  | Safety Measures                                                 | 82 |
| 3.10 | Monitoring for Adverse Events                                   | 83 |
| 3.11 | Preparing for Phase II Studies and Additional Safety Evaluation | 84 |
| 3.12 | Conclusions                                                     | 84 |
|      | References                                                      | 84 |

### **3.1 PREPARING HUMAN STUDIES IN LIGHT OF PRECLINICAL STUDIES**

The first study to assess for safety in human subjects begins with the initial development of the drug. The particular physiological target for the drug should be known (e.g., whether it is designed to bind to serotonin receptors or specific tumor receptors). This target is determined by the group or company designing the pharmaceuti-

cal, which develops a molecule that is based upon existing ones or based upon known physiological molecules. Thus, the initial safety assessment targets the most likely pharmacological effects of similar molecules. For example, a drug that is intended to bind to the serotonin reuptake sites would be expected to have a safety profile similar to that of other drugs that bind the serotonin reuptake sites. With this knowledge in mind, the initial approach toward evaluating the safety of the new drug would include an evaluation of all pharmacological effects of known related drugs.

In addition to the presumed effects of a new drug based on related drugs, extensive animal studies (as described elsewhere in this book) also provide important information regarding how to assess for safety in human subjects [1]. The Food and Drug Administration (FDA) guidelines for nonclinical safety studies required for the conduct of human trials was published in 1997 [2]. Initial data in mouse and rat models provides preliminary information regarding the physiological changes to blood counts, electrolytes, liver function, immune function, and cholesterol levels. The initial small-animal data also provides clues of potential clinical adverse effects such as disturbances in behavior, appetite, sleep, and activity levels. If such effects are observed, then special care must be taken to observe similar effects in human subjects, and the study must be designed accordingly to evaluate these effects and also to reduce the risk of adverse effects. Initial toxicology studies in animals typically require a small and large animal in addition to nonhuman primates. However, there may be reasons to forego such studies in the case of a drug that targets a process for which there is no appropriate animal model. For example, recent studies of radiopharmaceuticals designed to bind to amyloid plaque in patients with Alzheimer's disease did not require study in nonhuman primates since there is no good model. Therefore, the results are not very meaningful in evaluating the potential effect in human subjects.

Toxicology studies typically require the administration of several different doses, often at substantially higher doses than will eventually be given to human subjects. The drug is then given to the animals for a time period similar to that in which the drug will actually be given. The test article is then given with measurements made of the concentration in the serum of the animals throughout the study. At the end of the dosing regimen, animals are typically sacrificed in order to assess for drug-related changes. Of course, if any results of the animal toxicology studies suggest that the drug might, in fact, pose a substantial danger to human subjects, this might prevent its even being tested in human subjects.

Observations for mortality, morbidity, and the availability of food and water are conducted at regular intervals for all animals. Observations for clinical signs are usually conducted daily on main study animals. In addition, some type of functional observational battery (FOB) is conducted on surviving main study animals prior to initiation of dosing, throughout the trial, and then at completion. Body weights are measured and recorded daily for all animals. Food consumption is measured and recorded daily for main study animals. Ophthalmoscopic examinations are conducted on animals pretest and on main study animals at termination. Blood and urine samples for clinical pathology evaluations are collected from all main study animals at termination. Blood samples for determination of the plasma concentrations of the test article are collected at designated time points throughout the study.

Upon completion of the study, complete necropsy examinations are performed on all main study animals, organ weights are measured, and selected tissues are microscopically examined. A similar analysis is performed in a large-animal model such as dogs with similar evaluations of clinical and laboratory assessments in addition to necropsy evaluations. At the conclusion of these studies, if there are any alterations observed in the animals, reporting of these findings in the investigational new drug (IND) are required and also necessitates careful evaluation in human subjects. Animal studies provide valuable information to assist in the assessment of safety during the development of a test article for human studies. Thus, if there was a significant change observed in platelet function in dogs, then platelet function should be closely evaluated in humans. Often this requires evaluation for at least the duration of the effect observed in animals, and sometimes longer. Thus, if platelet effects were noticed to last up to 2 days in animals, this should be evaluated for at least 2 days and most likely longer in human subjects.

### **3.2 ESTIMATED BIODISTRIBUTION AND PHARMACOKINETICS IN HUMANS**

Similar to toxicology studies, pharmacological evaluation also follows from animal studies and estimates of drug effects. Serum and tissue evaluation of drug concentrations and metabolism is necessary for determining the overall pharmacological effects and to evaluate potential adverse effects of the test article. Biodistribution evaluation in mice and rats and nonhuman primates can be very valuable in predicting the distribution in humans.

Some drugs are easier to have their biodistribution and pharmacology quantified. For example, many radiopharmaceutical products, which combine a radioactive atom with a molecule that follows some aspect of body physiology, can have their biodistribution evaluated through a series of imaging studies in which the radioactivity emitted from the product is detected with scans performed at a number of time points after administration of the drug. This allows for an evaluation of the distribution of the drug as well as the dosimetry that refers to the radioactive exposure an individual receives from the product. Nonradioactive drugs typically require an evaluation of serum concentration of the drug at various time points after administration. It may also be necessary to evaluate urine and fecal samples to determine drug concentrations and help evaluate the mechanism of excretion from the body.

The distribution and concentration of the drug in the body also helps with dose determination. If it is found that a drug is excreted more rapidly in human subjects than in animals, it might be appropriate to increase the dose administered. Drug escalation trials are frequently performed with chemotherapy agents and psychiatric drugs to evaluate both safety and efficacy. The dose of the drug is increased at regular intervals, usually after at least three subjects have had no significant adverse reactions or toxicity. If the drug is determined to be relatively safe, or has minor safety concerns, then the next step is to appropriately design the phase I safety evaluation in human subjects to evaluate the metabolic and pharmacologic actions of the study drug, side effects, and possible early evidence of the drug's effectiveness in humans as described in 21 CFR part 312.21. Preclinical studies to evaluate safety in humans are an essential step in the drug development process and will provide

valuable information for the investigational new drug application and the investigational drug brochure.

### 3.3 INITIAL IND PROCESS

In order to begin any clinical investigation in one or more humans that involves a test article, the study sponsor or sponsor-investigator is required to file an IND application with the FDA. For clarification the “sponsor” means the person who initiates the clinical investigation but who does not conduct the investigation; the “sponsor-investigator” is an individual who both initiates and conducts a clinical investigation either alone or with others. This distinction is important since there are legally binding responsibilities associated with each role. Guidance from the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) at the FDA describes a number of important regulations and provides guidance regarding the IND process. This information must be consulted for the development of any new pharmaceutical products and the submission of an IND application in accordance with 21 CFR part 312.

Under current regulations, any use in the United States of a drug product not previously authorized for marketing in the United States requires submission of an IND to the FDA. The Code of Federal Regulations sections 21 CFR 312.22 and 312.23 contain the general principles underlying the IND submission and the general requirements for an IND’s content and format.

The IND process has a number of components that are summarized below. These sections are also important due to the need for the FDA to determine what safety data is most important to evaluate in human subjects. Most pharmaceutical companies are very aware of the requirements for the IND process. For investigators in the academic or clinical setting, the IND is more formidable and more comprehensive than investigative protocols. However, it is generally recommended that investigators request a telephone conference with the appropriate FDA office prior to submitting an IND so that the process can run as smoothly as possible. In this conference call, specific issues regarding the evaluation of safety in humans can be discussed so that the final IND and study protocols are more likely to be acceptable to the FDA and also the institutional review board. Ultimately, the IND application is designed to assist and guide investigators (either private or public) to study the safety and pharmacology of new drugs.

Once an IND is filed with the FDA, there is a 30-day period in which the FDA is able to review the IND information, request additional information, and provide a nonobjection letter to enable the investigators to begin human studies, or put the study on hold. If additional information is required, this can sometimes be in the form of additional toxicology studies in animals including genotoxicity. The FDA might also require additions to be made to the study protocol for more extensive safety evaluation. This review might include additional types of measures or adding more time points for evaluation. Once these issues are resolved, and the protocol and IND application has been amended accordingly, the FDA may issue a nonobjection letter to begin a phase I human study. Timely and open communication with the FDA is encouraged to facilitate the IND process. If an investigator does not respond, an IND application may be put on hold. This is generally an unfavorable



outcome since the FDA is no longer under any time restriction to process the application, which can often substantially prolong the initiation of a study. However, the Institutional Review Board (IRB) approval of the amended protocol must always be obtained prior to the initiation of the study.

### 3.4 OVERVIEW OF SECTIONS OF IND

There are a number of sections required for an IND submission as described below and summarized from 21 CFR 312.23 for studies with a new drug or for a new indication of an approved drug:

1. Section 1 is the cover sheet (Form FDA-1571) containing the name, address, and telephone number of the sponsor, the date of the application, and the name of the investigational new drug. The phase or phases of the clinical investigation to be conducted is identified. By signing this form, the sponsor (or sponsor-investigator) makes a commitment not to begin clinical investigations until an IND covering the investigations is in effect. There is also a commitment that an Institutional Review Board that complies with the federal requirements will be responsible for the initial and continuing review and approval of each of the protocols in the proposed clinical investigation. The investigator also states a commitment to conduct the investigation in accordance with all other applicable regulatory requirements. The name and title of the person responsible for monitoring safety evaluation and also the conduct and progress of the clinical investigation is provided. The sponsor (or the sponsor-investigator) or the sponsor's authorized representative then signs the cover letter form.

#### 2. *Table of Contents*

3.i. *Introductory Statement and General Investigational Plan* This section usually includes a brief summary of the pharmaceutical agent to be tested, its presumed pharmacological and clinical effects, and how this will be tested through this IND.

3.ii. *Summary of Previous Human Experience* This section includes currently performed human studies. For completely new pharmaceutical products, this will be very brief and simply state that no human studies have been performed. However, for pharmaceuticals that are now being tested for other indications, there may be a substantial amount of information regarding the safety of physiological effects of the drug. It is important to provide a history of the drug development that includes a comprehensive summary from the literature that includes study populations, sample size, adverse events, and a summary of study outcomes.

3.iii. *Withdrawal of Drug* This section reports any reasons why the drug, or comparable drugs that have been developed, has been withdrawn for medical or other reasons.

#### 3.iv. *Overall Plan for Investigational Year*

a. *Rationale for this study:* The reason for the study design, including reasons for specific safety testing, and evaluation of the pharmacology profile and initial efficacy are described here.

b. *Indications for this study:* This section typically establishes why this study is needed and describes the overall goals for developing the investigational product.

c. General approach for evaluating this drug: This section describes whether the proposed study is for safety, biodistribution, pharmacology, dose escalation, or some other study design.

d. Clinical trials to be conducted during the first year: a comprehensive description of the anticipated activity of the first year of the study.

e. Estimated number of patients to be given the drug.

f. Risks: This section identifies the risks and severity and seriousness of the drug risks based upon the toxicological data found in animal studies and in prior studies conducted with humans with the drug or related drugs.

4. This section is reserved for the FDA.

5. *Investigator's Brochure* The investigator's brochure is required for clinical trials involving sites other than the one of the primary sponsor of the IND. Thus, a single site study of a new drug may not require an investigator's brochure. However, if multiple sites are involved, then an investigator's brochure is required to provide information about the drug and to ensure consistency in the conduct of the study at each site. The investigational brochure is developed in accordance with 21 CFR 312.55.

5.i. This section includes the drug substance and the structural formula of the drug.

5.ii. This section provides a summary of the toxicological and pharmacological effects of the drugs in animals and to humans if known.

5.iii. This section describes the pharmacokinetics and toxicological effects of the drug in animals and humans if known.

5.iv. This section provides a summary of known information related to safety and effectiveness in humans from prior clinical studies.

5.v. This section includes risks and side effects on the basis of prior experience with the drug or with similar drugs and provides precautions for drug monitoring and safety requirements in the use of the investigational drug.

#### 6.1. *Protocols*

6.1.a. *Objectives and Purpose* Describes an overview of the reasons and goals for the study—that is, to evaluate safety, biodistribution, pharmacology, efficacy, and the like.

6.1.b. *Personnel and Qualifications* This section lists and describes the individuals and the sites/organizations involved in executing the study providing evidence of their individual qualifications.

6.1.c. *Patient Selection and Number* A basis for the number and type of subjects to be recruited for the study are described. The source of recruitment must be mentioned particularly if it involves patient populations rather than controls. Justification of the number of subjects is also essential for ensuring that the appropriate numbers of subjects are studied without putting too many individuals at risk.

6.1.d. *Study Design* This section details the overall study design and is similar to what would be included in a full IRB submission including evaluative measures, “the type of control group that will be used and the methods to be used to minimize bias on the parts of the subjects, investigators and analysis” (21 CFR 312.23), and standardized procedures that will be used for data collection, data analysis, and

statistical analysis. It is important to state that safety is a primary component of the study and to design the protocol and standardized measures accordingly. Human studies must include safety as a critical evaluation measure.

6.1.e. *Determination of Dose, Maximum Dose, and Duration* The dose, route of administration, and maximum dose and individual dose exposure are described, usually based on animal data or prior human studies.

6.1.f. *Observations and Measures* In this section, the specific types of measures, both physiological and clinical (i.e., safety) are described. It is also important to describe what measures will be regarded as safe and which as adverse effects giving clear ranges for various measures.

6.1.g. *Clinical Procedures and Minimization of Risk* All clinical procedures are to be described with a focus on monitoring how risk will be minimized during the study. This might refer to existing animal or prior human experience and might also describe how adverse events will be managed.

## 7. *Chemistry, Manufacturing, and Control Information*

7.i. *Chemistry* This section describes the actual chemical composition of the drug including both active and inactive ingredients “to assure the proper identification, quality, purity and strength of the investigational drug.” The molecular weight and structure is also typically provided. In IND phase I studies, the emphasis in this section is on the identification of the raw materials that comprise the new drug.

### 7.ii. *Manufacturing and Control*

a. *Drug Substance* This is usually an extensive section that details where and how the drug is to be produced. Drug production should be according to good manufacturing practices and includes a description of the facilities and chemical reactions required to produce the material. The standard operating procedures for production do not have to specifically be included, but this section essentially follows them, and they can be included in the IND appendices to clarify the production process of the drug. A comprehensive description of the facilities and the personnel trained to oversee the drug production processes are included. If the drug is manufactured off-site, additional drug standard operating procedures and accountability measures are documented. The scope of the investigation will guide the amount of information to be submitted with the initial IND application. However, as the drug development proceeds and the study moves from a limited clinical investigation to a larger drug production and subject enrollment, the manufacturing and control procedures will be reported accordingly.

b. *Drug Product* This lists all the components of the drug product including active and inactive components, possible alternatives, and materials that do not appear in the final product but are used during manufacturing. Acceptable limits and analytical methods for assuring product stability, purity, and sterility are also included.

c. *Placebo Product* In studies that use a placebo in a controlled clinical investigation, a brief description of the manufacture and control is provided.

d. *Labeling* This section provides copies of all labels and labeling to be provided to each investigational site.

e. *Environmental Analysis Requirements* Describes whether an analysis of environmental issues is required.

## 8. *Pharmacology and Toxicology Information*

8.i. *Pharmacology and Drug Disposition* This section describes in detail the pharmacological properties of the investigational product as currently understood. This might be based primarily on animal studies but can also include previous human experience. This section is divided into each of these sections for clarification of the information provided to the FDA. In particular, data on nonprimate and primate species can be included under separate subheadings. Information provided should include the known pharmacology and biodistribution in animals and the presumed pharmacology in human subjects.

8.ii. *Toxicology* Similar to the pharmacology studies, this section can be expanded into subsections relating data from nonprimate and primate species with regard to toxicology studies.

a. Results on laboratory and clinical data as well as postmortem organ analysis are usually provided. It is often preferable to have two nonprimate species and one nonhuman primate species included in the analysis.

b. Each toxicology study that supports the safety of the investigator and data will be reviewed.

8.iii. *Statements of Compliance or Noncompliance* These statements of compliance indicate that the studies were conducted in compliance with good laboratory practice or provide a statement that provides the reasons or explanation for non-compliance with good laboratory practice (part 58).

9. *Previous Human Experience with Investigational Drug* In human studies, this section can be extremely brief. For drugs that have been used in humans for other purposes and are now being used for new indications, there may be substantial human experience. This information can be extremely valuable in establishing safety, pharmacology, and toxicology profiles of the study drug or test article.

9.i. This section provides detailed information that is relevant if the drug has been studied or marketed in the United States or in other countries. Comprehensive descriptive information must be provided about any clinical studies that have been conducted with the drug. This information may include published studies. Other information that may be relevant to the proposed investigation can be referenced in a bibliography or literature review.

9.ii. This section is needed if the test article is a combination of drugs previously investigated or marketed.

9.iii. If the drug has been marketed outside of the United States, this information should be described here, including whether the drug has been withdrawn from marketing for potential safety reasons.

10. *Additional Information* This section can include miscellaneous aspects of drugs including those that might have drug dependence and abuse potential or drugs that have radioactive elements. In pediatric studies, plans to assess safety and effectiveness in this population are described. Additional information that would aid in the evaluation of the drug safety can be included.

## 3.5 IND EXEMPTION STATUS

With these requirements in mind, it is also important to evaluate whether an IND is actually necessary. Many times in the clinical setting a particular pharmaceutical

product is used in an off-label manner. However, such a use in a research study may potentially require an IND. For example, if a drug with the specific indication for hypertension is used to help heart failure and an investigator wishes to test the use of the drug for heart failure, an IND might be required. However, the FDA and CFR regulations allow for an investigator to use a pharmaceutical product without an IND if the following conditions are met.

To begin with, the drug which is to be used must already be lawfully marketed in the United States. The investigational study must not be intended to support a new indication for use or to support a significant change in the advertising of the product. Importantly, the study should not involve the use of a different route of administration or dosage level or use in a patient population that significantly increases the risks associated with the use of the drug product. As with all studies, the study must be submitted, reviewed, and approved by the IRB and be conducted in accordance with standard regulations.

If these conditions are met, then an IND may not be required and the study can be performed under an exempt status. The advantage of an IND exemption is that it allows academic researchers who do not have an interest in altering the indications or marketing of a particular pharmaceutical to conduct a clinical investigation. Frequently, an investigator or clinician is interested in the pharmaceutical for a clinical or purely research perspective. In such cases, they may not need to utilize the full IND process.

It should also be clearly stated that there is no such thing as an “IND exemption” in the sense that there is no letter or form that can be obtained from the FDA. The FDA can always be contacted if there is a question as to whether an IND is required. Some institutions have initiated programs through their research offices or IRBs to help investigators determine whether an IND is required for a specific study product and may even provide a letter attesting that an IND is not required. However, for any investigator working in conjunction with a sponsor or a pharmaceutical company, or for all completely new materials, an IND is required.

### **3.6 INSTITUTIONAL REVIEW BOARD ISSUES**

IRB review is not substantially different for initial prospective studies of drugs in human subjects than in other nonsafety studies. However, the use of a new pharmaceutical agent typically requires more careful consideration due to the potential for a higher level of risk to patient safety and the additional institutional risk. It is important that the IRB can determine exactly how safety will be evaluated and how adverse effects will be managed. The specific descriptive language in consent forms can be difficult to develop since the drug has never been used in human subjects. Care must be taken to explain this lack of human experience, and it must be clear to the IRB and to the subject who gives consent to participate in the study that the principal aim of the study is to test the drug’s overall safety. It is important to state that this is the first time this drug has been administered to human subjects.

The timing of IRB approval and submission of the IND to the FDA can sometimes be challenging. The FDA requires a 30-day period to review an IND. During that period, the FDA can request additional information regarding the preclinical

physiology, pharmacology, and toxicology status of the pharmaceutical. In addition, the FDA may also request changes to the study design and protocol that has been submitted to the IRB for review. Thus, while IRB approval can be obtained concurrently with FDA evaluation of the IND, sometimes there is a substantial amount of interaction with the regulatory entities that require changes to protocols depending on the recommendations of both the IRB and FDA. In other words, if the original study design will evaluate safety data such as serum electrolytes at time 0 and 4 hours after drug administration, the FDA may request an additional blood draw at 24 hours. Approval to add this additional blood draw would have to be obtained as an amendment to the IRB protocol and may also require an additional revision to the consent form. Sometimes the IRB will not provide approval of the study until the FDA provides its “nonobjection” letter. Thus, it is sometimes helpful to know the FDA response prior to full submission to the IRB.

Any revisions that an IRB requires in the protocol or consent must also be submitted to the FDA as amendments. The amendment process is much different, though. For an amendment to be considered active, all that is required is IRB approval and submission to the FDA. Once these two requirements are met, the amendment can be considered to be in effect and future subjects can be studied according to the amended protocol.

However, if the investigator is conducting the study with a sponsor, the sponsor will act as the liaison between the investigator and the FDA. In clinical trials, the study cannot be initiated until the FDA, the IRB, and the sponsor have granted approval in writing for the study to begin.

### 3.7 STUDY DESIGN

All clinical studies must be performed under the current good manufacturing practices and good clinical practices guidelines (these practices are described by the guidelines produced through the International Conference on Harmonisation [3]). Further, they must be conducted with the requirements of appropriate ethical conduct as set by the Nuremberg Code [4], the Declaration of Helsinki [5], the 1962 Kefauver–Harris Amendments to the Food and Drug Act, and the Belmont Report [6]. Many of these guidelines are formalized in the Code of Federal Regulations [7].

The focus of safety studies is primarily to assess for all possible adverse physiological and clinical effects of the drug over the time period that the individual is exposed to the drug. The time period of exposure depends on the dose, dosing regimen, and half-life of the drug. For radiopharmaceuticals that require administration only on a single day, subjects are usually evaluated for only 1 or 2 days. Prolonged follow-up is not usually necessary since it is unlikely for the drug to have some effect days after it is in the subject’s body. On the other hand, drugs that are administered over many days require evaluation over a similarly long time period. It may even be necessary to evaluate subjects months or even a year later. Some of the issues related to study design are described below in more detail.



### 3.8 SUBJECT SELECTION

The subjects usually recruited for phase I studies of safety are healthy controls. This provides the best evaluation of safety in the human body. Rigorous prescreening consisting of serum and urinalysis along with comprehensive medical histories can assist in recruiting a healthy control group that is not being treated for chronic medical or psychiatric conditions. In phase I studies conducted with an IND, healthy controls should have normal physiological measures and clinical status, thus making abnormal changes related to the new drug easier to detect. It is also ideal that subjects are not taking any medications, including over-the-counter drugs, since these could potentially interfere with the evaluation of physiological effects of the new drug or may interfere with its pharmacology. Such drug interaction assessments are usually made after the initial safety assessment in controls.

There are a number of circumstances in which healthy controls are not appropriate in the initial safety assessment of a drug. When a drug is specific for a target population, measuring safety in healthy individuals may not accurately reflect the safety of the drug in the manner in which it will be used. For example, a drug that targets Alzheimer's disease (AD) pathophysiology (i.e., the development of amyloid plaque) may not make sense to study in healthy individuals since they do not have the plaque formation in the first place. Thus, the drug will bind very differently in healthy controls compared to AD patients. If the physiological half-life of the drug in controls is 12 hours, but in AD patients is 1 week, then the safety profiles in the different populations may also be vastly different.

Another reason that controls might not be appropriate would be if the drug is expected to have substantial risks but is developed for a population in which such risks may be warranted. The classic example here is of chemotherapy agents. These drugs, which often have substantial effects on blood counts or electrolytes, cannot be tested in healthy individuals since the drugs are too dangerous with no benefit either to them or to healthy controls. In these cases, the target population is studied, usually with a dose escalation design in which the first small group of three to five subjects is studied with a relatively low dose. If the lower dose does not result in substantial adverse effects, then the next group of subjects is studied at an incremental increase in dose. The drug dose continues to increase until specific safety criteria thresholds are reached. Thus, if hematological toxicity reaches level 3, then the drug is not given in any higher doses. After the initial safety assessment, then an efficacy trial can be performed at the highest, safe, dose.

One problem with studying safety in disease populations is that many of these individuals are regularly taking one or more drugs that may affect safety and pharmacology of the new drug. It is not always feasible or safe to withdraw all medication to wash out existing drugs prior to the initiation of a new drug regimen. For example, many patients with cancer take a variety of medications designed to help with their cancer, their overall health and mental well-being, or other related problems. Care must be taken to exclude patients with specific medications that are expected to interfere with the test drug. This may be determined by the evaluation of other related drugs or simply based on knowledge of the physiological processes associated with the different drugs.

**TABLE 1    Laboratory Values Assessed for Safety Studies**

---

|                                                                       |
|-----------------------------------------------------------------------|
| Cell blood count with differential (CBC)                              |
| Coagulation values (prothrombin time and partial thromboplastin time) |
| Prolactin and cortisol levels                                         |
| Electrolyte panel, including Blood Urea Nitrogen (BUN) and creatinine |
| Thyroid function tests, including Thyroid Stimulating Hormone (TSH)   |
| Follicle Stimulating Hormone (FSH) and Leutinizing Hormone (LH)       |
| Urinalysis, routine                                                   |
| Liver function tests                                                  |
| Cholesterol                                                           |
| Urine toxicology screen                                               |
| Autoimmune screening panel                                            |
| Pregnancy test                                                        |

---

**3.9    SAFETY MEASURES**

Safety measures usually include both laboratory as well as clinical assessment. Laboratory values are typically evaluated at baseline and then at regular intervals spanning the subject’s exposure to the test drug. For single-dose drugs, the evaluation may last only a couple of days while for drugs given over many days, the laboratory values must range over a similar time period. Table 1 shows typical laboratory values that are assessed in such studies. The most commonly used values include complete blood count, electrolytes, and liver function, but additional studies should be evaluated as determined through the results of preclinical animal studies. Values are compared to determine if substantial changes are associated with the administration of the test drug. If abnormalities are observed, then changes should be followed until they resolve. Usually, clear guidelines are established prior to the study that indicate what will be considered a drug-related change. In many cases, this requires an assessment of change rather than absolute values since a patient may experience a 10% drop in platelet count but still be in the normal range. This may ultimately lead to the need for evaluation of platelet counts prior to receiving the drug and if the counts are already low, limiting the use of the drug may be necessary.

In addition to laboratory values, physiological measures of heart rate, blood pressure, respiratory rate, and oxygen saturation may be useful in evaluating the safety of the test drug. Electrocardiography (EEG) is also important, especially when the drug may have specific cardiac effects. As with laboratory values, such measures might be made across the period during which the drug is administered. For a single-use drug such as a radiopharmaceutical imaging agent, physiological measures might be performed from a period immediately prior to drug administration up to one hour after administration. Subsequent measures might be made several hours later, 1 day, or even 2 days later. As with laboratory values, specific ranges of change might be described to clearly delineate what will and will not be considered a safety issue with the test drug.

A clinical evaluation including physical exam and report from the patient are also necessary at predefined intervals. The physical exam often includes evaluation of the skin (especially for an allergic reaction), hair, eyes, mucosa, heart, lungs, abdomen, and extremities. In addition, a neurological exam of strength and sensa-



tion may be useful. In addition to the psychological measures and laboratory values, the subjective experience of a study participant is also valuable. An ongoing dialog or direct questioning with the study participant should include whether they have the onset of any specific symptoms such as headache, dizziness, lightheadedness, gastrointestinal discomfort, nausea, chest or abdominal pain, weakness, and shortness of breath, sleep disturbance, or changes in appetite or mood.

It should also be noted that these extensive safety evaluations might be lessened after the initial experience with the test drug. Thus, after the initial safety evaluation, a more streamlined approach to safety measurement might be needed. In the example above of a single-dose radiopharmaceutical, eventual monitoring might only be required up to the first 10 or 20 min after administration.

### 3.10 MONITORING FOR ADVERSE EVENTS

Adverse events and serious adverse events have a very specific definition. Such events must be carefully evaluated and assessed to determine if the test drug is, in fact, related to the event. *Unexpected adverse drug experiences* are defined as any adverse drug experience, the specificity or severity of which is not consistent with the current investigator brochure; or, if an investigator brochure is not required or available, the specificity or severity of which is not consistent with the risk information described in the general investigational plan or elsewhere in the current application. Such adverse events may be minor unexpected laboratory changes or other clinical responses that are not anticipated. It is important that these adverse events be regarded and documented very specifically. For example, hepatic necrosis would be unexpected (by virtue of greater severity) if the investigator brochure only referred to elevated hepatic enzymes or hepatitis. It should also be clear that the term “unexpected” refers to an adverse drug experience that has not been previously observed rather than from the perspective of such experience not being anticipated from the pharmacological properties of the drug.

*Serious adverse drug experiences* are defined as any adverse drug experience occurring at any dose that results in any one of the following outcomes: Death, a life-threatening adverse drug experience, inpatient hospitalization or prolongation of existing hospitalization, a persistent or significant disability/incapacity, or a congenital anomaly/birth defect. A serious adverse drug experience may also be defined based upon appropriate medical judgment, as an event that may jeopardize the patient or subject and require medical or surgical intervention to prevent one of the outcomes listed above. Examples of such medical events include allergic bronchospasm requiring intensive treatment in an emergency room or at home. It is extremely important for the investigator to anticipate which effects are associated with a patient’s clinical condition and which effects can be associated with the study drug or the rigors of the study procedures.

Reporting of adverse events is critical to any preclinical safety study. The requirements include that the sponsor (investigator) notify the FDA and all participating investigators in a written safety report of: (a) any adverse experience associated with the use of the drug that is both serious and unexpected or (b) any finding from tests in laboratory animals that suggests a significant risk for human subjects including reports of mutagenicity, teratogenicity, or carcinogenicity. Each notification

should be made as soon as possible and in no event later than 15 calendar days after the sponsor's initial receipt of the information and should be prominently identified as an "IND safety report."

The sponsor must also notify the FDA by telephone or by facsimile transmission of any unexpected fatal or life-threatening experience associated with the use of the drug as soon as possible but no later than 7 calendar days after the sponsor's initial receipt of the information. Each telephone call or facsimile transmission to the FDA should be transmitted to the FDA new drug review division in the Center for Drug Evaluation and Research or the product review division in the Center for Biologics Evaluation and Research that has responsibility for review of the IND.

### **3.11 PREPARING FOR PHASE II STUDIES AND ADDITIONAL SAFETY EVALUATION**

Once the safety, pharmacology, and toxicology data is available from the initial human studies, this information is reported to the FDA as part of the next step to begin phase II studies. The phase II study protocols can be submitted under the same IND as amendments. The data from the phase I studies is typically included as new information that helps support the protocol to study the use of the drug in more specific patient populations and also in an expanded sample size.

The phase II trials will typically continue to evaluate safety, but on a lesser scale, and with more specific evaluations. Thus, if in the phase I study, electrolytes, liver function, cholesterol, and complete blood counts were evaluated with only the liver function studies showing any change, the phase II study may only include evaluation of liver function tests. If no laboratory abnormalities are observed in the phase I study, then no additional laboratory values may be necessary. Whatever changes to the safety evaluation of the drug in the phase II studies, these will ultimately have to be acceptable to the FDA and the IRB review committees.

### **3.12 CONCLUSIONS**

Overall, the process of moving studies from the animal preclinical studies to human use studies is complex, but following the regulatory guidelines for the FDA IND process, the IRB, good manufacturing practices, and good clinical practices is imperative and relatively straightforward. The Code of Federal Regulations specifies the requirements for these studies and can be obtained in book or online formats. By following these requirements, the ability to propose and execute phase I studies can be readily performed by both industry and academic sponsors and investigators.

## **REFERENCES**

1. Schacter, B. Z. (2006), *The New Medicines: How Drugs Are Created, Approved, Marketed, and Sold*, Praeger, Westport, CT.
2. FDA. Center for Drug Evaluation and Research (CDER) (1997), Guidance for industry: M3 nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals; available at <http://www.fda.gov/cder/guidance/1855fnl.pdf>.

3. International Conference on Harmonisation (1994), *ICH harmonized tripartite guideline: Clinical safety data management; definitions and standards for expedited reporting: International conference on harmonization of technical requirements for registration of pharmaceuticals for human use*, Geneva, Switzerland.
4. Nuremberg Military Tribunal (1949), *Nuremberg Code*, U.S. Government Printing Office, Washington, DC.
5. World Medical Association (2002), *Ethical Principles for Medical Research Involving Human Subjects*, World Medical Association, Ferney-Voltaire, France.
6. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979), *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*, Department of Health, Education, and Welfare; available at <http://ohrp.osophs.dhhs.gov/humansubjects/guidance/Belmont.htm>.
7. *Code of Federal Regulations* (CFR) (2006), U.S. Government Printing Office, Washington, DC.

### ADDITIONAL SOURCE

FDA. Center for Drug Evaluation and Research (CDER) (1995), CBER. Guidance for industry: Content and format of investigational new drug application for Phase I studies of drugs, including well characterized therapeutic biotechnology-derived products; available at <http://www.fda.gov/cder/guidance/phase1.pdf>.



---

# *Predicting Human Adverse Drug Reactions from Nonclinical Safety Studies*

**Jean-Pierre Valentin,<sup>1</sup> Marianne Keisu,<sup>2</sup> and Tim G. Hammond<sup>1</sup>**

<sup>1</sup>*Safety Assessment, AstraZeneca, Macclesfield, Cheshire, United Kingdom*

<sup>2</sup>*Patient Safety, AstraZeneca, Södertälje, Sweden*

## **Contents**

|                                                                                                |     |
|------------------------------------------------------------------------------------------------|-----|
| 4.1 Background                                                                                 | 88  |
| 4.1.1 Reasons for Drug Attrition                                                               | 88  |
| 4.1.2 Frequently Used Definitions                                                              | 89  |
| 4.1.3 Data Availability                                                                        | 90  |
| 4.2 Assessment of Predictive Value of Nonclinical Safety Testing to Humans by Organ Systems    | 93  |
| 4.2.1 Cardiovascular System                                                                    | 93  |
| 4.2.2 Nervous System                                                                           | 95  |
| 4.2.3 Respiratory System                                                                       | 97  |
| 4.2.4 Gastrointestinal System                                                                  | 98  |
| 4.2.5 Hepatic System                                                                           | 99  |
| 4.2.6 Renal and Urinary System                                                                 | 100 |
| 4.2.7 Endocrine System                                                                         | 100 |
| 4.2.8 Hemopoietic System                                                                       | 100 |
| 4.2.9 Immunological System                                                                     | 101 |
| 4.2.10 Skin                                                                                    | 101 |
| 4.3 Special Considerations                                                                     | 101 |
| 4.3.1 Biologic and Biotechnology-Derived Pharmaceutical, Biopharmaceuticals, and Biotech Drugs | 101 |

|       |                                                  |     |
|-------|--------------------------------------------------|-----|
| 4.3.2 | Genotoxicity                                     | 103 |
| 4.3.3 | Genital System and Teratology                    | 104 |
| 4.3.4 | Safety Biomarkers                                | 104 |
| 4.4   | Summary and Future Challenges                    | 105 |
| 4.4.1 | New Targets and New Approaches to Treat Diseases | 105 |
| 4.4.2 | Science and Technology                           | 106 |
| 4.4.3 | Regulatory Requirements                          | 106 |
| 4.4.4 | Training and Development                         | 106 |
|       | References                                       | 107 |

## 4.1 BACKGROUND

### 4.1.1 Reasons for Drug Attrition

The reasons for drug attrition have evolved over the years; over the last decade, however, lack of safety (combining both nonclinical and clinical) remains the major cause of attrition during clinical development, accounting for approximately 35–40% of all drug discontinuation (see Table 1) [1–3]. More worrying is the fact that over the last few years despite a widened/increased testing battery there is no clear trend toward a reduction of the attrition due to safety reasons.

In this section, a brief summary of the nature, frequency, and consequences of adverse drug reactions (ADRs) in two clinical situations is presented. They are ADRs experienced by healthy volunteers and patients participating in clinical studies with potential new medical entities (NMEs) and those experienced by patients prescribed licensed medicines. A review of these two situations points to areas of success with the current practices for nonclinical safety testing but also identifies areas where further research might lead to new or better nonclinical safety testing.

Prior to reviewing the literature, it is worth considering some frequently used definitions. Differences can be found in the literature in how the concepts are defined. We have decided here to use the definitions published by the International Conference on Harmonisation (ICH) [4, 5], as those concepts are used (direct or with some adaptations) by many pharmaceutical companies and are recognized by regulatory agencies.

**TABLE 1 Evolution of Reasons for Drug Discontinuation in Clinical Development**

| Reasons                  | Percentage [2] | Percentage [3] | Percentage [1] |
|--------------------------|----------------|----------------|----------------|
| Portfolio considerations | 23             | 22             | 20             |
| Clinical efficacy        | 22             | 23             | 26             |
| Clinical safety          | 12             | 20             | 14             |
| Toxicology <sup>a</sup>  | 23             | 19             | 21             |
| Other <sup>b</sup>       | 20             | 16             | 19             |

<sup>a</sup>Includes general/safety pharmacology.

<sup>b</sup>Includes reasons such as clinical pharmacokinetics/bioavailability, nonclinical efficacy, nonclinical pharmacokinetics/bioavailability, formulation, patent, legal, or commercial, regulatory. Overall, safety reasons accounted for up to ~40% of all drug discontinuation.

#### 4.1.2 Frequently Used Definitions

An *adverse event* (AE) can be defined as any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product.

An *adverse drug reaction* (ADR) is an adverse event where a causal relationship between the AE and the medicinal product is at least a reasonable possibility.

*Serious adverse events/adverse drug reactions* are defined as those that might be significant enough that if related to the medicinal product lead to important changes in the way the medicinal product is developed or used. This is particularly true for reactions that in their most severe forms threaten life or function.

The *severity*, which is the quantification of the reactions/symptoms (mild, moderate, severe), is used to describe grades of discomfort. One often used definition has been suggested by Tangrea et al. [6]: (i) mild, slightly bothersome, relieved with symptomatic treatment; (ii) moderate, bothersome, interferes with daily activities, only partially relieved with symptomatic treatment; and (iii) severe, prevents regular activities, not relieved with symptomatic treatment. A single serious ADR is always significant and very often has a high impact, its impact depending on when in the development process it occurs, and on the perceived benefit–risk profile of the NME. It can lead to the discontinuation of a drug in development, a significant limitation in the use of a drug (precaution, contraindication), or even to the withdrawal of the drug from the marketplace. A nonserious ADR can be more or less severe in its intensity, and its impact will depend upon its frequency and intensity. Nonserious ADRs can lead to a high degree of noncompliance, if they are perceived as annoying even if the symptoms they cause are not medically serious.

Pharmacological classification can divide ADRs in humans into five types (Table 2). Type A reactions (approx. 75% of all ADRs belong to this category) result from an exaggeration of the drug's normal pharmacological action when given in the usual therapeutic dose; they are normally dose dependent. Conventional pharmacology studies, combining primary, secondary, and safety pharmacology studies, can therefore reasonably be expected to predict type A ADRs. Functional toxicological measurements may predict type C ADRs. Conventional toxicology studies address type D ADRs, whereas prediction of type B responses (traditionally deemed as not predictable) occurring only in “susceptible” individuals require a more extensive

**TABLE 2 Classification of Adverse Drug Reactions (ADRs) in Humans**

|        |                                                                              |                                                           |
|--------|------------------------------------------------------------------------------|-----------------------------------------------------------|
| Type A | Dose-dependent; predictable from primary, secondary, and safety pharmacology | Main cause of ADRs (~75%), rarely lethal                  |
| Type B | Idiosyncratic response, not predictable, not dose related, usually serious   | Responsible for ~25% of ADRs, but majority of lethal ones |
| Type C | Long-term adaptive changes                                                   | Commonly occurs with some classes of drug                 |
| Type D | Delayed effects, e.g., carcinogenicity, teratogenicity                       | Low incidence                                             |
| Type E | Rebound effects following discontinuation of therapy                         | Commonly occurs with some classes of drug                 |

Source: Adapted from Redfern et al. [7].

nonclinical and clinical evaluation to identify individuals at risk. The increased possibility to identify genetic risk factors will hopefully lead to a better understanding of this type of reactions [8]. The type E ADRs are rarely investigated nonclinically using functional measurements unless there is cause for concern.

#### 4.1.3 Data Availability

The first human studies (phase I and early phase II) are generally very safe as a carefully selected population limit the potential for adverse effects (Table 3). In fact, molecules with a significant potential to generate serious ADRs are probably never given to healthy volunteers and are only given to patients where a significant benefit might be expected (e.g., refractory cancer patients) only with great care. These studies are to be conducted under frequent and extensive surveillance for the emergence of potentially worrisome ADRs. While AEs occur in phase I/healthy volunteer studies, they are generally more related to the required changes in lifestyle (e.g., caffeine deprivation) and the experimental procedures (e.g., needle puncture) than to the drugs [9, 10]. Nonclinical safety testing probably contributes significantly to the maintenance of this good track record; this is supported by published reports showing that single-dose nonclinical safety studies could overall accurately predict the clinical outcome [11–13]. The common ADRs observed with a high incidence (10–30%) during this early phase is linked to the gastrointestinal (nausea) and central nervous (headache) systems. In addition, an ADR specific to the NME under investigation that is pharmacologically mediated can be detected at this stage.

Early phase II studies are usually so-called dose-finding studies in selected patients and give the first indication of what type A reactions are to be expected in the targeted patient population. During continued clinical development as the population of exposed patients increase both in numbers and diversity, an increasing number of patients report AEs, with a wide variation in the type, frequency, and severity. Nonserious ADRs are often mechanism, drug class, or disease related (Table 4). Such ADRs limit the utility of an NME by restricting its use to those patients who either do not experience or can tolerate the ADRs, and they do not usually pose a “safety” issue. Serious ADRs tend to be present only at low frequencies. Pharmacological mechanism related to serious ADRs can occur in sensitive individuals, those with unusual kinetics, and in the presence of kinetic or occasion-

**TABLE 3 Risks to Healthy Volunteers in Phase I Clinical Research Trials**

| Year    | Number of Volunteers | Moderately Severe AE | Potentially Life-threatening AE | Deaths | References |
|---------|----------------------|----------------------|---------------------------------|--------|------------|
| 1965–77 | 29,162               | 58 (0.2%)            | —                               | 0      | 14         |
| 1980    | —                    | —                    | —                               | 1      | 15         |
| 1983    | —                    | —                    | —                               | 1      | 16         |
| 1984    | —                    | —                    | —                               | 1      | 17         |
| 1986–87 | 8,163                | 45 (0.55%)           | 3 (0.04%)                       | 0      | 18         |
| 1986–95 | 1,015                | 43 (3%)              | 0                               | 0      | 9          |
| 2000    | —                    | —                    | —                               | 1      | 19         |
| 2006    | 6                    | —                    | 6 (100%)                        | 0      | 20         |

Source: Adapted from Redfern et al. [7].



**TABLE 4 Major Causes of Acute Functional Adverse Drug Reactions**

| Acute Adverse Drug Reaction                                                            | Example                                                                                                                             |
|----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| Augmented (“supratherapeutic”) effect of interaction with the primary molecular target | Pronounced bradycardia with a $\beta$ blocker; pronounced hypotension with an angiotensin II receptor antagonist                    |
| Interaction with the primary molecular target present in nontarget tissues             | Sedation caused by anti-histamines                                                                                                  |
| Interactions with secondary molecular targets                                          | Interactions with the hERG cardiac channel leading to QT interval prolongation (e.g., some antipsychotics and antihistamines drugs) |
| Nonspecific effects                                                                    |                                                                                                                                     |
| Pharmacologically active metabolites                                                   | Inhibition of the hERG channel transcription by desmethyamiodarone, metabolite of amiodarone                                        |

Source: Adapted from Redfern et al. [7].

ally dynamic drug interactions. In principle, such ADRs might be predictable from safety pharmacology testing, although it should be acknowledged that safety pharmacology testing is usually conducted in young adult healthy animals under conditions that may be suboptimal to detect such effects. Occasional nonpharmacological (type B) serious ADRs occur; these can be induced by direct chemical toxicity, hypersensitivity or immunological mechanisms. Serious ADRs always limit the use of an NME by requiring warnings, precautions, and contraindications or preclude regulatory approval (depending on the frequency, seriousness, and perceived benefit–risk balance of the NME). In addition to preventing the development of NMEs likely to induce certain types of serious ADRs, a key contribution that can be made by nonclinical safety testing is in the elucidation of the mechanisms responsible for these ADRs. Once the mechanism responsible for the ADR is known, it becomes possible to prepare soundly argued precautions and contraindications.

When medicines are on the market, the actual incidence of serious ADRs is difficult to judge. The main source today for judging the safety profile of a drug on the market comes from spontaneous reports (from health professionals and in some countries patients) to health authorities. Numerous publications discuss the limitations of spontaneous reporting systems because of, for example, the underreporting of serious ADRs and the difficulty to assess the frequency of a certain ADR due to lack of exposure data in such systems [21–23]. However, based on these data and on studies looking at ADRs in a defined population such as hospitalized patients it is clear that serious ADRs occur with sufficient frequency to be a serious concern [24]. One authoritative review concludes that between 1 in 30 and 1 in 60 physician consultations are caused by ADRs (representing 1 in 30–40 patients [25]). The same review concludes that 4–6% of hospital admissions could be the result of ADRs. Although there is debate about the number of deaths caused by ADRs, the figure of around 100,000 deaths per year in the United States is often quoted [7, 24]. The majority of the above references are about serious ADRs that could be predicted or avoided, only a subset are idiosyncratic [26, 27].

The frequency of a serious ADR can be very low (e.g., 0.25–1.0 cases of rhabdomyolysis per 10,000 patients treated with a statin [28]; however, when millions of patients are under treatment, this can generate substantial morbidity. Serious ADRs may also be due to clinical error (e.g., misprescribing contraindicated drugs) or to

patient self-medication error—especially in the era of mass media communication and information. Independent of the cause of a certain serious ADR, it is important to investigate the pharmacological mechanisms driving these events. For example, the elucidation of the connection between drug-induced Torsades de Pointes, QT interval prolongation, and hERG potassium channel blockade has been considered as a major advance in this area, leading to avoidance of NMEs that have the potential to cause QT interval prolongation due to cardiac arrhythmias and sudden death by the rapid development of nonclinical in vitro screening assays of medium to high throughput capabilities [29].

To better understand the main causes for ADR-related drug withdrawals, medicines withdrawn from either the United States or worldwide markets were reviewed [25, 30]. The principal reasons, presented in Table 5, highlighted the fact that several of these toxicities fall into the remit of nonclinical safety testing, with the usual suspects ranking top of the list such as toxicities related to the cardiovascular, hepatic, and nervous systems. The prominence of arrhythmias in Stephens' review [25] probably reflects the recent interest in Torsades de Pointes type of arrhythmias.

Limitations of the data set are now considered. In most cases, hard evidence regarding the predictive value of nonclinical testing is not readily available in the public domain [12, 31–38]. When an NME has a serious negative effect in a nonclinical test, there may be limited information in the public domain, as the result may have precluded clinical development, and therefore the information was never published. If there has been no effect in a nonclinical test, and likewise no effect on the corresponding variable in humans, these negative data may have been deemed not to be of general interest. Thus there is a huge amount of unpublished information that might be of value for better understanding certain types of adverse effects. In addition, the scientific and medical community is left with examples of side effects in humans that were not identified during nonclinical assessment.

One publication attempted to explore the predictive value of safety/general pharmacology assays to humans [12]. Some significant correlations were reported. For example, decreased locomotor activity in rodents was positively correlated with dizziness and sleepiness in humans; decreased intestinal transit in rodents was correlated with constipation and anorexia in humans; decreased urinary and sodium excretions in the rat were correlated with edema in humans; decreased blood pressure in dogs was positively correlated to flushes, dizziness, headache, and malaise in humans; increased heart rate in dogs was correlated to palpitation; and increased blood flow in dogs was correlated to flushes and headache. Rather more bizarrely,

**TABLE 5 Evolution of Main Safety Reasons for Drug Withdrawal over Last 40 Years**

| Worldwide Withdrawal<br>(121 medicines) [30] | Percentage | U.S. Withdrawal (95 medicines) [25]                     | Percentage |
|----------------------------------------------|------------|---------------------------------------------------------|------------|
| Hepatic toxicity                             | 26         | Cardiovascular toxicity incl. arrhythmias               | 19 (12)    |
| Hematological toxicity                       | 10         | Neuropsychiatric effects/abuse liability/<br>dependency | 12         |
| Cardiovascular toxicity                      | 9          | Hepatic toxicity                                        | 9          |
| Dermatological effects                       | 6          | Bone marrow toxicity                                    | 7          |
| Carcinogenicity issues                       | 6          | Allergic reactions                                      | 6          |

the findings of analgesia, decreased body temperature, and anticonvulsive activity in rodents were each correlated with thirst in humans; the finding of pressure reflex to vagal stimulation was correlated with sleepiness, malaise, and thirst in humans. This indicates the limitations of such methods and the fact that correlations between events do not necessarily imply causal relationships.

There are, however, numerous examples of drugs that cause AEs in humans, which would be detectable in nonclinical safety evaluation assays. Some notable examples of individual drugs showing untoward effects in nonclinical studies that are correlated in a quantitative sense with AEs in humans have been reported [7]. For example, (a) the sedative effects of clonidine in various animal species and humans; (b) the propensity of cisapride to prolong ventricular repolarization; (c) the respiratory depressant effects of morphine; (d) the nephrotoxic effect of cyclosporine; and (e) the gastrointestinal effects of erythromycin. These examples illustrated the very good agreement of effects across all species tested and across a narrow range of doses, concentrations, or exposures.

However, there are areas that should be carefully considered to ensure optimization of the assays and ultimately increase the predictive value of nonclinical testing. These include, but are not limited to: (a) species differences in the expression or functionality of the molecular target mediating the adverse effects; (b) differences in pharmacokinetic properties between test species and humans; (c) sensitivity of the test system; (d) optimization of the test conditions; (e) appropriately statistically powered study designs; (f) appropriate timing of functional measurements in relation to the time of maximal effect, maximal exposure, and/or maximal tissue concentration; (g) delayed/chronic effects of parent drug and/or metabolites; (h) difficulty of detection in animals in standard nonclinical safety studies (e.g., arrhythmia, headache); and (i) assessment of a suboptimal surrogate endpoint that predict with some degree of confidence the clinical outcome (e.g., QT/QTc interval prolongation as a surrogate of TdP). Overall careful consideration should be paid to the sensitivity, specificity, and overall predictivity of nonclinical assays prior to using them for building an integrated risk assessment.

## **4.2 ASSESSMENT OF PREDICTIVE VALUE OF NONCLINICAL SAFETY TESTING TO HUMANS BY ORGAN SYSTEMS**

### **4.2.1 Cardiovascular System**

Over the last few years, data have been generated to assess the value of nonclinical tests to predict the potential of NMEs to prolong the duration of the QT interval of the electrocardiogram and ultimately the proarrhythmic potential of these drugs. The published data converged in that an integrated risk assessment based upon data on the potency against hERG, an *in vivo* repolarization assay, and if necessary an *in vitro* action potential assay are in a qualitative sense predictive of the clinical outcome [39–42]. These data have been further supported by publications suggesting that a 30-fold margin between the highest free plasma concentration of a drug in clinical use ( $C_{\max}$ ) and the concentration inhibiting by 50% the hERG current ( $IC_{50}$ ) could be adequate to ensure an acceptable degree of safety from arrhythmogenesis with a low risk of obtaining false positives [43–45]. Most recently, Wallis et

al. [46] have reported that, based on a data set of 19 compounds, combining data from the hERG assay and the in vivo QT repolarization assay predicted in 90% of cases the clinical outcome of the “thorough QT/QTc study.” In addition, the author suggested that a robust electrocardiographic monitoring during the phase I clinical trials combined with the preclinical data predicted in all cases the outcome of the thorough QT/QTc study. These data support the view that a dedicated thorough QT/QTc study as required per ICH E14 [47] does not add any additional value to the integrated QT risk assessment. These preliminary data would require to be strengthened with a larger data set. Such initiative is ongoing under the auspice of the ILSI-Health and Environmental Sciences Institute (HESI). Along the same lines, in a review of 25 anticancer drugs, Schein et al. [48] showed that based on general clinical assessment and pathology rather than on functional changes both primates and dogs predicted cardiovascular toxicities in 90% of cases in which the toxicities were observed in humans (Table 6). The authors suggested that physiological measurement monitoring could have further improved the overall predictivity to the clinical outcome. Furthermore, Olson et al. [13] showed good concordance between cardiovascular findings in dogs and humans. The translation was less robust between humans and rodents possibly due to the technical challenges associated with monitoring of cardiovascular function in rodents. Furthermore, based on a data set of 88 drugs, Igarashi et al. [12] showed good concordance between some pharmacological findings in animal models and their associated clinical adverse drug reactions (e.g., blood pressure reduction versus flushes, dizziness, headache, malaise; increase in heart rate versus palpitation and increase in blood flow versus flushes and headache) [12].

**TABLE 6 Species and Assays Predictive Value of Humans' Safety Endpoints**

| Safety Endpoint | Specie and/or Assay | Number of Compound | Sensitivity (%) | Specificity (%) | Predictivity (%) | References |
|-----------------|---------------------|--------------------|-----------------|-----------------|------------------|------------|
| Injection site  | Dog                 | 25                 | 67              | 53              | 56               | 48         |
|                 | Monkey              | 23                 | 59              | 67              | 71               | 48         |
| Integument      | Dog                 | 25                 | 43              | 56              | 52               | 48         |
|                 | Monkey              | 23                 | 50              | 77              | 76               | 48         |
| Cardiovascular  | Species combined    | 150                |                 |                 | 80               | 13         |
|                 | Dog                 | 25                 | 70              | 60              | 64               | 48         |
|                 | Monkey              | 23                 | 50              | 54              | 57               | 48         |
| Respiratory     | Dog                 | 25                 | 80              | 20              | 32               | 48         |
|                 | Monkey              | 23                 | 59              | 38              | 47               | 48         |
| Nervous system  | Species combined    | 150                |                 |                 | 61               | 13         |
|                 | Rodent              | 21 (7)             |                 |                 | 33               | 55         |
|                 | Dog                 | 21 (7)             |                 |                 | 29               | 55         |
|                 | Monkey              | 21 (7)             |                 |                 | 66               | 55         |
|                 | Rodent              | 21 (13)            |                 |                 | 75               | 55         |
| Bone marrow     | Dog                 | 25                 | 100             | 0               | 80               | 48         |
|                 | Dog                 | 21 (13)            |                 |                 | 85               | 55         |
|                 | Monkey              | 23                 | 95              | 0               | 90               | 48         |
|                 | Monkey              | 21 (13)            |                 |                 | 100              | 55         |
|                 | Dog                 | 25                 | 100             | 25              | 28               | 48         |
| Lymphoid        | Monkey              | 23                 | 0               | 68              | 71               | 48         |

TABLE 6 *Continued*

| Safety Endpoint          | Species and/or Assay                   | Number of Compound | Sensitivity (%) | Specificity (%) | Predictivity (%) | References |
|--------------------------|----------------------------------------|--------------------|-----------------|-----------------|------------------|------------|
| Gastrointestinal         | Species combined                       | 150                |                 |                 | 83               | 13         |
|                          | Rodent                                 | 21 (13)            |                 |                 | 82               | 55         |
|                          | Dog                                    | 25                 | 100             | 0               | 92               | 48         |
|                          | Dog                                    | 21 (13)            |                 |                 | 92               | 55         |
|                          | Monkey                                 | 23                 | 81              | 0               | 80               | 48         |
|                          | Monkey                                 | 21 (13)            |                 |                 | 86               | 55         |
| Hepatic                  | Species combined                       | 150                |                 |                 | 52               | 13         |
|                          | Rodent                                 | 21 (6)             |                 |                 | 83               | 55         |
|                          | Dog                                    | 25                 | 100             | 8               | 56               | 48         |
|                          | Dog                                    | 21 (6)             |                 |                 | 100              | 55         |
|                          | Monkey                                 | 23                 | 100             | 27              | 71               | 48         |
|                          | Monkey                                 | 21 (6)             |                 |                 | Not tested       | 55         |
| Renal and urinary        | Species combined                       | 150                |                 |                 | 68               | 13         |
|                          | Rodent                                 | 21 (3)             |                 |                 | 100              | 55         |
|                          | Dog                                    | 25                 | 80              | 7               | 36               | 48         |
|                          | Dog                                    | 21 (3)             |                 |                 | 100              | 55         |
|                          | Monkey                                 | 23                 | 90              | 21              | 52               | 48         |
|                          | Monkey                                 | 21 (3)             |                 |                 | 100              | 55         |
| Neuromuscular            | Dog                                    | 25                 | 86              | 17              | 36               | 48         |
|                          | Monkey                                 | 23                 | 71              | 57              | 66               | 48         |
| QT interval prolongation | Human—hERG in vitro <sup>a</sup>       | 19                 | 82              | 75              | 79               | 46         |
| QT interval prolongation | Dog in vivo <sup>a</sup>               | 19                 | 83              | 86              | 85               | 46         |
| QT interval prolongation | Dog Purkinje fiber <sup>a</sup>        | 19                 | 20              | 100             | 50               | 46         |
| QT interval prolongation | Combine nonclinical assay <sup>a</sup> | 19                 | 90              | 88              | 89               | 46         |
| Torsades de pointes      | Rabbit—in vitro                        | 64                 | 65              | 89              | 75               | 72         |
| Visual function          | Zebrafish                              | 37                 | 71              | 78              | 59, 73           |            |
| Seizure                  | Zebrafish                              | 25                 | 76              | 63              | 72               | 58         |

<sup>a</sup>Within twofold the free therapeutic plasma concentration. (Values in parentheses indicate) number of compounds showing toxicity in humans. The sensitivity was determined as the ratio of true positive (true positive/false negative). A high sensitivity reflects a low level of false positives. The specificity was determined as the ratio of true negative/true negative plus false positive. A high specificity reflects a low level of false negatives. The predictivity was determined as the ratio of true positive plus true negative/the total number of compounds evaluated.

#### 4.2.2 Nervous System

Most of the adverse drug effects relating to the nervous system impact on the quality of life rather than the risk to life (e.g., lethargy, anorexia, insomnia, personality changes, nausea). There are, however, some serious life-threatening adverse effects involving the nervous system (e.g., loss of consciousness and convulsions). Some of them reflect the fact that the central nervous system (CNS) controls the other two vital organ systems for that CNS impairment could be fatal (e.g., decreased respira-

tory drive leading to respiratory arrest; decreased sympathetic outflow leading to cardiovascular collapse). The nervous system adjusts the function of the other acutely vital organ systems according to current and long-term requirements of the organism. Therefore, drug effects on cardiovascular and respiratory functions can be mediated via a direct action within the CNS or via sensory nerve endings located in the cardiovascular and pulmonary systems. Some CNS adverse effects can be indirectly life threatening. For example, drowsiness, cognitive impairment, motor coordination, dizziness, involuntary movement, and visual or auditory disturbances can all affect driving performance; moreover, depression or personality changes can lead to suicidal tendencies. As an illustration, the number of deaths in the United States between 1984 and 1996 in patients receiving terfenadine was 396, a small proportion of which were attributed to sudden death resulting from Torsades de Pointes. This overall low incidence of fatalities is nevertheless a significant improvement over the first generation of antihistamine drugs that have been suspected to be responsible for significant fatalities in car accidents resulting from their sedative effects [49, 50]. General pharmacological tests for effects on the nervous system are usually observational studies of rodent general activity or multidimensional functional assays of motor activity [51–53]. For a series of 84 new drugs (excluding anticancer agents) studied in Japan, an evaluation of their capacity to predict adverse reactions in humans showed a general nonspecific correlation. For example, decreased locomotor activity in rodents was positively correlated with dizziness and sleepiness in humans [12]. A degree of overprediction was reported, particularly from studies that used high doses. Similarly, in the study of 45 miscellaneous drugs by Fletcher [54], high-dose effects such as ataxia and convulsions in animals did not occur in humans, and subjective effects such as dizziness, headache, dry mouth and sweating in humans were not predicted by animal studies. The correlation was stronger for other effects on the central nervous system.

Where effects on the central nervous system have been assessed in conventional toxicity studies using both clinical monitoring and histopathological examination of the brain and nervous tissue, a reasonable degree of concordance has been shown. Evaluation of the effects of up to 25 diverse anticancer drugs in dogs, monkeys, and humans showed a reasonable degree of concordance (nearly 40%) in neurological and neuromuscular toxic effects ([48]; Table 6). Dogs and monkeys had similar predictive values and high doses were needed to achieve the best correlation, whereas specific symptoms correlated poorly. The earlier study of 21 anticancer drugs by Owens indicated only a moderate correlation between neurotoxic effects in humans and animals [55]. The correlation was strong for alkylating agents but less so for other classes of drugs studied. Interestingly, the study of 150 miscellaneous drugs by Olson and colleagues [13, 56] showed that, overall, the nonrodent data were better correlated with adverse neurological effects in humans than the rodent data. A key neurological safety liability is the potential for drug-induced seizure (see Table 7); in this regard it is encouraging to note that the proconvulsive potential of marketed drugs was detected in a zebrafish larvae assay with a significant level of sensitivity, specificity, and overall predictivity [58] (Table 6).

So, while the data indicate poor prediction of subjective neurological effects, the information on the significant toxicities of anticancer drugs indicates that the conventional approach using histopathological examination detects potentially serious neurotoxic effects.

**TABLE 7 Incidence of Adverse Drug Reactions Related to Each of the Major Physiological Functions**

| Physiological System                | Percentage of Compounds Reporting ADR with Incidence >3% (%) |
|-------------------------------------|--------------------------------------------------------------|
| Cardiovascular                      | 35                                                           |
| Arrhythmias                         | 3                                                            |
| Central & peripheral nervous system | 56                                                           |
| Seizure                             | 3                                                            |
| Gastrointestinal                    | 67                                                           |
| Motility                            | 44                                                           |
| Hepatic                             | 11                                                           |
| Immune                              | 0                                                            |
| Renal & urinary                     | 17                                                           |
| Respiratory                         | 32                                                           |

*Note:* Data extracted from BioPrint [57]. Based on a set of 1138 drugs annotated for human ADRs. Not all compound-ADR annotations include incidence data; figures have been corrected to account for missing incidence data.

Certain types of side effects to the CNS (e.g., agitation, hallucinations, headache) are reported with some medicines. Such effects are not easily identifiable in animal studies.

**Special Senses** Relatively few instances of visual, auditory, or vestibular disturbances are reported in early clinical studies with new drugs. As such, there is a paucity of data comparing effects on these functions between laboratory animals and humans. However, ophthalmoscopic examination is usually performed in toxicity studies and is routinely accompanied by histopathological examination of the eye, prior to dosing of humans with a new drug. Consequently, it is probable that any agent that provokes severe ocular damage in animals after relatively short periods of dosing would not progress to clinical studies. Moreover, agents that have potent cataractogenic properties or that are severely toxic to the retina would be identified in relatively short, repeat-dose studies. Emerging data suggest that performing an optomotor assay in either the zebrafish [59] or in rodents [60] could detect drug-induced impairment of visual function with a reasonable sensitivity and specificity toward the clinical outcome. Specific tests of auditory function are seldom done routinely. But careful clinical observation of animal comportment probably eliminates agents that produce acute and severe auditory or vestibular damage.

#### 4.2.3 Respiratory System

A large body of data has accumulated on experimental methodology for examination of the effects of environmental and occupational chemicals on the respiratory tract. This is because inhalation is a primary mode of human exposure to foreign materials [61]. However, for inhaled drugs, there is limited information available in the public domain; part of the difficulty is that any significant respiratory side effects would be considered as unacceptable and consequently not progressed into humans. So we are probably left only with the circumstance of having taken a compound into humans, seeing an adverse effect, and then attempting to find suitable animal



models that could have predicted it in order to screen against it in the future. Some companies working in this area have expressed concerns over the ability to predict potential for “cough” reactions or bronchospasm. The effects of drugs are evaluated preclinically in dedicated safety pharmacology studies [62]; until the implementation of the ICHS7A [62], this was usually done alongside cardiovascular assessment in anesthetized dogs. In the comparison of 104 investigational new drugs by Igarashi and colleagues [12] in which this approach was used, respiratory disturbance was not frequently reported in humans. However, when respiratory related ADRs were reported in humans, these were not predicted by safety pharmacology testing.

In toxicity studies, effects on respiration are usually evaluated by clinical observation and histopathology of lungs and air passages, although it is usually recognized that clinical observation are “inappropriate to assess drug effects on respiratory function” [62]. In the study of 45 drugs by Fletcher [54], both toxicology and pharmacology animal studies overpredicted respiratory effects in humans. Similarly, Schein and colleagues [48] noted that this form of screening in nonrodents predicted respiratory signs or respiratory pathology in four out of five cases, but with a high percentage of false positives (Table 6).

Nowadays respiratory function is usually assessed in rodents using whole-body plethysmography, which provides an indication of drug effect on ventilatory parameters. Dennis Murphy [63] has argued that drug effect on both ventilatory parameters and lung mechanics should be assessed prior to first administration to humans since changes in lung mechanic (e.g., bronchoconstriction) may remain undetected/unpredicted using whole-body plethysmography methods.

#### 4.2.4 Gastrointestinal System

Clinical gastrointestinal adverse reactions account for ~18% of total ADRs, and 20–40% in hospitalised patients. There is a high degree of underreporting [64]. Diarrhea alone accounts for about 7% of all drug ADRs. For example, more than 700 drugs are implicated in causing diarrhea [65]. The majority are functional in nature (nausea, vomiting, dyspepsia, abdominal cramps, and diarrhea or constipation) and fewer are pathological (e.g., ulceration) or enhanced susceptibility to infection (e.g., pseudomembranous colitis) [66]. Overall, ~80% of gastrointestinal (GI) ADRs are predictable type A pharmacological reactions [67], that is, predictable from the primary and/or secondary pharmacological targets. From 1960 to 1999 two drugs withdrawn due to GI toxicity—Indoprofen and Pirprofen nonsteroidal anti-inflammatory drugs (NSAIDs) [30]. NSAID use in the United States alone is estimated to be responsible for over 100,000 hospitalisations and 17,000 deaths per year [68]. More recently, a prospective analysis of 18,820 United Kingdom patients; 17 deaths attributed to GI ADRs; most attributed to NSAID use [69].

The review of safety pharmacology studies performed in Japan on 88 noncancer drugs showed a good correlation between rodent intestinal transport and general adverse effects such as anorexia and constipation in humans [12]. In the review of conventional toxicology studies that included histopathology of the gastrointestinal tract, Olson and colleagues [13] showed good concordance between gastrointestinal effects in animals and humans, particularly for nonsteroidal anti-inflammatory drugs, anti-infective, and anticancer agents. In that review, large animal data were



a better predictor than data obtained from rodents. The data also showed good correlation between animal toxicology studies and humans for a diverse set of 45 drugs [47] and may be a link to the fact that large number of drugs are associated with gastrointestinal ADRs, thus increasing the sensitivity of detection level.

The rodent, dog, monkey, and human GI toxicity data also showed a strong correlation in the study of 21 anticancer drugs by Owens [55]. Surprisingly, in the study of 25 anticancer drugs by Schein [48], the dog was superior to the monkey as a predictor of adverse GI effects in humans (Table 6). For example, monkeys were remarkably resistant to vomiting, an adverse event that was observed in humans with 21 of the 25 compounds. Gastrointestinal tract toxicity was a significant contributor to the remarkably good quantitative correlation of toxicity across species based on dose/body surface area for the 18 anticancer drugs studies by Freireich and colleagues [70]. This is not surprising since oncology drugs tend to be used at maximum tolerated doses at which gastrointestinal side effects are quite common.

It has been suggested that the GI tract of dogs is highly physiologically similar to that of humans in terms of motility patterns, gastric emptying, and pH, particularly in the fasted state [71]. This observation, coupled with the ability to use a formulation similar to that used in humans, makes the canine GI tract a most relevant model.

#### **4.2.5 Hepatic System**

Hepatotoxicity is an important adverse drug effect and a relatively common reason for termination of the development of an NME [13, 73]. At present, drug-induced hepatic injury accounts for more than 50% of cases of acute liver failure in the United States [74]. In conventional nonclinical studies of toxicity, the cornerstone of the assessment of hepatotoxic potential is measurement of circulating liver enzymes and hepatic histopathology [75]. A review of 38 chemicals, 24 of which were drugs that produce hepatic toxicity in humans, showed a concordance of 80% with findings in conventional toxicity studies [76]. Hepatic toxicity was not under-predicted in the study of 25 anticancer drugs in dogs and monkeys that used conventional hepatic enzyme measurements and histopathology ([48]; Table 6). The study of anticancer drugs by Owens [55] showed a similar good correlation.

Conversely, the study of data on 150 drugs exhibiting human toxicities showed that the concordance between hepatotoxicity found in animal studies and that observed in clinical practice was little more than 50% [13]. This larger study undoubtedly included agents that developed idiosyncratic responses, which are not usually detected in early clinical trials because of their rarity. This is a significant problem—in recent years there have been notable examples of hepatic toxicity of a poorly understood, idiosyncratic nature that have caused the withdrawal of marketed drugs despite extensive and essentially negative nonclinical testing and large clinical trials. The thiazolidinedione troglitazone, an antidiabetic drug, was associated with serious hepatic injury in patients despite its lack of hepatic toxicity in preclinical studies [77]. Another example is bromfenac, a nonsteroidal anti-inflammatory drug [74]. Hepatic failure also occurred in clinical trials with the nucleoside analog fialuridine as a result of mitochondrial disturbance and steatosis. Despite long-term treatment of monkeys, dogs, and rats with fialuridine, the only hepatic effects observed were

increases in apoptosis and nuclear atypia in rats [77]. Ximelagatran was found to cause liver enzyme elevations such as alanine aminotransferase (ALT) in 7.9% of patients treated in long-term studies (>35 days). Routine nonclinical animal data did not show any liver enzyme elevations. Ximelagatran and its metabolites were tested in an extensive study program including cell viability, mitochondrial function, formation of reactive metabolites, and reactive oxygen species without finding an explanation to the mechanism behind the ALT elevation [78].

It is probable that most NMEs that produce severe hepatotoxicity in animals are not tested in humans, so that the true level of concordance is likely to remain obscure. However, overall, the data seem sufficiently robust to conclude that overt liver damage observed in animal toxicity studies indicates potential risk of hepatic toxicity in humans. This underlines the prudence of a critical histopathological examination of the liver tissue in nonclinical studies and careful patient monitoring in response to any hepatic alerts from animal studies.

#### **4.2.6 Renal and Urinary System**

Renal toxicity is assessed by conventional histopathology, measurement of blood urea and electrolytes, and examination of urine volume and the sediment it contains. Concordance in the database of 150 drugs reviewed by Olsen and colleagues [13] was fair. A good correlation was noted among the 21 anticancer drugs reviewed by Owens [55], with rodents and dogs performing equally well. However, in a review of 45 drugs, renal toxicity was correctly predicted by animal studies in 3 instances but overpredicted in 22 others [54]. Similarly, the study of 25 anticancer drugs in dogs or primates correctly predicted renal toxicity in 9 cases, underpredicted in 1, and overpredicted in 14 [48].

#### **4.2.7 Endocrine System**

Endocrine changes during nonclinical studies are routinely assessed only by histological examination of endocrine organs, unless there are particular reasons to suspect endocrine effects. Olson and colleagues [13] noted only moderate concordance (60%) between animals and humans. As might be expected from the way in which the endocrine system responds to stimuli, these effects were not common in humans and generally occurred after phase I studies. The review by Fletcher [54] indicated that endocrine findings in nonclinical studies significantly overpredict effects in humans. Endocrine effects, particularly those involving the adrenal gland, are commonly reported in toxicity studies [79]. These findings often represent adaptive alterations to repeated doses of drugs and usually manifest as changes in glandular weight and cellular atrophy or hypertrophy. These changes might not have significant implications for human safety in single-dose studies, but they characterize possible endocrine effects that need to be assessed in clinical trials.

#### **4.2.8 Hemopoietic System**

Hemopoiesis is routinely assessed by examination of peripheral blood, bone marrow smears, and histopathology of the blood-forming and lymphoid organs. Theus and Zbinden [80] reviewed prior industry practice for the assessment of coagulation in

1984 and found substantial deficiencies. The screening practice that they proposed for animal studies is similar to that used in humans and has now been almost universally adopted for pharmaceutical testing.

There is substantial data on the concordance of adverse effects on hemopoietic tissue due to anticancer and antimetabolic drugs between animals and humans. The evidence indicates good correlation for both rodents and humans and dogs and humans for myelotoxicity, although the particular cell series affected sometimes differs [48]. Thrombocytopenias were correctly predicted for 13 of 18 anticancer drugs that produced this toxicity in humans. Moreover, in the series of 18 anticancer drugs studied by Freireich and colleagues, hemopoietic toxicity was one of the most significant contributors to the remarkably good quantitative correlation across species based on dose/body surface area [70].

A reasonable correlation between animals and humans was also noted for decreases in white blood cell counts in the study of 139 drugs by the Japanese Pharmaceutical Manufacturers Association [12]. Anticancer drugs and antibodies did predominate in this series, but the authors also detected a considerable number of false negatives and false positives in their data.

#### **4.2.9 Immunological System**

Specific tests of immune function are not routinely performed for conventional new drugs prior to their use in humans. An international collaborative study showed that examination of peripheral blood white cells, histological examination of thymus and spleen, and, in particular, careful histological examination of lymphoid tissue in the rat is a good primary method of identifying agents that are significant direct-acting immunotoxins [81]. New screens of immune function in animals are sometimes proposed for drug assessment [82, 83], but more sophisticated tests of immune function might be more appropriate and safely conducted in human studies. Coping with the potential impact of biotechnology-derived pharmaceuticals on immune status and immunogenicity is a special challenge for which careful attention to the principles of immunology is needed [84].

#### **4.2.10 Skin**

Of all tissues, skin shows the least concordance between effects in animal studies and human patients. A general lack of predictive reliability for skin reactions in humans has been noted in the reviews of anticancer and other drugs ([48, 54, 55]; Table 6). Adverse skin hypersensitivity effects have caused the development of a relatively large number of potential NMEs to be terminated [13, 73, 85].

### **4.3 SPECIAL CONSIDERATIONS**

#### **4.3.1 Biologic and Biotechnology-Derived Pharmaceutical, Biopharmaceutical, and Biotech Drugs**

The definitions of biology-related therapeutic agents and examples of agents are presented in Table 8; there are a number of important differences between biologics

**TABLE 8 Definitions of Biology-Related Therapeutic Agents**

| Term                                                                               | Definition                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Examples                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Biologic                                                                           | A therapeutic agent derived from a biological source or produced using a biological process. For the purposes of this chapter, this will include drugs obtained by extraction from human or animal tissues and body fluids or produced by biotechnological means (recombinant or hybridoma technology).<br>It will include cell-, virus-, and bacteria-based products as well as DNA-based therapeutics. It will NOT include certain small-molecule agents produced by fermentation (e.g., antibiotics) or plant-derived products (botanicals). | Plasma-derived proteins (albumin, clotting factors, immunoglobulin preparations)<br>Tissue-derived proteins (animal insulins, human pituitary-derived growth hormone)<br>Vaccines (live, attenuated, or killed whole-cell preparations, recombinant peptide vaccines, DNA, and viral vector vaccines)<br>Recombinant peptides and proteins<br>Monoclonal antibodies<br>Products derived from transgenic animals or plants<br>Gene therapy and related tissue or cell therapy products<br>Antisense drugs |
| Biotechnology-derived pharmaceutical, biopharmaceutical, biotech drug <sup>a</sup> | Any drug produced by biotechnological means, such as recombinant or hybridoma technology, are generally regarded to include recombinant peptides or proteins, antibodies, genetically modified tissue, and cell-based products as well as DNA-based products.                                                                                                                                                                                                                                                                                   | Vaccines (as above)<br>Recombinant proteins or peptides and derivatives (e.g., erythropoietin, clotting factors, insulins, or insulin analogs)<br>Monoclonal antibodies and antibody fragments (e.g., Fabs)<br>Transgenic animal products (e.g., rh albumin)<br>Gene therapy products (as above)                                                                                                                                                                                                         |

<sup>a</sup>These terms are basically interchangeable and broadly refer to the same type of therapeutic agents.

and conventional small molecules that need to be understood to design effective nonclinical and clinical programs for these products. They refer principally to the production process, molecular weight and composition, microheterogeneity, species specificity, pleiotropism, dose-level selection, metabolism, and catabolism [86, 87]. The available regulatory guidance on the nonclinical safety testing of biologics has recently been reviewed [88]. Although there are a number of product-class-specific guidelines that have been published over the years (for review see Ryle [87]), the most relevant general guidance on the nonclinical testing of biologics is the ICH S6 [89]. The guideline emphasizes the need for a flexible, case-by-case approach and appropriate species selection, with special attention paid to immunologically mediated effects in animals and their relevance to human patients. Some of the special considerations that need to be taken into account in the nonclinical testing of biologics include the source and quality of the test material (in particular batch-to-batch variability), species selection, immunogenicity (in particular after repeated dosing), dosing schedule, dose levels and study durations, immunomodulation, and the assessment of local tolerance. The serious adverse events observed with the CD28 agonist antibody TGN-1412 in a phase I study in early 2006 ([20]; Table 3) demonstrate how difficult it is to lay out any general examples of typical nonclinical

programs for biologics. Recent investigations on the mechanisms of the TGN-1412-mediated “cytokine storm” will enable the development of novel procedures to improve nonclinical safety testing of immunomodulatory therapeutics [90]. Meanwhile this serious adverse event has led to evolving regulatory expectations for this type of product [91].

#### 4.3.2 Genotoxicity

There has been much contention about the relevance of genotoxicity assays to the testing of pharmaceutical agents since their introduction more than 30 years ago [92]. However, extensive study has led to a better understanding of the chemical determinants that provoke genotoxic effects through electrophilic attack of biological macromolecules [93]. As a consequence of this understanding, mutagenic activity is often simply avoided in the drug discovery process with the exception of certain classes of drugs aimed at treating cancer (for review see Guzzie-Peck [86]). Nevertheless, prior to first human exposure, *in vitro* tests for mutations and chromosomal damage are routinely carried out according to internationally agreed technical guidelines that are based on a large body of historical data for diverse chemicals. However, it can be difficult to assess human risk when unexpected or unexplained activity in these bacterial or mammalian cell tests occurs. Such activity usually precludes dosing to healthy volunteers at least until further work elucidates the mechanism of activity and characterizes any hazard. Subsequently, *in vivo* assays of bone marrow micronucleus, peripheral blood cytogenetics, and liver unscheduled DNA (deoxyribonucleic acid) synthesis in rodents are usually done. The technical performance of these tests has also been the subject of international collaborative studies. *In silico* approaches are now used routinely and tend to supersede *in vitro* testing of genotoxic potential.

An area of growing regulatory concern is the assessment of potentially genotoxic impurities of pharmaceuticals. The ICH and the European Medicine Agency (EMA) have published guidance documents focusing on the safety evaluation of impurities in pharmaceutical drug substances and drug products [94–100]. The EMA guidance is based on a threshold of toxicological concern (TTC) derived from animal carcinogenicity data using multiple worst-case assumptions to estimate a daily dose associated with a lifetime cancer risk of 1 in 100,000, a risk level considered acceptable for genotoxic impurities in human medicines. Based on these assumptions, presentation of the TTC as a single figure infers an unwarranted level of precision and supports the adoption of a more flexible approach by regulatory authorities when evaluating new drug products; a range within five fold of the TTC limit would be sensible. Furthermore, the limit is based on 70 years continuous daily exposure, a scenario that is uncommon for most medicines and irrelevant to the preregistration clinical development phase. To address this later point a staged TTC has been developed that proposes limits based on shorter durations of treatment (e.g., up to 1-year). Based on recent history, this approach has been acceptable by some authorities but not others, and it is imperative that steps are taken to reach a common agreement between the pharmaceutical industry and regulatory authorities worldwide in order that new medicines can continue to be developed and delivered to benefit patients in a safe and timely manner [101].

### 4.3.3 Genital System and Teratology

Reproductive changes are seldom reported in early clinical trial studies, largely due to the exclusion of women of child-bearing potential from these studies [54]. In a report from the Federal Institute for Occupational Safety and Health, several reviews as well as studies on individual compounds have been analyzed with respect to the suitability of different study designs and endpoints to detect effects on male reproduction in animal species [102]. However, only a few studies were available that characterize the human situation. A considerable inter- and intraindividual variability was noted with respect to key parameters associated with fertility in men (e.g., sperm count, motility, morphology, and volume). Interspecies extrapolation factors were derived from the most sensitive endpoint in laboratory animals. Despite the small database and limitations of the studies that prevented any robust conclusion, it was felt that humans were generally not more susceptible to reproductive toxicants than laboratory animals as was originally assumed. For the purpose of hazard identification a subacute study exploring concentrations that produce significant general toxicity might be sufficient. If effects were found, for the purpose of risk assessment the no-adverse-effect level has to be identified by testing sensitive endpoints. Although a subchronic study was felt preferable, a subacute study may be sufficient. Similar observations emerged from a collaborative study in Japan of 16 drugs, 12 of which were associated with infertility in humans; results showed that histopathological endpoint was the most sensitive method for preclinical detection of drugs with antifertility properties [103]. A recommendation from the Federal Institute for Occupational Safety and Health report was to develop and validate a rabbit model allowing sequential sperm analysis and better observation of behavior [102].

Teratology, the study of abnormal prenatal development and congenital malformations induced by exogenous chemical and physical agents is being primarily assessed using *in vivo* approaches. Although the interspecies concordance and extrapolation to humans is in large part unknown, mainly because teratogenic compounds are either not progressed into humans or progressed under very restricted usage. Although, animal studies in mammals remain the gold standard, other alternative methods/assays are being explored. In this regard, it is worth mentioning three *in vitro* methods endorsed as scientifically validated by the European Centre for In Vitro Alternative Methods (ECVAM) in 2001 [104]: the embryonic stem cell test, the micromass test, and the whole-embryo culture. More recently, the use of the zebrafish opens some promising avenues [105].

### 4.3.4 Safety Biomarkers

The field of safety biomarkers (SBMs) is advancing rapidly. Our improved understanding of the molecular bases of organ toxicity suggests that monitoring specific molecular responses may provide improved prediction of human outcomes and in doing so provide “bridging SBMs” that may eliminate much of the current uncertainty in extrapolating from laboratory models to human outcome. Modern high-throughput technologies for proteins or endogenous metabolites offer a major opportunity to systematically identify sensitive and specific plasma or urine SBMs that could serve as an index of damage specific to each of the important



internal tissues and organs. SBMs can serve many decision-making purposes. Depending on where SBMs are used in the drug discovery/development process they demand different levels of validation and qualification. SBMs can be deployed for (i) target related toxicity, (ii) CD family-related toxicity, (iii) unexpected toxicity during GLP studies, and (iv) unexpected toxicity during clinical development. In accepting an SBM as qualified for regulatory decision-making, FDA has exhibited clearly that it will operate only within a broad scientific consensus. This means that:

- Data on an SBM from different investigative methods should be convergent and support a single hypothesis for its role in a certain organ toxicity.
- There are no or few data that are incompatible with the hypothesis.
- The data available in support of an SBM are persuasive to independent expert peer groups.
- The data should have originated from multiple and independent investigations in several laboratories.

A suitable way to reach this consensus is to share the interest in developing biomarkers in consortia formed by, e.g., pharma and diagnostic companies, academia, and regulators.

## 4.4 SUMMARY AND FUTURE CHALLENGES

### 4.4.1 New Targets and New Approaches to Treat Diseases

Advances in molecular biology and biotechnology allow for the identification of new molecular targets, leading to the discovery and development of newer pharmaceutical agents that act at these novel molecular sites in an attempt to ameliorate the disease condition. Moreover, new therapeutic approaches are being developed (e.g., gene therapy, biotech products) that offer new challenges to assessing their safety in humans. Inherent in the novelty of new targets and new approaches is the risk of unexpected and unwanted effects that may or may not be detected based on current scientific knowledge and with current techniques and assays.

One of the biggest challenges for the biotechnology and pharmaceutical companies in the twenty-first century will be to develop and deliver drugs that fit the individual patient's biology and pathophysiology ("personalized medicine") [106, 107]. In addition, new developments in therapeutic approaches, such as those involving biopharmaceuticals and gene therapy, offer promise in the treatment or prevention of diseases for which current approaches are ineffective.

At the same time health care costs are increasing dramatically in many countries, and aging populations often require treatment with multiple drugs. In addition increased drug development costs and recent high-profile issues relating to drug safety highlight the need for finding new medicines with acceptable safety profiles while minimizing development cost.

There is a need to identify/screen out drug candidates with poor safety profiles as early as possible in the drug discovery process, and nonclinical safety assessment functions have an important role to play in achieving this goal [108].

To increase the overall value of the nonclinical assays would require to better understand their predictive value for humans; to achieve this would require reviewing analysis data available in the public domain as well as proprietary information and sharing the outcome of such effort widely. This could ideally be achieved via consortia involving academic institutions, regulatory agencies, and the pharmaceutical industry. Such approaches have been recently initiated [109, 110].

#### **4.4.2 Science and Technology**

Nonclinical safety evaluation faces significant scientific challenges to keep pace, to adapt, and to incorporate new technologies in the evaluation of NMEs in nonclinical assays/models and identifying the effects that pose a risk to human volunteers and patients. Recent examples have included the use of electrophysiological techniques to evaluate the effects of NMEs on the hERG channel [111] and telemetry techniques to assess the effects of NMEs on the duration of the QT interval in unstressed animals (for review see McMahon et al. [29]), therefore enabling to establish integrating QT risk assessment reliable and predictive of the clinical outcome. The development and utilization of technologies and approaches that have a direct clinical correlate should also be encouraged; for example, the utilization of echocardiography to assess drug effect on ejection fraction can be applied in both a nonclinical and clinical setting. Furthermore, efforts should continue to construct databases relating to the predictive value of nonclinical assays to humans either through retrospective analysis or through purposely designed studies [13, 39–42, 57, 72].

#### **4.4.3 Regulatory Requirements**

Continued benefit–risk assessment is today at the center of drug development and drug life-cycle maintenance together with the requirement to submit risk management plans to regulatory agencies together with licensing applications [112, 113]. These requirements have helped pharmaceutical companies to develop cross disciplinary working to ensure identification, assessment, and better understanding of safety risks and to devise risk minimization activities. Wherever feasible conditional approvals based on fewer patients and with rigorous prospective safety follow-up should be considered. In its March 2004 Critical Path Report [114], the Food and Drug Administration (FDA) suggests that limited exploratory IND investigations in humans (phase 0) can be initiated with less, or different, nonclinical support, that is required for traditional IND studies because exploratory IND (e-IND) studies should present fewer potential risks than do traditional phase I studies that look for dose-limiting toxicities. The nonclinical program should be considered on a case-by-case basis depending on the specific objectives for a given e-IND.

#### **4.4.4 Training and Development**

Disciplines involved in nonclinical safety evaluation of NMEs face significant challenges of attracting, training, and certifying investigators to ensure the future of these disciplines. The paucity of training in certain biomedical scientific disciplines (toxicology, pathology, pharmacology, physiology) has had detrimental long-lasting effects such as (a) an impact on the development of intact animal models of human



function and disease; (b) an impact on skills to conceptualize biomedical hypothesis and experiments at the level of the intact animal; and (c) an impact on the process of nonclinical and clinical drug discovery and development. There is a clear need to ensure all parties involved in the training, education, and development of individuals working in these disciplines work together to ensure continuous supply of these key skills.

Developing organized, planned, and prospective methods for clinical safety data review to interpret and act on this data will require integrated databases, possibilities to search historical data, and possibilities to bring information together from multiple sources in a new way would enable assessing the concordance between preclinical and clinical safety data and ultimately refine the nonclinical safety testing strategies to select the safest candidate drugs.

## REFERENCES

1. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rates? *Nature Rev./Drug Disc.*, 3, 711–715.
2. Kennedy, T. (1997), Managing the drug discovery/development interface, *Drug Disc. Dev.*, 2, 436–444.
3. Lasser, K. E., Allen, P. D., Woolhandler, S. J., Himmelstein, D. U., Wolfe, S. M., and Bor, D. H. (2002), Timing of new black box warnings and withdrawals for prescription medications, *JAMA*, 287, 2215–2220.
4. Anon. (1996), ICH E2C(R1) Harmonised Tripartite Guideline, Clinical Safety Data Management: Periodic Safety Update Reports for Marketed Drugs. CPMP/ICH/288/95.
5. Anon. (2003), ICH E2D Harmonised Tripartite Guideline, Post-approval Safety Data Management: Definitions and Standards for Expedited Reporting. CPMP/ICH/3945/03.
6. Tangrea, J. A., Adrianza, M. E., and McAdams, M. (1991), A method for the detection and management of adverse events in clinical trials, *Drug Inf. J.*, 25, 63–80.
7. Redfern, W. S., Wakefield, I. D., Prior, H., Hammond, T. G., and Valentin, J. P. (2002), Safety pharmacology—A progressive approach, *Fund. Clin. Pharmacol.*, 16, 161–173.
8. Wilke, R. A., Lin, D. W., Roden, D. M., Watkins, P. B., Flockhart, D., Zineh, I., Giacomini, K. M., and Krauss, R. M. (2007), Identifying genetic risk factors for serious adverse drug reactions: Current progress and challenges, *Nature Rev./Drug Dis.*, 6, 904–916; available at: [www.nature.com/reviews/drugdisc](http://www.nature.com/reviews/drugdisc).
9. Sibille, M., Deigat, N., Janin, A., Kirkesseli, S., and Durand, D. V. (1998), Adverse events in phase-I studies: A report in 1015 healthy volunteers, *Eur. J. Clin. Pharmacol.*, 54, 13–20.
10. Rozenzweig, P., Brohier, S., and Zipfel, A. (1993), The placebo effect in healthy volunteers: Influence of experimental conditions on the adverse events profile during phase I studies, *Clin. Pharmacol. Thera.*, 54(5), 578–583.
11. Greaves, P., Williams, A., and Eve, M. (2004), First dose of potential new medicines to humans: How do animals help? *Nature Rev./Drug Disc.*, 3, 226–236.
12. Igarashi, T., Nakane, S., and Kitagawa, T. (1995), Predictability of clinical adverse reactions of drugs by general pharmacology studies, *J. Toxicol. Sci.*, 20, 77–92.

13. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000), Concordance of the toxicity of pharmaceuticals in humans and in animals, *Regul. Toxicol. Pharmacol.*, 32(1), 56–67.
14. Zarafonitis, C. J., Riley, P. A., Willis, P. W., et al. (1978), *Clin. Pharmacol. Ther.*, 24, 127–132.
15. Kolata, G. B. (1980), The death of a research subject, *Hastings Center Report*, 10, 5–6.
16. Daragh, A., Kenny, M., Lambe, R., and Brick, I. (1985), Sudden death of a volunteer, *Lancet*, I 93–94.
17. Anon. (1985), Editorial. Death of a volunteer, *BMJ*, 290, 1369–1370.
18. Orme, M., Harry, J., Routledge, P., and Hobson, S. (1989), Healthy volunteer studies in Great Britain: The results of a survey into 12 months activity in this field, *Br. J. Clin. Pharmacol.*, 27, 125–133.
19. McCarthy, M. (2001), Healthy volunteer dies in US physiology study, *Lancet*, 357, 2114.
20. Anon. (2006), Expert Scientific Group on Phase I Clinical Trials. Interim report. Duff, G. W. (chairman); accessed November 4, 2006 at: <http://www.dh.gov.uk/assetRoot/04/13/75/69/04137569.pdf>.
21. Alvarez-Requejo, A., Carvaajal, A., Vega, T. L., and Bégaud, B. (1994), Underreporting of adverse drug reactions in a Spanish regional centre of pharmacovigilance, *Drug Inf. J., Abstr.* 249 (Suppl. 1), S104.
22. Bégaud, B., Martin, K., Haramburu, F., and Moore, N. (2002), Rates of spontaneous reporting of adverse drug reactions in France, *JAMA*, 288(13), 1588.
23. Brewer, T., and Colditz, G. A. (1999), Postmarketing surveillance and adverse drug reactions: Current perspectives and future needs, *JAMA*, 281, 824–829.
24. Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998), Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies, *JAMA*, 279, 1200–1205.
25. Stephens, M. D. B. (2004), Introduction, in Talbot, J., and Waller, P., Eds., *Stephens' detection of new adverse drug reactions*, 5th ed., John Wiley & Sons, Chichester, UK, pp. 1–91.
26. Avery, A. A., Taylor, R. L., Partidge, M., Neil, K., et al. (2001), Investigating preventable drug-related admissions to a medical admissions unit, *Pharmacoepidemiol. Drug Saf.*, 10(S103), 243.
27. Bhalla, N., Duggan, C., and Dhillon, S. (2003), The incidence and nature of drug-related admission to hospital, *Phara. J.*, 270, 583–586.
28. McKenney, J. M. (2005), Pharmacologic options for aggressive low-density lipoprotein cholesterol lowering: Benefits versus risks, *Am. J. Cardiol.*, 96(4A), 60E–66E.
29. McMahon, N., Pollard, C., Hammond, T. G., and Valentin, J. P. (2007), Cardiovascular safety pharmacology, in Sietsema, W. K., and Schwen, R., Eds., *Nonclinical Drug Safety Assessment—Practical Considerations for Successful Registration*, FDA News, Washington, DC, 87–123.
30. Fung, M., Thornton, A., Mybeck, K., Wu, J. H., Hornbuckle, K., and Muniz, E. (2001), Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceuticals markets—1960 to 1999, *Drug Inform. J.*, 35, 293–317.
31. Calabrese, E. J. (1984), Suitability of animal models for predictive toxicology: Theoretical and practical considerations, *Drug Metab. Rev.*, 15, 505–523.

32. Garratini, S. (1985), Toxic effects of chemicals: Difficulties in extrapolating data from animals to man, *Ann. Rev. Toxicol. Pharmacol.*, 16, 1–29.
33. Grieshaber, C. K., and Marsoni, S. (1986), Relation of preclinical toxicology to findings in early clinical trials, *Cancer Treat. Rep.*, 70, 65–72.
34. Lumley, C. E., and Walker, S. R. (1985), The value of chronic animal toxicology studies of pharmaceutical compounds: a retrospective analysis, *Fund. Appl. Toxicol.*, 5, 1007–1024.
35. Lumley, C. E., Parkinson, C., and Walker, S. R. (1992), An international appraisal of the minimal duration of chronic animal studies, *Human Exp. Toxicol.*, 11, 155–162.
36. Monro, A., and Mehta, D. (1996), Are single-dose toxicology studies in animals adequate to support single dose of a new drug in humans? *Clin. Pharmacol. Ther.*, 59, 258–264.
37. Oser, B. L. (1981), The rat as a model for human toxicological evaluation, *J. Toxicol. Environ. Health*, 8, 521–642.
38. Zbinden, G. (1994), Predictive value of animal studies in toxicology, *Regul. Tox. Pharm.*, 14, 167–177.
39. Ando, K., Hombo, T., Kanno, A., Ikeda, H., Imaizumi, M., Shimizu, N., Sakamoto, K., Kitani, S., Yamamoto, Y., Hizume, S., Nakai, K., Kitayama, T., and Yamamoto, K. (2005), QT PRODACT: in vivo QT assay with a conscious monkey for assessment of the potential for drug-induced QT interval prolongation, *J. Pharmacol. Sci.*, 99, 487–500.
40. Miyazaki, H., Watanabe, H., Kitayama, T., Nishida, M., Nishi, Y., Sekiya, K., Suganami, H., and Yamamoto, K. (2005), QT PRODACT: Sensitivity and specificity of the canine telemetry assay for detecting drug-induced QT interval prolongation, *J. Pharmacol. Sci.*, 99, 523–529.
41. Omata, T., Kasai, C., Hashimoto, M., Hombo, T., and Yamamoto, K. (2005), QT PRODACT: Comparison of non-clinical studies for drug-induced delay in ventricular repolarization and their role in safety evaluation in humans, *J. Pharmacol. Sci.*, 99, 531–541.
42. Sasaki, H., Shimizu, N., Suganami, H., and Yamamoto, K. (2005), QT PRODACT: Inter-facility variability in electrocardiographic and hemodynamic parameters in conscious dogs and monkeys, *J. Pharmacol. Sci.*, 99, 513–522.
43. De Bruin, M. L., Pettersson, M., Meyboom, R. H. B., Hoes, A. W., and Leujkens, H. G. M. (2005), Anti-hERG activity and the risk of drug-induced arrhythmias and sudden death, *Eur. Heart J.*, 26, 590–597.
44. Webster, R., Leischmann, D., and Walker, D. (2002), Towards a drug concentration effect relationship for QT prolongation and torsades de pointes, *Curr. Opin. Drug Disc. Develop.*, 5, 116–126.
45. Redfern, W. S., Carlsson, L., Davis, A. S., Lynch, W. G., MacKenzie, I., Palethorpe, S., Siegl, P. K. S., Strang, I., Sullivan, A. T., Wallis, R., Camm, A. J., and Hammond, T. G. (2003), Relationship between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: Evidence for a provisional safety margin in drug development, *Cardiovas. Res.*, 58, 32–45.
46. Wallis, R. (2007), QT-understanding the complexities of defining the translation of animal data to humans. 7th Annual Meeting of the Safety Pharmacology Society, Edinburgh, September 20–21.
47. Anon. (2005), ICH E14: The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. London, 25 May 2005. CPMP/ICH/2/04. <http://www.emea.eu.int/pdfs/human/ich/000204en.pdf>.

48. Schein, P. S., Davis, R. D., Carter, S., Newman, J., Schein, D. R., and Rall, D. P. (1970), The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man, *Clin. Pharmacol. Ther.*, 11, 3–40.
49. Cimbura, G., Lucas, D. M., Bennett, R. C., Warren, R. A., and Simpson, H. M. (1982), Incidence and toxicological aspects of drugs detected in 484 fatally injured drivers and pedestrians in Ontario, *J. Forensic Sci.*, 27, 855–867.
50. Weiler, J. M., Bloomfield, J. R., Woodworth, G. G., Grant, A. R., Layton, T. A., Brown, T. L., McKenzie, D. R., Baker, T. W., and Watson, G. S. (2000), Effects of fexofenadine, diphenhydramine, and alcohol on driving performance. A randomised, placebo-controlled trial in the Iowa driving simulator, *Ann. Intern. Med.*, 132, 354–363.
51. Mattsson, J. L., Spencer, P. J., and Albee, R. R. (1996), A performance standard for clinical and functional observation battery examination of rats, *J. Am. Coll. Toxicol.*, 15, 239.
52. Irwin, S. (1968), Comprehensive observational assessment: 1a. A systematic, quantitative procedure for assessing the behavioural and physiologic state of the mouse, *Psychopharmacologia (Berl.)*, 13, 222–257.
53. Haggerty, G. C. (1991), Strategies for and experience with neurotoxicity testing of new pharmaceuticals, *J. Am. Coll. Toxicol.*, 10, 677–687.
54. Fletcher, A. P. (1978), Drug safety tests and subsequent clinical experience, *J. Roy. Soc. Med.*, 71, 693–696.
55. Owens, A. H. (1962), Predicting anticancer drug effects in man from laboratory animal studies, *J. Chron. Dis.*, 15, 223–228.
56. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000), Concordance of the toxicity of pharmaceuticals in humans and in animals, *Regul. Toxicol. Pharmacol.*, 32(1), 56–67.
57. Krejsa, C. M., Horvath, D., Rogalski, S. L., Penzotti, J. E., Mao, B., Barbosa, F., and Migeon, J. C. (2003), Predicting ADME properties and side effects: The BioPrint approach, *Curr. Op. Drug Dis. Dev.*, 6(4), 470–480.
58. Winter, M. J., Redfern, W. S., Hayfield, A. J., Owen, S. F., Valentin, J-P., and Hutchinson, T. H. (2008), Validation of a zebrafish larval locomotor activity screen for assessing the seizure-liability of early-stage development drugs, *J. Pharmacol. Toxicol. Methods.*, 57, 176–187.
59. Richards, F. R., Alderton, W. K., Kimber, G. M., Liu, Z., Strang, I., Redfern, W. S., Valentin, J-P., Winter, M. J., and Hutchinson, T. H. (2008), Validation of the use of WIK and TL strain zebrafish larvae for visual safety assessment, *J. Pharmacol. Toxicol. Methods.*, 58, 50–58.
60. Maung, K. P., Storey, S., McKay, J., Bigley, A., Heathcote, D., Elliott, K., Valentin, J-P., Hammond, T. G., and Redfern, W. S. (2008), Validation of an optometry system for measurement of visual acuity in Han Wistar rats, *J. Pharmacol. Toxicol. Methods.*, 58, 152.
61. Nemery, B., Dinsdale, D., and Verschoyle, R. D. (1987), Detecting and evaluating chemical-induced lung damage in experimental animals, *Bull. Eur. Physiopathol. Respir.*, 23, 501–528.
62. Anon. (2000), ICHS7A: Safety pharmacology studies for human pharmaceuticals. CPMP/ICH/539/00; accessed on November 16, 2000 at: <http://www.emea.eu.int/pdfs/human/ich/053900en.pdf>.
63. Murphy, D. J. (2005), Comprehensive non-clinical respiratory evaluation of promising new drugs, *Toxicol. Appl. Pharmacol.*, 1, 207(2 Suppl), 414–424.

64. Lewis, J. H. (1986), Gastrointestinal injury due to medicinal agents, *Am. J. Gastroenterol.*, 81, 819–834.
65. Chassany, O., Michaux, A., and Bergmann, J. F. (2000), Drug-induced diarrhoea, *Drug Saf.*, 22, 53–72.
66. Ghahremani, G. G. (1999), Gastrointestinal complications of drug therapy, *Abdom. Imaging*, 24, 1–2.
67. Gatenby, R. A. (1995), The radiology of drug-induced disorders in the gastrointestinal tract, *Semin. Roentgenol.*, 30, 62–76.
68. Whittle, B. J. (2003), Gastrointestinal effects of nonsteroidal anti-inflammatory drugs, *Fundam. Clin. Pharmacol.*, 17, 301–313.
69. Pirmohamed, M., James, S., Meakin, S., Green, C., Scott, A. K., Walley, T. J., Farrar, K., Park, B. K., and Breckenridge, A. M. (2004), Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18,820 patients, *BMJ*, 329, 15–19.
70. Freireich, E. J., Gehen, E. A., Rall, D. P., Schmidt, L. H., and Skipper, H. E. (1966), Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey and man, *Cancer Chemother. Rep.*, 50, 219–244.
71. Dressman, J. B. (1986), Comparison of canine and human gastrointestinal physiology, *Pharmacol. Res.*, 3, 123–131.
72. Lawrence, C. L., Bridgland-Taylor, M. H., Pollard, C. E., Hammond, T. G., and Valentin, J.-P. (2006), A rabbit Langendorff heart proarrhythmia model: Predictive value for clinical identification of torsade de pointes, *Br. J. Pharmacol.*, 149(7), 845–860.
73. Lumley, C. (1990), in Walker, S. R., Ed., *Animal Toxicity Studies: Their Relevance for Man*, Quay, Lancaster, pp. 49–56.
74. Lee, W. M. (2003), Drug-induced hepatotoxicity, *N. Engl. J. Med.*, 349, 474–485.
75. Amacher, D. E. (1998), Serum transaminase elevations as indicators of hepatic injury following administration of drugs, *Regul. Toxicol. Pharmacol.*, 27, 119–130.
76. Hayes, A. W. et al. (1982), Correlation of human hepatotoxicants with hepatic damage in animals, *Fund. Appl. Toxicol.*, 2, 55–66.
77. Schwartz, S., Raskin, P., Fonseca, V., and Graveline, J. F. (1998), Effect of troglitazone in insulin-treated patients with type II diabetes mellitus, *N. Engl. J. Med.*, 338, 861–866.
78. Kenne, K., Skanberg, I., Glinghammar, B., Berson, B., Pessayre, D., Flinois, J.-P., Beaune, P., Edebert, I., Diaz Pohl, C., Carlsson, T., and Andersson, T. B. (2007), Prediction of drug induced liver injury in humans by using in vitro methods: The case of ximelagatran, *Toxicol. In Vitro*, 22(3), 730–746.
79. Ribelin, W. E. (1984), The effects of drugs and chemicals upon the structure of the adrenal gland, *Fund. Appl. Toxicol.*, 4, 105–119.
80. Theus, R., and Zbinden, G. (1984), Toxicological assessment of the hemostatic system, regulatory requirements, and industry practice, *Regul. Toxicol. Pharmacol.*, 4, 74–95.
81. Dayan, A. D. et al. (1998), Report of a validation study of assessment of direct immunotoxicology in the rat, *Toxicology*, 125, 183–201.
82. Dean, J. H. (1997), Issues with introducing new immunotoxicology methods into the safety assessment of pharmaceuticals, *Toxicology*, 119, 95–101.
83. Dean, J. H., Hinks, J. R., and Remander, B. (1998), Immunotoxicology assessment in the pharmaceutical industry, *Toxicol. Lett.*, 102–103, 247–255.
84. Cavagnaro, J. A. (2002), Preclinical safety evaluation of biotechnology-derived pharmaceuticals, *Nature Rev. Drug Disc.*, 1, 469–475.
85. Lichfield, J. T. (1961), Forecasting drug effects in man from studies in laboratory animals, *JAMA*, 177, 104–108.

86. Crommelin, D. J. A., Storm, G., Verrijck, R., de Leede, L., Jiskoot, W., and Hennink, W. E. (2003), Shifting paradigms: Biopharmaceuticals versus low molecular weight drugs, *Int. J. Pharm.*, 266, 3–16.
87. Ryle, P. R. (2007), Special considerations in the preclinical testing of biologics: The ICHS6 guideline, in Sietsema, W. K., and Schwen, R. Eds., *Nonclinical Drug Safety Assessment—Practical Considerations for Successful Registration*, FDA News, Washington, DC, pp. 301–330.
88. Snodin, D. J., and Ryle, P. R. (2006), Understanding and applying regulatory guidance on the nonclinical development of biotechnology-derived pharmaceuticals, *Biodrugs*, 20, 25–52.
89. Anon. (1997), Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals. ICH Harmonised Tripartite Guideline S6. Geneva: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, CPMP/ICH/302/95.
90. Stebbings, R., Findlay, L., Edwards, C., Eastwood, D., Bird, C., North, D., Mistry, Y., Dilger, P., Liefoghe, E., Cludts, I., Fox, B., Tarrant, G., Robinson, J., Meager, T., Dolman, C., Thorpe, S. J., Bristow, A., Wadhwa, M., Thorpe, R., and Poole, S. (2007), “Cytokine Storm” in the Phase I Trial of Monoclonal Antibody TGN1412: Better Understanding the Causes to Improve PreClinical Testing of Immunotherapeutics, *J. Immunol.*, 179, 3325–3331.
91. Anon. (2007), Guideline on requirements for first time in man clinical trials for potential high-risk medicinal products, London, March 22, 2007. EMEA/CHMP/SWP/28367/2007.
92. Clive, D. (1985), Mutagenicity tests in drug development: Interpretation and significance of test results, *Regul. Toxicol. Pharmacol.*, 5, 79–100.
93. Benigni, R., and Zito, R. (2003), Designing safer drugs: (Q)SAR-based identification of mutagens and carcinogens, *Curr. Topics Med. Chem.*, 3, 1289–1300.
94. Guzzie-Peck, P. (2007), Genotoxicity testing and risk management, in Sietsema, W. K., and Schwen, R., Eds., *Nonclinical Drug Safety Assessment—Practical Considerations for Successful Registration*, FDA News, Washington, DC, pp. 197–272.
95. Anon. (2005), Guideline on the non-clinical investigation of the dependence potential of medicinal products, London, April 21, 2005. EMEA/CHMP/SWP/94227/2004.
96. Anon. (2006), EMEA Guideline on the Limits of Genotoxic Impurities. Committee for Medicinal Products for Human Use. The European Medicines Agency, London. CPMP/SWP/5199/02. EMEA/CHMP/QWP/251344/2006.
97. Anon. (2006), HMPC concept paper on the development of a guideline on the assessment of genotoxic constituents in herbal substances/preparations. Committee on Herbal Medicinal Products European Medicines Agency, EMEA/HMPC/413271/2006.
98. Anon. (2006), ICH Q3A(R2) Impurities in new drug substances, in *International Conference on Harmonisation Harmonised Tripartite Guideline*. Current Step 4 version 25; available at: <http://www.ich.org/LOB/media/MEDIA422.pdf>.
99. Anon. (2006), ICH Q3B(R2) Impurities in new drug products, in *International Conference on Harmonisation Harmonised Tripartite Guideline*. Current Step 4 version 2; available at: <http://www.ich.org/LOB/media/MEDIA421.pdf>.
100. Anon. (2005), ICH Q3C(R3) Impurities: Guideline for residual solvents, in *International Conference on Harmonisation Harmonised Tripartite Guideline*. Current Step 4 version.
101. Humfrey, C. D. N. (2007), Recent developments in the risk assessment of potentially genotoxic impurities in pharmaceutical drug substances, *Toxicol. Sci.*, 100(1), 24–28.



102. Mangelsdorf, I., and Buschmann, J. (2002), *Extrapolation from Results of Animal Studies to Humans for the Endpoint Male Fertility*. Project F1642, Federal Institute for Occupational Safety and Health, Dortmund, Germany.
103. Takayama, S., Akaike, M., Kawashima, K., Takahashi, M., and Kurokawa, Y. (1984), A collaborative study in Japan on optimal treatment period and parameters for detection of male fertility disorders induced by drugs in rats, *Regul. Toxicol. Pharmacol.*, 14, 266–292.
104. Bremer, S., Pellizzer, C., Adler, S., Paparella, M., and de Lange, J. (2002), Development of a testing strategy for detecting embryotoxic hazards of chemicals in vitro by using embryonic stem cell models, *ATLA*, 30(Suppl. 2), 107–109.
105. Gustafson, A. L., Weiser, T., Clemann, N., Hossaini, A., Janaitis, C., Bluemel, J., Delongas, J. L., and Hill, A. (2008), *Validation of zebrafish as a model for screening teratogenicity*. Annual Meeting of the Society of Toxicology, Washington, D.C., March.
106. Frueh, F. W., and Gurwitz, D. (2004), From pharmacogenetics to personalized medicine: A vital need for educating health professionals and the community, *Pharmacogenomics*, 5, 571–579.
107. Ginsburg, G. S., and Angrist, M. (2006), The future may be closer than you think: A response from the Personalized Medicine Coalition to the Royal Society's report on personalized medicine, *Personalized Med.*, 3(2), 119–123.
108. Lesson, P. D., and Springthorpe, B. (2007), The influence of drug-like concepts on decision-making in medicinal chemistry, *Nature Rev. Drug Disc.*, 6(11), 881–890.
109. FDA (2007), FDA's Response to the Institute of Medicine's 2006 Report. The Future of Drug Safety—Promoting and Protecting the Health of the Public. U.S. Department of Health and Human Services Food and Drug Administration (FDA), January.
110. Anon. (2006), The Innovative Medicines Initiative (IMI) Strategic Research Agenda. Creating Biomedical R&D Leadership for Europe to Benefit Patients and Society; available at: [http://www.efpia.org/4\\_pos/SRA.pdg](http://www.efpia.org/4_pos/SRA.pdg).
111. Bridgland-Taylor, M. H., Hargreave, A. C., Easter, A., Orme, A., Harmer, A., Henthorn, D. C., Ding, M., Davis, A., Small, B. G., Heapy, C. G., Abi-Gerges, N., Paulsson, F., Jacobson, I., Schroeder, K., Neagle, B., Alberston, N., Hammond, T. G., Sullivan, M., Sullivan, E., Valentin, J-P., and Pollard, C. E. (2006), Optimisation and validation of a medium-throughput electrophysiology-based hERG assay using IonWorks™ HT, *J. Pharmacol. Toxicol. Methods*, 54, 189–199.
112. Report of CIOMS Working Group VI. Management of Safety Information from Clinical Trials, 2005.
113. Anon. (2005), CHMP Guideline on Risk Management Systems for Medicinal Products for Human Use, EMEA/CHMP/96268/2005.
114. Anon. (2004), Innovation or stagnation, challenge and opportunity on the critical path to new medicinal products, March 2004.





# 5.1

---

## *History of Clinical Trial Development and the Pharmaceutical Industry*

**Jeffrey Peppercorn,<sup>1</sup> Thomas G. Roberts, Jr.,<sup>2</sup> and Tim G. Hammond<sup>3</sup>**

<sup>1</sup>*Division of Medical Oncology, Duke University, Durham, North Carolina*

<sup>2</sup>*Noonday Asset Management, L.P., Charlotte, North Carolina*

<sup>3</sup>*Department of Safety Pharmacology, AstraZeneca, Macclesfield, Cheshire, UK*

### **Contents**

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| 5.1.1 Introduction                                                 | 115 |
| 5.1.2 From Heroic to Evidence-Based Medicine                       | 116 |
| 5.1.3 Role of FDA and Rise of Pharmaceutical Industry              | 122 |
| 5.1.4 Protection of Human Research Subjects and Birth of Bioethics | 126 |
| 5.1.5 Academia/Industry Collaboration                              | 128 |
| 5.1.6 Base Study: Early-Stage Oncology Trials                      | 129 |
| 5.1.7 Conclusion                                                   | 131 |
| References                                                         | 131 |

### **5.1.1 INTRODUCTION**

For over 2000 years medicine was practiced based on lessons passed down from teacher to student drawn from ancient beliefs regarding illness in the human body and the actions that must be taken to restore health. The principles of Hippocrates and the writings of Galen served as a guide for therapy. There is no doubt that observation of what worked and what did not work, and refinement of practice

based on experience, has also shaped medical practice from its earliest days, but for centuries, theory held sway over empiricism.

The situation changed with the systematic application of data collection and observation of outcomes and evolved into the modern practice of evidence-based medicine, driven primarily by the results of clinical trials. The history of clinical trials begins with the recognition of disease as a discrete condition that is similar enough from patient to patient to allow the rational application of the experimental method to determine which therapies work and which do not. It continues with recognition of the need to control for biases that can emerge in clinical experimentation resulting in the establishment of the gold standard of the double-blind randomized controlled trial. In addition, this history involves the recognition of the fact that using human subjects to gain scientific knowledge and help future patients requires regulation and procedures to ensure that the rights of these subjects as individuals are protected. Finally, this history involves the development of government agencies to support and oversee the development of clinical research and the emergence of the pharmaceutical industry as the dominant force in the development of new therapeutic agents. Each of these elements, and other aspects of this history, could be explored in a separate book, and each discipline from oncology to cardiology to psychiatry to surgery has unique stories and features that are beyond the scope of this chapter. We present here an overview of some of the key developments and ongoing issues that should serve as an introduction to the rich history of clinical trials and the rise of the pharmaceutical industry.

### 5.1.2 FROM HEROIC TO EVIDENCE-BASED MEDICINE

Medical practice from the time of Hippocrates (460–361 BC) was based on theory, and chief among these was the humoral theory that held that the human body was composed of four humors: yellow bile, black bile, phlegm, and blood. Disease was understood to arise from an imbalance of humors; the role of the physician was to bring the individual back into balance, by adding or removing substances from the body. In essence, this was the ultimate form of personalized medicine, where detection of a fever did not lead to a search for an underlying disorder but indicated an imbalance unique to the individual that needed to be corrected. Balance could be restored by vomiting, diarrhea, or sweating or taking in substances such as mercury or arsenic. One of the most dramatic means of restoring balance, and of demonstrating the knowledge and power of the clinician, was through bloodletting, or phlebotomy.

This practice, depicted on ancient Greek vases dating to the time of Hippocrates, and described extensively in the writings of Galen (AD 129–200), physician to Roman Emperor Marcus Aurelius, was a dominant form of medical therapy for over 2000 years until the mid-nineteenth century. Galen wrote several books on bloodletting and established the use of bloodletting as a form of heroic medicine, stating “the first and most important indications for phlebotomy are ... the severity of the disease and the strength of the patient” [1]. Patients with severe disease might be bled to the point of syncope in an effort to restore health [2]. Though this may seem counterproductive based on our current understanding of illness, there are conditions such as heart failure or kidney failure where, in the absence of diuresis or

dialysis, phlebotomy may have helped some patients. Further, the impetus to “heroic medicine” where severe disease in an otherwise healthy patient is treated with a dramatic, but unproven therapy, still exists in medicine and society as evidenced by the use of bone marrow transplant for breast cancer as recently as the 1990s [3].

The centrality of phlebotomy to practice in Galen’s time is indicated by his instructions to subject patients to bloodletting, not only when they were sick but also for disease prevention. He wrote in *On Treatment by Venesection*, “the time for phlebotomy is not only when severe disease is established, but also when it is likely to occur” [1]. Galen’s teachings were translated into Arabic and preserved in the Arabic world while Europe was in the Dark Ages. This practice was later adopted by European monks and continued throughout the Renaissance and into the nineteenth century.

One of the chief American proponents of bloodletting, and the practice of heroic medicine, was Benjamin Rush (1745–1813). In the face of the yellow fever epidemic of 1793, when thousands were dying, Rush advocated not just bloodletting but copious bloodletting, regardless of the state of the patient [4]. Rush was a notable physician in Philadelphia (in addition being signatory to the Declaration of Independence and treasurer of the U.S. Mint), and when the epidemic struck he tried a variety of therapies including purging with calomel (mercury), blistering, wrapping patients in vinegar-soaked blankets, and cinchona bark before deciding that more drastic measures were needed [4]. He reportedly saw over 100 patients a day and was both practitioner and promoter of bloodletting, continuing to advocate the practice for widespread use through yellow fever epidemics in 1794 and 1797 [5]. Rush called on his students to “Venerate the Lancet. It is the Magna gratia Coeli, The great gift of Heaven” [4]. Bloodletting was perhaps most famously used in American history in the treatment of President George Washington. Despite four heroic bloodlettings totaling up to 2 liters of blood in less than 24 hours (about 40% of total volume for an average male), performed by colleagues of Benjamin Rush, Washington succumbed to what was likely bacterial epiglottitis and died on December 14, 1799 [6].

Thus, principles and theories established over 2000 years earlier continued to provide the basis of medical practice; while knowledge of biology and human physiology advanced, the rational use of therapy did not. This unfavorable situation would begin to change with the conduct of clinical experiments and, importantly, the distribution of information regarding new therapeutic interventions and methods of validation. Progress toward the age of clinical trials required both the willingness to try new things (i.e., experiment) and the appreciation of the need for systematic observation, recording, and analysis of outcomes.

One of the first documented clinical experiments was the trial of inoculation as a treatment for smallpox performed by Cotton Mather during the epidemic of 1721 (Table 1). Inoculation, the practice of introducing a small amount of infectious material into a healthy subject in order to convey a mild case of disease that would prevent development of a more severe life-threatening form, was practiced in China as early as 1000 bc. However, it was not accepted practice in 1721 when Mather convinced Dr. Zabdiel Boylston to inoculate Bostonians against a smallpox epidemic, and they decided to collect data. Their report of 2% mortality among inoculated patients versus 14.9% among naturally infected patients, represents one of the first known instances of clinical data collection to guide future clinical practice [7].

**TABLE 1** Timeline of Select Events in History of Clinical Research

---

|      |                                                                                                  |
|------|--------------------------------------------------------------------------------------------------|
| 1721 | Mather smallpox inoculation in one of earliest documented clinical experiments                   |
| 1753 | Lind study of citrus for scurvy compares multiple concurrent treatments                          |
| 1798 | Jenner conducts smallpox vaccination trial                                                       |
| 1834 | Trousseau conducts first blinded placebo-controlled trials                                       |
| 1836 | Louis pioneers systematic evaluation of therapy and numerical method.                            |
| 1846 | Morton publicly demonstrates value of ether for anesthesia                                       |
| 1847 | Semmelweis studies role of hygiene in preventing maternal death with historical controls         |
| 1870 | Lister conducts single-arm study of antiseptics with historical control                          |
| 1896 | Fibiger conducts first randomized control trial, of diphtheria serum for treatment of diphtheria |
| 1938 | U.S. Food Drug and Cosmetics Act requires regulation of claims of new medications                |
| 1946 | Brittish MRC conducts first modern randomized control trial, evaluates streptomycin for TB       |
| 1949 | Nuremberg Code states importance of voluntary informed consent for research                      |
| 1954 | Salk polio vaccine trial, landmark large randomized placebo control trial                        |
| 1964 | Declaration of Helsinki stresses concern for the interests of research subjects                  |
| 1978 | Belmont report produces a practical guide to ethical research conduct                            |

---

In a testament to the slow dissemination of medical knowledge in the eighteenth century, the same study was essentially conducted by Edward Jenner in England in 1796. Jenner found that exposure to small amounts of cowpox could prevent both fulminant cowpox and the more lethal smallpox; he coined the term vaccination, based on *vaccinia*, Latin for cowpox [8].

It is possible to see the origin of clinical trials in these early clinical experiments. Though these efforts lacked the structure and methodology of anything resembling a modern clinical trial, the notion of studying or demonstrating the effectiveness of an intervention, rather than applying it based solely on theory, as was the case with phlebotomy, marked a profound shift in the approach to patient care. Other pioneering examples include the demonstration of the value of citrus (vitamin C) for scurvy by James Lind in 1747 [9], demonstration of ether (anesthesia) by William Morton in 1846 [10], and demonstration of the value of hygiene for prevention of childbed fever by Ignaz Phillip Semmelweis in 1847 [9].

Lind conducted a study among 12 Brittish sailors assigned to different dietary therapies for treatment of scurvy, which is estimated to have killed over 1 million people in the seventeenth and eighteenth centuries, and showed rapid recovery of sailors eating citrus. He published “Treatise on the Scurvy” in 1753 [9]. Morton demonstrated the value of ether to prevent operative pain to a room full of skeptical surgeons in the Massachusetts General Hospital Amphitheatre on October 16, 1846 [10]. His work was publicized by Henry J. Bigelow in the November 1846 *Boston Medical and Surgical Journal* [5]. Semmelweis conducted a natural experiment of sorts, demanding handwashing by physicians and medical students between work with cadavers and work on the delivery wards and demonstrated a drop in the rate of childbed fever and death from sepsis from 10% of mothers to less than 2% after imposition of this change in hygiene [5]. It is notable that widespread adoption of the practice of anesthesia for surgery and handwashing in obstetrics lagged many years behind these clinical demonstrations and publication highlighting the importance of an established method for demonstrating efficacy in medicine and the importance of community reliance on evidence to guide practice [9].

Though case series and public demonstrations are a form of clinical research, the true origin of systematic clinical research dates to the era of the Paris Clinical School

and the work of Pierre Charles Alexandre Louis. As noted above, bloodletting, whether by leech or lancet, was used widely though the nineteenth century based on the inherited wisdom for the time of Galen. The first evidence-based challenge to this practice emerged when Pierre Louis, a young physician at La Charite Hospital in Paris in the 1820s, decided to evaluate the impact of bloodletting compared to other treatments on outcomes in pneumonia. The very concept of such a study represents a shift in the understanding of disease. As historian of science Charles Rosenberg has noted in *The Therapeutic Revolution*, the movement toward recognition of diseases as discrete entities as opposed to personal states of imbalance was a necessary conceptual step for the study of the impact of therapeutic interventions on populations of patients, and the collection of data to guide therapy for future patients [11]. This subtle difference between therapies as means to restore balance versus therapies as targets for specific processes was critical for the development of evidence-based medicine. The development of the randomized clinical trial required both innovation in methodology and a transformation in this concept of disease. However, this step was accompanied by other changes in medicine, perhaps equally important for the development of clinical trials, such as movement of the place of illness from home to hospitals, allowing evaluation of large numbers of like cases, improvements in dissemination of knowledge, and a skepticism toward unproven therapies [11].

Pierre Louis evaluated 77 cases of pneumonia, stratified by whether they were bled on days 1–4 of illness or days 5–9. He found that 44% (18 of 41) of patients undergoing early bleeding died, compared to 25% (9 of 36) undergoing delayed bleeding [12]. That there was no comparison to a “no bleeding” control group speaks to the importance of bleeding as a therapy during this period. Describing what he called the “numerical method,” Louis wrote about the need to compare treatments among comparable numbers of patients to determine which treatment should be used in the future [13]. This was a radical suggestion at the time and contradicted the established practice of medicine based on the physician’s clinical judgment on a case-by-case basis. The move toward standardized evaluation, by no means robust in Louis’ time, now seems obvious, yet echoes of the debates over evidence-based medicine versus clinical judgment remain today.

Louis’ work was published in 1836, entitled *Researches on the Effects of Blood Letting in Some Inflammatory Diseases* [14]. His findings sufficiently challenged the dogma that he conceded: “The results of my researches . . . are so little in accordance with general opinion, that it is not without a degree of hesitation I have decided to publish them.” He published patient-specific data in detailed tables so that his work could be reviewed and analyzed by others. Even in the face of his own empirical evidence, he questioned whether a more aggressive “dosing” might have achieved better results. “Should we obtain more important results if, as is practiced in England, the first bleeding were carried to Syncope? The practice deserves a trial, but great success cannot, I think be anticipated; since many cases, the history of which I have drawn up, and which were fatal, were bled to a sufficient extent” [14]. In the face of this evidence, among the first well-documented clinical evaluations in which we can see the origins of modern clinical research [12], he reluctantly concluded that “bloodletting has had very little influence on the progress of pneumonitis, or erysipelas of the face, and angina tonsillaritis, in the cases under my observation” [14].

Though Pierre Louis relied on large numbers of cases, rather than randomization, he had hit upon both the importance of empirical observation to determine the safety and efficacy of therapy and the need to control for confounding factors among patients undergoing different interventions. Further, he argued not only that such evaluation was valid, but that “therapeutics cannot advance without it” [14]. Louis relied on advances in mathematics and probability theory to develop a basis for clinical science and called for evaluation of a magnitude of harms and benefits of any intervention to determine its worth [15].

Of note, bloodletting declined based on the work of Louis and others but persisted for many decades. A Philadelphia physician reported in 1862 that among over 9500 cases, only one was treated with general bloodletting, 12 with cupping, and 3 with leeching [11].

The work of Pierre Louis inspired the great American clinician Oliver Wendell Holmes who argued in 1843 that clinical evidence must guide practice [16]. Physicians from Great Britain and the United States took up the scientific method of Pierre Louis and began to collect outcome data on common practice, essentially conducting retrospective observational evaluation of common therapeutic practices [13]. In the absence of such evidence for most common medical practices of that day, this dawning awareness of the need for clinical research initially led to therapeutic nihilism [11]. However, it also produced the necessary environment for development of advances in care.

When Joseph Lister pioneered antiseptic surgical techniques in 1865 and studied their impact in subsequent years, he evaluated them based on clinical evidence by comparing outcomes from cases prior to the technique with those after introduction of the technique, essentially conducting a single-arm study and using a historical control [13, 17]. Similar advances in the management of infectious disease were made through testing of new therapies and comparison with historical controls in studies of diphtheria.

The problem with historical controls is that improvements in outcome may be due to the intervention, due to other changes in practice, or due to changes in the patient population that may occur over time. Therefore a pioneering advance in the history of clinical trials was the use of concurrent controls, first conducted in the trial of diphtheria serum by Danish physician Johannes Fibiger in 1896 [18].

Fibiger (1867–1928) alternately assigned new patients with diphtheria to standard therapy versus standard therapy plus diphtheria serum depending on the day of treatment between May 1896 and May 1897 [19]. The primary outcome was mortality, and secondary outcomes included croup and fever. He also evaluated toxicity, primarily development of serum sickness. Among 484 patients, 8/239 (3%) treated with serum died versus 30/245 (12%) controls [19]. Though no formal statistics were used, this trial demonstrated the importance of large numbers, randomization (future studies would improve over alternating days as a means of randomization), and concurrent controls. As Fibiger wrote: “In many cases a trustworthy verdict can only be reached when a large number of randomly selected patients are treated with the new remedy and at the same time, and equally large number of randomly selected patients are treated as usual” [18].

If Fibiger took the first steps toward pioneering the randomized control trial, the final steps in establishing this methodology were taken by the British Medical Research Council (MRC) in 1946 in what is typically termed the first true random-



ized control trial. Interestingly, at this juncture the history of clinical trials merges with the history of the pharmaceutical industry. The MRC conducted a trial of the promising new drug streptomycin for treatment of pulmonary tuberculosis, and one of the ethical rationales for randomization was that the drug, produced by the American pharmaceutical company Merck, was in short supply in Britain [20]. Widespread availability of the drug in America precluded a randomized trial, but in Britain, one of the only ways to obtain the drug was through participation in the MRC trial. The trial randomized 97 patients using random numbers in sealed envelopes to streptomycin versus control in a double-blinded fashion [20]. The trial demonstrated both the value of randomization and the value of streptomycin for tuberculosis. There was “considerable improvement” of chest X rays among 55% in the streptomycin group versus 8% of the controls, and 7% deaths in the streptomycin group versus 27% among controls [20].

Since the 1940s when the methodology of the randomized clinical trial was definitively established and embraced by the medical community, the pace of advancement in medical therapy has been remarkable. A clinical trial system for systematic assessment of novel interventions has been established, and government agencies, academic institutions, and private industry have been organized around the goal of discovering new treatments and testing them in a scientifically rigorous fashion. Other important steps along this pathway included the development of the trial system with phase I trials to test first in human interventions or combinations focused on safety, phase II studies to expand evaluation of safety in a select patient population and to evaluate efficacy, and phase III trials to provide a randomized controlled comparison between an experimental intervention and a standard therapy. There are of course variants of this trial structure, such as phase I/II trials, randomized phase II trials, pilot studies, phase IV trials (typically aftermarket trials to provide real-world safety and efficacy data), and emerging phase 0 trials (very small trials using low doses to study method of action) [21], and a full discussion of trial design is beyond the scope of this chapter. The conceptual steps described above paved the way for our modern system of clinical research.

Two other aspects of modern clinical trials that deserve note in this history, however, are the development of blinding and the placebo-controlled trial. Both of these are related to controlling for bias. When a research subject knows he or she is receiving an intervention that is supposed to work, it may affect subjective and even objective outcomes through psychological factors. Blinding means that the subject does not know what intervention he or she is receiving. Double blinding, used in some trials, means the physicians providing the intervention and assessing the outcome also do not know which intervention the subject is receiving. One common form of blinding involves use of a placebo. A placebo, based on the Latin word for “I shall please,” is an inert substance that is not expected to have any direct therapeutic value but that can be used in a trial to make a subject believe he or she is receiving a therapy and help sort out psychological effects of being treated from true physiologic effects. In modern clinical trials, this is used in randomized studies when there is no appropriate standard of care, or when you want to compare adding a novel intervention to a standard intervention to the standard intervention alone.

The history of using placebo controls, or “blinded assessment,” in clinical research dates to the late eighteenth century when blinded assessment was used to determine

if “mesmerism,” a then popular form of therapy in which the practitioner directed a sort of psychic force (characterized as a form of animal magnetism) at a patient to treat illness [22]. Researching the history of this subject, Kaptchuk relates that the first experiment actually used blindfolds to prevent the subject from knowing if mesmerism was really being applied, and it was conducted in 1784 in the house of Ben Franklin. Later, blind assessment was combined with the use of placebo in trials intended to study the validity of homeopathy, a practice of treating disease with minute amounts of a substance linked to the disease (on the principle of “like cures like”) [22]. These trials conducted in 1834, by the French physician Armand Trousseau (later famous for describing a syndrome of migratory blood clots associated with abdominal cancer), were the first blinded placebo-controlled trials, leading Trousseau to conclude that homeopathic remedies were no more active than placebo [22]. Use of placebo controls did not become a standard part of clinical research until later in the nineteenth century when they were used in a series of German studies examining the health effects of a variety of natural and nutritional remedies [22]. The use of randomization combined with blinding and placebo controls appears to have first emerged in the Michigan tuberculosis trials in 1926 in which a gold-based intravenous therapy was compared to intravenous water [22]. In 1954, this methodology was famously used to establish the safety and efficacy of the Salk polio vaccine, in a placebo-controlled trial involving almost 2 million subjects [23].

The degree to which medical therapy has evolved from reliance on theory, and distrust of empiricism, as recently as the time of Benjamin Rush, to reliance on evidence-based medicine and the randomized clinical trial, is exemplified by debate over whether treatment in a clinical trial itself now represents the standard of medical care in some settings [24, 25]. This view has been widely advocated in oncology, given both the poor outcomes with standard therapy for many diseases, and the clear advances demonstrated in clinical trials, particularly in treating childhood malignancy. The concept of a “trial effect” has been postulated, whereby treatment within a clinical trial conveys therapeutic benefit over and above the benefit of the experimental intervention itself [26]. Though it is not clear that such a trial effect holds up when comparisons of patients treated within clinical trials and patients treated outside of trials are subjected to the type of rigorous standards used to assess any intervention [27], it is notable that the clinical trial has emerged not only as a means to demonstrate the value of novel therapies, but as *the* recommended option for the care of some patients [24].

### 5.1.3 ROLE OF FDA AND RISE OF PHARMACEUTICAL INDUSTRY

No discussion of the history of clinical trials and drug development would be complete without mentioning the role of the Food and Drug Administration (FDA) and its analogous regulatory authorities throughout the world. Prior to the twentieth century, the U.S. government did little to regulate the marketing of therapeutic products; fraudulent claims were commonplace and went unpunished. Over the last 100 years, the FDA’s role has evolved to encompass three fundamental assurances: safety, efficacy, and adequate and accurate labeling [28].

Most observers consider 1906 to be the birth of the modern FDA [29]. In that year, Congress passed the original *Pure Food and Drugs Act*, which was signed into



law by Theodore Roosevelt. The statute prohibited misbranded and adulterated foods, drinks, and drugs in interstate commerce. Specifically, the law required companies to disclose weights and measures of their products and to provide labels disclosing whether their products contained alcohol, morphine, opium, cocaine, heroin, euaine, chloroform, cannabis, chloral hydrate, or acetanilide. Several forces provided the impetus for the law, including an exposé of the harmful preservatives used in the meat-packing industry as well as the realization of an increasing incidence of drug addiction from the use of “patent medicines.” Congress originally designated the Bureau of Chemistry in the Department of Agriculture, founded in 1862, to enforce the law; it would not be until 24 years later, in 1930, that the agency name would be changed to the FDA.

It is important to note that the original Pure Food and Drugs Act focused on ensuring accurate disclosure about product content, but the act did not prohibit firms from making false therapeutic claims as long as the firms provided accurate content labeling. Congress enacted the Sherley Amendment, in 1912, to address this loophole. The amendment specifically prohibited firms from the labeling of medicines with false therapeutic claims if the labels were intended to defraud the purchaser. That this legal standard was difficult to prove weakened the practical impact of the amendment. Radiothor, a radium-containing tonic that could be fatal with chronic ingestion, and Lash-Lure, an eyelash dash that blinded some women, are just two examples of worthless or harmful products widely marketed at the time. Efforts to enhance the laws stalled in Congress for years until the occurrence, in 1937, of the unfortunate Elixir Sulfanilamide disaster.

Sulfanilamide had been shown, by 1937, to have dramatic, curative activity against streptococcal infections. In an attempt to capitalize on a recognized demand for the agent in liquid form, especially among children, Tennessee-based S.E. Massengill Co. sought to manufacture an elixir formulation. Chemists at the company found that sulfanilamide would dissolve in diethylene glycol and that this solvent had an attractive mixture of fragrance and flavor. The company was not required to, nor did it chose to, carry out toxicity testing on the new formulation. Tragically, the highly toxic analog of antifreeze killed more than 100 people, many of whom were children. In the wake of the tragedy, Congress and the president moved quickly to enact the *Food, Drug, and Cosmetic Act*. The new law required, for the first time, that drugs be safe for their intended use. Specifically, the law required premarket approval for all new drugs; firms had to demonstrate proof that their drugs were safe prior to their marketing. Furthermore, the law gave the FDA authority to identify drugs that would require a prescription from a physician.

It was not until 1962, almost 25 years after the original Food, Drug, and Cosmetic Act, that Congress amended the 1938 act by codifying a requirement that drugs demonstrate adequate evidence of efficacy in addition to safety. Congress indicated that the FDA should require “full reports of investigations which have been made to show whether or not such drug is safe for use and whether such drug is effective in use” [30]. The new law was in many ways the result of hearings held by Senator Estes Kefauver. The senator had worked for years to try to reform the agency, but his efforts languished until the thalidomide tragedy, which played out mainly in Europe, provided the catalyst for Congressional action. Congress also gave the FDA control over the regulation of drug development trials, including the requirement

for informed consent, the regulation of drug advertising, and the power to establish and enforce good manufacturing practices. As part of the law Congress required the FDA to assess the efficacy of all drugs introduced since 1938.

Over the next 40 years, Congress enacted several additional laws governing the FDA, including the Orphan Drug Act and the Federal Advisory Committee Act. Multiple sets of drug regulations were also developed during the latter half of the twentieth century to refine and interpret the laws. The 1990s, however, proved to be the next highly consequential period for the FDA's evolution. The AIDS crisis in the late 1980s and early 1990s provided the major impetus for reform. Activists targeted the FDA for failing to approve drugs quickly enough in the face of the crisis. Congress responded over the period of 1992–1997 by enacting two laws, the Prescription Drug User Fee Act of 1992 (PDUFA) and the FDA Modernization Act of 1997. Collectively, these laws have had a profound impact on the FDA's regulation of drugs intended to treat serious or life-threatening illnesses. Programs such as Fast Track Designation, Priority Review, and Accelerated Approval have each worked to expedite the development of drugs intended to treat serious illnesses such as cancer.

Of these three programs, the accelerated approval mechanism has had the largest impact [28]. The provisions under accelerate approval permit the FDA to approve agents to treat serious or life-threatening illnesses before the clinical benefit necessary to meet the standard for regular approval has been demonstrated. Specifically, the FDA can grant initial drug approval on the basis of a surrogate measure of clinical benefit (e.g., CD4 count or tumor shrinkage) if the treatment is intended to treat a serious or life-threatening illness and is reasonably likely to be superior to available therapies [31]. The FDA in turn receives agreement from the drug's sponsor to complete confirmatory trials in the postapproval period. If the postapproval trials fail to show clinical benefit, the FDA has a mechanism to remove the drug from the market in a timely manner. If history is any guide, the regulatory evolution of the FDA will continue into its next 100 years. Recent efforts have focused on enhancing the agency's ability to monitor the safety of drugs and devices in the postapproval period [32]. The agency has also become more active in collaborating with sponsors, academe, and clinical societies in order to improve the return on the nation's public and private investment in research and development. The FDA's Critical Path project is one such promising area of regulatory research. The program includes efforts to provide a structure for the inclusion of response or toxicity biomarkers into the drug development process, an area that is certain to receive additional attention in the coming years [33].

Amidst this regulatory framework, and in concert with the evolution toward evidence-based medicine, the commercial pharmaceutical industry has risen to become the dominant source in the development of new therapeutics. From the development of aspirin in 1897 to today's molecularly targeted cancer therapies, most diseases now have at least some symptomatic or curative medicinal treatments. In addition to development of the basic science that paved the way for rational development of new drugs, and the clinical science that created the foundation to determine their safety and efficacy, two pieces of legislation played a major role in paving the way for the modern pharmaceutical industry. Both pieces of legislation arose from a perceived need to provide a greater economic incentive for the development of new drugs. The Orphan Drug Act, passed in 1983, provided 7 years of

exclusive marketing and large tax credits to cover research and development costs for any drug designed to treat a rare disease. Similarly, the Drug Price Competition and Patent Term Restoration Act, passed in 1984, provided industry with extended patent protection to cover the significant time required for research and development and allowed generic versions of drugs to be produced and marketed without repeated expenditure for clinical trials of the generic formulation. Though other factors likely contributed, the passage of these laws in the early 1980s corresponded with an exponential rise in expenditure on research and development by the pharmaceutical industry.

In 2006, research and development (R&D) spending for the biopharmaceutical industry alone was as high as \$55.2 billion [34]. Member companies of the Pharmaceutical Research and Manufacturers of America (PhRMA) spent approximately \$2 billion on R&D in 1980; over the last quarter century, they have grown their R&D expenditures at an astonishing compound annual growth rate of more than 12%. This growth in private R&D spending has dramatically outstripped the growth in spending by the National Institutes of Health (NIH). In 2006, the NIH budget amounted to \$28.6 billion, a little more than half of the amount spent by biopharmaceutical companies [35]. This growth reflects underlying secular changes in the process of drug development. Once the purview of a few fine chemical companies located in the Upper Rhine Valley in Switzerland and in the greater Philadelphia area, the effort to discover and develop new drugs has transformed into a commercially focused, capital-intensive enterprise fraught with risk and uncertainty and producing many more failures than successes. In fact, fewer than 1 in 10 drugs entering clinical trials will ever gain marketing approval from the FDA [36], and at least 1000 molecules are screened for each that enters clinical testing [34]. For some diseases such as cancer, the ratio of winners to losers is even more unfavorable [37]. Of the thousand of agents in some stage of clinical or preclinical development, as few as 20 new molecular entities will reach the FDA regulatory standard for approval. The high failure rate and demanding regulatory standard together drive the cost of new drug development, which by some estimates exceeds \$800 million for each successful approval [38].

The effort to manage the investment in drug development has led to the formation of another thriving entity—the contract research organization (CRO). With more than 1000 CROs in operation and with industrywide revenue exceeding \$17 billion, for-profit CROs have, in large part, taken over academia's historic role in organizing new drug development. Clinical researchers from the academe remain crucial to some aspects of development, but CROs now often take the lead in identifying study centers, recruiting patients, acquiring and monitoring data, and even performing relevant statistical analyses. As of December 2007, there were more than 11,000 clinical trials actively recruiting patients for a treatment clinical trial according to [clinicaltrials.gov](http://clinicaltrials.gov). The work required to plan and execute all of these trials has overwhelmed the resources in both the academe and in the pharmaceutical industry, especially as the pharmaceutical industry completes substantial downsizing efforts. A series of high-profile incidents involving research subjects sustaining harm along with potential conflicts of interest have drawn criticism over the industry's increasing reliance on CROs. It is possible that at some time in the future, CROs will be subject to additional regulatory scrutiny and action; for now, their role in clinical drug development appears secure.

#### 5.1.4 PROTECTION OF HUMAN RESEARCH SUBJECTS AND BIRTH OF BIOETHICS

The history of clinical trials and the pharmaceutical industry is closely tied to the history of bioethics, the concerns of which now shape the regulation of clinical research and pharmaceutical development. Though medical ethics traces its routes back at least to the teachings of Hippocrates and the dictum “*primum non nocere*” or “first do no harm,” the rise of modern bioethics began in the twentieth century with the reaction to the atrocities of Nazi science. It is important to understand that the origin of modern bioethics as it concerns research ethics (the ethics of clinical research) emerged from a reaction to scandal, rather than an *a priori* guide to how to conduct clinical science. There is a strong and continued emphasis on protection of human subjects from research that at times appears to come into conflict with the need to recruit patients to clinical trials to advance care for future patients. By recognizing the origins of the concerns of human subjects’ protections, and the continued need for emphasis on respect of research subjects as persons, it is possible to conduct clinical research efficiently, while securing the ethical foundation of such science.

Nazi science was characterized by use of human beings as research subjects without their consent, to explore the limits of human endurance, the effects of exposure to extremes and a variety of toxins, and characterized by complete disregard for the well-being of the subject [39]. The crimes of Nazi science were publicly exposed at the Nuremburg Medical Trial of 1946–1947 in which 23 German doctors were tried for experiments performed on victims in concentration camps from 1933 to 1945 [40]. The Nuremburg Code, published in 1949, was developed to not only ensure that such atrocities were never repeated but to advance the ethical conduct of all studies involving human subjects. The first principle established by Nuremburg was the requirement for voluntary informed consent on the part of the research subject [41]. This principle remains at the heart of human subjects protection and is a core component of all further codes of ethical research, though how to define and obtain meaningful informed consent remains a subject of debate. Additional principles related to the need for clinical research to be conducted in the interest of humanity, based on good preclinical science, designed to minimize harm to subjects, with consideration of the welfare of subjects and their right to decline participation ensured throughout the research [41]. The Nuremburg Code was a starting point, but further guidance was needed to establish broadly based guidelines for research ethics.

In the wake of involvement by physicians in Nazi experiments, the World Medical Association (WMA) established the Committee on Medical Ethics in 1952. Based on a series of meetings and debates, the WMA published the Declaration of Helsinki in 1964 [42]. The declaration emphasized the rights of research subjects to be informed of the potential risks involved in research and of the voluntary nature of research participation. This document recognized a distinction between research with therapeutic intent that might benefit the subject and research with scientific goals only, but emphasized that under all circumstances “the interest of science or society could never take precedence over considerations related to the well-being of the subject” [42]. Ethicists have debated the practicality of imposing such a principle on research that by definition involves unknown risks to subjects, but the

principle of respect for the rights and well-being of the subject that the Declaration establishes is clear.

No history of clinical trials would be complete without consideration of the significance of the Tuskegee syphilis trial. This study was initiated in the 1930s by the U.S. Public Health Service to evaluate the impact of untreated syphilis in a cohort of African American men. Though there were few known effective therapies for syphilis at the time of initiation of the study, the deadly natural history of the disease had already been demonstrated in Scandinavian subjects [43]. In the context of that time, this question of the course of untreated syphilis in black patients clearly arose out of racism. Further, the subjects were tricked into participation in an observational study under the false premise that they would be receiving therapy. For example, subjects undergoing a lumbar puncture for research purposes were solicited to receive a “special treatment” from the nurse [43]. Though this study was initiated prior to the Nuremburg Code and Declaration of Helsinki, it was continued by the U.S. Public Health Service and published in major journals until 1972, when the scandal was exposed by the press [43].

Failure to consider the well-being of research subjects was also not exclusive to racist research. In 1966, H. Beecher, a professor of anesthesiology at Harvard Medical School wrote a landmark exposé of 22 clinical studies supported by top academic, government, and industrial institutions, and reported in leading medical journals, that violated the rights of the participants, frequently through failure to obtain informed consent [44]. In response to Beecher’s publication, the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was established, and the report of the commission, the Belmont Report was released in 1978 [45]. The first principle, respect for persons, called for respect of subjects’ autonomy and the need for special protections for those with decreased autonomy (e.g., in children, subjects with neuropsychiatric disorders, or imprisonment). The second principle, beneficence, called for proactively taking steps to promote the well-being of research subjects, in addition to seeking to avoid harm and respecting their wishes. Finally, the Belmont Report recognized the principle of justice, which called for research to be conducted among populations who could benefit from the research. The report emphasizes that all clinical research must be clearly distinguished for routine clinical practice, that voluntary informed consent is required, and that research must be carefully designed to maximize potential benefits and minimize potential harms to the research participants [45].

Many guidelines and national and international codes of research ethics have been developed and promoted since the publication of the Nuremburg Code. The core principles in 13 major codes and declarations were evaluated and analyzed recently by Emmanuel and colleagues; they elucidated seven ethical principles that should guide all clinical research [46]. These principles of value, scientific validity, fair subject selection, favorable risk–benefit ratio, independent review, informed consent, and respect for potential and enrolled subjects are proposed as universal requirements of clinical research, grounded in the underlying principle of respect for persons [46].

The need for regulation of clinical research should not obscure the fact that, for many interventions, testing of safety and efficacy within a clinical trial may be the most ethical way to provide a novel intervention to a patient for whom outcomes with standard therapy are inadequate. A prime example was high-dose chemother-

apy and bone marrow transplantation for breast cancer. Early-phase single-arm studies among patients with advanced disease demonstrated results that appeared substantially better than those seen among historical controls treated without transplant [47]. In fact, advocates for breast cancer patients and some clinicians questioned the ethics of conducting randomized trials for a “proven therapy” [48]. In a tawdry portion of the history of clinical research, the story was complicated by the falsification of data by a prominent researcher, Bezwoda, who presented fabricated randomized clinical trial data at major scientific meetings and in publication [49], further promoting the use of this intervention prior to confirmation in ongoing randomized trials, and in fact, likely slowing accrual to those trials [3]. In the end, well-designed randomized control trials demonstrated no improvement of outcomes for women with advanced breast cancer from high-dose chemotherapy and transplant compared to less aggressive and toxic forms of therapy [50].

### 5.1.5 ACADEMIA–INDUSTRY COLLABORATION

In addition to general concerns over human subjects protection in clinical research, there has been recent increased attention to the role of the pharmaceutical industry in clinical research. In brief, this concern stems from the fact that the pharmaceutical industry now plays the dominant role as the sponsor of clinical research and the fact that most elements of the pharmaceutical industry are for-profit. As noted above, in 1992, pharmaceutical industry investment in research exceeded the operating budget of the NIH for the first time. Increased funding appears to have translated into increased sponsorship of published clinical research. An increase in documented pharmaceutical sponsorship of clinical research over time has been noted in stroke trials [51], oncology trials [52, 53] and in randomly selected trials from five medical journals [54]. The majority of studies in many areas of medical research are now supported in some way by the pharmaceutical industry.

One interesting aspect of the association between pharmaceutical sponsorship and clinical research has been the observation that pharmaceutical industry involvement correlates with publication of positive clinical trials. This association was first noted by Davidson et al. in 1986 [55]. Davidson’s initial observation was supported by a systematic review reported in 2003 by Bekelman et al. [56]. Pooled analysis of 37 studies, which collectively included 1140 clinical trials, demonstrated that industry sponsorship correlated with positive study outcomes or “pro-industry” conclusions with an odds ratio of 3.60 (95% confidence interval, 2.63–4.91). The reason for this association is unclear, but potential explanations include biased trial design, or interpretation of results, superior or safer selection of agents to take forward into later phase trials, or failure to publish negative studies [52]. These possibilities support the movement toward clinical trial registries and further evaluation of this association.

Concerns over potential bias in research sponsored by the pharmaceutical industry have heightened the tension over closer academia–industry collaborations. It is not surprising that as the pharmaceutical industry became the dominant funding source for clinical trials and the academic medical centers remain the dominant site for clinical research that academia and industry would by necessity become partners. Several studies have addressed conflicts of interest in academia–industry relationships. In 2000, Boyd and Bero reported that 7.6% of faculty investigators reported



financial ties with industry with the percent of involvement growing over time, likely representing an underestimate based on variable and voluntary reporting requirements [57]. In an effort to manage potential bias and conflicts of interest in academia–industry collaboration, the International Committee of Medical Journal Editors issues guidelines for clinical research, but as of 2002, it was not clear that these guidelines were widely implemented, leaving this an ongoing area of interest and evaluation [58].

Ultimately, well-regulated collaboration between academic centers and the pharmaceutical industry should continue to yield advances in therapy to the betterment of society. In fact, the model of academia–industry collaboration in oncology stands out as an example of what can be achieved when the resources and efficiencies of industry partner with the intellectual resources of the academic world, and increasingly with clinical care in both academic and community-based research centers.

### 5.1.6 CASE STUDY: EARLY-STAGE ONCOLOGY TRIALS

Over the past two decades, the growth of clinical research in oncology has exceeded the growth in all other areas of medicine. A great deal of scrutiny of the cancer-related research infrastructure has accompanied this growth, providing an excellent case study on which to highlight many of the issues raised in this chapter. As of early December 2007, there were 6214 clinical trials actively recruiting patients with cancer that were registered with the National Institutes of Health (NIH; [clinicaltrials.gov](http://clinicaltrials.gov)). Most (80%) of these trials were early-phase, or developmental trials (phases I and II), with 1653 phase I trials and 3323 phase II trials. The number of drugs and biologics in clinical trials for the treatment of cancer is now greater than the combined total of the next two most represented therapeutic classes, anti-infectives and immunologics [36]. Cancer drug development has transformed from a low-budget, government-sponsored enterprise to a high-stakes multi-billion-dollar industry with hundreds of biotech and pharmaceutical companies seeking approval and adoption of their products [59].

The structure regulating clinical trials is complex, with multiple agencies and groups sharing responsibility and oversight. At the federal level, the FDA regulates industry-sponsored research, while the Office of Human Research Protection (OHRP) oversees human research sponsored by the Department of Health and Human Services (HHS). In the wake of highly publicized lapses in the oversight of clinical research and instances of questionable ethics of some investigators, some of which led to tragic outcomes, the OHRP was recently elevated from the NIH to the Office of the HHS Secretary. In addition, individual or centralized institutional review boards (IRBs), ethics committees, clinical investigators, and sponsors all add layers of oversight to clinical cancer trials to ensure that they are performed safely and ethically.

Phase I trials represent the first testing of an experimental agent in humans, acting as a point of translation of years of preclinical work into the clinic [37]. As already outlined, the major objectives during phase I are to characterize the agent's toxicity profile and to determine a schedule and dose appropriate for further testing. Phase I cancer trials differ from phase I trials in other areas of medicine in two important ways. First, phase I trials in other areas of medicine typically enroll

healthy participants, whereas phase I trials in oncology almost always enroll patients who have cancer and who have exhausted standard treatments. Second, investigators and patients seek to realize therapeutic benefit in phase I cancer trials, usually as a secondary endpoint. Unlike phase I trials in other areas of medicine, treating physicians almost always enroll patients in phase I cancer trials with therapeutic intent, and patients often expect to benefit [60, 61].

The ethical basis of phase I cancer trials has been questioned, in part because they involve potentially vulnerable cancer patients near the end of life [62]. Some ethicists have raised concern that patients who chose to participate may experience significant risks with little chance to benefit. Others have pointed out that patients who participate have unrealistic expectations about the probability of benefit or the goals of research, despite having gone through the informed consent process [63]. Still others have been concerned that cancer patients may have judgment that is clouded by their illness and therefore cannot make truly voluntary decisions.

Much of the debate has now evolved to focusing on the estimated risk–benefit ratio in phase I trials. Over the 5-year period from 1986 to 1991, three groups published response rates in meta-analyses of phase I clinical trials [64–66]. Rates of objective response (usually as defined by tumor shrinkage by greater than 50%) ranged from 4 to 6%. Toxic death rates were reported to be around 0.5%. A major limitation to the relevance of these studies is that they are outdated and do not include assessment of the risks and benefits associated with phase I trials of the newer targeted agents. Researchers at Harvard recently reported trends in the risks and benefits to patients participating in phase I clinical trials submitted for presentation at the American Society of Clinical Oncology over the period of 1991–2002 [37]. The overall respective response and toxic death rates of 3.8 and 0.54% were similar to those published in prior meta-analyses. What was striking, however, were the trends over time. The overall toxic death rate for 213 published studies from this sample decreased over time, from 1.1% over the first 4 years of the study (1991–1994) to 0.06% over the most recent 4-year period (1999–2002). After adjusting for characteristics of the investigational agents and the experimental trials, the odds of patients dying from an experimental treatment fell by more than 90% from 1991 to 2002. Interestingly, the odds of a patient dying from a targeted or biologic treatment were four times lower than the odds of dying from a traditional cytotoxic agent. Rates of objective response also declined over time but by proportionally much less.

There has been no explicitly articulated standard about what determines a socially acceptable risk–benefit ratio [67]. Agrawal and Emanuel recommended comparing risk–benefit ratios in phase I cancer trials to socially accepted determinations already used for cancer such as FDA approval standards. By this construct, the risk–benefit ratio of many phase I cancer trials may not be clearly worse than those used by the FDA in its approval of anticancer drugs or by medical oncologists in their treatment decisions. In the Harvard study [37], some phase I trials produced as high or higher rates of response than those used to support FDA approvals. FDA approvals for irinotecan in colon cancer and topotecan in ovarian cancer were based on phase II response rates of 10–15%. In comparison, the response rates among trials in the top decile of phase I trials analyzed by the Harvard group exceeded 13%. Response rates in phase I trials are also comparable to those in the third-line or greater treatment of many solid tumors. Making these considerations more complex is the



understanding that patients confronting death may have higher tolerances for risk than agencies overseeing their clinical research. George Zimmer, an English professor and phase I trial participant, wrote that “the enemy is not pain or even death, which will come for us in any eventuality. The enemy is cancer, and we want it defeated and destroyed. This is how I wanted to die—not a suicide and not passively accepting, but eagerly in the struggle” [68].

### 5.1.7 CONCLUSION

Clinical trials have evolved from a time when empiricism was viewed with derision by those who believed medical therapy should be based on theory and a priori understanding of health and disease to the foundation of evidence-based medical practice. Recognition of diseases as discrete entities, rather than personal states of imbalance, transition of health care from home to institutions such as hospitals, and the development of methods for valid comparisons of therapy and control of bias were all critical factors in development of the modern clinical trial. Clinical trials have been instrumental both in preventing the use of ineffective treatments from bloodletting, to bone marrow transplant, to breast cancer and in providing a platform for development of novel therapeutics that have revolutionized medicine and improved outcomes for patients. The progress in clinical trials has been built on the efforts of many diligent investigators and countless patients who have become research subjects. Along the way, we have learned to design better trials and to take better care of patients participating in these trials. Institutions and industry have emerged to regulate and promote clinical research and the development of complex relationships between government, industry, clinical research organization, academia, and community practices to advance medical knowledge safely and efficiently is ongoing. The optimal design of clinical trials remains contested in many areas, ethical questions regarding the quality and requirements of informed consent abound, and debates regarding the balance of evidence-based practice versus clinical wisdom continue.

We are entering an era of targeted therapy based on understanding of the molecular biology of disease and the ability to rationally design drugs for selected targets. In addition, we are finding that understanding the discrete disease entity alone is not sufficient for optimal treatment, and the specifics of the individual patient, from co-morbidities, physiologic status, and pharmacogenomic differences can play a major role in the outcome of any given therapy. Now more than ever before, clinical trials have the potential to transform medicine and improve health, and the dedication of researchers, and most importantly patients who are willing to participate in clinical research are needed to write the next chapter in the history of medicine.

### REFERENCES

1. Brain, P. (1986), *Galen on Bloodletting*, Cambridge University Press, Cambridge, United Kingdom.
2. Haller, J. S., Jr. (1986), Decline of bloodletting: A study in 19th-century ratiocinations, *South Med. J.*, 79(4), 469–475.

3. Antman, K. H., Rowlings, P. A., Vaughan, W. P., et al. (1997), High-dose chemotherapy with autologous hematopoietic stem-cell support for breast cancer in North America, *J. Clin. Oncol.*, 15(5), 1870–1879.
4. Kopperman, P. (2004), “Venerate the Lancet”: Benjamin Rush’s yellow fever therapy in context, *Bull. Hist. Med.*, 78, 539–574.
5. Garrison, F. (1929), *History of Medicine*, 4th ed., W.B. Saunders, Philadelphia.
6. Morens, D. M. (1999), Death of a president, *N. Engl. J. Med.*, 341(24), 1845–1849.
7. Best, M., Neuhauser, D., and Slavin, L. (2004), “Cotton Mather, you dog, dam you! I’ll inoculate you with this; with a pox to you”: smallpox inoculation, Boston, 1721, *Qual. Saf. Health Care*, 13(1), 82–83.
8. Gross, C. P., and Sepkowitz, K. A. (1998), The myth of the medical breakthrough: Smallpox, vaccination, and Jenner reconsidered, *Int. J. Infect. Dis.*, 3(1), 54–60.
9. Bender, G. (1966), *Great Moments in Medicine*, Northwood Institute Press, Detroit.
10. Campagna, J. A. (2005), The end of religious fatalism: Boston as the venue for the demonstration of ether for the intentional relief of pain, *Surgery*, 138(1), 46–55.
11. Vogel, M., and Rosenberg, C. (1979), *The Therapeutic Revolution*, University of Pennsylvania Press, Philadelphia.
12. Best, M., and Neuhauser, D. (2005), Pierre Charles Alexandre Louis: Master of the spirit of mathematical clinical science, *Qual. Saf. Health Care*, 14(6), 462–464.
13. Bull, J. P. (1959), The historical development of clinical therapeutic trials, *J. Chronic Dis.*, 10, 218–248.
14. Louis, P. (1836), *Rechearches on the Effects of Bloodletting in Some Inflammatory Diseases*, C.G. Putnam, Boston.
15. Pernick, M. (1983), The calculus of suffering in 19th-century surgery, *Hastings Center Report*, 13.
16. Holmes, O. W. (1843), The contagiousness of puerperal fever. *N. Engl. Quart. J. Med. Surg.*, 1, 503–530.
17. Tan, S. Y., and Tasaki, A. (2007), Joseph Lister (1827–1912): Father of antisepsis, *Singapore Med. J.*, 48(7), 605–606.
18. Fibiger, J. (1898), Om Sreumbehandling af Difteri, *Hospitalstdende*, 4(6). In Hrobjartsson, A., Gotzsche, P. C., and Gluud, C. (1998), The controlled clinical trial turns 100 years: Fibiger’s trial of serum treatment of diphtheria, *BMJ*, 317(7167), 1244.
19. Hrobjartsson, A., Gotzsche, P. C., and Gluud, C. (1998), The controlled clinical trial turns 100 years: Fibiger’s trial of serum treatment of diphtheria, *BMJ*, 317(7167), 1243–1245.
20. Yoshioka, A. (1998), Use of randomisation in the Medical Research Council’s clinical trial of streptomycin in pulmonary tuberculosis in the 1940s, *BMJ*, 317(7167), 1220–1223.
21. Anon. (2006), Drive for drugs leads to baby clinical trials, *Nature*, 440(7083), 406–407.
22. Kaptchuk, T. (1998), *Bull. Hist. Med.*, 72(3), 389–433.
23. Lambert, S. M., and Markel, H. (2000), Making history: Thomas Francis, Jr, MD, and the 1954 Salk Poliomyelitis Vaccine Field Trial, *Arch. Pediatr. Adolesc. Med.*, 154(5), 512–517.
24. Gelber, R. D., and Goldhirsch, A. (1988), Can a clinical trial be the treatment of choice for patients with cancer? *J. Natl. Cancer Inst.*, 80(12), 886–887.
25. Antman, K., Schnipper, L. E., and Frei, E., 3rd. (1988), The crisis in clinical cancer research. Third-party insurance and investigational therapy, *N. Engl. J. Med.*, 319(1), 46–48.
26. Braunholtz, D. A., Edwards, S. J., and Lilford, R. J. (2001), Are randomized clinical trials good for us (in the short term)? Evidence for a “trial effect,” *J. Clin. Epidemiol.*, 54(3), 217–224.

27. Peppercorn, J. M., Weeks, J. C., Cook, E. F., et al. (2004), Comparison of outcomes in cancer patients treated within and outside clinical trials: Conceptual framework and structured review, *Lancet*, 363(9405), 263–270.
28. Roberts, T. (2006), Food and Drug Administration role in oncology product development, in Chabner, B. A., ed., *Cancer Chemotherapy and Biotherapy: Principles and Practice*, 4th ed., Lippincott Williams & Wilkins, Philadelphia, pp. 502–515.
29. Chabner, B. A., and Roberts, T. G. (2007), The FDA in 2006: Reasons for optimism, *Oncologist*, 12(3), 247–249.
30. Drug Amendments of 1962, Pub. L. No. 87–781, 76 Stat. 780 (1962), codified as amended at 21 U.S.C. §321.
31. Johnson, J. R., Williams, G., and Pazdur, R. (2003), End points and United States Food and Drug Administration approval of oncology drugs, *J. Clin. Oncol.*, 21(7), 1404–1411.
32. Hennessy, S., and Strom, B. L. (2007), PDUFA reauthorization—drug safety’s golden moment of opportunity? *N. Engl. J. Med.*, 356(17), 1703–1704.
33. Woosley, R. L., and Cossman, J. (2007), Drug development and the FDA’s Critical Path Initiative, *Clin. Pharmacol. Ther.*, 81(1), 129–133.
34. PRaMoA (2007), *Pharmaceutical Industry Profile 2007*, PhRMA, Washington, DC.
35. Loscalzo, J. (2006), The NIH budget and the future of biomedical research, *N. Engl. J. Med.*, 354(16), 1665–1667.
36. Mathieu, M. (2006), *Parexel’s Pharmaceutical R&D Statistical Sourcebook 2006/2007*, Parexel International Corporation, Waltham, MA.
37. Roberts, T. G. Jr, Goulart, B. H., Squitieri, L., et al. (2004), Trends in the risks and benefits to patients with cancer participating in phase 1 clinical trials, *JAMA*, 292(17), 2130–2140.
38. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22(2), 151–185.
39. Annas, G., and Grodin, M. (1992), *The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation*, Oxford University Press, New York.
40. Leaning, J. (1996), War crimes and medical science, *BMJ*, 313(7070), 1413–1415.
41. Anon. (1949), The Nuremberg Code, in *Trials of War Criminals before the Nuernberg Military Tribunals under Control Council Law No. 10*, U.S. Government Printing Office, Washington DC, pp. 181–182.
42. World Medical Association (1996), Declaration of Helsinki (1964), *BMJ*, 313, 1448–1449.
43. Brandt, A. M. (1978), Racism and research: The case of the Tuskegee Syphilis Study, *Hastings Cent. Rep.*, 8(6), 21–29.
44. Beecher, H. K. (1966), Ethics and clinical research, *N. Engl. J. Med.*, 274(24), 1354–1360.
45. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1978), The Belmont Report: Appendix. Vol 1, U.S. Government Printing Office, Washington, DC, Chap. 9.
46. Emanuel, E. J., Wendler, D., and Grady, C. (2000), What makes clinical research ethical? *JAMA*, 283(20), 2701–2711.
47. Canellos, G. P. (1997), Selection bias in trials of transplantation for metastatic breast cancer: Have we picked the apple before it was ripe? *J. Clin. Oncol.*, 15(10), 3169–3170.
48. Mello, M. M., and Brennan, T. A. (2001), The controversy over high-dose chemotherapy with autologous bone marrow transplant for breast cancer, *Health Aff. (Millwood)*, 20(5), 101–117.

49. Weiss, R. B., Rifkin, R. M., Stewart, F. M., et al. (2000), High-dose chemotherapy for high-risk primary breast cancer: An on-site review of the Bezwoda study, *Lancet*, 355(9208), 999–1003.
50. Farquhar, C., Marjoribanks, J., Bassier, R., et al. (2005), High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy for women with metastatic breast cancer, *Cochrane Database Syst. Rev.*, 3, CD003142.
51. Dorman, P. J., Counsell, C., and Sandercock, P. (1999), Reports of randomized trials in acute stroke, 1955 to 1995. What proportions were commercially sponsored? *Stroke*, 30(10), 1995–1998.
52. Peppercorn, J., Blood, E., Winer, E., et al. (2007), Association between pharmaceutical involvement and outcomes in breast cancer clinical trials, *Cancer*, 109(7), 1239–1246.
53. Djulbegovic, B., Lacey, M., Cantor, A., et al. (2000), The uncertainty principle and industry-sponsored research, *Lancet*, 356(9230), 635–638.
54. Buchkowsky, S. S., and Jewesson, P. J. (2004), Industry sponsorship and authorship of clinical trials over 20 years, *Ann. Pharmacother.*, 38(4), 579–585.
55. Davidson, R. A. (1986), Source of funding and outcome of clinical trials, *J. Gen. Intern. Med.*, 1(3), 155–158.
56. Bekelman, J. E., Li, Y., and Gross, C. P. (2003), Scope and impact of financial conflicts of interest in biomedical research: A systematic review, *JAMA*, 289(4), 454–465.
57. Boyd, E. A., and Bero, L. A. (2000), Assessing faculty financial relationships with industry: A case study, *JAMA*, 284(17), 2209–2214.
58. Schulman, K. A., Seils, D. M., Timbie, J. W., et al. (2002), A national survey of provisions in clinical-trial agreements between medical schools and industry sponsors, *N. Engl. J. Med.*, 347(17), 1335–1341.
59. Chabner, B. A., and Roberts, T. G., Jr (2005), Timeline: Chemotherapy and the war on cancer, *Natl. Rev. Cancer*, 5(1), 65–72.
60. Meropol, N. J., Weinfurt, K. P., Burnett, C. B., et al. (2003), Perceptions of patients and physicians regarding phase I cancer clinical trials: Implications for physician-patient communication, *J. Clin. Oncol.*, 21(13), 2589–2596.
61. Daugherty, C., Ratain, M. J., Grochowski, E., et al. (1995), Perceptions of cancer patients and their physicians involved in phase I trials, *J. Clin. Oncol.*, 13(5), 1062–1072.
62. Miller, M. (2000), Phase I cancer trials. A collusion of misunderstanding, *Hastings Cent. Rep.*, 30(4), 34–43.
63. Henderson, G. E., Churchill, L. R., Davis, A. M., et al. (2007), Clinical trials and medical care: Defining the therapeutic misconception, *PLoS Med.*, 4(11), e324.
64. Von Hoff, D. D., and Turner, J. (1991), Response rates, duration of response, and dose response effects in phase I studies of antineoplastics, *Invest. New Drugs*, 9(1), 115–122.
65. Decoster, G., Stein, G., and Holdener, E. E. (1990), Responses and toxic deaths in phase I clinical trials, *Ann. Oncol.*, 1(3), 175–181.
66. Estey, E., Hoth, D., Simon, R., et al. (1986), Therapeutic response in phase I trials of antineoplastic agents, *Cancer Treat. Rep.*, 70(9), 1105–1115.
67. Agrawal, M., and Emanuel, E. J. (2003), Ethics of phase 1 oncology studies: Reexamining the arguments and data, *JAMA*, 290(8), 1075–1082.
68. Daugherty, C. K., Siegler, M., Ratain, M. J., et al. (1997), Learning from our patients: One participant's impact on clinical trial research and informed consent, *Ann. Intern. Med.*, 126(11), 892–897.

# 5.2

---

## *Adaptive Research*

**Michael Rosenberg**

*Health Decisions, Inc., Durham, North Carolina*

### **Contents**

|         |                                                         |     |
|---------|---------------------------------------------------------|-----|
| 5.2.1   | Types of Adaptive Techniques: Strategic and Operational | 138 |
| 5.2.2   | Strategic Adaptations                                   | 140 |
| 5.2.2.1 | Drug Performance and Rising Dose Escalation Studies     | 140 |
| 5.2.2.2 | Adaptive Dose Finding                                   | 142 |
| 5.2.2.3 | Sample Size Reestimation                                | 144 |
| 5.2.2.4 | Adaptive Randomization                                  | 145 |
| 5.2.2.5 | Seamless Designs: Rolling One Phase into Next           | 147 |
| 5.2.2.6 | Other Strategic Adaptations                             | 148 |
| 5.2.3   | Operational Side of Adaptive Research                   | 148 |
| 5.2.3.1 | Enrollment                                              | 150 |
| 5.2.3.2 | Site Performance                                        | 150 |
| 5.2.3.3 | Adaptive Site Monitoring                                | 151 |
| 5.2.4   | Implementation Issues                                   | 153 |
| 5.2.4.1 | Data Capture and Validation                             | 153 |
| 5.2.4.2 | Planning                                                | 154 |
| 5.2.4.3 | Process Optimization                                    | 155 |
| 5.2.4.4 | Decision Making                                         | 156 |
| 5.2.4.5 | Adaptive Data Monitoring                                | 156 |
| 5.2.4.6 | Regulatory Considerations                               | 156 |
| 5.2.5   | Promise of Adaptive Methods                             | 158 |
|         | References                                              | 159 |

Adaptive research denotes an approach to clinical trials that incorporates what is learned during the course of a study or development program into how it is completed, without compromising validity or integrity. Adaptive components need

not be confined to the frequently encountered but unduly narrow vision of enabling changes in a study's design, valuable and interesting as such changes are. Rather, adaptive methods may encompass potential changes in all program-related resources and activities, including changes in logistical, monitoring, and recruitment procedures, and sometimes even personnel and travel requirements. The goal of adaptive methods is making better and more timely decisions to allocate all study resources more efficiently, reduce costs and timelines, and better achieve informational goals compared to traditional study and program approaches.

Efficient management is particularly important in activities as complex as clinical research, which involves a range of activities that includes patient recruitment, randomization, supply chain logistics, and flow of information. Additional complexity commonly arises when pharmaceutical studies are conducted at multiple sites, often in different countries, cultures, and languages. Effective management of clinical trials requires continuous monitoring and measurement of numerous activities. The essence of adaptive research is to continuously measure progress in the many aspects of a complex study, learn from such measures, and, based on what is learned, act expeditiously to make changes to improve the remainder of the study and even an entire development program.

On a pragmatic level, adaptive studies not only require the ability to measure outcomes of interest continuously but also to make data and summarized information about those measurements available in a timely manner to different audiences according to study role. This is essential for effective study management. In a clinical context, this means not just continuously tracking trial data collected on case report forms but also generating performance metrics that enable refinements in operations. This learn-as-you-go approach contrasts with the traditional black-box methodology of clinical trials in which data and particularly operational indices are often lacking altogether or available too late to enable study personnel to respond: Clean data are generally not available until after a study is completed, and study performance metrics such as recruitment rate, reasons for screen failures, and the like are often lacking entirely.

Interest in adaptive methods has mounted as a result of the soaring cost of clinical research and numerous trial failures, including particularly costly and well-publicized failures of major late-stage trials. In addition to greater efficiency, adaptive methods provide a number of appealing advantages, such as a more nuanced view of product performance that may enable earlier strategic decisions about appropriate target populations, earlier warnings about ineffective trials, and a broader view of research programs as continuous, integrated activities rather than a staccato, linear series of separate trials with inevitable delays in between.

Adaptive methods seek to bring "gold-standard" trial methodology up to date by taking advantage of the many technological advances during the 60 years since the trials that established comparative clinical trials as a means of assessing a pharmaceutical product's performance in the late 1940s [1]. Particularly salient are markedly more efficient approaches to data capture and validation, the growing power of affordable desktop computers, and the ability of the Internet to move data throughout the world easily, quickly, and cheaply.

Advances in data capture and validation open possibilities for much earlier understanding of study performance and trends. Computational advances stimulated the development of statistical methodology enabling midcourse "looks" at study progress based on data collected to date. The most common example is

sequential analyses, which preserve design integrity but inform decisions that improve the course of the study. The simplest result of such an interim analysis is early stopping for futility. Although statistical methodology is beyond the scope of this chapter, both books and software have become available that focus on adaptive research. See, for example, Chow and Chang [2] and Hu and Rosenberger [3].

An additional benefit of adaptive approaches is enabling a more nuanced perspective into candidate performance, following the realization over time that a simple “yes–no” answer as to a drug’s efficacy is likely an oversimplification. Adaptive trials maintain the same high standards of scientific integrity and reliability as the standard methodology while dramatically improving operational and economic efficiency and informational breadth, depth, and quality.

Adaptive research also allows clinical researchers to employ the same basic management principles as typical modern businesses, using real-time data and analysis to inform decisions that continually optimize operations. Figure 1 contrasts the continuous acquisition of knowledge in adaptive studies with a conventional trial that does not generate clean, meaningful data until much later. Adaptive methods require continuously updated operational performance metrics; the conventional approach often lacks the real-time data essential to make such metrics available to improve trial management.

Preserving study integrity—the ability to perform unbiased clinical evaluations—is paramount with all clinical evaluations, including adaptive ones. While commonplace in other industries, managing in response to changing data is fairly new in clinical research because excluding bias has been achieved primarily by denying access to data, including much of the data that would be useful for trial management. Like conventional trials, adaptive research still relies on techniques to exclude bias, including blinding. However, adaptive trials also employ additional planning and special operational procedures that prevent those performing the study from accessing unblinded results data. Only designated individuals have access to the information required to make decisions about specific adjustments during the trial. Firewalls

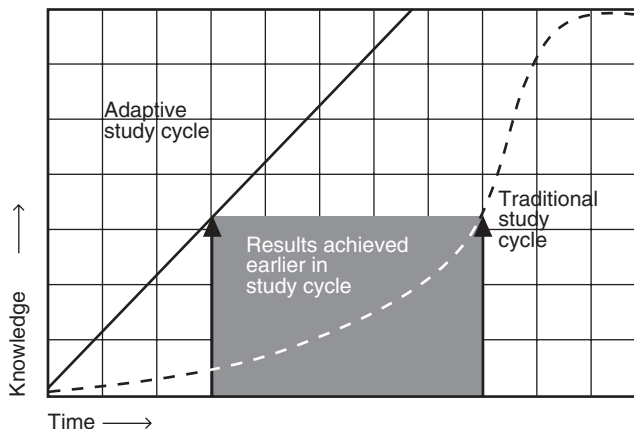


**FIGURE 1** With the traditional approach (dashed line), most knowledge is acquired at the end of the trial. An adaptive approach uses newer techniques to capture and analyze data continuously, to support both scientific inferences and performance measures—knowledge that managers can use to support decisions to tightly manage the trial. The difference between the two approaches represents the ignorance handicapping the managers of a traditional trial. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)



must be incorporated from the outset to ensure that decision makers cannot jeopardize, whether knowingly or not, the study's scientific integrity. Fortunately, sophisticated computer access control and data encryption techniques provide useful tools for controlling the dissemination of information and potential sources of bias.

Developing protocols for adaptive trials demands more attention than conventional planning because multiple scenarios must be considered and specific plans included for addressing each. For example, a study that involves midstudy elimination of dosing arms (pruning, a common technique in dose-finding trials) normally includes safeguards to ensure that investigators are blinded as to which dose is being eliminated. Maintaining the blind may require logistic changes such as different packaging of study supplies. Additional examples of measures to exclude bias are included in Section 5.2.4.

### 5.2.1 TYPES OF ADAPTIVE TECHNIQUES: STRATEGIC AND OPERATIONAL

Any study involves both a strategic design and operational plans. Adaptive techniques similarly fall into the two broad categories of strategic adaptations and operational components. Strategic adaptations refer to changes in the study's design, such as the number of subjects to be included or how patients are to be allocated to treatment arms during the study. Operational adaptations focus on how the study is run: recruitment, enrolling patients, improving data quality (measured by number of queries generated), assuring timely responses to data discrepancies, detecting site performance problems early, and efficient means of allocating resources such as field monitors. Although much recent discussion focuses on the strategic components of adaptive research, the operational components can be at least as beneficial to timelines and budgets. Moreover, operational adaptations have the advantage of not requiring regulatory approval. Operational adaptive methods enable far tighter trial management through continuous monitoring and refinement of the many elements involved, resulting in savings in time and expenses of 20% or more compared with traditional management techniques. In the light of ever increasing study budgets, the savings from operational adaptations have strong appeal.

There are several prerequisites for both strategic and operational components of adaptive research: the continuous, timely collection of data and generation of meta-data (data *about* the data, such as performance metrics that might include rates of enrollment, screen failures, queries, and the like); the ability to rapidly turn a stream of raw data into meaningful information; and the ability to present information in different forms to meet the needs of study staff performing different functional roles. All these are essential to enabling specific informed actions to improve trial operations. Indeed, at many points during a trial, the measures of how the study is progressing may demand greater attention than the actual data collected as part of the clinical evaluation; the need for trial data may be intermittent (e.g., at an interim analysis), while the need for metadata to inform management decisions is continuous.

Efficient electronic data collection (EDC) is an absolute requirement for adaptive research. Unfortunately, not all EDC is efficient in practice. EDC for adaptive purposes must be able to capture data in electronic form shortly after it is generated and quickly transfer it to a central database. Many EDC systems currently used in

the pharmaceutical industry rely on hand keyboard entry (“Web-based” EDC systems), which result in unnecessary delays in data acquisition. Such delays typically range from several days to several weeks, reflecting the lack of enthusiasm among clinical personnel for performing the tedious chore of keyboard data entry. A second and more important shortcoming of many EDC systems is their inability to track metadata, or the metrics that enable effective study management. EDC systems that fail to provide an integrated means of addressing important dynamic components of adaptive trials such as supply management and patient randomization can induce as many problems as they solve. The fact that Web-based EDC systems have existed for more than 7 years and yet the majority of clinical trials continue to be done with paper and pen reflects the industry’s experience that Web-based EDC is expensive, difficult to implement and maintain, and often fails to deliver bottom-line benefit. Alternatives to Web-based EDC for adaptive trials are included in Section 5.2.4.

While substantial improvements in efficiency may be gained through strategic adaptive methods, the optimal approach to clinical trials incorporates both strategic and operational elements. Combining both elements promises to go beyond improving the mechanics of individual studies to improve how studies and development programs are managed. Like the navigators of old who planned long voyages based on inaccurate maps and poor or nonexistent navigational equipment, planners of clinical trials today must start out with plans based on guesses. In clinical trials, the guesses are about key trial parameters such as the size of the treatment effect, variability of data, dropout rates, and even the appropriate range of dosages to test. Modern drug companies, like the early explorers, have often learned too late that their best planning efforts have landed them not at the intended destination with rich economic rewards but in unforeseen surroundings with bleak prospects. Adaptive techniques represent a GPS (Global Positioning System) for present-day clinical trials: We must still make initial guesses, but the capacity to make multiple midcourse corrections, on many levels, is an integral part of the study. Rather than relying on a series of discrete decision points, an adaptive approach substitutes a process of continuous assessment and response. As a result, study staff is no longer condemned to making guesses and then sticking with them until the study is done, only then discovering how each guess compares to reality.

Clinical development is evolving from the traditional model of discrete phases punctuated by pauses to a continuous process. The traditional model is one of defining safety and kinetics (phase I), then defining optimal doses (phase II), then pivotal studies required for regulatory approval (phase III). The emerging integrated model is largely made possible by an adaptive approach that renders the process continuous, with faster decision making along the way, the flexibility to shorten (or lengthen) certain components in response to data generated, and minimizing or eliminating the gap between studies. Indeed, this so-called learn–confirm model can extend beyond development to postmarketing, maximizing learning about a product throughout its life cycle, minimizing postmarket problems, and maximizing opportunities.

The following sections that discuss specific adaptive techniques are arranged according to the chronology of development, starting with the earliest clinical studies in a product’s life cycle, efforts to define the safety envelope, and the balance between efficacy and safety. We then progress through larger studies that focus on

better definition, selection of dosing for large-scale confirmatory (pivotal) studies, and finally pivotal studies. An important difference in how study design is approached, especially in regard to adaptive techniques, between earlier studies to define safety and dosing and the pivotal studies that serve as the basis for marketing approval, is that regulatory bodies grant a greater degree of autonomy in the early (learning) evaluations (subject, of course, to the broad constraint of not jeopardizing the safety of subjects involved in evaluations). Thus, the learning phase currently offers greater opportunity to apply adaptive approaches. The use of adaptive methods in the learning phase allows earlier, more accurate determination of promising and less promising products, reduced development time and expense, and reduced wasting of resources on candidates that eventually fail to reach market.

## 5.2.2 STRATEGIC ADAPTATIONS

### 5.2.2.1 Drug Performance and Rising Dose Escalation Studies

The earliest phases of development involve confirmation of pharmacokinetic (PK) and pharmacodynamic (PD) assessments from animal models and determination of approximate dosing ranges for human administration. Such studies are usually performed in “normal” (disease-free) populations and may, to a limited degree, explore these parameters in individuals for whom the product is ultimately targeted. The keys for these early studies are rapid acquisition and assessment of data and timely decision making, minimizing the interval between when a dose is administered and when a decision can be made with respect to the next dosing.

For PK and PD studies as well as early studies that rely on these parameters (e.g., dose escalation), assessments are normally batched to reduce the expense involved in running samples. However, the balance between the reduced costs of batched assessments—always a choke point—and the potential economic gains from earlier availability of assessment results, may warrant a reexamination of traditional reliance on batching samples. Each assay and product under evaluation differs, but researchers should at least consider the alternative of obtaining data from individual tests earlier.

Rising-dose escalation studies illustrate the critical nature of rapid, accurate data in achieving high velocity in clinical studies and programs. The interval between dosing, assessment, and a decision about progressing to the next higher dose often consumes 6 weeks. Adaptive techniques, with their focus on the rapid collection and assessment of data, as well as rapid dissemination of information for appropriate action, may reduce this interval to days or even a single day. This can be accomplished by (1) exploiting the availability of data very quickly after drug administration and collection of safety data, (2) using a system that efficiently collects and summarizes data, and (3) disseminating and displaying the information via a universal mechanism such as the Web.

Figure 2 shows a system that meets these requirements and enables next-dose decisions to be made within 2 days of the last dose; this rate-limiting step is dictated by a 48-hour primary observation period after administration. The requirement for rapid data availability mandates use of a digital pen (Fig. 3) for data collection along with processes and software that enable the data for each individual to be

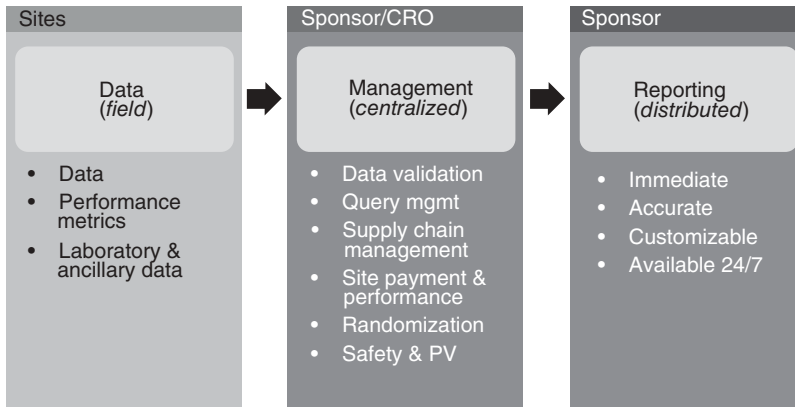


**FIGURE 2** Immediate, clean data requires the ability to collect, digest, and report a variety of data and performance metrics. While data collection systems are common, these may take extended periods to enter data. Most systems currently in use lack the “middleware” (central box) necessary for rapid data collection and validation as well as the ability to report data rapidly. In addition, few systems routinely generate performance metrics, such as indicators of recruitment progress and issues, essential to running a tight study that minimizes timelines. (Figure copyright 2007, Health Decisions, Inc. Used by permission.)



**FIGURE 3** A digital pen used for data capture. This method electronically reads in data, eliminating the requirement for manual data entry. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)

displayed within minutes of when it is collected. Summaries that show data on all subjects are displayed over the Web, enabling timely access by decision makers regardless of location or time.

The benefits of this system are significant. First, the usual bottleneck the delays getting clean data is eliminated, making each subject and the collective experience of all subjects immediately available to appropriate members of the study team. A stream of raw data is transformed into meaningful information, summarized with drill-down capabilities, and presented in a manner that supports decisions about next steps. Second, decisions are more timely, nuanced and better informed because accumulated information can be tracked and analyzed along the way, rather than

waiting until the end to examine all data. Third, the ability to make this information immediately available to all decision makers regardless of location drives timelier and more efficient study management across all offices and sites associated with a study. Such an approach can easily cut four or more weeks from each decision cycle. With an early-phase study, this can translate to a reduction of 50% or more in study timeline.

### 5.2.2.2 Adaptive Dose Finding

Conventional dose-finding studies typically utilize a modest number of equal-sized treatment arms, generally three to four, often with a comparator arm. Each arm is administered a different dose for a given period, data are examined and, hopefully, justify a decision to proceed with the most promising doses into pivotal testing (the “confirm” part of the learn–confirm model). Planning for this approach involves selecting the most promising doses, then bracketing them; in effect, this builds on earlier dose escalation studies that gave a sense of the same information but with fewer subjects.

Dose-finding studies are conducted with the goal of winnowing dosing choices to a small enough number to allow proceeding to larger, more expensive pivotal trials. Compared to earlier studies, these involve more subjects receiving each dose, yielding more data, enabling greater understanding and greater certainty about efficacy and safety. Such studies are often planned (“powered”) with enough subjects to achieve statistical differentiation between dosing arms for the efficacy outcome. Although reassuring to decision makers, this approach may not be necessary to achieve the objective of rejecting some of the dosing arms. Adaptive techniques make possible earlier decisions to reject less effective or less safe doses, thus conserving resources and time. In addition, the ability to examine data as it accumulates can lead to superior understanding of the data and contribute to better decisions about doses to carry forward.

Under an adaptive scenario, information is made available as data are generated and go/no-go decisions can be made not at some predetermined interval but whenever the information generated yields sufficient knowledge to justify the step. In many instances, less desirable treatment arms (outliers) will be apparent early on. A key benefit of the adaptive approach is the ability to accumulate adequate information early in the trial to justify eliminating less promising arms, allowing the concentration of resources (notably study subject) to arms that show greatest promise. In such early learning-phase trials, regulators generally allow sponsors the latitude to forego a rigidly defined decision point in favor of making decisions on dosing arms when the sponsors are comfortable that sufficient information has accumulated (sponsor’s risk). Adaptive methods provide an opportunity for sponsors to make earlier, better informed judgments without pushing every arm to a sample size that provides statistically significant results. Rather, there is the flexibility to discontinue a treatment arm as soon as it becomes apparent that it is less desirable than others.

The adaptive approach changes the way dose-finding evaluations are performed. Importantly, since some arms can likely be eliminated early, there is the luxury of starting with a larger number of arms than would be possible with the traditional approach. For example, in a study of a neuroprotective agent administered a single time by IV (intravenous) soon after stroke, researchers established a procedure for

selecting 1 of 16 possible doses to be given each patient, with the dose in each instance selected based on data observed to that point in the trial. In the actual conduct of the trial, 15 different doses were tried [4].

The ability to test more arms at this stage means that earlier development often need not be as thorough. And because arms are cut off early, there is little or no extra expense in gathering data on a greater number of dosing arms and then focusing resources where they are most needed—in differentiating the final two or three dosing arms rather than in accumulating additional information about arms that data already show lack promise. Another benefit of earlier termination of less promising arms is that fewer patients are exposed to less efficacious and less safe doses. In addition, the ability to examine data as they are generated—where each new patient, visit, and evaluation adds to an existing storehouse of information, and trends may emerge from the changing data—often provides a far more nuanced perspective on product performance than a single cross-sectional view at the conclusion of an evaluation.

Figure 4 illustrates how an adaptive dose-finding study is conducted. If the informational goal is defined by 80 patients, then we begin with a number of dosing arms, in this case 8. Early on, data will likely make it apparent that many of these arms

| Time (mo)      | 2  | 4    | 6    | 8    | 10   | 12   | 14 | 16               | TOTAL             |
|----------------|----|------|------|------|------|------|----|------------------|-------------------|
|                |    |      |      |      |      |      |    | cost/pt \$15,000 |                   |
| ADAPTIVE       |    |      |      |      |      |      |    |                  |                   |
|                | 10 |      |      |      |      |      |    |                  | 10                |
| 50/mo          | 10 | 12.5 | 12.5 | 12.5 |      |      |    |                  | 48                |
|                | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 |    |                  | 81                |
|                | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 |    |                  | 81                |
| comparator     | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 |    |                  | 81                |
| total enrolled | 50 | 50   | 50   | 50   | 50   | 50   |    |                  | 300               |
| #/arm/mo       | 10 | 12.5 | 12.5 | 12.5 | 16.7 | 16.7 |    |                  |                   |
| cost           |    |      |      |      |      |      |    |                  | \$4,500,000       |
| NON-ADAPTIVE   |    |      |      |      |      |      |    |                  |                   |
|                | 10 | 10   | 10   | 10   | 10   | 10   | 10 | 10               | 80                |
| 50/mo          | 10 | 10   | 10   | 10   | 10   | 10   | 10 | 10               | 80                |
|                | 10 | 10   | 10   | 10   | 10   | 10   | 10 | 10               | 80                |
|                | 10 | 10   | 10   | 10   | 10   | 10   | 10 | 10               | 80                |
|                | 10 | 10   | 10   | 10   | 10   | 10   | 10 | 10               | 80                |
| comparator     | 10 | 10   | 10   | 10   | 10   | 10   | 10 | 10               | 80                |
| total enrolled | 50 | 50   | 50   | 50   | 50   | 50   | 50 | 50               | 400               |
| #/arm/mo       | 10 | 10   | 10   | 10   | 10   | 10   | 10 | 10               |                   |
|                |    |      |      |      |      |      |    |                  | TOTAL \$6,000,000 |

**FIGURE 4** Comparison of adaptive (top) and traditional dose-finding studies. Dropping less promising arms early on (pruning) allowed this study to achieve its informational goals four months sooner, at a cost \$1.5 million less, by monitoring incoming data from the study's outset and focusing resources on the most promising arms. After the first observation period, one arm (T3) is terminated because of poor results for efficacy, safety or both. The remaining arms are continued to the next observation period, at which time an additional arm (T2) is similarly cut. The remaining dose arm and the comparator (P) may under certain circumstances then be rolled into the pivotal evaluation, utilizing data already gathered. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)



are less promising and can be cut off earlier than is customary. Although Figure 3 shows discrete, evenly spaced decision points, these are in practice more likely to be irregular, with the most obvious decisions coming early in the study. The result of this approach in this example shows how the desired number of patients is reached earlier and at a lesser cost than would be the case if we had taken a traditional approach of using, say, four dosing arms that enroll 80 patients each. If each patient costs \$15,000, the adaptive study will be completed 25% faster (saving 4 months) and cheaper (saving \$1.5 million). The cost savings are even greater in the light of the institutional costs of maintaining a company for an extra 4 months, a factor that is particularly compelling for developing companies where such institutional costs can be a key determinant of success and, in some cases, survival.

### 5.2.2.3 Sample Size Reestimation

The progression of a compound into large-scale testing demands major commitments of time and money. A key driver of both study duration and cost is the number of patients required for successful completion of the study. The appropriate sample size is determined based on the magnitude of difference in outcome measure between the product being evaluated and a comparator, acceptable levels of possible error (risk of false-positive or false-negative results), variance of data, and rates of subject compliance and drop-out. With luck, initial estimates of these factors may be off only modestly, but they will still be off. If the estimates err more than modestly, the consequences for sample size requirements can be enormous. For example, reducing the estimate of treatment effect by one-half can quadruple the sample size for a fixed sample test and will require the maximum sample size for a group sequential test [5].

Underestimation of sample size may result in a study that fails to detect a difference between the test drug and the comparator, even if one is present; overestimation of required sample size wastes time, money, and other resources. The severe penalty for underestimation means that, in practice, sample size estimates err on the high side, with the consequence that time and money will often be spent unnecessarily in overcompensating for the possibility of falling short of the sample size necessary for detecting a difference. It is therefore not surprising that much research is devoted to approaches to adjusting sample size during clinical trials and a variety of methods are in use [6].

The benefit of an adaptive approach is that it enables an informational goal to be met precisely, arriving at a defined goal of informational content rather than by surrogate measures, neither undershooting nor overshooting. In this context, reestimation means adjusting each of the initially estimated parameters based on data that has actually been observed in the study rather than continuing to rely on the earlier guesses made without benefit of such observations. Reestimation can be employed multiple times for multiple course corrections, although careful attention must be paid to statistical techniques to ensure that design integrity is preserved. Sample size reestimation can be viewed as an extension of well-accepted group sequential trial designs that have evolved over the past two decades.

The value of sample size reestimation is illustrated by a recent oncology study, where this technique allowed the study to be completed 9 months sooner than originally planned. An interim analysis demonstrated an effect size ( $\delta$ ) considerably



stronger than originally anticipated. The savings from the resulting sample size reduction also extended beyond recruitment itself, eliminating the cost of treatment, supplies, monitoring, and follow-up for each patient that might have been included in excess of the number required to meet the study's informational goals. The oncology study's use of sample size reestimation contributed to a savings of \$16 million in development costs. Even greater financial benefits flowed from a 9-month reduction in time to market.

#### 5.2.2.4 Adaptive Randomization

Adaptive randomization can provide greater efficiency by altering the probability that a new subject entering the study will be allocated to a treatment arm that accumulating trial data shows to be more desirable and by ensuring that the desired statistical power is preserved [3]. Early on, when little is known about patient response to different treatments, a subject will be equally likely to be randomized into any treatment arm; during the course of the study, as outcome information accumulates, the randomization ratio can be continuously changed to favor the more beneficial or safe outcome (response-adaptive randomization), to balance covariates (risk factors that modify the probability of an outcome) across different treatment arms (covariate-adaptive randomization), or to balance an undesired deviation from the intended allocation ratio (treatment-adaptive randomization) [2]. There are also sophisticated combinations, such as covariate-adjusted response-adaptive randomization, a procedure that takes into account previous patient responses to treatments, previous patient covariates, and the covariates of the patient to be randomized [3].

In contrast, normal fixed allocation schedules ensure that each patient entering the study will have the same probability of allocation to each arm throughout the life of the study regardless of whether study data shows some arms to be unpromising by reason of lesser efficacy or safety. Continuing with fixed allocation ratios despite evidence of lesser efficacy or safety raises ethical concerns. When data shows an imbalance in covariates between treatment groups, continuing with a fixed allocation procedure can undermine the ability to draw valid inferences about differences in treatment effect. Moreover, continuing with a fixed allocation in the unfortunate event that an imbalance develops in the size of actual treatment groups can undermine statistical power and thus jeopardize the validity of the trial as a whole.

With response-adaptive randomization, allocation ratios are most commonly changed based on favorable outcome (assuming safety not to be an issue). There are different algorithms for achieving this goal. The most common is the randomized "play-the-winner" scheme, which assumes that the outcome for the previous patient is known before the next patient is assigned a treatment group. If the treatment for the preceding patient has a positive outcome (a success), an additional "ball" representing that patient's treatment group is added to the randomization pool; if the preceding patient has a negative outcome, no "ball" is added to the urn. If randomization were based on a bowl of black and white Ping-Pong balls, the study would begin with the same number of white and black balls. Over time, however, if one treatment is more beneficial than others, the randomization pool would progressively be seeded in favor of that treatment [2].

Response-adaptive randomization offers the additional advantage of collecting more data on patient response to the drug under test in the doses that are likely to reach market. This may result in greater understanding of the drug's behavior in the formulation that physicians will prescribe and patients will receive after approval. Physicians and patients will both benefit from superior prescribing information. When deciding whether to approve the drug, regulators may also benefit from the availability of more and better information on the doses that will be prescribed in the event of approval.

Covariate-adaptive randomization seeks to balance covariates across treatment groups by weighting the randomization procedure to increase the number of patients with certain covariates in the treatment group or groups in which these covariates have turned out to be underrepresented [2, 3]. (In practice, covariates can also be accounted for through mathematical techniques such as multivariable analysis.)

Treatment-adaptive randomization balances the number of patients assigned to each treatment group by using any of several weighting schemes based on adjusting the number of hypothetical balls in an urn in favor of the treatment group with lagging membership or creating the algorithmic equivalent of a coin that is biased in favor of that treatment group [2].

Adaptive randomization can be combined with a Bayesian approach to study design. This allows patients to be randomized to a specific arm in direct proportion to the degree of promise it shows relative to other arms. Thus, as experience accumulates and successful treatments are retained while unsuccessful ones dropped, the allocation ratio is progressively slanted in favor of the successful treatment. When a certain balance has been achieved, the study is declared to be completed. For example, if a study begins with a 1:1 randomization scheme but then adds each successful outcome back to the randomization pool over time and drops the unsuccessful outcomes (failures), the randomization pool becomes successively seeded in favor of the successful arm. A Bayesian approach allows a study to be defined as completed when a predetermined proportion is reached, say, 95% of patients are randomized into a particular arm, reflecting the experience of far greater success with that arm. The Bayesian approach is intuitively appealing in that it incorporates each new piece of information on success or failure of treatment into how the study proceeds from that point onwards. The Bayesian approach is also more broadly appealing in that it provides a way to minimize patient exposure to less successful arms on a continuous basis. In practice, however, Bayesian approaches may be complicated by the complex definitions of success and the requirement to consider safety, which is difficult to measure on a dichotomous scale.

The mechanics of adaptive randomization and a Bayesian approach require a centralized, real-time randomization system in addition to the other prerequisites for conducting adaptive studies, such as efficient data capture and cleaning, previously discussed. Besides supporting adaptive randomization, centralization and the immediacy of the electronic process also provide the ability to stop enrolling patients immediately when the target population size is reached, eliminating unnecessary effort and expense that are inevitable in less centralized systems. Even more important, the ability to stop patient enrollment instantly reduces the exposure of additional patients to less desirable treatments.

### 5.2.2.5 Seamless Designs: Rolling One Phase into Next

With growing appreciation of drug development as a continuous process, the notion of defining best doses and then rolling straight into pivotal studies utilizing the identified best doses becomes compelling. The strongest reasons for doing so are (1) minimizing the delay, often 12 months or more, between dose-finding studies and initiation of pivotal studies; (2) the efficiencies gained from going through the trial startup process a single time rather than twice; (3) a head start recruiting investigators and test subjects for the pivotal study; (4) the potential to adapt the population of the confirmatory phase based on data on responsive subgroups from the learning phase; and (5) the ability to combine data from the learning phase and the confirmatory phase in conducting a final analysis of trial data.

However, it is important to recognize as well that the break between phases can be an important time to analyze data and perfect the final design of a separate pivotal study. Study managers should weigh the possible need for an interval to allow greater analysis and more refined planning against the potentially huge savings in time and expense if phases can be combined. Additionally, managers should weigh the need to take advantage of an interim period for discussions with regulatory authorities to assure their acceptance of the plan as a basis for approval in the event of a successfully executed study.

A large study that is progressively refined from the phase II dose-finding stage and then continues into the phase III confirmatory study is among the most complex adaptive strategies but also the most rewarding. Executing this strategy is demanding because it requires a great deal of advance planning to anticipate and deal with different possible outcomes. In addition, the final statistical analysis may be extraordinarily complex.

The use of adaptive methods to combine phase II and phase III studies begins with establishing a number of dosing arms and pruning those down to the manageable two to three (including a comparator) for a confirmatory stage. Once the final doses have been identified, the study is then quickly expanded and, assuming a second pivotal study is warranted, it is implemented immediately (Fig. 3).

Simulations can be extremely useful in modeling possible outcomes and their ramifications. Adopting the strategy of rolling a phase II study into phase III requires confidence that all the methodologies and infrastructure essential for the adaptive approach are in place and functioning well—quick and accurate data capture, rapid data validation, the prompt generation of meaningful information, readiness for continuous decision making, and the capacity to manage logistics of the study, such as supply chain management, with great efficiency based on timely data. The benefits of rolling a phase II study into phase III are commensurate with the effort: This approach can easily reduce development time by a year or more and save many millions of development dollars.

Eliminating the gap between phases is only one of the benefits of rolling directly from a phase II study into phase III, conducting, in effect, one continuous study instead of two separate ones. This has the added advantage of incorporating in the phase III study the knowledge gained from relevant arms in the early dose-finding portions. There are also benefits for patient recruitment and data collection. Patients already screened in the dose-finding phase can be enrolled in the pivotal portion of the study, and existing patient data collected in the dose-finding phase can contrib-

ute to greater efficiency in the pivotal phase. However, it is important to be mindful that the protocol must not change from that specified at the initiation of the study; this places a premium on the challenging task of anticipating all outcomes when planning the study. Despite the challenges, the risks associated with combining phases II and III are low. The worst-case scenario for such an adaptive study, with no successful adaptations carried out, is the equivalent of conducting the same study conventionally, without the use of adaptive methods.

Combining data from the two phases requires procedures to control the type 1 error rate for the comparison of the test drug with the control. In addition, the final statistical analysis for the combined data from the learning and confirmatory phases will be more complex than usual [2]. It is also important to think through the plan for the seamless trial to ensure that the issues that might be analyzed between two separate studies are adequately analyzed in advance.

One of the greatest advantages of combining phases is that the entire startup process is handled once instead of twice. A single protocol can be developed, reviewed, and approved in a single process for both phases. Recruitment of investigators and patients is simplified because some of the investigators for the pivotal phase will already be familiar with the study. However, combined studies also require that many other investigators be set and ready to go when the final dosing decision is made. This raises a host of other issues such as making appropriate provisions for addressing an investigational review board (IRB) and making consent forms and study materials available in a timely manner. Considerably greater attention than usual is required to ensure that all requisite components mesh.

#### **5.2.2.6 Other Strategic Adaptations**

Another important class of adaptive techniques lays particular emphasis on extending the period of assessment and decision making. For example, it is possible to plan an adaptation that allows changing the test hypothesis from superiority to noninferiority, and to redesign multiple endpoints to update correlations or change the hierarchical order. It is also possible to establish a decision rule with criteria for determining whether to refocus the study on a subpopulation that is specified in advance. Noninferiority studies may be handled quite differently, especially in regard to hierarchical outcomes: These may, for example, declare success for an initial noninferior phase and then continue the study with the hope of generating data to support a superiority claim. The advantage of this approach is the potential for earlier marketing of the product. Obtaining data to support a noninferiority claim requires a more modest sample size and may yield results that allow the product to enter the market with a noninferior approval while the study continues to accumulate information that may support a superiority claim.

The range of techniques for strategic adaptations continues to evolve with the refinement of existing methods as well as the addition of new approaches and tools.

### **5.2.3 OPERATIONAL SIDE OF ADAPTIVE RESEARCH**

The fundamental principles of adaptive research apply not only to the sophisticated and innovative techniques for strategic adaptations but also to many of the activities

common to virtually all clinical studies, whether or not they involve strategic adaptations. Implementing efficient data capture, rapid data cleaning, generation of a range of performance metrics, and readiness for informed decision making based on such information can produce dramatic gains in efficiency. For historical reasons, the pharmaceutical industry makes little use of such capabilities today. Nonetheless, there is no reason why the industry should not take advantage of what amount to the same management techniques already used by most contemporary businesses to bring clinical studies into the modern era.

Since the necessary changes for operational adaptations need not affect study or program design, they do not require regulatory approval. Operational adaptations can therefore be implemented immediately, and their benefits are at least as profound as those flowing from strategic adaptations. One clear illustration is a large, complex phase III evaluation of an Alzheimer's drug candidate, where efficient collection of data and performance metrics enabled the completion of patient enrollment in record time and the closing of the database within 2 weeks of the last patient visit. As a result, this study saved 1.6 years and \$32 million in direct costs measured against the sponsor's internal projections of 5 years and \$100 million [7, 2000; this article can be found online at <http://www.healthdec.com/media/articles/AnAlzheimersDrugGoesonTrial.pdf>].

As previously noted, the same capabilities involved in making such gains in efficiency are required for all types of adaptive studies. In a broader sense, such capabilities represent the application of principles of tight management to the complex realities of clinical studies. Other industries have shown the way. The principle of just-in-time inventory brought new efficiencies to the automobile industry; the same principle, managing operations based on continuous, real-time information about important business processes, has been widely adopted in manufacturing and other highly competitive industries. While pharmaceutical development is considerably more complex and knowledge based than manufacturing, intelligent management can apply the same general principles to clinical studies while preserving study integrity and validity through careful operational controls and information management, preserving blinding, randomization, and other hallmarks of clinical research. The need for specific measures to exclude the possibility of bias does not, in an age of sophisticated access control systems, require near total ignorance for all study personnel of all study operations until the very end, when it is too late to take advantage of data and performance metrics for effective trial management. The need to remedy major shortcomings in the efficiency of current development practices is evident; so is the availability of methods with the potential to solve current problems and make dramatic, rapid improvements in efficiency.

Surprisingly, one critical requirement for tight study management is frequently overlooked: the timely reporting of performance metrics. Effective management is impossible without timely, accurate information about performance; in clinical studies, achieving tight management through operational adaptations requires the same capabilities as strategic adaptations: rapid data collection from the field, rapid data cleaning, timely analysis and summarization, and, importantly, presentation of information in different forms meaningful to staff performing different functional roles. The study manager, for example, may be centrally interested in knowing why certain sites are enrolling faster than others and will therefore want to track frequency of screen failures and the distribution of different reasons for them. The

field monitor may wish to know how to help a site decrease its query rate, allowing her to spend less time in managing minutiae of the study and, thanks to more accurate data collection and the reduced incidence of queries, more time helping sites conduct the study efficiently and achieve database lock faster. The head of R&D may be most interested in the projected dates for completion of enrollment and database lock. In summary, good performance metrics enable greater understanding of study progress, far tighter control, more effective allocation of resources such as monitoring time, faster enrollment, and, in the larger scheme of things, shorter timelines and lower costs.

#### **5.2.3.1 Enrollment**

Closely monitoring the progress of enrollment and the incidence of reasons some patients cannot be enrolled allows continuous tuning of enrollment strategies and criteria. Comparison of site performance, especially in larger studies, generally reveals a wide range of performance. The study manager's job is often to use such information to dig deeper, determine the reasons some sites lag in enrollment while others excel, and intervene as necessary to improve enrollment across the entire study. In some cases, close scrutiny of enrollment performance at different sites may reveal difficulties repeatedly encountered despite best efforts. In that event, case study managers may confront a decision about taking measures to overcome sub-optimal enrollment that is not due to operational inefficiencies, such as adding sites.

Comparison of site performance usually does reveal that certain sites are enrolling more effectively than the rest. The reasons may vary; the ability to track a variety of enrollment metrics in real time enables a manager to determine what the reasons are. Analysis begins with examining the frequency of screen failures and the reasons for them. Real-time access to performance metrics of the individual sites allows early identification of effective and ineffective recruitment activities for the particular study and its population. The lessons, both good and bad, should be shared immediately with all sites to enhance recruitment efforts studywide.

Detailed performance metrics on recruitment can also quickly determine whether any specific inclusion/exclusion criteria are disproportionately hindering recruitment of the desired study population. If this is the case, it is imperative to find out quickly. It may be appropriate to consider different recruitment strategies to obtain access to a more suitable population. Failing that, possible adjustments in inclusion/exclusion criteria may require consideration. Understanding the options depends on having timely, accurate enrollment data from investigative sites. The ability to manage enrollment effectively is impaired in direct proportion to the degree to which the availability of such data is lacking or delayed.

#### **5.2.3.2 Site Performance**

Much higher levels of efficiency are possible when study managers and monitors can track site performance continuously and in sufficient detail to identify and address problems. Most studies currently lack the ability to track such metrics. As a consequence study management becomes a passive affair left to the vagaries of the site monitor. Because monitors often lack training and experience managing



sites, there may be no one at all providing effective management. Furthermore, depending on any individual as an information and management filter invariably introduces subjectivity and uncertainty, no matter how well trained that individual may be.

Query rates and query response times on submitted data are easily measured and provide strong indications of site performance in general. Query metrics are important on several levels. Most importantly, the existence of numerous queries late in a study delays database lock and all downstream events, including analyses, progression, and submissions. In addition, instituting rapid data entry, validation, and return of queries produces faster, more meaningful feedback to sites on their performance, enabling each site to reduce errors, often quite significantly. By contrast, systems that rely on paper and pen, with double key data entry, generally take a month or more—often much more—to return queries to sites about data previously submitted. During the interval, each site, unaware of the problems in the data it collected, will continue to make the same errors, increasing the volume of errors and the time required to correct them and inevitably delaying database lock.

Key site performance metrics that should be continuously tracked include patients screened, screen failures and reasons, good clinical practice compliance issues, query rates, query resolution, the number and age of outstanding queries, and specific case report form (CRF) fields, forms, and validation ranges that are generating the most queries at each site. Number of adverse events, both serious and nonserious, should also be tracked.

Continuous attention to site performance indices has the additional benefit of allowing technology-enabled distributed management with reduced requirements for a monitor to go to each site to determine how things are going. Rather, the generation and tracking of such parameters enables all members of the study team to examine performance metrics. The metrics enable each member of the team to focus attention where it is most needed. The availability of detailed metrics also opens the door to performance-based site management, allowing study management to institute incentives for strong performance and disincentives for poor performance. Performance-based management stands in stark contrast with the present approach, with each site knowing it will be paid the same amount for each patient regardless of performance. Although this is currently the standard approach to “management” of investigational sites, it hardly deserves the name.

### 5.2.3.3 Adaptive Site Monitoring

Monitoring is one of the most expensive components of clinical research, typically accounting for one-third of study costs. Yet the objective of monitoring, while important, is fairly modest: To ensure that the data in the database is accurate. The high costs of monitoring investigational sites are often assumed to be inevitable sacrifices to the cause of ensuring data accuracy. An adaptive approach, however, can use the stream of timely information from the field to allow many functions now performed during site visits to be handled centrally, enabling far more rapid and standardized assessment of site performance based on performance metrics while also reducing costs.

The highest levels of monitoring efficiency are made possible by use of electronic data capture methods, such as the digital pen, that can serve as source documenta-



tion. Source document verification itself accounts for approximately 80% of monitoring time. The digital pen and similar devices, detailed in Section 5.2.4, are used by the clinician to write in a familiar way on a form laid out with a grid that enables software to match data entries to appropriate fields. Captured data is stored in the digital pen and transmitted to a central database, where the pen is inserted in a dock attached to a computer. With this approach, the digital data is not a transcription of the source data; it is the source data itself. It is unnecessary for the monitor to compare electronic data to an original piece of paper. Data first recorded in electronic form, including data values and an image of each completed form (reviewed if necessary to resolve any ambiguities), is checked and corrected promptly through queries to site personnel after both automated and human review. Use of EDC technology capable of serving as source data provides great cost savings in the source data verification process by eliminating the time and expense of comparing electronic data to stacks of printed material.

Much of the work that monitors currently do can be done much better and more accurately by systems that harness technology. Data capture that obviates the need for traditional source data verification is one example. Effective use of such technology in conjunction with adaptive techniques can reduce monitoring costs by two-thirds or more while enabling far faster and more effective verification of data accuracy.

Adaptive approaches to site monitoring reflect a fundamental change in the way studies are conducted. Increasingly, the tedious, repetitive manual tasks such as comparing two values on different pieces of paper will be replaced by electronic tools and automated processes. The evolution to such processes will be further improved with the adoption of electronic medical records, ultimately resulting in a monitoring process requiring no manual steps. As a consequence, the monitoring function will be changed from that of the box checker of today into that of a manager and consultant whose main job will be to assure site performance. The monitor will be able to focus on optimizing each site's performance against trial objectives rather than constantly comparing the content of one database field to the content of the same field on a paper CRF. Monitors can spend less time on minutiae and more time observing, analyzing, and reacting to study trends. The job of the monitor will increasingly shift to anticipating and addressing issues before they can develop into problems that require extensive cleanup.

Performance metrics can help monitors identify problems at a sufficiently specific level to suggest both causes and potential solutions. The number of visits needed at sites can be determined based on each site's performance metrics rather than equal numbers of visits at similar intervals for all sites. Sites with more and bigger problems can receive more site visits; sites with outstanding performance, fewer. For example, with the tracking of the number of unmonitored fields, the accumulation of a sufficient number can trigger a site visit. But while a good monitor might verify between 500 and 1000 fields per day, depending on the study, a less experienced monitor might do half that. The less experienced monitor could schedule a visit when, say 500 fields accumulate, while a more effective monitor could wait longer. Using an arbitrary, uniform schedule, the less experienced monitor might have a difficult time, while a more experienced monitor would waste time at the site and traveling.

## 5.2.4 IMPLEMENTATION ISSUES

An adaptive development program requires a different approach to study design, data capture, job functions, business processes, and planning. When first moving into adaptive research, the simplest approach is to start with baby steps while refining processes and increasing capabilities. Phase II adaptive dose finding and sample size reestimation are generally the most straightforward types of adaptive trials to undertake.

### 5.2.4.1 Data Capture and Validation

Adaptive research is critically dependent on the timely availability of accurate information. Information to effectively manage any given study may include specific decision criteria related to strategic adaptations, but it always involves continuous assessment of operational performance measures. Careful attention to operational issues is a prerequisite for achieving strategic adaptive goals. For example, slow availability of outcome information will delay and possibly cripple efforts at adaptive dose finding, and slow recruitment inevitably undermines any strategic adaptive decisions because a shortage of patients guarantees a shortage of data. Thus, the trial will be inefficient no matter how sophisticated or sound the planned adaptations.

Efficient data capture is a prerequisite for adaptive studies. In contrast to Web-based EDC and its attendant delays for data reentry, two newer forms of data capture have demonstrated superiority for adaptive studies. The first is an electronic pen incorporating an optical sensor that records keystrokes as data are entered on paper CRFs. After completion of CRFs, the pen is docked and the data transmitted to a central location, where keystrokes are converted to numbers and letters and the data recorded in the study database, along with an exact electronic image of each CRF, available for review by data management personnel in the case of ambiguous or missing data. Together with complementary systems that enable automated validation, the electronic pen provides the capability of having information from a patient visit—summarized and interpreted information, as opposed to raw data—on a sponsor's desktop before a patient even leaves the investigational site.

The second new form of data capture consists of fax-back systems. These offer another way to allow the simplicity of pen-and-paper data entry coupled with efficient software to read incoming forms. Fax-back tools, first introduced more than a decade ago, have improved markedly in recent years, adding the ability to read data electronically and store both the data and an image of each completed form in the study database.

In practice, digital pen systems have proven to be the most accurate and timely data-capture method, with query rates of approximately 1% (1 query issued per 100 fields) compared with approximately 5–10% for fax-back systems. Both methods are significant improvements over older pen-and-paper, double-key manual data entry, which often has query rates on the order of 20%. However, the most important determinant of query rates is not the data-capture method itself but the “middleware”—the software and/or systems that bring data into the trial system, validate it, and return queries. The main reason for the superior efficiency of the electronic pen is its linkage with the sophisticated middleware that enables queries to be returned

to investigational sites within minutes of their submitting data. The middleware not only markedly reduces the feedback time but also enables tracking of query rates by interviewer, question, and a multitude of other variables that help identify sources of error so that they can be immediately corrected. This is a boon for adaptive research since the ability to capture data, analyze it, and rapidly respond to performance metrics such as query rates is the essence of an adaptive approach.

A data-capture device such as the digital pen allows the clinician an intuitive, familiar means of recording and transmitting data to the study database, where it can be transformed electronically to information meaningful for individual study roles through “widgets” (Fig. 5) or the Web. And, as previously noted, if the CRF can be used as source data, the work of monitors is greatly diminished, allowing them to focus on management rather than rote tasks.

Advanced data management systems also incorporate sophisticated procedures that automate validation and query generation. These go well beyond the simple checks such as data ranges to incorporate algorithms that consider trends, consistency across visits as well as within a visit, and head off future problems by identifying studywide weaknesses as well as individual site or interviewer problems. These processes similarly help assure rapid availability of clean data as well as providing rapid feedback and management assistance that assure better site performance in the future.

#### 5.2.4.2 Planning

Since potential adaptations and related decision criteria must be specified in advance, the effort required to plan an adaptive study is greater than for a conventional trial.

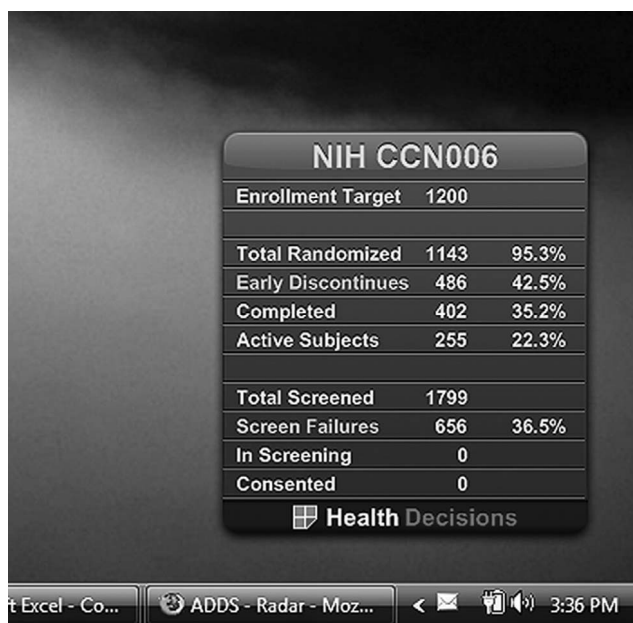


**FIGURE 5** Desktop “widget” that provides immediate project status, updated in real time. (Figure copyright 2006, Health Decisions, Inc. Used by permission.)

Protocols employing adaptive techniques must specify through predetermined decision rules precisely what will occur in different circumstances and with different values emerging in trial data. This approach is a requisite element to ensure that scientific integrity is not compromised, knowingly or otherwise. Simulation during planning can effectively be used to explore a variety of “what if” scenarios to determine sample size adequate to control variance and other elements affecting statistical power. The simulations are also useful in analyzing decision rules for planned adaptations and operational issues such as ensuring adequate supplies in the event of different adaptive changes.

Planning of adaptive trials must include specific arrangements to exclude bias. This includes measures to eliminate the possibility of unauthorized parties inferring from adaptations actually made in the course of a trial the data that brought about the adaptations, effectively unblinding the study. Such back solving of study data can be prevented by measures such as limiting access to information about the specific rules governing planned adaptations and defining decision criteria in terms of ranges rather than precise values. The latitude allowed by regulatory authorities for blinding and other study elements are, however, substantially greater during early testing than during confirmatory studies.

When unblinded assessments are required (such as sample size reestimations or, in some instances, pruning treatment arms), a firewall must be in place to ensure that those running the study do not gain access to information that may affect how the study is conducted. For example, an adaptive trial protocol may require independent (internal or external, focusing on safety and/or efficacy) analyses that require access to unblinded data. Decisions may also involve a broader range of business issues such as whether to discontinue a study. While such decisions may be delicate, it is critical that safeguards be incorporated to prevent contamination of the ongoing study; this is equally true whether required by regulatory stipulations (confirmatory phase) or not (learning phase), as compromised data creates opacity in a situation that demands clarity. These situations are normally handled by compartmentalized groups who are not involved in the study itself, often by a steering committee that may have other groups (such as safety) reporting to it. The ultimate power to decide, subject to the specified decision criteria, whether to implement adaptive changes often rests with the steering committee.

Simulation can also be used not only to explore the likelihood of different scenarios that affect study design but also for optimizing business processes and defining job functions to support the development program’s objectives. Although the effort required is greater, the thorough analysis of possible future scenarios provides an invaluable opportunity to moderate or eliminate potential choke points. Over time, the longer view and greater flexibility that are hallmarks of adaptive research can substantially improve allocation of a development program’s financial resources and accelerate attainment of the program’s goals.

#### **5.2.4.3 Process Optimization**

Despite the current variability in the conduct of clinical trials, almost all trials require repeatedly performing the same essential processes. These processes can be defined much more clearly and precisely than is customarily done. The analysis required to define a process precisely will often suggest ways to improve the process.

It is process improvement in combination with new technology that brings huge rewards. The gains from inserting expensive new technology into an inefficient process are minimal.

Optimizing processes for adaptive research requires minimizing the individual variations in trial processes, including variations among the approaches taken by different project managers. The organization of the trial should be implemented in a system rather than according to each project manager's idiosyncrasies. The test of whether the project manager's role is defined precisely enough is whether it would be easy to plug in a replacement.

Furthermore, optimizing a study or program based on analysis of data as it is collected is not just a matter of inserting more frequent interim analyses into existing processes. The trial processes must provide a continuous flow of up-to-date, clean data. To the extent that current processes render data unavailable—whether because it has not been entered, validated, passed on from person A to person B, or made accessible in a useful form—study managers may be prevented from making key decisions or deceived by incomplete data into making decisions that turn out to be erroneous.

#### **5.2.4.4 Decision Making**

In adaptive trials, learning and decision making no longer wait until the very end, with the sudden transformation from having no data and knowledge to having a deluge of data and the need to extract knowledge all at once. Progress must be considered weekly, emerging trends identified continuously, and decisions taken as necessary to optimize the trial.

The point of gaining a greater understanding of data at an earlier time is to inform decisions to optimize the course of a trial. Greater efficiency in an adaptive study is the result of active managers making timely decisions directing specific changes. Having information earlier achieves little if managers fail to understand it and, where appropriate, to take action.

#### **5.2.4.5 Adaptive Data Monitoring**

The challenges of adaptive data monitoring require careful attention in implementing strategic adaptations. Besides attention to the exclusion of bias through organizational arrangements such as independent data management committees and independent statistical services, there must also be careful attention to arrangements for timely, accurate analysis of data to facilitate all potential adaptive decisions. Statistical methods must be selected and approved, and analytical procedures and tools created and tested to avoid delay or confusion in the implementation of pre-defined decision rules [8].

#### **5.2.4.6 Regulatory Considerations**

Regulators have been receptive to proposed methods of conducting clinical trials more efficiently while preserving validity and integrity. Understandably, they have not granted blanket approval for clinical researchers to try all new methods for strategic adaptations. However, they have shown a willingness to consider adaptive

techniques. Regulators have also expressed greater receptiveness to certain adaptive techniques and have indicated some general preferences for the use of those techniques. As detailed guidance for conducting adaptive studies is awaited, it is possible to provide a brief summary of regulatory preferences.

1. Two of the easiest adaptive components to apply are sample size reestimation and pruning in dose-finding studies. Regulatory groups allow sponsors a great deal of latitude as to how they conduct early studies, and the issue of defining doses suitable for final (pivotal) testing is basically at the sponsor's own risk. Sample size reestimation can be used in this early stage as well as more powerfully in late (pivotal) testing; the benefits of sample size reestimation in the early phases are sufficiently powerful as to merit serious consideration for all such studies.

2. All possible adaptations in a trial must be specified in advance, as well as the criteria for deciding whether to carry out the adaptations. Adaptive studies should never be thought of as providing license to make any changes that happen to occur to trial managers along the way; it would be a grievous mistake to think that adaptive techniques provide a less stringent substitute for well-planned, carefully executed studies, or that regulators regard adaptive techniques so indulgently.

3. Simulations are often helpful in understanding possible outcomes and are of interest to regulators. Simulations allow assigning probabilities to different potential developments in the trial and analyzing the repercussions of each development. Newer software and powerful computers facilitate exploration of scenarios, often forcing consideration of study variables that might otherwise go unattended. Simulations increase understanding of the effects of specific adaptations on the conduct and results of the trial. Both the results of simulations and the software used to conduct simulations for trial planning should be submitted to regulators. It is advisable to request a special protocol assessment to review an adaptive design and, at that time, to provide simulations demonstrating that the type 1 error is preserved.

4. If decisions on adaptations require considering unblinded data, regulators may prefer that the data is analyzed and reviewed by independent statisticians and an independent data management committee rather than by the sponsor. It is important, however, to recognize and limit the power of outside bodies that may fail to fully appreciate business consequences of managerial decisions. Not surprisingly, academic groups are often less attuned to the business consequences of decisions made about clinical studies.

5. Criteria for decisions on adaptations should be specified and managed in such a way as to minimize the possibility of unauthorized parties inferring important blinded information from the specific adaptations that are carried out. For example, if sample size is to be readjusted based on the magnitude of the treatment effect, knowledge of the sample size chosen might enable unauthorized parties to back solve for the treatment effect. Such concerns can be allayed in at least two ways. First, the specific technique and criteria that will be used for sample size reestimation can be kept confidential. For example, there should be restricted access to knowledge about the type of sample size reestimation that is planned—whether reestimation will be based on the size of the treatment effect alone, or the size of the placebo effect, or the variability observed in trial data, or based on multiple parameters. Second, the decision criteria themselves should be specified in ranges;



thus, when the sample size is readjusted to a specific value, the most that can be inferred by those who learn of the adjustment is that a parameter fell somewhere within a range.

6. Regulators will want detailed information about the measures taken to ensure the exclusion of bias. There must not only be a “firewall” between those who are allowed access to specific information and those who are not, but the firewall must be one that regulators agree is likely to be effective.

## 5.2.5 PROMISE OF ADAPTIVE METHODS

Adaptive techniques represent a major advance in pharmaceutical development made possible by advances in communications and technology. In many aspects, adaptive techniques represent a step into the type of modern management that is the norm in other large industries. Modernizing trial management requires evolutionary changes built on time-honored principles—but also requires greater flexibility in an industry known for its conservatism and “not invented here” attitude. Adaptive methods have been developed over time, based on solid theoretical foundations, specifically to address many of the limitations and inefficiencies of the current approach to clinical research.

The changes required for the shift to adaptive methods are profound and extend beyond technology to encompass work processes and the functioning of individuals performing different study roles. Technological tools and related processes are meant to provide a stream of current information that sheds light on study progress and bottlenecks. The new technologies and processes do not replace managers. Rather, they enable managers truly to manage, by providing a wealth of current status information and communications that drive effective functioning studywide.

Committing to an adaptive development program offers the potential for great rewards. Because of these high rewards, adaptive research is often assumed to be accompanied by high risk. However, a properly conceived and managed adaptive program actually reduces risk. This is by design—reducing risk is a major objective of adaptive techniques. Indeed, utilizing adaptive techniques entails no risk of conducting less efficient studies—the worst that could happen in an adaptive trial is that no adaptive changes would be made, leaving study performance at the accustomed level. On the other hand, every adaptive change that is made saves time, reduces costs, improves the amount and quality of information produced, or provides some combination of these benefits. Indeed, the greater risk lies in *not* utilizing adaptive techniques. Entire programs are condemned to suboptimal decision making and the inevitable consequences: unnecessary expenditures, avoidable delays, and windows of opportunity slammed shut before new drugs can make their way through the process of clinical evaluation.

The existing conventional approach to clinical studies has spawned many horror stories that reflect the risks of the current approach to clinical research. Every experienced pharmaceutical researcher has a horror story to tell. Virtually all such stories include an account of a small initial problem that grows to disturbing proportions before it is recognized and addressed. Adaptive processes can reduce the instance of such horror stories and the risks they embody by providing earlier, better indicators of incipient problems.



The potential of adaptive research to improve operations industrywide, for companies large and small, is simply too great to ignore. The potential gains in efficiency—the savings in both time and money—are dramatic. Adaptive tools and techniques are here and ready. They address many of the specific drawbacks and weaknesses of traditional methods. With concentrated attention by the industry and regulators, adaptive methodology can make important contributions to a new and more productive era in clinical research.

## REFERENCES

1. Doll, R. (1998), Controlled trials: The 1948 watershed, *BMJ*, 317, 1217–1220.
2. Chow, S.-C., and Chang, M. (2007), *Adaptive Design Methods in Clinical Trials*, Chapman & Hall/CRC, Boca Raton, FL, pp. 47–67, 55–59, 58–60, 171.
3. Hu, F., and Rosenberger, W. (2006), *The Theory of Response-Adaptive Randomization in Clinical Trials*, J Wiley, Hoboken, NJ, pp. 4, 6–7.
4. Berry, D. A., Mueller, P., Grieve, A. P., Smith, M., Parke, T., Blazek, R., Mitchard, N., and Krams, M. (2000), Adaptive Bayesian designs for dose-ranging drug trials, in Gatsonis, C., Kass, R. E., Carlin, B., Carriquiry, A., Gelman, A., Verdinelli, I., and West, M., Eds., *Case Studies in Bayesian Statistics V*, Springer, Verlag, New York, pp. 99–181.
5. Jennison, C., and Turnbull, B. (2006), Efficient group sequential designs when there are several effect sizes under consideration, *Statist. Med.*, 35(6), 917–932.
6. Chuang-Stein, C., Anderson, K., Gallo, P., and Collins, S. (2006), Sample size reestimation: A review and recommendations, *Drug Info. J.*, 40(4), 475–484.
7. Schoenberger, C. (2000), An Alzheimer's drug goes on trial, *Forbes Mag.*, March 20, pp. 94–96.
8. Abrams, K., Myles, J., and Spiegelhalter, D. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Wiley, Chichester, UK, pp. 202–224.



# 6

---

## *Organization and Planning*

**Sheila Sprague<sup>1</sup> and Mohit Bhandari<sup>2</sup>**

*<sup>1</sup>Department of Clinical Epidemiology & Biostatistics and <sup>2</sup>Division of Orthopaedic Surgery, Department of Surgery, McMaster University, Hamilton, Ontario*

### **Contents**

|       |                                                                      |     |
|-------|----------------------------------------------------------------------|-----|
| 6.1   | Protocol                                                             | 162 |
| 6.1.1 | Protocol Format                                                      | 162 |
| 6.1.2 | Protocol Amendment Procedure                                         | 167 |
| 6.1.3 | Prestudy Requirements                                                | 167 |
| 6.2   | Finance                                                              | 169 |
| 6.2.1 | Reviewing an Offer to Participate in Clinical Trial                  | 169 |
| 6.2.2 | Budget Considerations                                                | 169 |
| 6.2.3 | Investigator Agreements                                              | 170 |
| 6.2.4 | Sponsor Interactions                                                 | 170 |
| 6.2.5 | Regulatory Documentation (Essential Documents)                       | 171 |
| 6.3   | Patient Selection                                                    | 171 |
| 6.3.1 | Research Ethics Board Approval (Institutional Review Board Approval) | 171 |
| 6.3.2 | Feasibility                                                          | 171 |
| 6.3.3 | Sample Size                                                          | 172 |
| 6.3.4 | Recruitment Methods                                                  | 173 |
| 6.3.5 | Screening                                                            | 173 |
| 6.3.6 | Informed Consent                                                     | 173 |
| 6.3.7 | Patient Confidentiality                                              | 174 |
| 6.3.8 | Intention to Treat Analysis                                          | 174 |
| 6.4   | Treatment Schedules                                                  | 174 |
| 6.4.1 | Supplies                                                             | 174 |
| 6.4.2 | Adherence to Protocol                                                | 175 |
| 6.4.3 | Follow-up                                                            | 175 |
| 6.4.4 | Monitoring Visits                                                    | 175 |
| 6.4.5 | Composite Endpoints                                                  | 176 |

|                                                                             |     |
|-----------------------------------------------------------------------------|-----|
| 6.5 Patient Response Evaluation                                             | 176 |
| 6.5.1 Outcome Measures                                                      | 176 |
| 6.5.2 Case Report Forms                                                     | 176 |
| 6.5.3 Adverse Event Reporting                                               | 176 |
| 6.5.4 Data Queries                                                          | 177 |
| 6.5.5 Adjudication                                                          | 177 |
| 6.6 Design Considerations                                                   | 179 |
| 6.6.1 Study Design (Observational Studies and Randomized Controlled Trials) | 179 |
| 6.6.2 Trial Organization and Responsibilities                               | 180 |
| 6.6.3 Data Management                                                       | 182 |
| 6.6.4 Quality Control                                                       | 182 |
| 6.6.5 Study Close-out Activities                                            | 182 |
| Bibliography                                                                | 183 |

## 6.1 PROTOCOL

The study investigator is responsible for the review of all protocols and for the assessment of the feasibility of the clinical site to conduct the clinical research study. The investigator is also responsible for reviewing the protocol and providing input to the design so that the protocol can be properly conducted at the investigator's clinical site. The investigator is also responsible for ensuring that the research ethics board or the institutional review board approval is obtained before initiating the protocol.

Each sponsor's protocol received by a clinical site should describe the purpose, the objectives, the study design, the methods, and the planned analysis of a clinical study. The protocol must be sufficiently detailed so that the study can be conducted accurately and properly. Since the protocol often outlines the ethical, clinical, and regulatory responsibilities of both parties (the sponsor and the clinical site), it usually acts as a contract between the investigator and the sponsor.

Each protocol must be reviewed by the investigators, research coordinators, and other relevant clinical and research associates to ensure that the following takes place: (1) The study is safe; (2) the study is based on good medical science; (3) the study is ethically acceptable; (4) the objective of the study is clear and the study design, sample size, procedures, and statistical analysis will enable the study objectives to be met; (5) the protocol satisfies the clinical sites standard operating procedures (SOPs) and the good clinical practice (GCP) guidelines; (6) the protocol is financially and practically feasible for the clinical site and the investigator; (7) the investigator and clinical site have a sufficient number of potential study participants to meet the protocol enrolment goals; (8) the investigator has the sufficient staff to successfully conduct the trial; and (9) the investigator has sufficient resources and equipment to successfully complete the study.

### 6.1.1 Protocol Format

We will describe several protocol formats that are commonly used by industry-sponsored trials and by investigator-initiated trials (non-industry-funded trials).

**Industry-Sponsored Trials** The first format is the one that our institution (McMaster University/Hamilton Health Sciences) recommends in their standard operating procedures (SOPs).

**Cover Page and Table of Contents** The protocol should begin with a cover page that outlines the status of the trial, the revision date and number, the trial's full title, the study number, and the sponsor. There should also be space for signatures and dates signed for the author, the sponsor approval, and the investigator approval, as the protocol often acts as a contract between the investigator and the sponsor. The sponsor contact names and telephone numbers should also be listed to respond to investigator queries or any adverse events. A detailed table of contents, including a list of all appendices, should follow the cover page.

**Introduction** The introduction contains a brief summary of the relevant background information on the study design and the protocol methodology. Sufficient background information should include the development of the study agent or device, a description of the disease process to be studied, and the current medical treatments available. The introduction should provide enough detail for the readers to clearly understand the rationale for the study. Relevant information regarding pharmacological and toxicological properties of the test article and previous efficacy and safety experience should be included.

**Study Objectives** The objectives of the study should be clearly stated, and there should be a description on how the objectives are related to the design of the study. Both primary and secondary objectives should be included.

A brief overview of the study design, indicating how the study objectives will be met, should be provided. This part of the protocol includes a description of the study type (i.e., randomized controlled trial, prospective cohort study, phase III, multicenter, double blind, placebo controlled), details on the specific treatment groups, information on the sample size in each group required for the total trial and at each clinical site, how the study participant identification numbers are assigned, and the type, sequence, and duration of the study periods. A brief description of the methods and procedures to be used during the study should also be included. There should also be a discussion on the study design that details the rationale for selecting the study design. Critical decision points and any atypical features of the study design should be described. The treatment schedule, the dosages for the study agents, and the follow-up schedule and follow-up requirements should be discussed.

**Study Participant Population** The next section of the protocol describes the study participant population. The study's inclusion criteria (the criteria each study participant must satisfy to participate in the study) are described. These criteria can include the following: age, sex, race, diagnosis, diagnostic test result requirements, concomitant medication requirements, severity of symptoms and signs of disease, and the ability and willingness to perform study requirements and to provide informed consent. The criteria must be sufficiently detailed to provide the investigative site with the information needed to recruit eligible study participants and to ensure a homogenous study population. The exclusion criteria describes the items that would eliminate a potential study participant from being enrolled in a study. Exclusion criteria can include, but are not limited to, the following: previous medical history, pregnancy, childbearing potential, current or past therapy, severity of disease, current medical conditions, a minimum time since the last clinical study, and substance abuse problems. There should be a discussion detailing the rationale for the

inclusion and exclusion criterion to obtain the homogenous study population required.

***Concomitant Medications*** The listing of the required therapies including the dose, frequency, and duration should be detailed in this section. Required therapies are medications that are required in addition to the test treatment. The investigator must also determine whether the sponsor or the site is providing these medications and determine the accountability procedures required by the sponsor. There must also be a list of allowed medications and disallowed medications. Allowed medications are the medications that are permitted during the study, the dosage, the frequency, and the conditions under which they are permitted. Disallowed medications are medications that the study participant is not permitted to use during parts of or the entire study. The study participant and investigator are often told to document the allowed medications and the disallowed medications (if there is a protocol deviation) on the patient diary cards and on the case report forms.

***Study Plan and Methods*** The study plan and the methods portion of the protocol details the plan of action, all procedures to be followed, and the methods to be used throughout the study. The activities for each follow-up visit are clearly outlined, as well as any specific criteria that must be met before the study participant moves to the next stage of the study. For difficult testing methodologies, a separate section may be included to describe in detail the methods of the testing. If the study participant must have specific results from this testing to enter, or to continue in the study, these criteria are described. Types of study activities that would be included in this section include: medical history, type of physical examination, blood or urine testing, electrocardiograms (ECGs), diagnostic testing such as pulmonary function tests, symptom measurement, dispensing and retrieval of medication, diary card exchange, study subject number assignment, adverse event review, and the like. Each visit should be detailed in separate sections in the protocol, such as 6.1 Visit 1, 6.2 Visit 2, and so forth. The investigator should ensure that this section provides clear methods, procedures, and timings of activities.

***Study Medication Supplies*** The study medication supplies portion of the protocol describes the test treatments including placebo with formulation and dosage information. It also details the packaging, dispensing, retrieval, and accounting of these supplies. All ingredients and the formulation of the investigational test article and any placebos that are used in the study are described. The precise dosing required during the study is discussed. The method of packaging, labeling, and blinding is described. The method of assigning treatments to study subjects and the test article and study subject numbering system is detailed. If applicable, the method of blinding to make the test treatments indistinguishable should be explained. The method of test article coding, code storage, and code access is described. If a third party such as a pharmacist will be dispensing or blinding the test article, then specific instructions on how and when the blinding will occur will be detailed in this section.

The coding and study subject test article assignments for the study are detailed. Generally, all study staff and study participants are blinded to the identity of the test treatments. This section describes when a study subject treatment code assignment is given and if, and when, it will be changed. A description of where the test

article randomization code will be kept by both the investigator and the sponsor or delegate is required and the procedure for breaking this code, if necessary, should also be included. There should also be a section describing the study subject kit packaging, which should detail the administration and dosage of study medication, as well as how the elements of the kit are distributed. Generally, the medication is labeled visit 1, visit 2, and so forth for easier dispensing. Any specific instructions for the administration of the test article treatments, such as label removal, test spray, or how and when the test article is to be returned should be detailed in this section. Explicit details regarding the dosage of the test treatment are explained. Instructions for broken test article applicators or lost test articles are also described. If extra treatments are provided for emergencies, the circumstances regarding their use should be described. Explicit instructions for the storage of the test treatments at the investigative site are detailed, as well as if special conditions are required, such as refrigeration. Federal regulations and the International Conference on Harmonisation (ICH) guidelines require a complete accounting of all test materials received, disbursed, and returned.

***Discontinued-Study Participants*** The research protocol should outline how investigators will handle study participant withdrawal, dropouts, or other discontinued participant subjects. It also details how to handle these situations within the study context and whether these study subjects are to be replaced.

***Adverse Events*** Investigators should include a listing of expected adverse events from previous clinical studies with the test article. They may also include the policies and procedures for the reviewing and the reporting of adverse events that occur during the clinical study.

***Concomitant Illness*** The protocol should describe the procedure for documentation of any concomitant illness, which a study participant develops during the study. A concomitant illness is an illness, disorder, or any other medical condition that is not considered a consequence of the study medication being administered.

***Variables and Evaluations*** The investigators should describe how clinical study efficacy and safety variables will be evaluated. The studies primary and secondary variables will be identified and discussed. Descriptions of evaluations are given here including how they are recorded, measured, or calculated.

***Statistical Analysis*** It is important to provide details on how the study results will be analyzed and reported. Specifically, sample size determination, objectives or hypotheses being tested, levels of significance, statistical tests being used, and the methods used for missing data. The method of evaluation of the data for treatment failures, noncompliance, and study participant withdrawals is presented. If an interim analysis will be performed, the rationale and conditions are clearly described. A description of how the safety data and adverse events will be tabulated and evaluated is included.

***Informed Consent*** This section of the protocol describes the procedures and responsibilities for obtaining informed consent for the study.



**Study Monitoring** A description of study monitoring policies and procedures, including the right for the sponsor or method center's representatives, federal government, and/or other regulatory authorities to verify and audit the study data, is presented in this section.

**Deviations from Protocol** The method for handling protocol deviations is described.

**Discontinuation of Study** This section describes the conditions under which the study would be discontinued.

**Confidentiality of Data** The importance of confidentiality of data and the fact that the data is owned by the sponsor or by the coordinating methods center are detailed.

**Publication of Study Data** The conditions for publication of study data are outlined.

**Research Ethics Board Approval** This section states that the investigator is required to obtain research ethics board (REB) or institutional review board (IRB) approval for the protocol, informed consent form, and any protocol amendments.

**Investigator's Statement** An outline of the investigator's responsibilities and agreements are included in the protocol.

**Investigator-Initiated Trials** For investigator-initiated randomized controlled trials, the Canadian Institute of Health Research requires the following format.

**Need for Trial** The problem to be addressed and the principal research questions to be addressed are listed. There should also be a strong justification as to why a trial is needed now. Evidence from the literature, professional and consumer consensus, and pilot studies should be cited if available. References to any relevant systematic reviews should be provided and the need for the trial in the light of these reviews should be discussed. If you believe that no relevant previous trials have been done, provide details of the search strategy used for existing trials. A description of how the results of this trial will be used should be provided.

**Proposed Trial** This section of the protocol begins by stating the proposed trial design, including whether the trial is open-label, double or single blinded, and so forth. The planned trial interventions for both the experimental and control groups and the proposed practical arrangements for allocating participants to trial groups are discussed. For example, the randomization method and if stratification or minimization are to be used provide justification for using either methodology. Factors that will be stratified or minimized should be listed. The proposed methods for protecting against other sources of bias include blinding or masking. If blinding is not possible, please explain why and give details of alternative methods proposed or implications for interpretation of the trials results. The planned inclusion/exclusion criteria should be listed.

There should be a justification for the proposed duration of treatment period and the proposed frequency and duration of follow-up. The rationale for the proposed primary and secondary outcome measures and how the outcome measures will be measured at follow-up should be provided.

For the proposed sample size, include both the sample size for the control and treatment groups, a brief description of the power calculations detailing the outcome measures on which these have been based, and provide event rates, means, and medians, as appropriate. Provide a justification for the size of difference that the trial is powered to detect and mention whether the sample size calculation takes into account the anticipated rates of noncompliance and loss to follow-up.

A description of the planned recruitment rate, how recruitment will be organized, over what time period recruitment will take place, and what evidence there is that the planned recruitment rate is achievable should follow. There should be a discussion on any likely problems with compliance and loss to follow-up; the evidence that the compliance and loss to follow-up figures are based on should also be included.

The details of the planned analyses are described, including any planned subgroup analyses and the proposed frequency of analyses (including any interim analyses).

The investigators should specify whether the trial addresses any economic issues. This is not a requirement; however, it is important to justify the inclusion/exclusion of any health economic studies and give details of any study proposed.

The final section of the protocol on the proposed trial should provide an accurate budget, budget justification, and the length of the trial.

*Details of Study Team* A description of the study team must be provided that explains the overall trial management, including the role of each applicant proposed, the steering committee, the methods center, and whether a data safety and monitoring committee will be established and its composition. The role of international collaboration should be discussed, including the nature of and the need for any international collaboration. The proposed participating centers should be listed along with their experience with previous trials and estimated recruitment rates.

### **6.1.2 Protocol Amendment Procedure**

Protocol amendments can be suggested by either the sponsor or the investigator and can be made after the protocol has been finalized. A draft protocol with the proposed amendments is prepared and reviewed internally and by the investigators before it is approved. The investigator site must obtain REB or IRB approval of the protocol amendment before it is implemented.

### **6.1.3 Prestudy Requirements**

Before a study can be initiated at a clinical site, the following activities must be completed:

- Receive final protocol from sponsor or methods center.
- Receive amendments from sponsor or methods center.
- Finalize budget with sponsor or methods center.
- Distribute protocol amendments to relevant study team members.
- Receive investigator's brochure from sponsor or methods center.
- Read investigator's brochure.
- Complete qualified investigator undertaking, FDA 1572A, or investigator agreement, as required.
- Prepare informed consent form.
- Submit documentation to the REB or IRB.
- Receive REB approval and Health Canada REB application form (if applicable).
- Receive case report form (CRF) books from sponsor or methods center.
- Design source documents, as necessary.
- Send all required documentation to sponsor or methods center.
- Receive clinical supplies from sponsor or methods center.
- Inventory supplies and return documentation of receipt to sponsor or methods center.
- Plan subject recruitment strategy.
- Prepare regulatory documentation binder/file.
- File regulatory documentation received to date.
- Set up necessary local resource utilization for study (e.g., pharmacy, laboratory, etc.).
- Set up a contract with the sponsor or methods center.
- Conduct in-services with each involved department.

The investigator or delegate must send the following documents to the sponsor or methods center before the study can begin:

- Final signed protocol.
- Final signed amendments.
- Signed Qualified Investigator Undertaking Form, FDA form 1572, or investigator agreement.
- Current curriculum vitae (CVs) and medical licenses of principal investigator and subinvestigators.
- Signed financial disclosure document.
- Budget approval documentation.
- REB or IRB approval letter(s) approving protocol, consent form, advertisements, and any other relevant documents.
- Copy of REB or IRB approved consent form.
- Copy of current REB or IRB membership list or letter from REB or IRB.
- Copy of laboratory license, normal ranges, and CV of director, if required.

Finally, before each study is initiated, appropriate research staff members are informed of the requirements of the study and their role and responsibilities regarding the conduct of the study.

## 6.2 FINANCE

### 6.2.1 Reviewing an Offer to Participate in Clinical Trial

Participating in clinical trials requires a substantial commitment of both time and effort, and participation often continues for months or even years. There are usually clear financial incentives to participating in clinical trials; however, other incentives include a chance to collaborate with other clinical investigators and opportunities to improve knowledge about the disease and treatment being investigated. Another advantage, which may be offered in some trials, is the exposure to new investigative techniques or access to special equipment or facilities.

The scientific, practical, and financial implications need to be considered before agreeing to participate in a clinical trial. The first item to assess is the study question and the study methodology. It is important to ensure that it is a relevant question and that the study methodology is sound and will meet the goals of the trial. The eligibility criteria must be carefully assessed to ensure that your clinical site has sufficient eligible patients to be recruited for a clinical study. It is often necessary to perform a survey of potentially eligible patients over a 4-week period to provide an accurate estimate of recruitment rates.

It is important to assess the impact of the trial on your patients, as your primary obligation is to protect the welfare your patients. It is important to consider if the patients will be required to have any investigations or procedures that are not part of standard care, and whether these will be painful and will possibly put the patients at risk. You need to estimate how much time the patient will devote to the study and how much, if any, compensation the patient may receive.

The time required for the overall coordination, the patient enrolment and follow-up, and the data collection also needs to be carefully considered. It is best to have an experienced clinical research coordinator carefully review the protocol and case report forms to provide an accurate estimate on the time and resources required to successfully participate in a clinical trial. Tasks often take longer than anticipated, so it is often good to add in a little extra time to account for the unanticipated. It is also important to know whether the study will be published, who will write the manuscripts, and what the authorship policies are. After carefully weighing these items, you can make an informed decision to participate in a clinical trial.

### 6.2.2 Budget Considerations

The investigator should consider all potential costs and prepare an accurate budget for running the clinical trial at his or her clinical site. These costs include: (1) administrative assistant time, (2) research associate/nurse time, (3) supplies, (4) expenses incurred by the patient, (5) extra medical and hospital costs (i.e., pharmacy, radiology, laboratory tests), and (6) physician time. It is also important to include

departmental and institutional overhead costs in the budget and to have your institution's grants and contracts office, as well as the REB or IRB, review your budget.

### 6.2.3 Investigator Agreements

The investigator is responsible for the review or preparation and approval of the contracts with the sponsor, and the investigator's institution's legal advisor or delegate is responsible for legal review of these contracts. When the sponsor provides the contract, the investigator and the institution's legal advisor or delegate reviews the contract to ensure that it contains the responsibilities assigned to the investigator's site, indemnification language, budget information, reporting requirements, and any other information required by the legal council. If the sponsor does not provide a contract, the investigator's site will provide an agreement to the sponsor that covers the same elements. When completed, the authorized institutional official for the investigator site and the investigator review all agreements. Usually three originals are prepared and signed: one for the investigator, one for the investigator's site, and one for the sponsor. It is important to allow sufficient time for both the sponsor and investigator's site to review and amend the contract or agreement. If possible, this process should begin at the same time as the ethics review.

### 6.2.4 Sponsor Interactions

The investigator is responsible for complete and proper study communication with the sponsor. A number of study visits are typically required by industry-sponsored trials. In investigator-initiated trials, which often have less funding available, fewer site visits are conducted.

The first visit is referred to as the *prestudy inspection visit*, and the purpose of this visit is for the sponsor to ensure that the clinical site is equipped to perform the study properly. The investigator should provide a tour of the facility and describe the patient population base and the methods that will be used to enrol participants into the study.

If the prestudy inspection visit is successful and the investigator agrees to participate in the trial, there is an *initiation visit*. The purpose of the initiation visit is for the sponsor to ensure that the investigator and the research associates have a correct understanding of the protocol activities and the methodologies. This is also an excellent opportunity for the investigator or research associates to ask any questions about the trial. In addition, the procedures for reporting adverse events, test article storage, and laboratory tests are reviewed.

Once the study is active and research participants are being enrolled, *ongoing study visits* occur. The frequency of these visits may vary, and the purposes are to ensure that the study protocol is being followed and that all documentation in the regulatory files is up-to-date. Adverse events and protocol deviations are also reviewed and outstanding issues are discussed.

After patient enrolment is complete, a *close-out visit* usually occurs. The purpose of this visit is to review study progress, to discuss how the test article is to be returned to the sponsor, and to determine how any corrections or outstanding issues will be resolved.

Appropriate research staff must communicate regularly with the sponsor, and all critical communication, including telephone calls and emails, must be documented. The investigator or delegate must notify the sponsor about the enrolment of the first research participant, recruitment progress, and if any adverse events occur. The investigator or research associate can also contact the sponsor if there is a question regarding a patient’s eligibility status or about the protocol. All communication should be recorded in the regulatory binder.

**6.2.5 Regulatory Documentation (Essential Documents)**

Most countries take measures to regulate the development and marketing of medications and medical products. The investigator is ultimately responsible for obtaining, maintaining, and storing all required clinical study documentation at the site, although this task can be delegated to the research staff. An essential document is defined as any document that permits the evaluation of the conduct of a study and the quality of the data produced. Examples of essential documents include the study protocol, study manuals, and research ethics communications. When a study is being planned, it is important to collect and organize all regulatory documentation in an organized fashion. As the study progresses, copies of all documents are stored in the regulatory binder.

**6.3 PATIENT SELECTION**

**6.3.1 Research Ethics Board Approval (Institutional Review Board Approval)**

An ethics committee is an independent body of medical professionals and lay members. The responsibility of an ethics committee is to ensure the safety, well-being, and human rights of the research participants. Ethics committees review the protocol and consent forms to ensure that the trial is justified, safe, and that the patients are properly informed. All research involving human subjects should be referred to the local IRB or REB. The responsibilities of the IRB/REB are listed in Table 1.

**6.3.2 Feasibility**

Assessing the study’s feasibility is closely related to reviewing an offer to participate in a clinical study. The eligibility criteria must be carefully assessed to ensure that

**TABLE 1 Responsibilities of the Ethics Committee**

|                                                                                                                                                                                     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| To review an application for ethical approval of a research protocol in a reasonable time                                                                                           |
| To consider the qualifications of the investigator                                                                                                                                  |
| To review each ongoing trial at an institution                                                                                                                                      |
| To recommend modifications to the patient information and consent form, when appropriate                                                                                            |
| To review payments to trial participants                                                                                                                                            |
| To determine that, when necessary, the trial protocol addresses ethical concerns such as consent by a patients legal representative and studies where prior consent is not possible |
| To perform duties in accordance with written operating procedures                                                                                                                   |
| To retain all relevant records for at least 3 years after completion of a trial                                                                                                     |

your clinical site has sufficient eligible patients to be recruited for a clinical study. It is often necessary to perform a survey of potentially eligible patients over a 4-week period to provide an accurate estimate of recruitment rates. If you do not have a sufficient number of patients to make participation worthwhile, completing the trial successfully is not feasible.

As mentioned previously, the time required for the overall coordination, the patient enrolment and follow-up, and the data collection also needs to be carefully considered. It is important to be sure that you and your research staff have sufficient time and resources to successfully complete the trial requirements.

### 6.3.3 Sample Size

A sample size calculation before the conduct of a trial can help to estimate an appropriate sample size and to false-negative ( $\beta$  error) results. Before calculating the sample size, the investigator should clearly state primary and secondary outcome parameters. The primary outcome parameter is the one that the investigators consider to be the most important. Subsequently, any other measures are designated to be secondary outcome parameters. The initial distinction between primary and secondary outcome measures is important because the amount of outcome parameters impacts the threshold significance level that needs to be used to determine if the result is significant or not. A significance level of  $p = 0.05$  is used by convention for the main outcome parameter. That means that a chance of 5% is being accepted to conclude that there is a significant difference between two groups, when in fact there is none (type I error or  $\alpha$  error). For any additional secondary outcome parameters, adjustments of the significance level need to be made depending on the number of analyzed parameters.

The magnitude of the difference in the primary outcome parameter that the investigators consider clinically relevant should be the basis for the sample size calculation. Alternatively, this difference can be simply hypothesized. The sample size calculation will reveal how many participants per group are necessary in order to show if that difference truly exists or not.

In addition to the hypothesized difference in the primary outcome parameter and the significance level (usually  $\alpha = 0.05$ ), the acceptable power of the study and the anticipated standard deviations of the primary outcome parameter in the two groups need to be established before proceeding with the sample size calculation. A study power of 0.8 is a conventionally accepted standard, which means that the investigators are willing to accept a 20% probability that there is no difference between two groups when a difference actually exists ( $\beta$  error). Any increase in study power and decrease of the  $\alpha$  level of significance will result in a higher sample size requirement. The anticipated standard deviations in the two groups can be either determined by performing a preliminary pilot study or from data in the literature. If no data is available, they only can be estimated.

Even at best, a sample size calculation is based upon the best available “estimate” of treatment difference between treatment groups. To improve the reliability of an a priori sample size calculation, investigators can conduct a pilot study of 20–50 patients to gain an estimate of the treatment effect in their proposed study population.



### 6.3.4 Recruitment Methods

Patient recruitment strategies vary depending on whether the study is investigating a chronic or an acute condition. For a chronic condition, research staff can screen pre-existing patient data, schedule appointments for potentially eligible patients, and identify patients who may be willing to participate in the study immediately after receiving approval from the ethics board. With a chronic condition, a known patient population can be recruited early and all at once. Patients with an acute condition can only be recruited when they present to the participating physicians; so enrolment will be staggered.

### 6.3.5 Screening

It is important to screen to identify potential patients who are eligible for participation in the clinical trial. Screening refers to identifying potentially eligible participants by applying the clinical trial's eligibility criteria. Areas that can be used for screening include hospital admission records, operating room schedules, and clinic appointment schedules. A screening log is used to maintain the screening information on each potential study participant that is screened, including if they met the inclusion and exclusion criteria, if informed consent was provided, and if they qualified and were entered into the study.

### 6.3.6 Informed Consent

Informed consent is the process by which a patient voluntarily confirms his or her willingness to participate in a clinical trial. Prior to giving consent, the study investigator or delegate must inform the patient of all aspects of the trial (Table 2).

**TABLE 2 What Patients Need to Know Before Participating in Clinical Research**

---

|                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| The purpose of the research                                                                                                                             |
| The trial treatment(s) and the probability for random assignment to each treatment                                                                      |
| The trial procedures to be followed, including all invasive procedures                                                                                  |
| The subject's responsibilities                                                                                                                          |
| Any aspects of the trial that are experimental                                                                                                          |
| The reasonably foreseeable risks and benefits                                                                                                           |
| Any alternative treatment(s) that may be available                                                                                                      |
| Any compensation available to the patient in the event of a trial-related injury                                                                        |
| The anticipated payment or expenses, if any, to the patient for participating in the trial                                                              |
| The patient's participation is voluntary and the patient may refuse to participate or withdraw from the trial at any time without prejudice             |
| Who will have access to their original medical records                                                                                                  |
| The records identifying the patient will be kept confidential                                                                                           |
| If the results of the trial are published, the patient's identity will remain confidential                                                              |
| The patient will be informed if information becomes available that may be relevant to the patient's willingness to continue to participate in the trial |
| The person to contact for additional information on the trial                                                                                           |
| The foreseeable circumstances or reasons under which the patient's participation in the trial may be terminated                                         |
| The expected duration of participation                                                                                                                  |
| The approximate number of patients involved in the trial                                                                                                |

---

**TABLE 3 Three of the Most Fundamental Rights**

---

|                                                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. The patient's participation in the research trial is voluntary.                                                                                                                                                  |
| 2. The patient has the right to refuse to participate or withdraw from the trial without providing a reason. Refusing to participate or withdrawing from the trial will not affect his/her subsequent medical care. |
| 3. The patient will be informed of any new findings that may affect his/her willingness to continue participating in the trial.                                                                                     |

---

Informed consent must be documented in a written form, signed, and personally dated by the patient or by the patient's legally acceptable representative and by the person who conducted the informed consent discussion. Conduct the informed consent discussion in a quiet room and allow adequate time for questions. During the discussion, it is vital to communicate in nontechnical language and to take into account any language barriers.

Patients who participate in research trials have many rights and they need to be informed of these rights during the informed consent discussion. Three of the most fundamental rights are listed in Table 3.

**6.3.7 Patient Confidentiality**

The principal investigator and all study staff are responsible for maintaining patient confidentiality throughout the clinical trial. Confidentiality refers to the prevention of disclosing a research participant's identity and medical information to nonauthorized individuals. The study participant's involvement in a clinical trial must be kept private between the principal investigator, the appropriate study staff, and the primary care physician. The information is only to be shared when there is written permission by the study participant. All research participants' names and data obtained from medical records must be kept confidential. Patients are identified through a number that is assigned to them at study enrollment. When providing source documentation (medical records) to the sponsor, all personal identifiers must be removed. Finally, all data must be stored in a secure area, such as a locked cabinet or a password-protected computer in a locked office.

**6.3.8 Intention to Treat Analysis**

Intention to treat analysis is a method of data analysis on the method of the intention to treat a research participant (i.e., the planned treatment regimen) rather than the actual treatment regimen given. It has the consequence that participants allocated to a treatment group should be followed up, assessed, and analyzed as members of that group regardless of their compliance with therapy or the protocol, or whether they crossed over later to the other treatment group.

**6.4 TREATMENT SCHEDULES**

**6.4.1 Supplies**

The investigator or delegate is responsible for the accurate and complete accountability of the clinical supplies (i.e., drugs, surgical equipment, laboratory equipment,

etc.) used in a clinical trial and for proper storage of the supplies according to the sponsor's written instructions. The study protocol or the investigator's brochure will provide information about the study test article, its proper use, and the required storage conditions during the trial. The study staff is responsible for assigning study identification numbers and test article numbers to the supplies and maintaining the records of the use of supplies. The test articles are often blinded, and the blind must be maintained as instructed in the study protocol. A note in the file must be prepared to describe the situations when and how the blind is broken.

#### **6.4.2 Adherence to Protocol**

It is vital that all procedures and activities described in the study protocol are followed throughout the entire trial. Adhering to the protocol ensures reliable results, high-quality research, and efficient trial progress. Any deviations from the protocol need to be reported to the sponsor as soon as possible after they occur. Most trials have specific case report forms for recording and reporting protocol deviations.

#### **6.4.3 Follow-up**

It is the responsibility of the participating center to ensure that all patients are followed up as outlined in the study protocol. Follow-up visits should be scheduled within the time frames indicated in the protocol. The research participants must be made aware of all procedures that will be performed and the time each one will take (i.e., blood tests, radiographs, gait analysis). It is also necessary to communicate with any hospital departments that are assisting with the follow-up (i.e., laboratory medicine, radiology, physiotherapy). A plan of action should also be in place to prepare for any participants who are withdrawn or who do not attend their follow-up appointment.

#### **6.4.4 Monitoring Visits**

The purpose of a monitoring visit is to verify that the rights and well-being of patients are protected. The monitor will ensure that patients are being enrolled according to the protocol, and that the consent procedure has been carried out correctly and that the patients fulfill the eligibility criteria. In addition, the monitor will check the case report forms for legibility, accuracy, and completion of all data points. Some of these are then compared with the source documentation, such as notes in the medical charts, patient diaries, questionnaires, and test results, to ensure accuracy and completeness of reporting. The monitoring visits also allow time to discuss any questions about the study protocol or issues that have arisen that are not covered in the protocol.

The monitoring visits typically occur after the first patient has been enrolled and then every 2–6 weeks, depending on the rate of patient enrollment. To prepare for a monitoring visit, the research staff should check that all the administrative paperwork is in order, assemble the source documents, ensure that the case report forms are clean and all queries are resolved, and notify all departments involved in the study of the monitor visit as the monitor may need to check these areas.

### 6.4.5 Composite Endpoints

A composite endpoint is an endpoint that is defined in terms of two or more primary clinical endpoints at the patient level. The rationale for including a composite endpoint as opposed to a primary endpoint is that diseases often need multidimensional characterization, low event rates in component endpoints, may need to account for mortality, and the treatment effect on primary endpoints may be small. For example, a cardiovascular study could have three primary endpoints: cardiovascular death, stroke, and myocardial infarction. The composite endpoint would be time to the first occurrence of cardiovascular death, stroke, and myocardial infarction.

When event rates are low, the use of composite endpoints in clinical trials allows investigators to reduce sample size and the duration of follow-up. These advantages come at a price: The interpretation of the effect on the intervention is complicated, and the combined endpoint can be profoundly misleading. It is important to report the event rate for each component of the outcome when reporting on trials using composite endpoints.

## 6.5 PATIENT RESPONSE EVALUATION

### 6.5.1 Outcome Measures

It is important to document all trial-related activities to provide a lasting record of how the trial was conducted and of the data that was collected. Study outcomes may be based on safety, efficacy, or another trial objective and must be clearly defined in the protocol. Outcomes are typically recorded on the case report forms.

### 6.5.2 Case Report Forms

Data collected during a trial should be tailored to answer the hypotheses that are proposed within the protocol. To ensure that the correct data are collected, case report forms (CRFs) are designed. Patient data from source documents, such as the patient's medical records, are used to complete the CRFs. The completed forms are sent to the coordinating center responsible for the collection of the trial data. The coordinating center will then ensure that the final data set is complete, accurate, and complies with good study methodology.

The completion of CRFs will vary from study to study. All CRFs must be completed accurately and legibly and be reported in a timely manner. They must correctly reflect the data found in the source documents (medical records). All questions must be answered in an attempt to limit the number of data queries. Any changes or corrections must be dated and initialed and should not obscure the original entry. Finally, the CRFs must be completed in accordance to the trial-specific instructions. Table 4 describes data that should be included in a case report form.

### 6.5.3 Adverse Event Reporting

A serious adverse event (SAE) is any adverse occurrence or response to a drug/intervention, whether expected or not, that requires in-patient hospitalization or prolongation of existing hospitalization, that causes congenital malformation, that

**TABLE 4 Data Often Collected on Case Report Forms**


---

|                                                      |
|------------------------------------------------------|
| Patient identification (i.e., study ID and initials) |
| Patient demographic details (i.e., age, sex, height) |
| Adherence to protocol inclusion and exclusion        |
| Baseline medical history                             |
| Diagnosis                                            |
| Medication prior to procedure                        |
| Treatment details                                    |
| Tracking of adverse events and other key outcomes    |
| Discharge details                                    |
| Follow-up visits                                     |

---

results in persistent or significant disability or incapacity, that is life threatening, or that results in death. The principal investigator is ultimately responsible for ensuring that all serious adverse events are correctly reported to the sponsor and the REB/IRB in the required time frame (often 24 hours). Typically, the following information is required: type of event, whether it is expected or unexpected, a description of the SAE, action taken, the outcome, and whether the research participants remain on the study protocol. In addition, the local study investigator is asked whether the SAE should warrant continuation of the study, changes to the protocol, or revisions to the information or consent form.

An adverse event (AE) is any untoward event that significantly affects the research participant's well-being and does not fit the criteria of a SAE. It is often optional to report adverse events to the REB or IRB; however, it is necessary to report them to the sponsor.

Before the trial begins it is a good idea to familiarize yourself with the safety profile of the product and the SAE and AE reporting procedures so that you know what to do when an SAE or AE occurs. During the trial, it is important to inform patients of any potential AEs and encourage them to report all AEs. It is necessary to follow all patients who have experienced an SAE or AE until their condition is stable or resolved.

#### **6.5.4 Data Queries**

Although every effort is made to ensure the correct completion of case report forms, errors can and do occur. To guarantee data quality, a system must be in place to check and query all data. Data queries are a necessary part of a clinical trial, as they help ensure the quality of the data and the integrity of the trial. Data queries are generated by the data management team in response to missing, inconsistent, or illegible data, or to protocol deviations (Table 5). It is best to respond to data queries as quickly as possible. It is also sometimes necessary to discuss data queries with the study monitor or project manager if clarification is required.

#### **6.5.5 Adjudication**

A blinded, central adjudication of outcomes can be used in clinical trials as a way of reducing bias and random error in determining outcome events. This process may be especially important in clinical trials when the intervention cannot be blinded,

TABLE 5 Purposes of Data Queries

|                                                                        |
|------------------------------------------------------------------------|
| To clarify or confirm data                                             |
| To request missing data                                                |
| In special cases, to request additional data not specified in the CRFs |
| As a teaching medium for the correct completion of study documents     |

such as in many surgical trials, or when the diagnosis of the primary outcome has low observer agreement. In addition to determining outcomes, central adjudication has also been used to assess eligibility of patients, protocol violations, and co-interventions.

There are many factors for clinical investigators to consider when deciding whether to use a central adjudication process for determining outcomes in a trial. Most importantly, the investigators must weigh the expected benefit of adjudication for accurate determination of outcomes against the substantial investment of resources involved and practicality undergoing this process. The investigators must also consider the potential for the adjudication process itself to bias the results of the trial.

To centrally adjudicate outcomes for a trial, it takes a considerable amount of administrative and expert time to collect the relevant information, prepare the information for review, review each case, and participate in consensus meetings. There may also be challenges involving the availability, validity, or usefulness of some documentation sources. If the treating physician is also required to make a judgment about whether an outcome occurred, it may be possible to compare the committee's judgment with the physician's to determine whether central adjudication is necessary.

Once the investigators or sponsors decide to centrally adjudicate outcomes, they must make several other decisions about the process, including who the adjudicators will be, what material must be evaluated, which judgments must be made, how to train the adjudicators, establishing a set of standardized decision rules, what the committee size will be, whether decisions will be made in pairs or by the full committee (if in pairs, whether to assign cases randomly), how to reach consensus on a decision, and how to monitor the accuracy of the process.

The agreement among adjudicators on a particular case can be affected by the number of decisions necessary, the number of choices for the outcome, and the complexity of the judgments. Disagreements can result from forgetting a decision rule, encountering a case for which there is no relevant decision rule, not having enough information, making an error, or the outcome is difficult to determine given all of the relevant information.

Outcomes assessment/adjudication committees review important endpoints reported by trial investigators to determine whether or not they meet protocol-specified criteria. Members of the adjudication committee may request radiographs, chart notes, operative reports, and other pertinent material to guide their decision making about a defined outcome. All attempts should be made to blind the committee to treatment allocation. This includes careful masking of all X rays and reports. Such committees are most desirable when the assessment of outcomes requires an element of judgment or subjectively (i.e., fracture healing), or when the intervention cannot be blinded. Ultimately, the adjudication committee members work together to limit bias in the outcomes assessment of a clinical trial.

## 6.6 DESIGN CONSIDERATIONS

### 6.6.1 Study Design (Observational Studies and Randomized Controlled Trials)

The types of study designs used in clinical research can be classified broadly according to whether the study focuses on describing the distributions or characteristics of a disease or if it focuses on elucidating its determinants (Table 6). Descriptive studies describe the distribution of a disease, particularly what type of people have the disease, in what locations, and when the disease occurred. Cross-sectional studies, case reports, and case series represent types of descriptive studies.

Case reports are an uncontrolled, descriptive study design involving an intervention and outcome with a detailed profile of one patient. Expansion of the individual case report to include multiple patients is a case series. Although descriptive studies are limited in their ability to make causal inferences about the relationship between risk factors and an outcome of interest, they are helpful in developing a hypothesis that can be tested using an analytic study design.

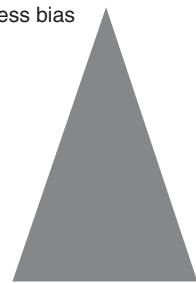
Analytic studies focus on determinants of a disease by testing a hypothesis with the ultimate goal of judging whether a particular exposure causes or prevents disease. Analytic design strategies are broken into two types: observational studies, such as case-control and cohort studies, and experimental studies, also called clinical trials. The difference between the two types of analytic studies is the role that the investigator plays in each of the studies. In the observational study, the investigator simply observes the natural course of events. In the trial, the investigator assigns the intervention or treatment.

One type of observational study is the case-control study, which starts with the identification of individuals who already have the outcome of interest, cases, and are compared with a suitable control group without the outcome event. The relationship between a particular intervention or prognostic factor and the outcome of interest is examined by comparing the number of individuals with each intervention or prognostic factor in the cases and controls. Case-control studies are described in greater detail later.

In the cohort study design, the cohort represents a group of people followed over time to see whether an outcome of interest develops. Ideally, this group meets a

**TABLE 6 Hierarchy of Evidence**

| Study Design                                              |           |
|-----------------------------------------------------------|-----------|
| Randomized controlled trials                              | Less bias |
| Controlled trials (e.g., controlled before-after studies) |           |
| Case-control studies and cohort studies                   |           |
| Cross-sectional studies                                   |           |
| Expert opinion, case reports, and case series             | More bias |





level of certain predetermined criteria representative of a population of interest and is followed with well-defined outcome variables. Usually, this group is matched with a control population selected on the presence or absence of exposure to a factor of interest. The purpose of this type of study is to describe the occurrence of certain outcomes with time and to analyze associations between prognostic factors and those outcomes.

Randomized controlled trials classically are held as the standard against which all other designs should be measured. In a randomized controlled trial, patients are assigned to a treatment group or a control group. The control group usually receives an accepted treatment or no treatment at all, whereas the treatment group is assigned the intervention of interest. Randomized controlled trials are thought to represent the highest quality of evidence based on their methodological strengths of randomization of patient assignment and blinding of intervention and outcome.

### 6.6.2 Trial Organization and Responsibilities

We will provide an overview of the organization of a multicenter trial with emphasis on the trial committees, methods and coordinating center, and participating sites. Figure 1 details how a trial may be organized.

The complexity of a trial with multiple participating centers and hundreds of participating investigators requires key organizing committees to overlook the conduct of the trial, to assure patient safety, and to limit bias in outcomes assessment.

The steering committee is responsible for the overall design and conduct of the trial. The members of this committee may or may not be direct participants of the proposed trial. Often the committee consists of the principal investigators, a biostatistician, a trial methodologist, and other key individuals deemed important to the design and conduct of the study. The steering committee communicates with the trials coordinating center, data safety monitoring board, outcomes adjudication committee, and participating sites on a regular basis. At the completion of the trial, the steering committee maintains responsibility for the data analysis and manuscript preparation on behalf of all study investigators and participating sites.

All clinical trials require safety monitoring, but not all trials require monitoring by a formal committee external to the trial investigators. Data Safety and Monitoring Boards (DSMBs) or Data Monitoring Committees (DMCs) have generally been established for large, randomized multicenter studies that evaluate interventions intended to prolong life or reduce risk of a major adverse health outcome.

Outcomes Adjudication Committees review important endpoints reported by trial investigators to determine whether they meet protocol-specified criteria. Outcomes adjudication is important when the primary outcome of a study requires judgement and is prone to bias in its assessment, especially if the assessment cannot be blinded.

The foundation of a large multicenter surgical trial is the methods and coordinating center. All the day-to-day activities, including centralized randomization, data management, and overall coordination of the trial, occur at this site. The coordinating center can be a contract research organization that “monitors” the trial for a group of clinical investigators. Alternatively, it can be the site of a principal investigator.

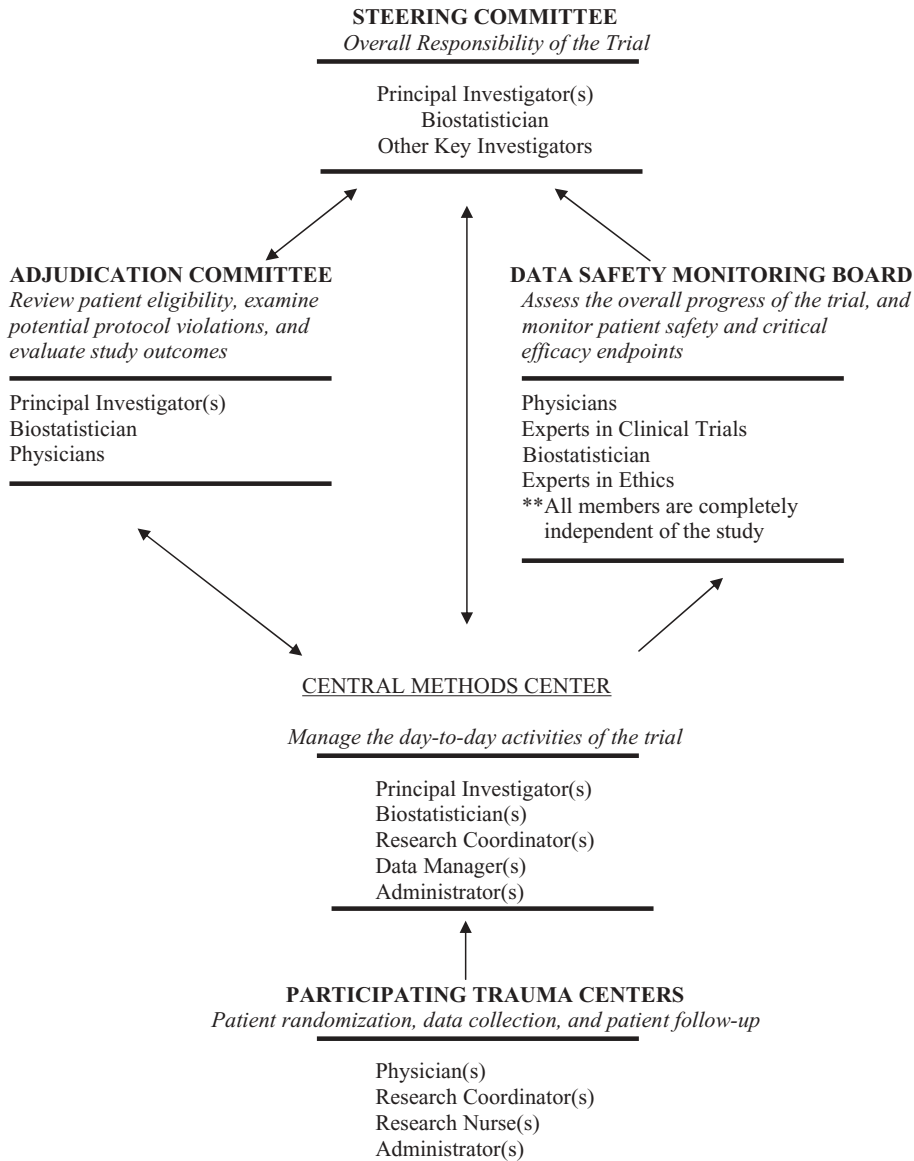


**FIGURE 1** Trial organization.

Each participating site often has more than one investigator enrolling patients for a multicenter trial. In this situation, one investigator from each site is designated as the “site principal investigator (PI).” This site PI serves as the primary contact for the methods and coordinating center and acts on behalf of the participating center. Each participating center also has a dedicated research coordinator who manages the day-to-day issues of patient enrollment, data collection, and follow-up planning. The research coordinator and the site PI work together to ensure compliance, data quality, and effective communication with the methods center.

Participating centers typically identify patients through direct within-center referral or referral from other medical centers. Clinical centers should receive complete sets of all data forms prior to joining the study. The research coordinator for each participating center will ensure that all forms are complete and faxed to the methods center as soon as completed. Each site should be given a follow-up schedule that details the type of information and forms to be completed.

**6.6.3 Data Management**

Collection of data by the site investigator is often the first step in a complex procedure leading to a clean database and final study report generation. Failure to adhere to data collection timelines can significantly delay this procedure. Case report form data pages and additional information is logged, and then the data is entered or scanned into a trial database. Any data queries are faxed or sent to the center for completion or resolution and the database is subsequently amended. Data is usually double-verified to check for accuracy of the data entry.

**6.6.4 Quality Control**

Quality control methods are operation systems and processes established to ensure the quality of a clinical trial, the accuracy and integrity of data, and the compliance with regulations. The clinical investigator is responsible for ensuring that quality control methods are conducted for each study at their clinical site. It is important for sites to have standard operating procedures (SOPs) to help ensure quality and consistency for both general research procedures as well as for procedures for each clinical study. The purpose of the SOPs is to describe in detail who will do each research task and how each task will be completed. Most research institutions have SOPs available for review.

**6.6.5 Study Close-out Activities**

A number of activities need to be completed by the study staff before the close-out of a study. Table 7 lists several of these items. It is important that everything is organized and complete for the close-out visit.

**TABLE 7 Items to Be Completed for Close-out of Clinical Trial**

---

|                                                                                 |
|---------------------------------------------------------------------------------|
| All monitoring visits                                                           |
| All case report forms                                                           |
| Any corrections on the case report forms                                        |
| Test articles from research participants                                        |
| Regulatory binder                                                               |
| Return all study material not used                                              |
| Completion report to the REB or IRB                                             |
| Arranges for secure storage of CRFs, source documents, and regulatory documents |

---

## BIBLIOGRAPHY

- Bhandari, M., and Schemitsch, E. H. (2004a), Randomized trials: A brief history and modern perspective, *Tech. Orthop.*, 19, 54–56.
- Bhandari, M., and Schemitsch, E. H. (2004b), Planning a randomized trial: An overview, *Tech. Orthop.*, 19, 66–71.
- Bhandari, M., and Schemitsch, E. H. (2004c), Beyond the basics: The organization and coordination of multicenter trials, *Tech. Orthop.*, 19, 83–87.
- Bhandari, M., and Tornetta, P. (2004), Issues in the hierarchy of study design, hypothesis testing, and presentation of results, *Tech. Orthop.*, 19, 57–65.
- Bristol, D. R. (1989), Sample sizes for constructing confidence intervals and testing hypotheses, *Stat. Med.*, 8, 803–811.
- Canadian Institute of Health Research, <http://www.cihr-irsc.gc.ca/>; accessed Dec. 2005.
- Flather, M., Aston, H., and Stables, R. (2001), *Handbook of Clinical Trials*, ReMedica Publishing, London.
- Food and Drug Administration (1998), Statistical principles for clinical trials, *Fed. Reg.*, 63(179), 49583–49598.
- Food and Drug Administration (2001), Guidance for Clinical Trial Sponsors: On the establishment and operation of clinical trial data monitoring committees. Center for Biologics Evaluation and Research, November.
- Guyatt, G. H., and Rennie, D., eds. (2001), *User's Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*, American Medical Association Press, Chicago.
- Hartz, A., and Marsh, J. L. (2003), Methodologic issues in observational studies, *Clin. Orthop.*, 413, 33–42.
- Jackowski, D., and Guyatt, G. (2003), A guide to measurement, *Clin. Orthop.*, 413, 80–89.
- McMaster University, Hamilton Health Sciences, and St. Joseph's Healthcare (2004), Standard Operating Procedures for Clinical Research, May.
- Montori, V. M., Busse, J. W., Permyer-Miralda, G., Ferreira, I., and Guyatt, G. H. (2005), How should clinicians interpret results reflecting the effect of an intervention on composite endpoints: Should I dump this lump? *ACP J. Club*, 143(3), A8.
- Sprague, S., Hanson, B., and Bhandari, M. (2003), Informed consent: What your patients need to know about entering clinical research studies, *Can. J. Diag.*, 20(10), 29–31.
- Zlowszki, M., Bhandari, M., Driver, R., Obremskey, W. T., and Kregor, P. (2004a), Beyond the basics: Internet-based data management, *Tech. Orthop.*, 19, 88–93.
- Zlowszki, M., Bhandari, M., Brown, G., Cole, P., and Swiontkowski, M. F. (2004b), Planning a randomized trial: Determining the study sample size, *Tech. Orthop.*, 19, 72–76.



# 7

---

## *Process of Data Management*

**Nina Trocky<sup>1</sup> and Cynthia Brandt<sup>2</sup>**

*<sup>1</sup>The University of Maryland Baltimore School of Nursing*

*<sup>2</sup>Center for Medical Informatics, Yale University School of Medicine,  
New Haven, Connecticut*

### **Contents**

|       |                                                      |     |
|-------|------------------------------------------------------|-----|
| 7.1   | Introduction                                         | 186 |
| 7.2   | Data Management Study Plan: Overview and Definitions | 186 |
| 7.3   | Quality Plan                                         | 187 |
| 7.3.1 | Quality Control                                      | 188 |
| 7.3.2 | Quality Assurance                                    | 189 |
| 7.3.3 | Continual Quality Improvement                        | 190 |
| 7.4   | Data Management Team Structure                       | 192 |
| 7.5   | Case Report Form Design and Guidelines               | 194 |
| 7.5.1 | Overview and Definitions                             | 194 |
| 7.5.2 | CRF Design and Development                           | 194 |
| 7.5.3 | Measurement Basics                                   | 196 |
| 7.5.4 | Standards for Data                                   | 196 |
| 7.5.5 | Testing and Reviewing CRFs                           | 197 |
| 7.5.6 | Summary: Well-Designed CRFs                          | 197 |
| 7.6   | Data Acquisition and Handling Guidelines             | 198 |
| 7.6.1 | Definitions and Overview                             | 198 |
| 7.6.2 | Data Acquisition                                     | 198 |
| 7.6.3 | Data Handling                                        | 199 |
| 7.7   | Summary                                              | 201 |
|       | References                                           | 201 |

## 7.1 INTRODUCTION

“Collecting data that are accurate, honest, reliable and credible is one of the most important and most difficult objectives of conducting clinical research” [1]. The process of data management is composed of various discrete and interrelated activities that are intrinsically linked in an individual and building in a successive fashion. Once the processes and activities are identified, they can then be implemented and managed. Once initiated and implemented, they can be measured and monitored.

There is a need to have a structured and orderly approach to manage these complex processes, and a data management plan may help provide this structure. These plans may be referred to by various names such as data management plan (DMP), data study plan (DSP), data management study file (DMSF), and others that may be specific to the workplace or environment. In general, these plans do provide direction to conform practice to regulate the complex processes involved. Although they all will be based upon regulatory and sponsor requirements as well as good clinical practice (GCP) guidelines, the actual implementation may vary due to interpretation differences. Successful trial management demands careful planning and deliberate execution.

For the purposes of this chapter we attempt to provide a conceptual framework and structure for the components that may be included in a data management study plan (DMSP). The DMSP will be individualized to each clinical study that is implemented. This chapter is organized by components that might be included in a DMSP. The main sections to be addressed in this chapter include the quality plan, data management team structure, case report form (CRF) design and guidelines, and data acquisition and handling guidelines.

## 7.2 DATA MANAGEMENT STUDY PLAN: OVERVIEW AND DEFINITIONS

The DMSP can provide the framework for how and where the processes and procedures are defined and outlined for individual protocols or studies. The goal of the plan is to help standardize data management practice and minimize practice variations, assure compliance with regulations and institutional policies and procedures, and facilitate consistency of data management while improving study performance.

The components of the DMSP define the road map and the specifics of what, how, who, when, and why and help define the activities at the institutional level. There will be specific elements to support the activities of data management such as data collection, entry, handling, and developing discrepancy resolution guidelines. The entire DMSP should be designed in concert with quality tools that support the incremental review, periodic measurement, and adjustments to process deviations. The quality plan is a key section of the DMSP, and quality concepts and specific quality actions are included at every section in the DMSP.

The end product of the DMSP is quality data that can support testing of the hypothesis and compliance with the regulations and policies. It is important to identify the activities and personnel involved in data management prospectively. Once the trial is initiated, the defined activities must be monitored and measured



against pre-established benchmarks and thresholds to minimize deviations to the plan. The structure of the DMSP must be generic and applicable to all protocols or studies implemented while the content or specific procedures and definitions will be individualized. The DMSP should be used in concert with institutional standard operating procedures (SOPs) and is influenced by all appropriate regulatory and institutional requirements and good clinical practice (GCP) guidelines. It may even reference a manual of operations. The DMSP should emphasize quality and consistency from the design of the protocol through publication. It provides guidelines for ensuring consistent data collection, accurate data transfer, subject safety, regulatory compliance, and safety monitoring and reporting. In general, a DMSP might include the following components:

- Protocol
- Scope of work and study contract
- Data management plan
- Data validation testing guidelines
- Data entry and tracking guidelines
- Data coding guidelines
- Data handling and transfer guidelines
- Quality control and quality assurance plan
- Data management team structure
- Serious adverse events reporting and reconciliation guidelines
- Training guidelines
- Data import and export guidelines
- Study report guidelines
- Database management system specification guidelines
- Annotated case report forms
- Data storage and archiving guidelines

This chapter will describe the four sections to include in the DMSP:

- Quality plan
- Data management team structure
- Case report form design and guidelines
- Data acquisition and handling guidelines

### 7.3 QUALITY PLAN

The definition for good clinical practice is: “A standard for the design, conduct, performance, monitoring, auditing, recording, analysis, and reporting of clinical trials that provides assurance that the data and reported results are credible and accurate, and that the rights, integrity, and confidentiality of trial subjects are protected” [2]. A quality plan is a key element in the DMSP, and quality processes need to be incorporated into all components of data management activities. The purpose

of a quality plan is to decrease the degree of uncertainty in the measurement and reporting of data as well as to increase the reliability and validity of data utilized to test the study hypothesis. It should describe the quality control and the quality assurance procedures to be utilized for the study activities including the percentage of subjects for whom source documents will be monitored. Three components of the quality plan deserve particular attention: quality control, quality assurance, and continual quality improvement. Together these three approaches will measure and monitor the characteristics and traits of the data and related processes that support the clinical data management operations. They build upon each other to complement and encompass more and more processes that can be measured and analyzed and if needed, corrected. The quality components are built upon the stabilizing elements of the standard operating procedure, training processes, and basic trial conduct such as data collection and analysis activities. It is this combination of interwoven and interdependent components that facilitate the proper planning and execution of data management activities, which facilitates optimizing results.

### **7.3.1 Quality Control**

Guidance for Industry E6 Good Clinical Practice: Consolidated Guidance defines quality control (QC) as “The operational techniques and activities undertaken within the quality assurance system to verify that the requirements for quality of the trial-related activities have been fulfilled” [3]. Trial output is ultimately affected by the operational checks and balances throughout the data management processes that are performed at every step of the clinical trials process and applied to each stage of data handling that ultimately affect the trial output [1]. Identification earlier in the process will reduce errors by creating a robust checks and balance system. Implementing QC procedures will help to assure that trial procedures and processes have occurred as well as that the activities have been performed in accordance with some threshold or standard. A deviation or error may be noted soonest and adjusted when there is an established, ongoing measurement and data processing review. Since data management processes are performed by a variety of individuals, QC checks must be incorporated into the daily work flow of each data management team member. Quality control in clinical trials or clinical research should generally encompass all the individual procedures that involve protection of human subjects from research risk and affect reliability of the data in order to assure internal consistency of the trial.

Quality control checks may be thought of as snapshots that provide limited insight into the “health” of the data management activities. These snapshots will be used to decide whether to delve further in order to evaluate or assess data integrity components. QC activities may consist of written SOPs and work instructions to be followed such as the process to resolve data discrepancies or clarifications, data validation procedures, or personnel training requirements. QC checks can set pre-defined thresholds such as levels for acceptance testing for an electronic data capture tool.

The DMSP should include study-specific QC activities or steps for critical activities in the entire data management process from protocol and case report form development and subject accrual to data collection, archiving, and cleaning. QC steps can be performed in real time or midstream and some examples are:

1. Comparing programmed edit checks to the standard study-specific edit checks prior to sign off or implementation of an electronic data collection form for a study. An example could be an allowable data range that could vary across studies or populations such as age- or gender-related laboratory values.
2. Performing source document verification (visually comparing the original data source) of all the critically defined data fields just entered into a data collection instrument or computer system following data entry.
3. Following randomization of a new study subject, verifying the steps and the randomization assignment by double checking the work instruction or standard operating procedure.
4. Following the ordering of an investigational agent, a second qualified individual double checks the dosing based on the protocol investigational agent schema.
5. Reviewing and closing a univariate discrepancy as soon as the electronic database notifies the data entry operator of an expected value.

The quality assurance (QA) focus will take advantage of the cross-functional relationships of the data management team by directing efforts on searching for, identifying, and resolving patterns of weakness within one or more processes [1]. Building upon errors or inconsistencies identified at the operator level through the quality control checks, the quality assurance component of the quality plan can focus on measuring performance outcomes rather than just individual process variations.

### 7.3.2 Quality Assurance

Both QC and QA activities are intimately tied to data management. In contrast to QC, QA activities apply to actions that have been performed in the past or apply to past events. Guidance for Industry E6 Good Clinical Practice: Consolidated Guidance defines QA activities as “all those planned and systematic actions that are established to ensure that the trial is performed and the data are generated, documented (recorded), and reported in compliance with GCP and the applicable regulatory requirements” [4]. These activities can assure that predefined data quality requirements have been met and that compliance with certain standard operating procedures or specific data-handling guidelines are measured.

Two components of QA would be audits of performance and audits of systems. As the DMSP attempts to provide the structure for data management activities to occur in an organized and efficient order, QA audits will serve to measure process and outcomes of the data management activities. Performance audits would examine how closely an individual or a research site adheres to a procedure or a process. For example, a performance-focused audit surrounding data entry could include a second independent data entry operator (DEO) reviewing a sample of fields or CRFs completed by the initial DEO. Compliance with data entry guidelines or discrepancy resolution guidelines might be evaluated. An audit of a system could

involve a review of the procedure in order to identify a deficiency in the SOP versus a deficiency in the process of annotation.

These audits should be independent of all trial-related activities that should be used to examine and measure the degree of compliance to a certain procedure or measurement of a performance indicator such as queries generated per site or per study monitor.

Audit reports may address timeliness, completeness, compliance, and consistency of data as compared to the expected results. Additionally, audit reports may be used to communicate the need for corrective action plans to address deficiencies, inconsistencies, and other operational limitations that would adversely influence overall study conduct. A QA activity may include sampling a portion of data in order to assess overall confidence in the quality of a process, an outcome, or a specific task. QA activities may also be employed to ensure QC procedures have been effective in identifying and correcting errors. Some examples of QA activities that might be included in a DMSP are:

1. Assessing the data management study plan for completeness (required sections have been completed and maintained current).
2. Auditing a sample of the personnel-training file against the role-based requirements.
3. Sampling a number of serious adverse events to determine if the events were reported as required by the study protocol.
4. Running an ad hoc query to length of time each research site takes to resolve open discrepancies.
5. Reviewing the cycle times for the last three studies built by the clinical programmers to determine if each was within the defined time period (e.g., 5 days).
6. Audit query resolution processes of each in-house clinical monitor to the specific SOP.

### **7.3.3 Continual Quality Improvement**

Correcting variations in practice or compliance deficiencies must be addressed, but ongoing surveillance with the goal of improving processes is also important. The continuous quality improvement process is a natural companion to QC and QA tools. Adopted from the principles of total quality management (TQM), continuous quality improvement (CQI) serves as an additional tool to correct inefficient or ineffective processes. CQI is a method of evaluation that is composed of structure, process, and outcome evaluations that focus on improvement efforts [5]. CQI can help institutions enhance existing programs and improve effectiveness of processes because the variation in practice can be viewed from multiple angles. Effective CQI activities depend on the collaborative work environment, which is how data management operations are developed.

Drilling down to identifying a process to improve, even incrementally, can be achieved by incorporating the CQI methodology. The emphasis is on identifying the root cause(s) of the problem and then designing an appropriate intervention to eliminate or reduce the reasons for the problem. So, first a problem or issue must

be identified. Next a corrective action plan is developed. Then the plan is set in action. Finally, the results of the actions are monitored. One relatively convenient model to employ is called the FOCUS PDSA model [6]. The relatively simple steps may be utilized to quickly frame the problem, identify the reasons for the problem, and create an effective intervention. The steps are outlined below and will begin with an understanding and agreement on what the problem is and why it exists:

- Find a process to improve or a problem to correct.
- Organize the interdisciplinary team to discuss process or the problem.
- Clarify what is known and gather supporting documentation.
- Understand why the process variation occurs or the problem exists.
- Select a process improvement activity based on the above analysis.

Once the process or the problem is thoroughly understood and the remedial actions are defined, the next step is to continue through a process improvement plan as:

- *Plan* Create a timeline of resources, activities, training, and target dates to establish the completion date. Develop a data collection plan, tools for measuring outcomes, and define thresholds for determining when targets have been met.
- *Do* Implement interventions and collect data.
- *Study* Analyze the results of data collected and evaluate reasons for variation, if any.
- *Act* Act on what is learned and determine next steps. If the intervention was successful, determine if the current processes or procedure were not followed and so new procedures are not needed. Additional training may be required.

Conversely, practice variations may have resulted because procedure is antiquated or nonexistent. Finally, review of the PDCA system may be necessary as it was not robust enough to propel change. In this case it may be necessary to repeat the PDCA cycle.

Continual quality improvement builds upon the narrow or individually focused efforts of the QC checks and the broader procedure and process focused audits, organizational and system checks of the QA activities. The CQI model enhances the quality plan by directing collaborative energies toward processes not individuals. Data management is process driven. CQI efforts, likewise, focus on critical processes that require movement from one person to another and to another but from a systems perspective. CQI can be thought of as a shared effort that enables people to work together across organizational boundaries to improve shared processes [1].

A robust quality plan will force the data management team to evaluate their practices to measure how well they match their assumptions, formal guidelines, and predefined measurements of performance. In fact, it is the outputs from this plan that will form the team's quality profile. The determination and measurement of quality cannot rest in the hands of a select few. Rather it must be a shared responsibility and expectation among all participants on the data management team.

## 7.4 DATA MANAGEMENT TEAM STRUCTURE

It is important for data management team members to understand the scope of their work perfectly. The team members need to understand the deliverable or a specific work component (a portion of a larger deliverable) that they are expected to complete or assist in completing. If there are quality criteria to meet, the team members should know these quality requirements. More importantly, the data management team members should clearly understand the dependencies and relationships that exist between themselves and the other departments such as regulatory or information technology. The DMSP will include a reporting or organizational chart detailing responsibilities and lines of communication. By delineating the structure and composition of the team, energies may then be directed toward problem solving, task effectiveness, and maximizing resources to achieve the team's purpose. Sound team building recognizes that it is not possible to fully separate one's performance from those of others.

As the data management team structure is comprised of various professionals from diverse disciplines, what then are the primary activities of the data management process? In general, the primary aspects of data management consist of:

1. Handling, processing, and analyzing information for the purpose of supporting clinical research activities
2. Developing protocol-specific case report forms (CRFs)
3. Reviewing and approving programmed validation rules and edit checks in the database systems
4. Collecting data from the medical record or source documents and transcribing it into protocol-specific CRFs
5. Reviewing and evaluating data for inconsistencies, omissions, and errors.

The process of clinical data management knits together various key players who provide specialized skills and knowledge.

Depending on the work environment and setting, there may be different types of personnel performing one or more data management activities. For example, a quality specialist may perform audits in one setting and in another setting may serve as an independent review for final CRF design approval. Various disciplines and personnel comprise the data management team. As such, individual setting or environments may have their own unique set of responsibilities and the position descriptions may vary. Regardless, the key element is that we understand that various disciplines must all be interrelated to manage and process data. Some examples of personnel who are involved in the clinical data management activities are:

- *Data Manager* Core or central person who coordinates the various activities for a specific project assuring project goals and objectives are met. May develop protocol-specific CRFs, define programmed edit checks for electronic CRFs, identify critical safety data fields, and coordinate final approvals of CRFs. The data managers also may develop data entry guidelines and have other responsibilities depending upon the environment.

- *Medical Writers* Develop and prepare study protocol, safety reports, and for interim analyses.
- *Statistician* Review clinical and safety databases (or data structures and CRFs) to assure study endpoints can be measured and protocol-specific data are collected, determine safety and efficacy endpoints, conduct data safety monitoring board interim analyses. Develop or work with database programmers to create specialized statistical programs to support data integration, data reporting, and safety analysis.
- *Regulatory Specialists* Confirm human subject protection and subject safety are preserved, develop and maintain standard operating procedures or work instructions, conduct training and education classes, and maintain regulatory files; reviews initial and continuing protocol submission to the Institutional Review Board (IRB).
- *Programmers* Build study-specific data collection instrument (tool), revise multistudy clinical research database system or create study-specific database, modify eCRF or data system based on protocol amendments or sponsor safety reports. Work with statistician, study coordinator, or data manager to create data queries for reports and analyses.
- *Investigator* Responsible for study conduct at the research site, approves data collected and entered into the CRF, assures compliance with all regulatory, scientific principles, ethical, legal, and GCP standards, assures approved reports are prepared and safety information is reported to IRB, sponsor, and other investigators.
- *Research Nurse or Study Coordinator* May perform data collection and entry into CRFs, supervise data entry personnel, audit data entered by data entry personnel, coordinate site recruitment, subject screening, and enrollment, and monitoring activities, maintain regulatory binder and essential document file, coordinate training of site personnel, records maintenance, provide required documents/reports to regulatory, sponsor, institutional entities, notifies IRB/sponsor immediately of a serious adverse event (SAE).
- *Monitor* Principal link between sponsor and investigator, assure rights and well-being of human subjects protected, oversees progress of the trial and ensure study is conducted and data are handled in accordance to protocol, GCP, ethical and regulatory requirements, implements SOPs, provide site staff education, verify data is accurate, complete, and verifiable, assures CRFs are completed accurately.

Along with delineating the personnel and their roles and responsibilities in the DMSP, the plan should also include standard operating procedures (SOPs) that detail processes for data handling to assure accuracy, reliability, safety, security, and privacy. These processes will be further outlined in specific guidelines such as data entry guidelines, data-coding guidelines, and data-handling and transfer guidelines and case report form design guidelines. The data management team structure serves to clearly define role delineation from a functional framework as well as a division of responsibilities. Therefore, development of the case report forms must rely on the team members to integrate and apply study-specific guidelines, sponsor, protocol and regulatory requirements, clinical guidance, subject safety, and endpoint



measurements. It is from the case report forms that the data collection processes may occur.

## 7.5 CASE REPORT FORM DESIGN AND GUIDELINES

### 7.5.1 Overview and Definitions

Case report forms are questionnaires or instruments that are used to collect required data about cases or subjects enrolled in a study in a structured and standardized manner to facilitate reliable, consistent, and clean data for analysis. The CRFs, or data collection instruments, should be designed by investigators in order to measure and define the specific data needed to support or disprove the hypothesis or goal of the study. General purposes of the CRFs are to: (1) meet the objectives of research; (2) obtain most complete and accurate information possible; and (3) do this within the limits of available time and resources.

Ideally, CRF templates are designed to help standardize the information and data collected and facilitate collaboration and reuse for future similar studies. CRFs may be created and used in different modalities or formats including paper or electronic documents. The appearance and functionality of the CRFs will reflect the modality and how they will be used and the capabilities of the environment where they will be implemented.

### 7.5.2 CRF Design and Development

**CRF Design Best Practices** When designing CRFs, the content and structure of the data items (or questions) to be included in the CRF should be considered first rather than the modality or appearance of the CRF. It is best to design the CRFs at the same time that the protocol is being created in order to assure that data specified in the protocol is collected and consistent with the hypothesis of the study and feasible to collect. This will help to keep questions, prompts, and instructions clear and concise and help to assure that the CRF design will fit with the data flow from the perspective of the person completing it. The flow of study procedures and typical organization of data in a medical record (or data source) should be taken into account as well.

Other design issues include planning for reusability, collaboration, and standardization when possible. It is helpful to create and maintain a library of standard forms. If there are previously performed studies or pilot work that has been used for measuring similar outcomes or measurements, CRFs and previous items may be used and adapted as necessary using experiential information. Finally, design the CRF with the primary safety and efficacy endpoints in mind as the main goal of data collection and pretest and review them prior to finalization and approval.

Broadly conceptualizing the CRFs in the study. As discussed in previous chapters of the book, clinical studies generally are designed in a highly structured fashion, and many are divided into chronological periods of varying duration, with different evaluations and tests performed or scheduled for subjects at different time periods. Investigators need to determine how and when the data will be collected and

entered and what the reporting requirements will be. This information can assist in determining the design of the CRF that will fit with the data flow of the study procedures and hopefully decrease the tendency to measure data inappropriately and redundantly. Commonly used time points in a clinical study include pretreatment, screening, baseline, treatment, follow-up, and evaluation or summary points. It is important to remember that the time point a CRF is collected is important data as well as the individual data items as CRFs may be used at numerous times or only once during a study. An overall study data schedule will need to be examined along with the individual CRFs that will be collected, prior to study initiation in order to make sure that the proper data is getting collected at the right times and at the correct frequency (repeats) appropriate for analysis to test the study hypothesis(es).

**CRF Level Design** Once the overall study schedule of procedures or visits has been determined CRF designers should then look at form level issues that will facilitate or hinder data collection or entry and data analysis. First consider the type of data the CRF is used to collect and who will be using and completing it. The goal is to make it easy to use. For example, if a care provider such as a nurse or physician will be using the CRF for data collection, questions may be arranged and grouped differently than if the subjects will enter data into the form. The education level, age, language, and culture all must be considered when designing CRFs for subject entry. Likewise the skill level of the data collector in the research setting is also important to know. Questions may need to be worded for less clinically educated data collectors and more or less help made available to provide explanation of complex items. The following information should be described in the DMSP:

1. Designing for the type of data and the flow of subjects in the study.
2. Designing for who will be completing the CRF in order to assure that it makes sense to that person.
3. Ensure that a data item is collected once and only in one place to avoid conflicts in data. This will help to avoid referential and redundant data points within the CRF.
4. Grouping the same type of data together on the same forms.

For considerations of layout, format, and usability, we will focus primarily on issues related to eCRFs. For overall CRF layout and design, it is important to consider the data flow and the dependency of the items on each other when putting the items in an order or group on the CRF. The order of the items should be determined and interrelationships and dependencies between the questions clarified. For example, if the answer to question 1 determines whether the next few items will be relevant or not and skipped should be made clear on paper CRFs and programmed into electronic CRFs. These “skip patterns” can greatly facilitate either confusion or correct use and entry of data depending on formatting and programming for eCRFs.

**CRF Modality** Final considerations to overall CRF design will need to be made knowing the modality for use of the form. For example, will it be implemented electronically on a personal digital assistant (PDA) or on a paper form to be

scanned, faxed, or mailed? Once the modality(s) have been determined, different tools can be used to make the CRFs most visually conducive to clear and accurate data entry. The font size should be easily readable, copy and fax well, and the use of different styles for visual emphasis and clarity of reading should be used to help the form filler identify specific sections and complete the CRF.

### **7.5.3 Measurement Basics**

It is not practical to discuss in detail the large corpus of material on measurement theory that describes how to create questionnaires and how to design scales, as there are numerous textbooks and a large body of research on this subject. A few key practical steps that can help investigators and study staff to design data collection instruments or CRFs are contained in Sudman and Bradburn [7].

When designing items to measure standard concepts, it is best to identify standardized and validated previously used items to include in the CRFs. The use of items from standardized instruments will help assure that the items are appropriately defined. When standard validated questions do not exist, the investigator must create new questions or revise existing questions, taking into consideration question formats, phrasing, responses, and categories. Instructions for completion/prompts should be simple, clear, and sufficient for whoever is completing the form. Investigators should consult personnel trained in the development of measurement instruments, as this is complex and error prone.

When creating electronic CRFs, it is best to enter and store raw and original data rather than calculated or summary data. For example, the collection of birth date, rather than age, is more useful. Free text input are less optimal to forced selection of text choices, but many CRFs will require free text entry. Other data types may be based upon standard vocabularies such as those used to code common procedures or diagnoses, and a table look-up will help accurate data entry and collection. For more complex data types such as images, electrocardiograms (EKGs), and the like, different electronic solutions are available such as picture archiving and communication (PAC) systems and radiology information systems.

### **7.5.4 Standards for Data**

Researchers performing epidemiological and clinical research frequently participate in collaborations for reasons such as: joint-funding initiatives, to increase sample size in the case of rare diseases or outcomes, and to pool expertise and resources. When designing CRFs for collaborative research, it is important to have the entire collaborative team review and test pilot versions of CRFs. The use of standard CRFs or CRF templates with standard questions and data items will facilitate the ability of data to be merged for collaboration postcollection if the design of the CRFs cannot be done pre-data collection. Identification of useful, validated CRFs, however, is difficult for various reasons. For example, sample CRFs may not be available; there may be copyright and use restrictions; there may be multiple, different, versions of the same CRF, and the assumptions underlying the use of a particular CRF may not have been clearly documented and items may not be validated. There are efforts underway to standardize different CRFs for use in clinical research from both researchers and regulatory agencies and organizations. Unfor-

tunately, the majority of questionnaires and CRFs in existence are custom designed by individual investigators and do not achieve wide enough use that they would be expected to become candidates for standardization by different organizations.

### 7.5.5 Testing and Reviewing CRFs

Once the set of CRFs have been designed, they should be reviewed and approved by the entire study team. It is important to have the clinical study team review and test the CRFs before submitting for approval by the various regulatory committees and boards. Data management team members should review pilot and test data for accuracy (of asking, coding, set up) and analyzability. Each member of the team can evaluate the CRFs from a different perspective and purpose. This will then enable comprehensive feedback and review of the CRFs. Specific things to focus on when reviewing CRFs include:

1. The language on CRF is consistent or understandable by the persons who will be using and completing the CRF.
2. The units of the measures collected are consistent with the units used at different sites.
3. For the different types of CRFs (paper, electronic), the coding must be consistent with the database system that will be used to store the data. For example, it is important to check that the correct choice responses and data types are being collected (e.g., one or two decimal points).
4. The statistical team members can confirm that the appropriate analytical data points are complete and in a format needed by entering pilot test data.

As multiteam involvement is helpful during development of CRFs, testing and review after development can facilitate good CRFs that will work for everyone to increase the usability and data quality and reduce the time needed for cleaning data.

### 7.5.6 Summary: Well-Designed CRFs

It is important to develop and design CRFs when developing the protocol and make sure that only data items are measured and collected that can contribute to measurement of the outcomes and success or failure of the study. The balance between asking too many questions or too many data items or CRFs collected must be made to avoid unnecessary costs and burdens to the subjects and data managers. Too much and/or too little can both contribute to problems in conduct and performance and analysis of the study. By focusing on the overview of data and CRFs in the study, the CRF layout and ergonomics, the usability and CRF modality, along with appropriately designed measures, and standards for data items and databases, and finally by testing and reviewing CRFs with the entire study team, the result will be well-designed CRFs that can be used to collect the right data efficiently to produce data ready to be analyzed. By addressing these CRF design concepts in the DMSP, this will facilitate collection of clean data and will result in a reduction of queries from data managers to clarify items.

## 7.6 DATA ACQUISITION AND HANDLING GUIDELINES

### 7.6.1 Definitions and Overview

Data acquisition is the process by which the data is collected [8]. Data handling is a broad concept that we are defining as the multiple processes and activities that are utilized to transport data throughout the clinical study process [8]. The data-handling process covers the entire life cycle of the data from acquisition to archiving. The DMSP should include study-specific procedures for the data acquisition and collection and data-handling processes all necessary in assuring optimal performance, consistency, compliance, and integrity.

### 7.6.2 Data Acquisition

Data acquisition is the actual collection of protocol-specific data. This collection process can frequently be one of the most difficult and error-prone aspects of a study. Collection of data can occur by one or more methods, all ultimately being deposited or stored in the CRFs. Some methods by which data are acquired and collected include:

- Electronic or computer based—This can include eCRFs on personal computers (PC), laptops, PDAs, or other electronic devices used for data entry. Electronic data capture (EDC) can also be included in this modality type. This can be defined as automatic data collection by a device. An example is the glucose sensor monitors that automatically create electronic glucose data from the subject.
- Through in-person or telephone interviews.
- Paper-based CRFs—These can be scanned, faxed, data entered into a computer, and converted into electronic documents.
- Study subject self-administered diaries and questionnaires.
- Medical chart review and abstraction.
- Electronic imports or transfers of data from other systems (such as laboratory data).
- Sound based, telephone interactive voice response systems (IVRS) that connect telephone users using speech recognition technologies to computer systems [9].

The acquisition of various data utilizing one or more data collection methods presents a challenge. Incremental and systematic checks and monitoring should occur to determine if the data to be collected actually was collected. In addition, was data collected in the format that was expected. And, finally, was data acquired in the time frame expected.

Traditionally, paper has been used to collect and even to store clinical data for research studies. This is primarily due to the convenience of paper and the familiarity of staff and subjects with paper. The move toward more technical approaches is continuing, and for this chapter we will focus on electronic methods of data acquisition or collection for clinical studies. The different modalities that are used for data

collection into a CRF will have varying features that need to be described and included in a DMSP.

Many vendors and data management organizations are moving toward data entry into electronic case report forms often referred to as eCRFs. There are several methods that eCRFs can use to help ensure accurate data entry. The methods used should be described in the DMSP and can include:

- Use predetermined ranges of allowable and expected values.
- Provide simple data type checks, for example, ensuring valid numbers or dates in numeric or date fields. For questions that require text responses, string length checking and validation.
- Preset list(s) of choice values should be defined when possible for a restricted list of responses allowed for data entry selection.
- Skip inapplicable questions and set default values based on the response to specific question(s).
- Compute values for a question calculated based upon other values entered in the CRF.
- Intraform validation checks—In this case one value on a CRF is compared to previously entered data on the same CRF to see if it meets predefined criteria.
- Cross-form validation—In this situation, checks are made that compare the response to a question on one CRF to response(s) on a previously entered CRF, or other previously entered data.
- Spell checking—The software should perform automatic spell checking for certain text fields against online. Domain-specific dictionaries can be used for free text entry to check for spelling errors.
- Support entry and query of missing and approximate values.

Data acquisition guidelines should be included in the DMSP as these will be specific to the study and sponsor requirements. These guidelines will complement the data entry, validation, transfer guidelines, also included in the DMSP, which are based upon the project's scope of work, the study phase, system capacity, and other requirements. Following the acquisition of data, it is the handling processes that address the movement, the storage, archiving, and ultimate security of the data. Appropriate planning and safeguards must be established to address both paper and electronic data. For example, data handled electronically requires the added assurance that the integrity of recorded data are not altered, erased, lost, or accessed by unauthorized users.

### **7.6.3 Data Handling**

Individual data-handling processes may be applicable for a given environment and study, and the DMSP should incorporate appropriate sections for the processes and procedures used in the specific study. Many of the sections to be included in the DMSP for data handling will require the description of validation and testing of the processes and activities involved along with other appropriate ongoing quality

checks. For example, for studies that use a clinical data management system for managing study data, there should be sections in the DMSP on quality checks to be performed including validation and testing of the system and of software programming created for the study. Other examples include quality checks of the data processing and cleaning activities (discrepancy identification and resolution). This might involve describing the queries and reports that will be used for monitoring data entry and data-cleaning activities such as data entry completion and error rate monitoring, missing data reports, discrepancy, and discrepancy resolution reports. Based upon the different capabilities of the software systems used for the study, all or some of these activities can be performed and described in the DMSP.

A systematic approach to how the study will handle data changes or changes to data collection instruments or CRFs should be included in the DMSP. This approach should be an agreed upon process in affect prior to CRF development that details how changes in data collection instruments or CRF versions will be documented and handled. This section should also describe how all changes to data would be documented and copies of previous data stored in electronic audit trails. There should be included clear documentation in the audit trail of who, what, when, and why the change was made.

More data is now available in an electronic format and may be available for import and integration into the clinical study data management system. Examples include laboratory data and subject diary data. It is preferable for this data to be electronically transferred and imported/integrated into existing data management systems rather than be rekeyed into the system. In this case, it will be necessary to include a section in the DMSP describing the approach that will be used to perform the import and check the results in a standard way. Different studies also may require data to be transmitted to outside agencies. Data transfer (both in and out) may be specified to be compliant with standard formats or special formats required by the receiver. The format or use of data format standards such as HL7 [10] and CDISC [11] should be described in the DMSP along with methods to test and validate accuracy of the transfer(s).

**Security and Privacy of Records** The DMSP must describe [or reference the institutional standard operating procedure(s)], how the security and privacy of clinical study data will be assured throughout the data management process. This must include both electronic data and other forms of data (paper) if relevant. All the computer systems to be used must have methods to prevent unauthorized access, preserve subject confidentiality, and prevent retrospective tampering/falsification of data. Under the FDA's Title 21 Code of Federal Regulations (21 CFR Part 11) [12], access must be restricted to authorized personnel, the system must prevent malicious changes to research data through selective data locking, and an audit trail must exist. Additionally, compliance with the Health Insurance Portability and Accountability Act (HIPAA) must be assured. The HIPAA Privacy Rule describes how protected health information (PHI) may be used and disclosed. Under HIPAA, research is defined as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge" [13]. The conduct of research and clinical data management, in specific, must incorporate appropriate safeguards to assure compliance with this law.



Data systems also need to be adequately backed-up and recoverable in the event of catastrophic system failure. The physical and electronic security must be assured to keep the data both safe and secure, yet it must also be available to authorized users with password and authentication. Users should have assignment of role-based privileges. Database management systems should have the ability to store and/or generate deidentified data for purposes of analysis and data sharing. All systems must have robust electronic audit trails and allow for archiving and data locking.

## 7.7 SUMMARY

The process of data management is complex, error prone, and vital to reliable accurate data. A data management study plan is a structure by which these complex processes and activities can be organized to assist in a consistent performance of the actions. Beginning the DMSP with a study-specific quality plan and applying the quality processes throughout the entire data management process should result in an increase in the reliability of clinical data, decrease variation and errors, and improve on the timeliness of good clinical quality processes and outcomes that can be defined and measured.

## REFERENCES

1. Bohaychuk, W., and Ball, G. (2001), *Conducting GCP-compliant Clinical Research*, Wiley, New York, p. 127.
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996), Guidance for Industry E6 Good Clinical Practice: Consolidated Guideline; accessed Feb. 6, 2006; [www.fda.gov/cder/guidance/959fnl.pdf](http://www.fda.gov/cder/guidance/959fnl.pdf); glossary section Ref 1.24, p. 8.
3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996), Guidance for Industry E6 Good Clinical Practice: Consolidated Guideline; accessed Feb. 6, 2006; [www.fda.gov/cder/guidance/959fnl.pdf](http://www.fda.gov/cder/guidance/959fnl.pdf); glossary section Ref 1.47, p. 11.
4. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996), Guidance for Industry E6 Good Clinical Practice: Consolidated Guideline; accessed Feb. 6, 2006; [www.fda.gov/cder/guidance/959fnl.pdf](http://www.fda.gov/cder/guidance/959fnl.pdf); glossary section Ref 1.46, p. 11.
5. Donabedian, A. (1988), The quality of care. How can it be assessed? *Journal of the American Medical Association*, 260(12), 1743–1748.
6. Walley, P., and Gowland, B. (2004), Completing the circle: From PD to PDSA, *Int. J. Health Care Quality Assurance*, 17(6), 349–358.
7. Sudman, S., and Bradburn, N. M. (1987), *Asking Questions. A Practical Guide to Questionnaire Design*, Jossey-Bass, San Francisco, pp. 281–282.
8. Society for Clinical Data Management (2005), Good Clinical Data Management Practices, Version 4, October.
9. Interactive voice response systems; accessed Feb. 6, 2006; [http://www.iec.org/online/tutorials/speech\\_enabled/](http://www.iec.org/online/tutorials/speech_enabled/).
10. Health Level Seven, HL7 Standards; accessed Feb. 6, 2006; <http://www.hl7.org/>.

11. Clinical Data Interchange Standards Consortium (CDISC); accessed Feb. 6, 2006, <http://www.cdisc.org>.
12. U.S Food and Drug Administration (2006), Title 21 Code of Federal Regulations (21 CFR Part 11); accessed Feb. 6, 2006; [http://www.fda.gov/ora/compliance\\_ref/part11/](http://www.fda.gov/ora/compliance_ref/part11/).
13. Department of Health and Human Services (2003, April 3), Health Information Privacy. Research; accessed February 12, 2009; <http://www.hhs.gov/ocr/privacy/hipaa/understanding/special/research/>.

# 8

---

## *Clinical Trials Data Management*

**Eugenio Santoro<sup>1</sup> and Angelo Tinazzi<sup>2</sup>**

<sup>1</sup>*Laboratory of Medical Informatics, Department of Epidemiology, “Mario Negri” Institute for Pharmacological Research, Milan, Italy*

<sup>2</sup>*Merck Serono, Global Biostatistics, Geneva, Switzerland*

### **Contents**

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| 8.1 Clinical Data Management Aspects                                    | 204 |
| 8.1.1 Regulatory Framework for Clinical Data Management                 | 204 |
| 8.1.2 From Clinical Protocol to Data Acquisition Tools                  | 206 |
| 8.1.3 Database Design                                                   | 206 |
| 8.1.4 Data Processing                                                   | 208 |
| 8.1.5 Electronic Data Capture Principles                                | 211 |
| 8.1.6 Data Standards                                                    | 213 |
| 8.1.7 Infrastructure Requirements                                       | 214 |
| 8.1.8 Implementation of Clinical Study Data Management System           | 214 |
| 8.1.9 Computer System Validation                                        | 215 |
| 8.1.10 Future: EHR/EDC Integration                                      | 215 |
| 8.2 Web-Based Clinical Trials                                           | 216 |
| 8.2.1 Web-Based Clinical Trials                                         | 216 |
| 8.2.2 Tools for Participating in Web-Based Clinical Trial               | 216 |
| 8.2.3 Why a Clinical Trial Website?                                     | 216 |
| 8.2.4 Examples of Clinical Trial Websites and Web-Based Clinical Trials | 221 |
| 8.2.5 Advantages and Limitations of Web-Based Clinical Trials           | 222 |
| References                                                              | 223 |

8.1 CLINICAL DATA MANAGEMENT ASPECTS

8.1.1 Regulatory Framework for Clinical Data Management

In recent years the medical community has been involved in a great debate concerning ethical issues that include confidentiality and privacy, especially for data included in medical records or collected in clinical studies for research purposes. Furthermore, the improvements in information technology (IT), which made full paper replacement possible and new solutions for the electronic management of data available, have led the regulatory agencies to better clarify the requirements for processing personal data. For these reasons, new rules have been identified for the clinical data management process including compliance with the privacy and data protection laws when personal data are processed, compliance with good clinical practice (GCP) [1] when medical products or medical interventions are tested, and compliance with other regulatory directives when medical products are submitted to specific regulatory agencies (Table 1).

TABLE 1 Regulatory and Guideline References

|                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>International</i>                                                                                                                                        |
| E6: Guideline for Good Clinical Practice—Consolidated Guideline, International Conference on Harmonisation, EU Implementation CPMP/ICH/135/95/Step5         |
| <i>European Community</i>                                                                                                                                   |
| The rules governing Medical Products in the EU, Volume IV, 1998, Annex 11: Computerized Systems                                                             |
| Directive 95/46/EC—Data Protection                                                                                                                          |
| Directive 2002/58—Privacy and Electronic Communications                                                                                                     |
| Directive 1999/93/EC—Community Framework for Electronic Signatures                                                                                          |
| <i>Food and Drug Administration</i>                                                                                                                         |
| 21 CFR Part 11—Electronic Records; Electronic Signature; 2003                                                                                               |
| 21 CFR Part 11—Protection of Privacy                                                                                                                        |
| General Principles of Software Validation; Final Guidance for Industry and FDA Staff; January 2002                                                          |
| Guidance for Industry: Computerized Systems Used in Clinical Trials; September 2004                                                                         |
| <i>Country (Italy)</i>                                                                                                                                      |
| DRP nr. 318—Use of Personal Data and Privacy Protection                                                                                                     |
| DL 30/06/2003 nr. 196—Personal Data Protection                                                                                                              |
| <i>Other Guidelines</i>                                                                                                                                     |
| GAMP Forum, Good Automated Manufacturing Practice—Supplier Guide for Validation of Automated Systems in Pharmaceutical Manufacture; December 2001           |
| ACDM/PSI—Computer System Validation in Clinical Research: A Practical Guide; ACDM/PSI; 1998                                                                 |
| PICS/S Good Practice for Computerised Systems in Regulated GxP Environment, Pharmaceutical Inspection Co-operation Schema Final Guidance, rev. 2; July 2004 |

**Data Privacy** People involved in the data management process should be familiar with basic data privacy issues and should follow the principles established by their organization to ensure the privacy of research subjects and the compliance with the GCP and other international and/or local regulations [2].

Data privacy is related to the standards surrounding the protection of personal data, defined as “any information about a person who can be identified directly or indirectly” (e.g., patient names, initials, addresses, and genetic code). The privacy of any subject who participates in a clinical trial must be protected from the ethical and legal points of view. The data should be protected from accidental loss, alteration, and unauthorized access. For this reason, the implementation of appropriate security measures are required.

To guarantee data privacy, personal data must be handled separately from the clinical data and made anonymous. In addition, a written, signed, and dated informed consent should be asked and obtained from the owners of the data. The concept of data privacy is enforced in the GCP: “The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with applicable regulatory requirement(s)” [1].

Sometimes the identification of an individual cannot be fully masked. In fact, the data management staff usually use several sources of information such as primary medical and hospital records, genetic data, economic data, and adverse drug event reports. In these cases, to ensure proper assignment of data in a clinical database, data collection instruments should be designed with the need for the minimum research subject identifiers (in general, a subject identification number and gender can be used to solve any discrepancies that might arise from transcription errors).

In addition, the application of local laws, such as the Italian one, integrates the statements of the European Union (EU) directive with a technical attachment addressed to the electronic data management [3]. These include detailed requirements related to the use of the user name and password for a secure and diversified data access that is based on the user’s profile, and the use of backup and restoration procedures in order to ensure data integrity.

**Good Clinical Practice and Other Directives** The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Good Clinical Practice Guidelines [1] add an additional layer on the requirements for data management of clinical data (Table 2). One of these is related to the quality control process that should be applied to each stage of data handling in order to ensure that all data have been correctly processed. Another important requirement states that any change or correction to a case report form (CRF) should be dated, signed, and explained (if necessary), and should not obscure the original entry. This must be done by maintaining an “audit trail” and should be applied to both written and electronic changes or corrections. Similar requirements have been proposed by other regulatory guidelines. For example, the Food and Drug Administration (FDA) 21 CFR Part 11 [4] establishes the requirements for electronic records and electronic signatures to be trustworthy, reliable, and equivalent to paper records and handwritten signatures. These requirements, for example, imply data encryption, digital signature standards, and other safety requirements (i.e., device checks).

**TABLE 2 Main FDA 21 CFR Part 11 and GCP Requirements**

| Requirement                                                                      | Part 11   | GCP                        |
|----------------------------------------------------------------------------------|-----------|----------------------------|
| Validation of Computer System                                                    | 11.10 (a) | 5.5.3a                     |
| Data Protection (to enable accurate record retrieval)                            | 11.10 (c) | 2.10; 4.9.1; 5.5.3f        |
| Limiting Access (limiting system access to authorized people)                    | 11.10 (d) | 2.11; 5.5.3d               |
| Audit Trail (secure, computer generated, time-stamped audit trails)              | 11.10 (e) | 4.9.3; 5.5.3c              |
| Authority Checks (to ensure that only authorized individuals can use the system) | 11.10 (g) | 2.11; 4.1.5; 4.9.3; 5.5.3e |
| System Documentation                                                             | 11.10 (k) | 5.5.3b                     |
| Record Retention (to generate accurate and complete copies of records)           | 11.10 (b) | 4.9.7                      |

### 8.1.2 From Clinical Protocol to Data Acquisition Tools

The clinical data management activities should be planned early in the clinical study process. Ideally, they should occur concurrently with the development and writing of the protocol, when the decisions about the study data to be collected and the type of data flow are usually taken. It is a crucial phase for the development of proper data management tools, and it is therefore suggested that statisticians, clinicians, data managers, study monitors, study coordinators, and database programmers be involved in it. The design of the CRF is the first step in translating the protocol requirements into data and is one of the first results of this multidisciplinary work.

A CRF can be defined as a collection of forms including a series of data items, each aimed at obtaining a single response, and composed by a text (item label) and a response field (item value); while the first represents the question to be answered, the second is the space given to the investigator to record its response [5].

### 8.1.3 Database Design

Clinicians often have difficulty in deciding what patient information is relevant to a clinical trial. In their current practice, they are able to retrieve a summary of the relevant information from individual patients' charts, but are probably unable to give a whole list of the pertinent data items. When these are identified by the multidisciplinary team, they must be organized in a structured way so that they may be easily collected and analyzed. The CRF structure is then used to design the structure of the database where the patients' data will be stored once entered through a data entry application.

Computer science and various software technologies have revolutionized the way medical information is stored, accessed, and retrieved. Today, the relational database model is often considered the founder of many different types of clinical information systems, providing the closest fit to functional requirements of the protocol [6]. In brief, this model comprises data arranged in "tabular" data structure, consisting of several columns (items/variables) and rows (observations). The key aspects of a relational database are the relationships between tables (entities) through the definition of primary and secondary keys (e.g., the patient's ID to link the patient's

demography data to his adverse events), because they ensure database integrity. This means, for example, that a row cannot be added in the “adverse events” table representing an adverse event for a specific patient, unless a row for the same patient exist in the “demography table.” In addition, a table will have the required fields for representing the entities (type of data collected), and each field will contain data of a specific type, such as characters, numbers, dates, and times. The interaction between the case report form design process and the database design process is fundamental and unavoidable: The schema identified to collect the data may influence their meaning and the structure of the database in which they are stored.

**Vertical versus Horizontal File Structure** In designing the database tables containing data about a specific “entity” (such as demography, laboratory data, etc.), the critical question that must be addressed is the following: For multiple occurrence of data items at different time points, is it easier to handle multiple records or one record with repeated data items? There are two options to consider: vertical and horizontal file structures (Fig. 1). The vertical file structure defines the concept of multiple records, while the horizontal file structure is associated to the concept of repeated data items.

To utilize a vertical file structure, a method to distinguish between the multiple occurrences of the data must be identified. This can be done by using the visit date and the visit time. Therefore, in a vertical file structure, the single record contains information on the patient’s measurements at a defined date and time. The advantages of this structure include the utilization of cross-tabulation features for reporting and analysis and the minimization of the amount of missing data and space needed for storage in the database. Several statistical packages such as SAS System require this structure for data processing. In addition, programming efforts are reduced due to the need for processing multiple records instead of multiple fields.

When a horizontal file structure is used, the information collected must be fixed and prespecified. The names of the database fields should have suffixes that help

#### Vertical

| Site | Patient | Visit # | Visit Date | Temp | SBP | DBP |
|------|---------|---------|------------|------|-----|-----|
| 5    | 1       | 1       | 10/11/2005 | 37   | 120 | 70  |
| 5    | 1       | 2       | 10/12/2005 | 37.1 | 125 | 71  |

#### Horizontal

| Site | Patient | Temp1 | Temp2 | SBP1 | SBP2 | DBP1 | DBP2 |
|------|---------|-------|-------|------|------|------|------|
| 5    | 1       | 37    | 37.1  | 120  | 125  | 70   | 71   |

**FIGURE 1** Vertical vs. horizontal data structure.



the identification of the visit time (e.g., in Fig. 1 SBP1 and SBP2 identify the patients' systolic blood pressure taken during visit 1 and visit 2, respectively). This could be a problem if data for additional visits must be collected. In addition, the limit on the number of fields or in the record length for a single table may introduce other problems that cannot easily be solved. Although this structure allows easy comparison of similar fields over time (e.g., to find the worst value), it is not a recommended approach. Furthermore, it requires more programming efforts without any benefit in program flexibility.

### 8.1.4 Data Processing

Clinical data management deals with complex medical data. Managing this data requires familiarity with medical terminology, anatomy, physiology, and pharmacology, as well as practical knowledge of how data are collected in the health care setting and documented in medical records.

The data processing process includes the following steps: data entry, data validation, data modification, medical coding, and database locking [7].

**Data Entry Methods** When a paper-based CRF is used to collect patient data, the investigator is asked to access the source patient's documentation (e.g., the medical record), fill in the study forms, and deliver them to the coordinating center (or clinical data center), usually after a data-monitoring process is carried out. Once received, the CRF undergoes an initial inspection and then is entered into the clinical database. Two options are usually available for data entry: single data entry and double data entry. With double data entry two people enter the same data independently. Any discrepancies between the first and second entry are usually solved by a third person (e.g., the data manager in charge of the study) or by the person performing the second entry (interactive or online verification). The main regulations, such as those of the FDA and ICH, do not formally require double entry or any other specific data entry process, although some have questioned the need for duplicate data entry [8]. In general, the single-entry process with proper manual review, or with double data entry of key data (e.g., the main study endpoints) will work better than a sloppy double-entry process.

**Data Cleaning and Checking** Data are usually collected and entered in the database following detailed rules and conventions that govern how the data are to be recorded (i.e., CRF completion instructions, data entry guidelines, data entry conventions, etc.). However, many errors can be made (voluntary or involuntary) during the data collection and data entry processes including misspelled names, duplication of digits, partial data collection, and inconsistencies between forms.

For this reason, a data cleaning (or data validation) process should be defined. The data-cleaning process is based on a list of activities that are planned to assure the validity and accuracy of the data, including a manual review of the data before their entry, and aggregate descriptive statistics that reveal unusual patterns in the data. However, the most powerful tool provided by the data-cleaning process is the automatic data-checking process to identify errors, protocol violations, and data completeness, inconsistencies, and duplication.

**CONTROL ON PATIENT DEMOGRAPHY**

1. Age <> Missing AND between 18-75
2. Sex <> Missing AND In(1,2)
  - If SEX=2 THEN Inclusion Criteria nr. 3 (Pregnancy) <> Missing

**DRUG ADMINISTRATION**

1. (Actual Dose/Intended Dose) between 0.95 and 1.05 (5% dose modification allowed)

**FIGURE 2** Example of a data validation plan specifications.

Example of data-checking include range checks on the variables that may reveal values that fall outside the accepted limits (e.g., age, for a specific eligible criteria of the protocol, must range between 18 and 75), and cross checks between variables, which may reveal inconsistent answers (e.g., if a patient is dead the date of death is required). Other additional types of checks such as checking whether mandatory answers are not missing, checking for invalid dates and invalid date sequences, checking that complex multifile (or cross panel) rules have been followed (e.g., if a specific adverse event occurs, other data such as concomitant medications or procedures might be expected), and the comparison of the values entered to lists of valid patterns of value, either text/character (e.g., the smoking habits could be coded as “smoker,” “nonsmoker,” and “ex-smoker”) or numerical (e.g., the same habits could be coded as 1, 2, and 3) are usually programmed and performed.

A complete validation plan for the data cleaning should be defined, for example, using a pseudocomputer language which will be easily understood and translated into a computer language (Fig. 2).

The data-checking process may be performed as a batch or as an interactive process. In the first case, data checking takes place after data entry. In the second case, it is performed during the data entry, and the data entry person receives a warning message immediately if an error is made.

**Data Modification** Data may be changed as a result of a data-cleaning procedure. In this case, both the site and the data center must retain a record of all changes to the data. Data changes (with the new and the old values) are usually recorded and documented through a data (or query) clarification form that is signed by the site investigator and kept with its original study record. Under some circumstances, data-cleaning conventions (often called self-evident changes) may specify data that can be modified without a site’s acknowledgment. Examples include appropriately qualified personnel correcting obvious spelling errors and converting values when the units are provided. For this reason it is important for the site to receive and maintain a copy of each version of such data conventions.

However, an audit trail should always reflect all deletions, updates, or additions to data that are carried out after the initial entry. In addition, it should include the date and time of the changes, the reasons for the changes, and the name of the user who was in charge of making them. An electronic audit trail can provide a clear documentation of this information (Fig. 3).

| Pt#  | Vis# | User Id | Timestamp              | Item  | Old value | New Value | Reason for Change              |
|------|------|---------|------------------------|-------|-----------|-----------|--------------------------------|
| 1001 | 1    | DM_01   | 12/05/2004<br>11.46.52 | AEOUT | 1         | 2         | Query following Medical Review |
| 1001 | 1    | DM_01   | 12/04/2003<br>10.32.46 | AEOUT |           | 1         |                                |

**FIGURE 3** Example of an electronic audit trail. An adverse event occurred at visit 1 in patient 1001; the corresponding data was initially entered on 12/04/2003 by user DM\_01. Following the data validation process, a query was issued; the investigator reply implied the change of the adverse event outcome (Item AEOUT) from 1 to 2; the modification was performed by user DM\_01 on 12/05/2004.

**Medical Coding** Wherever possible, data should be collected in a coded format, while the text format should be used to collect summaries and data that do not need to be codified. This makes data entry and the search and data analyses procedures easier. Free text can be coded using an ad hoc dictionary (with a simple list of codes) or a standard dictionary (with thousands of entries).

Data coding can also avoid possible misinterpretations. For example, consider patients with “abnormal transaminases.” The investigators from different sites could indicate this condition in the CRF with the phrases “has ALT,” “has SGPT,” and so forth. When data are analyzed, if we need to identify all cases of “abnormal transaminases,” we should search for all possible combinations of words related to the abnormal transaminases concept.

This adds further implications in data analysis. Consider, for example, the statistical tables illustrated in Figure 4 analyzing the safety profile of a particular medical intervention, where worst (maximum intensity) adverse event toxicity by patient is reported. If we consider the first table we may conclude that seven patients had abnormal transaminases of grade 1 or 2. Instead, if we consider the second table based on the same set of patients, we may conclude that five patients had abnormal transaminases. The discrepancy is due to the investigator’s use of different terms to identify the same type of adverse event occurring in the same patient (this occurs in two patients as explained in Fig. 4).

The coding process can be automatically performed during data entry by matching the text collected on the CRF to the terms included in the standard dictionary.

Several medical dictionaries are currently available. Among them, the *Medical Dictionary for Regulatory Activities* (MedDRA) [9] has become the standard dictionary for coding a patient’s medical history, medical and surgical procedures, and adverse events. In addition, it includes the standard terminology used by regulatory agencies and biopharmaceutical industries within the ICH regions, through all phases of clinical development (including postmarketing). Other well-known medical dictionaries are the *World Health Organization Drug Dictionary* (WHODRUG) [10] and the *Systematized Nomenclature of Medicine—Clinical Terms* (SNOMED-CT) [11], which are used to code medications, the International Classification of Diseases (ICD) oncology for tumor classification, the ICD-9 for classifying morbidity and mortality, and the National Library of Medicine’s Unified Medical Language System (UMLS) for classifying general medical terms.

| Investigator term | Worst CTC Grade by Patient |      |     |      |     |     |     |      |
|-------------------|----------------------------|------|-----|------|-----|-----|-----|------|
|                   | 1                          |      | 2   |      | 3   |     | Any |      |
|                   | N                          | %    | N   | %    | N   | %   | N   | %    |
| ALT               | 1                          | 5.2  | 2   | 10.5 |     |     | 3   | 15.7 |
| SGPT              | 2                          | 10.5 | 2   | 10.5 |     |     | 4   | 20.3 |
| ....              | ...                        | ...  | ... | ...  | ... | ... | ... | ...  |

| MedDRA Primary Preferred Term | Worst CTC Grade by Patient |     |     |      |     |     |     |      |
|-------------------------------|----------------------------|-----|-----|------|-----|-----|-----|------|
|                               | 1                          |     | 2   |      | 3   |     | Any |      |
|                               | N                          | %   | N   | %    | N   | %   | N   | %    |
| Alanine Aminotransferase      | 1                          | 5.2 | 4   | 21.1 |     |     | 5   | 25.2 |
| ....                          | ...                        | ... | ... | ...  | ... | ... | ... | ...  |

| Pt# | Vis# | Event Description | MedDRA Primary Preferred Term | Max CTC Grade | Max CTC Grade per patient according to MedDRA term |
|-----|------|-------------------|-------------------------------|---------------|----------------------------------------------------|
| 1   | 1    | ALT               | Alanine Aminotransferase      | 2             | 2                                                  |
| 1   | 2    | SGPT              | Alanine Aminotransferase      | 1             |                                                    |
| 2   | 2    | SGPT              | Alanine Aminotransferase      | 2             | 2                                                  |
| 3   | 1    | ALT               | Alanine Aminotransferase      | 1             | 2                                                  |
| 3   | 2    | SGPT              | Alanine Aminotransferase      | 2             |                                                    |
| 4   | 1    | ALT               | Alanine Aminotransferase      | 2             | 2                                                  |
| 5   | 1    | SGPT              | Alanine Aminotransferase      | 1             | 1                                                  |

**FIGURE 4** Implication of medical coding on statistical analysis.

**Database Locking** The study databases must be locked to ensure their integrity for the generation of results, analysis, and submission to the regulatory agencies. Locking a study database is fundamental to prevent inadvertent or unauthorized changes once the final analysis and reporting of the data have begun. Furthermore, database closure is a critical issue in preserving the integrity of blinded randomized clinical trials when the blindness needs to be broken. Therefore, any clinical trial must have a well-defined process for closing its database and clear change-control procedures for reopening the database, if necessary.

Before locking a database, a database quality control should be considered in order to verify that study procedures for collecting and managing the data were correctly applied. This is usually performed by checking the contents of the database against the paper CRF in a sample of patients (e.g., 10% of CRFs).

### 8.1.5 Electronic Data Capture Principles

Electronic data capture (EDC) is the process of the collection of data into a persistent electronic form: This includes data entry (e.g., keyboard EDC, pen-based systems, voice recognition) and automated (or direct) data acquisition (e.g., bar code scanners, blood pressure cuff devices) [12].

The EDC processes emerged in the 1970s but languished for 20 years without having significant impact on the clinical trials arena. However, in the 1990s the development of tools for clinical trials research became more focused. With these applications data were captured and entered directly into the PC client at the investigator sites where the database application was installed; the data stored at the site were then periodically transmitted to the central server located at the study data

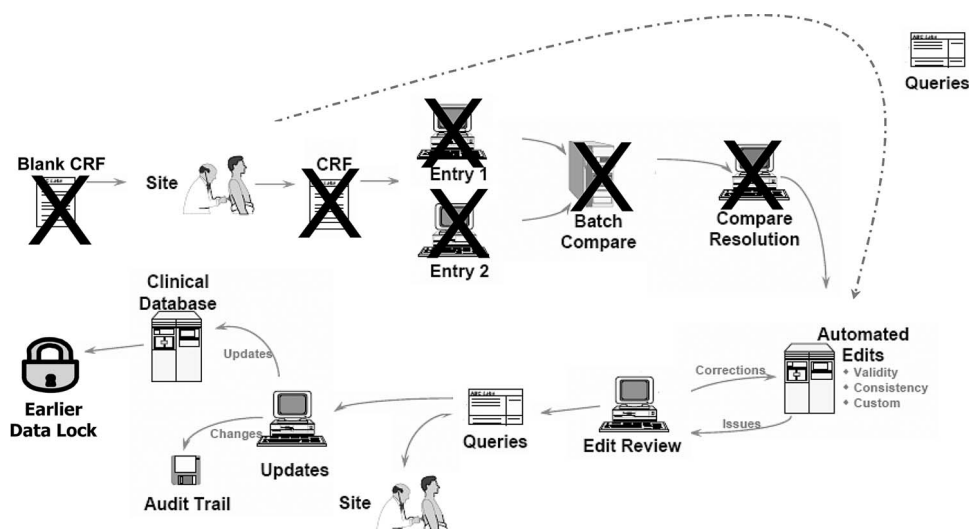


**FIGURE 5** Impact of EDC adoption in clinical data management. EDC adoption made the data management process faster by removing time-consuming steps in clinical data flow from source data to clinical trial data.

center (Fig. 5). Other, more sophisticated EDC systems have been developed during recent years, including those using the Internet and web technology (see following sections), and today they are commonly used by many research organizations to conduct clinical trials. The regulatory agencies themselves are today ready to accept the submissions for drug registration in which EDC tools are used.

The adoption of an EDC system has a great impact in the clinical data management process and introduces new features:

- No formal data entry exists because source data are collected and entered directly into the clinical database without first being captured onto paper; this means that transcription errors are eliminated and a double data entry is therefore not required.
- The data cleaning and editing can be performed during data entry (through the online checks), and the investigator can immediately clarify any discrepancy.
- The source data verification process cannot be eliminated; however, the number of queries issued for the clinical monitors, as well as the time they usually spend at the investigator's site, are reduced.
- The type of training and skills required of the data entry staff are different from those required when a paper-based system is used; in this case, in fact, the data entry personnel should receive training on the specific system being used in the study as well as on study-specific issues (i.e., eCRF completion guideline).

For these reasons, designing a good EDC system needs special attention to make moving from paper to EDC systems and computer programs efficient while maintaining data integrity.

### 8.1.6 Data Standards

One of the main drawbacks to the adoption of the EDC systems has been the proliferation of software to capture trial data; moreover, without the ability to share data across systems, the value of the EDC systems has been somewhat restricted to clinical researchers.

To solve this issue, the representatives from pharmaceutical industries, academic research institutions, health authorities and other research entities [i.e., clinical research organizations (CROs)], met in 1997 to develop a system of shared standards, known as CDISC (Clinical Data Interchange Standards Consortium, <http://www.cdisc.org>). Among the CDISC specifications, the Operational Data Model (ODM) describes a standard model to combine both the study data definition and the actual subject data. This standard specification allows two different systems to communicate and share the data if the source system can produce CDISC ODM format and the receiving system can read CDISC ODM format. The portability and sharing of the data is made possible by the use of the eXtended Markup Language (XML), which has been adopted by CDISC to describe the data hierarchy (see example in Fig. 6).

Other CDISC models have been developed. One of the most interesting is the Study Data Tabulation Model (SDTM) through which most common clinical data domains, variables, and their attributes are defined (i.e., name, type, length, standard definitions, code lists, etc.). The adoption of such a standard for data structures and conventions permits the sharing and application of the same, unique data structure and data-cleaning procedure to several different studies; moreover, pharmaceutical companies dealing with regulatory submissions or institutions running meta-analysis projects would be able to better integrate data from multiple studies.

In July of 2004 the U.S. Food and Drug Administration (FDA) announced the adoption of the CDISC SDTM as the standard format for submitting study data for

```
<ClinicalData StudyOID="S054T231" MetadataVersionOID="101.01">
  <SubjectData SubjectKey="1000" TransactionType="Insert">
    <StudyEventData StudyEventOID="101.Screen">
      <FormData FormOID="101.DEMOG">
        <ItemGroupData ItemGroupOID="101.DM">
          <ItemDataString ItemOID="101.USUBJID">002</ItemDataString>
          <ItemDataString ItemOID="101.SEX">F</ItemDataString>
        </ItemGroupData>
      </FormData>
    </StudyEventData>
  </SubjectData>
</ClinicalData>
```

**FIGURE 6** Example of an ODM XML file. An XML file portion describing demography data (DM) of a "Female" patient enrolled in the study S054T231 with the code 002.

registrative purposes; with such a decision the FDA intends to reduce the time for data evaluation and avoid the reorganization of large amounts of data submitted in varying formats.

### **8.1.7 Infrastructure Requirements**

Setting up an organization for clinical trials data management (i.e., data center, central coordination center) requires the evaluation of a number of issues including the selection of the optimal staff for supporting the various steps of clinical data management flow (i.e., data manager, database programmer, medical coding expert, etc.). Concerning the IT aspects, an appropriate hardware and software selection should be taken into account. The key aspect to consider is the physical security of original data sources (e.g., case report forms, electronic data files, and other original data documents), including the system used to store them. Original paper and electronic documents (servers) should be warehoused in secure rooms, or file cabinets, with controlled access. Direct access to database servers should be restricted to individuals who are responsible for monitoring and backing up the system; all other access to database servers should be controlled by logical security and should occur across a secure network protected by password access and appropriate system controls. Mechanisms should be implemented to capture and prevent unauthorized attempts to access the system (i.e., firewall); if such an attempt takes place, the administrator should be notified immediately.

The investigator sites should be considered part of the infrastructure, especially if data are stored locally at the study site before being sent to a central server (as in the case of use of “offline” systems).

### **8.1.8 Implementation of Clinical Study Data Management System**

An organization, either a company or academic research institute, may decide either to develop an ad hoc system or acquire a commercial clinical study data management system (CSDMS). Both options have some drawbacks: While developing an ad hoc system requires more effort in the system validation process and additional dedicated computer staff, buying an existing commercial system may require a lot of time to choose and evaluate it among many candidates available in the market, and, at least at the beginning, a lot of money.

The choice of the most appropriate CSDMS should take into consideration several aspects, including the availability of tools to compare double data entry sources, design user-friendly data entry forms, easily program checks for data entry validation, support the generation of ad hoc queries, and handle missing values. In addition, multiuser access regulated by user name and password, and the availability of tools to transfer data to other software packages (such as SAS System for performing statistical analysis), to simplify the coding process, to implement data dictionaries, and to perform the data backup and recovering, can be considered good add-ons.

In recent years the market has seen the launch of several solutions. These software programs help users in designing a complete data management system, without having to know any concept of data models. Oracle-based applications, such as Oracle Clinical RDC and PhaseForward Inform/Clintrial, represent the bigger portion of the market and, due to their high costs, are suitable for an environment



in which many trials are conducted. However, a number of alternative solutions, including free open-source CSDMS such as TrialDB (<http://ycmi.med.yale.edu/trialdb>) and Open Clinica (<http://www.openclinica.org>) are now available [13, 14].

### 8.1.9 Computer System Validation

The GCP requirements emphasize the importance of validation of systems, process, and data. In particular, they ask to “ensure and document that the electronic data processing system(s) conforms to the sponsor’s established requirements for completeness, accuracy, reliability, and consistent intended performance (i.e., validation)” [CFR GCP 5.5.3 a]. Similar requirements have been proposed by other regulatory guidelines. For example, the FDA 21 CFR Part 11 [4] emphasizes the need for a validation approach for any computerized system that is used to store electronic records or signatures so that they can be trustworthy, reliable, and essentially equivalent to paper records and to a handwritten signature (Table 2). In this context, the term “computer system” is used to describe the combination of hardware (servers, local area network, client PCs), software (operating system and software applications), procedures, documentation [standard operating procedures (SOPs), guidelines, manuals], and people involved in the process.

Data management software purchased off the shelf should have been validated by the vendor who originally developed it. These validations are usually referred to as “design-level validation” and do not need to be repeated by the end user.

However, the documentation of the design-level validation specifications and testing should, ideally, be available; it should ensure the completion and documentation of functional level testing, and include the documentation of the effect of any known limitation, problem, and defect on functions used for the study [15].

Additional validation activities need to be performed by the customer before the system can be used, including testing the data entry screens to ensure that the data are correctly mapped into the database structure, testing the data verification functions, the validation of any generic integrity constraints or data-checking routines running during data entry, and testing the audit trail and import/export (from/to other data formats) procedures.

### 8.1.10 Future: HER/EDC Integration

The electronic health records (EHR) systems usually include most of the data that are to be collected in a clinical study. Some examples are the demographic data, the medical history, the clinical events, and the concomitant treatments. For this reason many researchers have proposed to consider them as a primary source of information from which to automatically extract, when needed, the data requested in a study protocol [16–19].

However, this solution presents some limits. For example, some specific information related to the study treatment may not be included in the patient’s EHR or may not be properly collected. Similarly, some specific laboratory tests may not be usually performed. In addition, patient confidentiality must be addressed and correctly ensured.

Although some organizations have experimented EHR/EDC combinations, this technology is still young, and, as of today, solutions for health care providers and

hospitals are not available. Some projects to integrate the EHR and the EDC systems are still ongoing. The most important is the BRIDG (Biomedical Research Integrated Domain Group, <http://www.bridgproject.org>) project, which is a collaboration between CDISC, representing research data standards, and HL7 (Health Level Seven, <http://www.hl7.org>), representing health care data standards.

## 8.2 WEB-BASED CLINICAL TRIALS

### 8.2.1 Web-Based Clinical Trials

Telecommunication technology for the management of clinical trials has been used since the early 1990s when PC/modem-based randomization and data-monitoring systems were developed. One of the first applications is the system developed by the Gruppo Italiano per la Sopravvivenza nell'Infarto Miocardico (GISSI) for the management of the GISSI-3 trial, an Italian multicenter large-scale clinical trial in patients with myocardial infarction [20]. A computer/modem-based system allowed investigators from 200 coronary care units to randomize 20,000 patients in the trial using an automated randomization procedure running on a 24-h basis, and provided study reports and reminders [21].

Since from the mid-1990s the Internet and the World Wide Web have been proposed by several pharmaceutical industries, CROs, and international research groups as tools to support clinical research. Speed in communication, strong interaction among people involved in trial conduction, cost reduction, data quality improvement, and the use of simple and standard tools (such as the browser) to interface with study databases are the key elements for the Web's success [22–24].

### 8.2.2 Tools for Participating in Web-Based Clinical Trial

A potential investigator center who wishes to participate in a Web-based clinical trial (or e-trial) needs a personal computer, a Web browser, an electronic mail system, and access to the Internet via an Internet service provider (IP). Sometimes a printer is needed for printing documents (such as the reminder of the randomized treatment allocation) that need to be archived in the patient's file. Standard software such as Acrobat Reader may also be needed to read documents (such as operating manuals or CRFs) distributed in pdf format by the coordinating center.

A specific release of a browser (Internet Explorer 6.0 or 7.0 are more frequently suggested for this use) is often requested, as well as setting it up to accept cookies for the user's identification by the coordinating center's server.

For security and data safety reasons an Internet connection with a static IP address is suggested because this is one of the best methods to identify the computer of an investigator participating in a trial.

### 8.2.3 Why a Clinical Trial Website?

There are many reasons for a clinical trial to have a website. First, a clinical trial website provides trial personnel (coordinators, investigators, monitors, sponsors, committee members) with a powerful communication, organization, and monitoring

tool as well as with tools for the decentralization of trial activities (e.g., remote randomization and data entry). Second, the use of such a website can reduce research costs and the time required for trial completion. In addition, the remote data entry allows the building of a more accurate study database because data are entered at the same place where they are collected.

A clinical trial website is also an ideal vehicle for the dissemination of findings and information related to the trial or similar trials conducted in the past by the same group. For these reasons, a clinical trial website is developed mainly for multicenter and international clinical trials, and for those carried out by pharmaceutical companies.

**Communication among Trial Personnel** A study website can be used to provide the investigators with secretarial support through automatic trial report generation and delivery (Table 3). Examples of online reports are the list of randomized patients, the visit calendar, the list of patients lost to follow-up, and the list of CRFs and queries that are still outstanding [25, 26]. Automated tools may help investigators to organize their work. Web-based systems developed by some study groups include tools to handle the study drugs (how and when new drugs should be ordered, how drugs dispensed to patients should be registered in the database, how to handle the drug inventory, and when expired drugs should be destroyed) [27, 28] and tools to download study materials (protocol, study CRFs, and informed consent forms)

**TABLE 3 Examples of Online Reports and Information Available on a Study Website**

| Information and Reports                        | Examples                                                                                                                                                    |
|------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>For Site Investigator</i>                   |                                                                                                                                                             |
| Reports on the study's progress                | List of randomized patients<br>List of planned visits<br>List of patients lost to follow-up<br>List of outstanding CRFs/queries<br>List of uncompleted CRFs |
| Reports about the management of the study drug | Drug inventory monitoring<br>List of drugs dispensed to the patients<br>List of expired drugs                                                               |
| Study materials                                | Informed consent form<br>Study protocol<br>Case report forms<br>Study brochure                                                                              |
| <i>For Study Coordinating Center</i>           |                                                                                                                                                             |
| Geographical distribution of enrolled patients | Frequency distribution by center<br>Frequency distribution by state<br>Frequency distribution by country                                                    |
| Geographical distribution of recruited centers | Frequency distribution by state<br>Frequency distribution by country                                                                                        |
| Epidemiological data of enrolled patients      | Frequency distribution by gender<br>Frequency distribution by age<br>Frequency distribution of the study drug within specific subgroups of patients         |
| Statistics on participating centers            | List of the centers' quality indicators                                                                                                                     |

[26]. Automatic and daily procedures can aggregate data that can be hosted on the study website for the study coordinating center staff's use and to plan new strategies to improve data quality, patient recruitment, and center performance. These include epidemiological patient data, geographical distribution of patients enrolled in the trial, and information about the quality of data provided by the participating centers.

A directory on the website of the investigators, committees, sponsors, and monitors with their email addresses, as well as a directory of participating centers and regional coordinators, can also help improve communications, mainly in multicenter large-scale clinical trials [22].

The clinical trial website may be used as a virtual community in which investigators are continuously updated with the latest news on their trial. For this reason a news section of the website can provide information on trial status, trial newsletters, notification of national and international investigator meetings, reports from study committee meetings, results of similar, concomitant trials, and answers to common questions regarding the application of the trial protocol.

**Dissemination of Study Information to Public** A study website can provide detailed information about the clinical trial presented in a way that the general public can understand. They may include a geographical distribution of the study participating center and a summary of the background, aim, and design of the ongoing trial. The entire protocol (apart from any confidential aspects) can also be made available. Even if some protocols are already included in the public registries of clinical trials such as the National Library of Medicine's ClinicalTrials.gov (<http://www.clinicaltrials.gov>), in a clinical trial website more detailed information is provided. This information is useful particularly for those clinical trials allowing online recruitment by patients themselves. In this case the website provides the patients with tools to screen their eligibility, collect their data for enrollment into the study, and download the informed consent form.

**Online Randomization** Randomization is the method of randomly dividing subjects into two or more groups. In the case of two groups, one is allocated to the trial (tested) treatment and the other one to the standard (current) treatment. It is necessary to ensure that any baseline differences between groups are due to chance alone and to prevent selection bias. Several methods for randomizing have been used over the years, including minimization, biased coin, stratification, permuted block, and random number generators on computers. Web-based randomization systems allow investigators to directly include patients in a trial 24 hours a day. They are particularly useful in multicenter clinical trials where a central coordinating center serving 24 hours a day as a randomization center may not be feasible. Web-based randomization systems utilize the client/server technology and may be used directly by the investigators through the browser [29, 30]: The client software running on the investigator's PC provides a friendly interface for data collection; the server software (installed on the server located at the randomization center) processes and archives data and selects treatment allocation. Data checking is performed by the server software or at the client site. A typical online randomization system follows the steps illustrated in Table 4 to randomize a patient [21]. In particular, the checking of data validity and consistency, and the checking of the eligibility criteria may

**TABLE 4 Steps for Typical Online/Web-based Randomization**


---

|                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Automatic identification of the investigator and/or the participating center                                                                                                       |
| Entry of the patient's recruitment data (e.g., patient's initials, gender, age, systolic blood pressure, and those data useful to check if a patient can be enrolled in the trial) |
| Automatic data checking and validation                                                                                                                                             |
| Automatic check that the trial's eligibility criteria are met                                                                                                                      |
| Running of the randomization algorithm                                                                                                                                             |
| Data incorporated in the central database                                                                                                                                          |
| Notification to the investigator of treatment allocation and patient identification code                                                                                           |

---

reduce the risk of protocol errors and of noneligible patient enrollment. Web-based randomization systems are usually integrated in the study website [25] and developed in-house by the coordinating center staff. As an alternative to this service there are several online randomization programs (some free of charge and some commercial) that can generate random allocations [31].

**Online Data Collection and Validation** A clinical trial website often includes a Web-based data entry system that allows investigators to enter clinical data online as soon as they are available. Remote data entry to a central database is particularly useful in multicenter clinical trials where participating centers can be geographically separated by great distances across cities and countries [25].

With respect to traditional paper-based clinical trials, data comes directly from the investigator who is responsible for entering them in the data entry system. In this way steps in data collection are reduced, paper source documents are eliminated, and the quality of collected data is improved thanks to a real-time data validation.

The investigator is requested to fill out several electronic forms (often in HTML format) that have a layout and contents that are not too different from those of the traditional paper-based case report. Client/server technology is often used to develop Web-based data entry systems. The architecture of typical Web-based data entry systems is illustrated in Figure 7. The client application runs on the investigator's computer with the application itself running on a central Web server. The browser interacts with the Web server to submit data collected through a Web page; the Web server sends data to the database server where they are saved. Similarly, if the investigator wants to access data saved in the database server, a request is submitted to it through the Web server, which formats the database query results into HTML and sends it back to the browser as a Web page.

One of the main advantages of remote data entry is the possibility to perform real-time data validation (also called "online edit checks") before they are saved into the database server. When an invalid (or incompatible) entry occurs, the investigator is alerted and invited to make the appropriate corrections. The real-time data validation includes the edit checks already illustrated in discussion on data cleaning and checking in Section 8.1.4: checks on missing data, checks on answers to closed questions (e.g., those with "yes" or "no" as possible answers), and checks on values that must be comprised within a specific range, within a list of (pre)codified answers, or within a medical standard dictionary. Real-time data validation could also be applied to several data fields in order to perform a cross check of data entry. For example, the investigator recording the date of a serious adverse event (e.g., a myo-

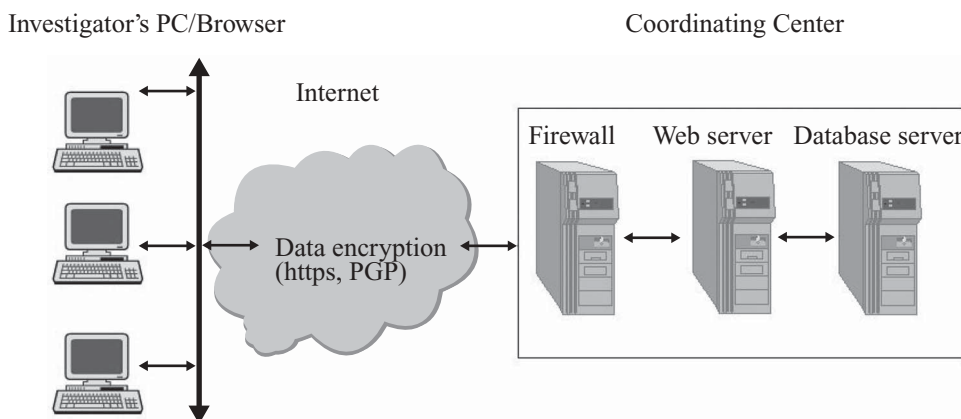


**FIGURE 7** Architecture of typical Web-based data entry system.

cardial infarction) and entering a date that comes before the start of the trial could be prompted immediately, allowing for immediate correction.

Real-time data validation is usually performed on the Web server through Java or Microsoft .NET, which allows for a strong interaction with each data field. In other cases, the data validation process is performed at the end of the data entry process ("offline edit checks") with an automatically generated summary of all the invalid entries that is emailed to the investigator or posted on the trial website for future reminder.

**Data Protection and Security Issues** Security is a key issue when developing a Web-based clinical trial [22, 32, 33]. This issue becomes crucial and requires more attention when a Web-based data entry system is used to allow remote entry in a centralized database of the patients' sensitive information. Patient's (and investigator's) data that are collected through an electronic form or sent by electronic mail must be protected against any kind of interception by unauthorized users. Similarly, patients' data (and the entire database server) need protection from unauthorized access (by Internet or intranet users) once they are stored on the central database in order to avoid a fraudulent use of them or an invalidation of the trial data and results.

To address the issue of secure Internet transmissions, encryption algorithms are usually used for the data exchanged by electronic mail or data that transit on the Web during a connection between a PC/client and a Web server. The data are coded so that only the user that has the decryption key can read and interpret them. The Pretty Good Privacy (usually available among the facilities provided by many electronic mail software) is one of the most used methods for electronic mail data encryption. On the other hand, a secure Web connection is usually provided by systems such as secure socket layer (SSL) and secure hypertext transfer protocol (S-HTTP). These tools are the same usually used to ensure secure transmissions for the applications of banking and trading online. The last versions of Web browsers support these protocols and so they can be used for a connection to a secure Web server (URLs of secure Web servers start with "https://" instead of "http://").

To protect the database server from unauthorized access, firewalls (a mix of hardware and software that is located between the Web server and the Internet and that protect the local area network where the server is hosted) and tools to check the IP address of the user's computer are used. A login and a password are also assigned to each user to enter the website for data entry and editing. Based on these data, a user's profile can be created and used to enable the user (investigator, monitor, staff of the coordinating center) to enter only the website's sections for which he or she is authorized. Sometimes, a more sophisticated user authentication is reached by using digital signatures.

In addition, the database server must be placed in a secure location and antivirus and updated antivirus software must be installed on each computer on the network. When data are particularly important, experts suggest implementing a backup system (with hardware and software components) that could replace the original one in case of failure.

#### 8.2.4 Examples of Clinical Trial Websites and Web-Based Clinical Trials

Several clinical trial web sites have been developed in the last few years. Some, such as the GISSI website (Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Micocarico Acuto, <http://www.gissi.org>) [20], ALLHAT website (Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial, <http://allhat.sph.uth.tmc.edu>) [34], and STAR website (Study of Tamoxifen and Raloxifene, <http://www.nsabp.pitt.edu/STAR/index.html>) [35] concern large-scale multicenter (and international) clinical trials and are used to disseminate information to the public related to ongoing or concluded clinical trials. Others, such as the Southwest Oncology Group's SELECT website (and Vitamine E Cancer Prevention Trial, <http://www.crab.org/select>) [26], provide online patient enrollment, randomization, and reminder notices to the participating centers.

Some have been designed to carry out all the aspects of a clinical trial in ophthalmology [36], orthopedics [30], obstetrics [37, 38], urology [39], and in other medical fields [40]. At present, the most important experience is the INVEST (International Verapamil/Trandolopril Study, <http://invest.biostat.ufl.edu>) study, a phase IV, international, randomized clinical trial conducted by the University of Florida on the efficacy of the calcium antagonist strategy versus the  $\beta$  blocker strategy for the treatment of hypertension in patients with coronary disease [25, 28, 41]. The trial is the first trial fully conducted online. Investigators from 869 sites located in 14 countries have used the clinical trial website to randomize 23,000 patients between 1997 and 2000, to remotely collect and enter patients' clinical data and adverse reactions, to handle the drug inventory at each site, and to disseminate reports and reminders to all the investigators, monitors, and member committees of the trial. Most of the tools described in this chapter have been used to protect the database server and to be compliant with the GCP, including a firewall, user IDs and passwords, data encryption, electronic signature, and the duplication of critical services to be used in case of system failure. The use of the remote data entry, the automatic data validation, and the automatically generated and delivered reports have reduced the time for the completion of trial and the total trial costs.

Most of these Web-based clinical trial systems are developed "in-house" using the current technology available on the market (database management systems,



HTML editors, Java, ASP, PHP, etc.). In order to reduce the amount of memory needed to store the data, some of them [33, 42] used the Entity-Attribute-Value (EAV) model [43, 44] to design the database. Others also allow an offline use (e.g., without an Internet connection); in this case a Web-based stand-alone application is installed on the investigator's personal computer, and a data synchronization engine permits the centralized and the local databases to be kept up-to-date during a daily Internet connection [42, 45].

Further, a number of clinical trials are usually carried out by pharmaceutical industries and CROs through the use of Web-based clinical trial systems. In these cases the Web version of commercial clinical trial information systems (such as Oracle Clinical and Clintrial) are usually used.

### 8.2.5 Advantages and Limitations of Web-Based Clinical Trials

Clinical trial websites may provide a powerful means to conduct clinical trials [46]. Investigators can download study materials, interact more frequently with researchers, and have their problems solved in real time. They can use automatic tools to schedule their work or to enter the clinical data directly. When remote data entry and real-time data validation are implemented, a clean database can be available in a short time for statistical analyses, with reduction of transcription errors and missing data, elimination of data entry costs and reduction of printing costs. For example, researchers of the INVEST study found an 80% reduction in monitoring costs and a 50% reduction in total trial cost [25]. Other cost reductions may be obtained with the implementation of automatic randomization systems and Web-based procedures for organizing drug distribution and accountability, which would allow investigators to order new drugs only when necessary and to destroy them when expired. However, the main advantages in using clinical trial websites are the availability of a friendly and homogeneous interface (the World Wide Web) and of standard software (TCP/IP, browser, electronic mail, Acrobat Reader), which do not require training time, and the ability to centralize study information and coordinate multiple trial processes in real time. Other advantages include security and backup of a whole trial at a single location, simplified data monitoring, and dissemination of study information and results to the public and the scientific community.

Some limitations must be outlined. First, clinical trial websites need the implementation of further security tools (with respect to paper-based or electronic, but not Web-based, clinical trials) in order to prevent unauthorized users from accessing the system and the patient data. This issue is perceived as a threat on data confidentiality by investigators, study centers, and patients and may inhibit them from participating in a Web-based clinical trial. In addition, setting up and maintaining a Web-based clinical trial system requires experienced computer professionals, and further hardware and software to duplicate the key functions in case of system failure. This is particularly important in case of the use of randomization and data entry systems and for the database backup and restoring procedures.

Another limit is the availability of Internet connections. At present, these are still a problem, especially in developing countries where communication facilities are rare. The problem exists also in developed countries, however, where direct connections from hospitals or medical centers are not always available where data are collected.

Other limits include the reluctance on the part of investigators to spend time entering information (in particular when alternative paper-based methods to provide the same data are available), lack of live support personnel (that may lead to the loss of some patients), and high setup cost. Furthermore, every step of the data entry process may be protracted in certain hours of the day due to peaks in Internet traffic. For this reason, many trial groups suggest that Internet connections should be provided by reliable Internet service providers that can assure easy and fast access to the Internet.

Finally, a potential selection bias in patient recruitment should be considered if Internet access is one of the criteria to recruit centers in a trial.

## REFERENCES

1. Guideline for Good Clinical Practice (2003), Consolidated Guideline, International Conference on Harmonisation, EU Implementation CPMP/ICH/135/95/Step5; available at <http://www.ich.org/LOB/media/MEDIA482.pdf>; accessed Jan. 9, 2008.
2. European Community (1995), Directive 95/46/EC of the European Parliament and the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, Brussels, Belgium, European Community Commission; available at [http://ec.europa.eu/justice\\_home/fsj/privacy/index\\_en.htm](http://ec.europa.eu/justice_home/fsj/privacy/index_en.htm); accessed Jan. 9, 2008.
3. Italian Law DL196/2003 (2003), personal data protection.
4. Food and Drug Administration (2003), 21 CFR Part 11, Electronic Records; Electronic Signatures; Final Rule, Rockville, MD, *Fed. Reg.*; available at <http://www.fda.gov/cder/guidance/5667fnl.htm>; accessed Jan. 9, 2008.
5. Spilker, B., and Schoenfelder, J. (1991), *Data Collection Form in Clinical Trials*, Raven Press, New York.
6. Kroenke, D. M. (2006), *Database Processing: Fundamentals, Design, and Implementation*, 10th ed., Pearson Education, Upper Saddle River, NJ.
7. Society for Clinical Data Management (2007), Good Clinical Data Management Practices; available at <http://www.scdm.org>; accessed Jan. 9, 2008.
8. Gibson, D., Harvey, A. J., Everett, V., and Parmar, H. K. B. (1994), Is double data entry necessary? The CHART trials, *Control. Clin. Trials*, 15, 482–488.
9. International Federation of Pharmaceutical Manufacturers and Associations, MedDRA—The Medical Dictionary for Regulatory Activities; available at <http://www.meddrasso.com>; accessed Jan. 9, 2008.
10. World Health Organization, WHO Drug Dictionary—WHODRUG; available at <http://www.unc-products.com>; accessed Jan. 9, 2008.
11. International Health Terminology Standards Development Organisation, Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT); available at <http://www.ihtsdo.org>; accessed Jan. 9, 2008.
12. The eClinical Forum and PhRMA EDC/eSource TaskForce (2006), The future vision of electronic health records as e-source for clinical research, Version 1.0; available at <http://www.cdsc.org/pdf/Future%20EHR-CR%20Environment%20Version%201.pdf>; accessed Jan. 9, 2008.
13. Brandt, C. A., Deshpande, A. M., Lu, C., Ananth, G., Sun, K., Gadagkar, R., Morse, R., Rodriguez, C., Miller, P. L., and Nadkarni, P. M. (2003), TrialDB: A web-based Clinical Study Data Management System, *AMIA Annu. Symp. Proc.*, 794.

14. Hanover, J., Golden, J., and Swenson, M. (2004), Clinical Trial Automation and Data Management Solutions. Document Nr. IDC30946, Volume 1; available at <http://www.healthindustry-insights.com/HII/getdoc.jsp?containerId=30946>; accessed Jan. 9, 2008.
15. Food and Drug Administration (2002), General principles of software validation; final guideline for industry and FDA staff; available at <http://www.fda.gov/cdrh/comp/guidance/938.pdf>; accessed Jan. 9, 2008.
16. Weiner, M. G., Lyman, J. A., Murphy, S., and Weiner, M. (2007), Electronic health records: High-quality electronic data for higher-quality clinical research, *Inform Prim. Care*, 15(2), 121–127.
17. Murphy, E. C., Ferris, F. L.3rd, and O'Donnell, W. R. (2007), An electronic medical records system for clinical research and the EMR EDC interface, *Invest. Ophthalmol. Vis. Sci.*, 48(10), 4383–4389.
18. Gerdson, F., Müller, S., Jablonski, S., and Prokosch, H. U. (2005), Standardized exchange of medical data between a research database, an electronic patient record and an electronic health record using CDA/SCIPHOX, *AMIA Annu. Symp. Proc.*, 963.
19. Powell, J., and Buchan, I. (2005), Electronic health records should support clinical research, *J. Med. Internet Res.*, 7(1); available at <http://www.jmir.org/2005/1/e4/>.
20. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardio (2004), GISSI-3: Effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction, *Lancet*, 343, 1115–1122.
21. Santoro, E., Nicolis, E., and Franzosi, M. G. (1999), Telecommunication technology for the management of large scale clinical trials: The GISSI experience, *Comput. Methods Programs Biomed.*, 60, 215–223.
22. Santoro, E., Nicolis, E., Franzosi, M. G., and Tognoni, G. (1999), Internet for clinical trials: Past, present, and future, *Controlled Clin. Trials*, 20, 194–201.
23. Santoro, E. (2002), Internet and cardiovascular research: The present and its future potentials and limits, *Minimally Invasive Ther. Allied Tech.*, 11, 73–75.
24. Paul, J., Seib, R., and Prescott, T. (2005), The Internet and clinical trials: Background, online resources, examples and issues, *J. Med. Internet Res.*, 7(1); available at: <http://www.jmir.org/2005/1/e5/>.
25. Marks, R., Bristol, H., Conlon, M., and Pepine, C. J. (2001), Enhancing clinical trials on the Internet: Lessons from INVEST, *Clin. Cardiol.*, 24(11 Suppl), 17–23.
26. Shaw, P. A., Goodman, P. J., and Brace, J. (2001), The web based management of the selenium and vitamin e cancer prevention trial, *Controlled Clin. Trials*, 22, 57S.
27. Long, D. T., Workman, J., Beck, R., and Moke, P. (2001), A web based procedure for drug distribution and accountability, *Controlled Clin. Trials*, 22, 80S.
28. Cooper-DeHoff, R., Handberg, E., Heissenberg, C., and Johnson, K. (2001), Electronic prescribing via the Internet for a coronary artery disease and hypertension megatrial, *Clin. Cardiol.*, 24(11 Suppl), 14–16.
29. Kiuchi, T., Ohashi, Y., Konishi, M., Bandai, Y., Kosuge, T., and Kakizoe, T. (1996), A world wide web-based user interface for a data management system for use in multi-institutional clinical trials—Development and experimental operation of an automated patient registration and random allocation system, *Controlled Clin. Trials*, 17, 476–493.
30. Dorman, K., Saade, G. R., Smith, H., and Moise, K. J. (2000), Use of the World Wide Web in research: Randomization in a multicenter clinical trial of treatment for twin-twin transfusion syndrome, *Obstet. Gynecol.*, 96(4), 636–639.
31. Bland, M. Directory of randomisation software and services; available at <http://www-users.york.ac.uk/~mb55/guide/randcery.htm>; accessed Jan. 9, 2008.

32. Marshall, W. W., and Haley, R. W. (2000), Use of a secure Internet web site for collaborative medical research, *JAMA*, 284, 1843–1849.
33. Oliveira, A. G., and Salgado, N. C. (2006), Design aspects of a distributed clinical trials information system, *Clin. Trials*, 3(4), 385–396.
34. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group (2002), Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), *JAMA*, 288, 2981–2997.
35. Vogel, V. G., Costantino, J. P., Wickerham, D. L., Cronin, W. M., Cecchini, R. S., Atkins, J. N., Bevers, T. B., Fehrenbacher, L., Pajon, E. R., Jr., Wade, J. L., 3rd, Robidoux, A., Margolese, R. G., James, J., Lippman, S. M., Runowicz, C. D., Ganz, P. A., Reis, S. E., McCaskill-Stevens, W., Ford, L. G., Jordan, V. C., Wolmark, N., and the National Surgical Adjuvant Breast and Bowel Project (NSABP) (2006), Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: The NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial, *JAMA*, 295(23), 2727–2741.
36. Kuchenbecker, J., Dick, H. B., Schmitz, K., and Behrens-baumann, W. (2001), Use of Internet technologies for data acquisition in large clinical trials, *Telemed. J.E. Health*, 7(1), 73–76.
37. Kelly, M. A., and Oldham, J. (1997), The Internet and randomised controlled trials, *Int. J. Med. Inform.*, 47(1–2), 91–99.
38. The GRIT Study Group (1996), When do obstetricians recommend delivery for a high-risk preterm growth-retarded fetus? Growth Restriction Intervention Trial, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 67(2), 121–126.
39. Lallas, C. D., Preminger, G. M., Pearle, M. S., Leveillee, R. J., Lingeman, J. E., Schwoppe, J. P., Pietrow, P. K., and Auge, B. K. (2004), Internet based multi-institutional clinical research: A convenient and secure option, *J. Urol.*, 171(5), 1880–1885.
40. Tighe, F. P., and Cohen, J. (2001), Web Data Management. Experience with 20,000 case report forms in 14 ongoing studies, *Controlled Clin. Trials*, 22, 51S.
41. Pepine, C. J., Handberg, E. M., Cooper-DeHoff, R. M., Marks, R. G., Kowey, P., Messerli, F. H., Mancina, G., Cangiano, J. L., Garcia-Barreto, D., Keltai, M., Erdine, S., Bristol, H. A., Kolb, H. R., Bakris, G. L., Cohen, J. D., Parmley, W. W., and INVEST Investigators (2003), A calcium antagonist vs a non-calcium antagonist hypertension treatment strategy for patients with coronary artery disease. The International Verapamil-Trandolapril Study (INVEST): A randomized controlled trial, *JAMA*, 290(21), 2805–2816.
42. Clivio, L., Tinazzi, A., Mangano, S., and Santoro, E. (2006), The contribution of information technology: Towards a better clinical data management, *Drug Dev. Res.*, 67, 245–250.
43. Brandt, C. A., Morse, R., Matthews, K., Sun, K., Deshpande, A. M., Gadagkar, R., Cohen, D. B., Miller, P. L., and Nadkarni, P. M. (2002), Metadata-driven creation of data marts from an EAV-modeled clinical research database, *Int. J. Med. Inform.*, 65(3), 225–241.
44. Chen, R. S., Nadkarni, P., Marenco, L., Levin, F., Erdos, J., and Miller, P. L. (2000), Exploring performance issues for a clinical database organized using an entity-attribute-value representation, *J. Am. Med. Inform. Assoc.*, 7(5), 475–487.
45. Santoro, E., Clivio, L., and Mangano, S. (2007), GCPBASE: A web-based tool for remote data capture in a clinical trial, *Tech. Health Care*, 15(5), 355.
46. McAlindon, T., Formica, M., Kabbara, K., Lavalley, M., and Lehmer, M. (2003), Conducting clinical trials over the internet: feasibility study, *BMJ*, 327(7413), 484–487.



# 9.1

---

## *Clinical Trials and the Food and Drug Administration*

**Tarek M. Mahfouz and Janelle S. Crossgrove**

*Raabe College of Pharmacy, Ohio Northern University, Ada, Ohio*

### **Contents**

|         |                                                     |     |
|---------|-----------------------------------------------------|-----|
| 9.1.1   | Food and Drug Administration: History and Structure | 228 |
| 9.1.1.1 | Brief History of FDA and U.S. Drug Regulation       | 228 |
| 9.1.1.2 | Structure and Organization of FDA                   | 229 |
| 9.1.2   | New Drug Development and FDA                        | 230 |
| 9.1.2.1 | Preclinical Studies and IND                         | 231 |
| 9.1.2.2 | Clinical Trials and New Drug Application (NDA)      | 233 |
| 9.1.3   | NDA and Biological License Application (BLA)        | 234 |
| 9.1.4   | Timeline                                            | 234 |
| 9.1.5   | Cost and Probability of Success                     | 236 |
| 9.1.6   | FDA Meetings and Drug Sponsors                      | 237 |
| 9.1.7   | Regulation of Drugs and Biological Products by FDA  | 238 |
| 9.1.8   | Expanded Access and Accelerated Approval            | 240 |
| 9.1.9   | Orphan Drugs                                        | 240 |
| 9.1.10  | Pediatric Drugs                                     | 241 |
| 9.1.11  | OTC Drug Products                                   | 241 |
| 9.1.12  | Behind-the-Counter Drugs                            | 242 |
| 9.1.13  | Drugs for Counterterrorism                          | 242 |
| 9.1.14  | Globalization and Harmonization: FDA and ICH        | 242 |
|         | References                                          | 243 |

### 9.1.1 FOOD AND DRUG ADMINISTRATION: HISTORY AND STRUCTURE

The Food and Drug Administration (FDA) is the agency within the Department of Health and Human Services (HHS) that is responsible for protecting the health of Americans by enforcing the Federal Food, Drug, and Cosmetic Act and other related public health laws. FDA inspectors visit manufacturing facilities to ensure that products are made correctly and labeled truthfully. The FDA also protects consumers in several other ways. It investigates the biological effects of widely used chemicals, it tests medical devices, radiation-emitting products, and radioactive drugs, and it tests food for substances such as pesticide residues. It also monitors biologics such as blood and its products, insulin, and vaccines. In addition, components such as dyes and other additives used in cosmetics, drugs, and foods are all subject to FDA scrutiny.

The most important and critical role of the FDA, however, is in the development of new drugs and medical devices, such as pacemakers. New drugs and medical devices must receive FDA approval before they can be marketed. As the FDA states:

In deciding whether to approve new drugs, FDA does not itself do research, but rather examines the results of studies done by the sponsor (that is the entity manufacturing or importing the new drug). For a new drug to be approved for marketing by the FDA, the agency must determine that the new drug produces the benefits it's supposed to without causing side effects that would outweigh those benefits [1].

In this chapter, we explain the different steps a new drug goes through before it is marketed and the involvement of the FDA in each step. Special attention will be given to FDA regulation of clinical trials. In order to understand how the FDA has evolved into the regulatory body that exists today, we begin with a brief history of the FDA and an overview of the structure and organization of the agency.

#### 9.1.1.1 Brief History of FDA and U.S. Drug Regulation

In 1820, and for the three decades that followed, the *United States Pharmacopoeia* (USP) served as a reference for physicians and pharmacists who worked in the extraction or compounding of drugs and drug components that were available at the time. No other laws against food and drug adulteration were in place at that time. In 1848 Dr. M. J. Bailey testified in Congress that more than one half of the imported medicinal products were so much adulterated that they were not only ineffective but could be dangerous [2]. This led to the Drug Importation Act, which required inspection and destruction of drugs that did not meet acceptable standards [3]. The law seemed to be successful at first as very few bad products came into the United States in the year after its passage [2]. However, the law did not achieve its goal, mainly because political influences and bribery affected inspectors who were also not working with fixed standards [2].

In 1862, chemist Charles M. Wetherill was appointed to head the chemical division in the U. S. Department of Agriculture, a precursor to today's FDA, charged with maintaining a safe food supply. In 1901, the chemical division became the Bureau of Chemistry and was headed by Dr. Harvey W. Wiley who was a knowl-



edgeable physician on the issue of food and drug adulteration and was renowned for his “poison squad” experiments [2]. Dr. Wiley’s experiments showed that some of the most commonly used preservatives at the time, such as borax, were not safe for human consumption and might lead to permanent stomach and bowel impairments. It was Dr. Wiley’s and others efforts that led to the passage of the Food and Drugs Act by the Congress in 1906, and the Bureau of Chemistry was charged to enforce that law. Later, the Sherley amendment in 1911 prohibited the labeling of medications with false therapeutic claims. In 1927, Congress authorized the formation of the Food, Drug, and Insecticide Administration from the regulatory wing of the Bureau of Chemistry, and 3 years later the name was shortened to the Food and Drug Administration. In response to the deaths of over 100 people (mostly children) from the use of the untested compound diethylene glycol to solubilize sulfanilamide, the Food, Drug, and Cosmetic Act was passed in 1938. This new act required that new drugs be tested by their manufacturers for safety and the results be submitted to the government for marketing approval via the new drug application (NDA) [3]. Importantly, the new law authorized the FDA to conduct unannounced inspections of drug manufacturing facilities. In 1940 [4], the FDA left the Department of Agriculture for the Federal Security Agency, and in 1953 the FDA became part of the Department of Health, Education, and Welfare (HEW). In 1968 the FDA became part of the Public Health Service within HEW. When the education function was removed from HEW to create a separate department, HEW became the Department of Health and Human Services in 1980.

### 9.1.1.2 Structure and Organization of FDA

The FDA consists of eight centers:

1. *Office of the Commissioner (OC)* The office of the commissioner is formed of several components, such as the history office and the office of combination products, and is responsible for the implementation of the FDA’s mission.

2. *Office of Regulatory Affairs (ORA)* The office of regulatory affairs is the office responsible for all field activities of the FDA. It ensures the compliance with FDA’s standards by [5]:

1. Monitoring the clinical trials that are conducted before a product is submitted for approval by the FDA and conducting domestic and foreign inspections of drug manufacturing facilities by the consumer safety officers.
2. Analyzing product samples to determine their adherence to FDA standards.
3. Reaching out to consumer groups and health authorities to explain the FDA’s policies and encourage compliance with the agency standards by public affairs specialists. The public affairs specialists also respond to public health emergencies caused by natural disasters and product problems.

3. *Center for Biologics Evaluation and Research (CBER)* The CBER regulates biological products [6] such as blood and blood products, vaccines, and protein-based drugs such as monoclonal antibodies and cytokines. CBER also helps advancing and licensing of products to diagnose, treat, or prevent outbreaks from exposure to bioterrorist pathogens by helping the products to rapidly meet the regulatory

requirements. CBER also works on developing procedures and protocols to advance and make available promising experimental products when there is no approved medication for the treatment of victims of terrorism [6].

4. *Center for Devices and Radiological Health (CDRH)* The CDRH regulates [7] medical devices ranging from contact lenses and blood sugar monitors to implanted heart valves. It makes sure that new medical devices are safe and effective before they are marketed, and it monitors these devices throughout their life cycle using a nationwide postmarket surveillance system. CDRH also ensures that radiation-emitting products, such as microwave ovens, TV sets, cell phones, and laser products meet radiation safety standards.

5. *Center for Drug Evaluation and Research (CDER)* The CDER regulates [8] prescription and over-the-counter (OTC) drugs. Its mission is to ensure that all prescription and OTC drugs are safe and effective. To do this, CDER evaluates all new drugs before marketing; and after a drug is marketed CDER serves as a consumer watchdog to be sure these drugs continue to meet the standards. CDER also monitors TV, radio, and print drug ads to ensure they are truthful and balanced. As a measure to counter terrorism, CDER facilitates development of new drugs and new uses for already approved drugs that could be used as medical countermeasures.

6. *Center for Food Safety and Applied Nutrition (CFSAN)* The CFSAN regulates about 80% [9] of all food consumed in the United States. The remaining 20%, which includes meat, poultry, and some egg products, are regulated by the U.S. Department of Agriculture.

7. *Center for Veterinary Medicine (CVM)* The CVM [10] regulates the manufacture and distribution of food additives and veterinarian drugs or devices that will be used on animals including both pets and animals from which human foods are derived. CVM assures that animal drugs and medicated feeds are safe and effective. In approving veterinarian drugs (whether prescription or OTC) for food animals, CVM determines that no unsafe residues or metabolites will result when the drug is used in the approved manner, and that all safety factors are considered when setting the approved levels of use.

8. *National Center for Toxicological Research (NCTR)* The NCTR [11] conducts peer-reviewed scientific research focused toward understanding critical biological events that lead to toxicity and toward developing methods and incorporating new technologies to improve the assessment of human exposure, susceptibility, and risk.

### 9.1.2 NEW DRUG DEVELOPMENT AND FDA

New drug development is a multistage process that starts by identifying a drug target, usually an enzyme or a biological process whose function is crucial to treating a disease or ameliorate a medical condition. Many isolated or newly synthesized compounds are screened for activity at the target site using high throughput in vitro assays. If the interaction is strong enough to be of any medical significance, these compounds are then tested in vivo in model animals to evaluate their pharmacological activities and acute toxicity potentials. These initial testing steps are called the preclinical studies. The FDA does not monitor these preclinical studies directly,

assuming that good laboratory practices (GLP) [12] have been followed. The sponsor's primary goal in these studies is to determine if the new compound is reasonably safe for initial use in humans and if it exhibits pharmacological activity that justifies commercial development. If the compound proves to be a promising candidate for further development, the sponsor must collect sufficient data to establish that the new compound will not expose humans to unreasonable risks when used in early-stage clinical studies. The results of these preclinical studies, therefore, are important and they must be included in the investigational new drug (IND) application submitted to the FDA for review before clinical trials can begin.

#### 9.1.2.1 Preclinical Studies and IND

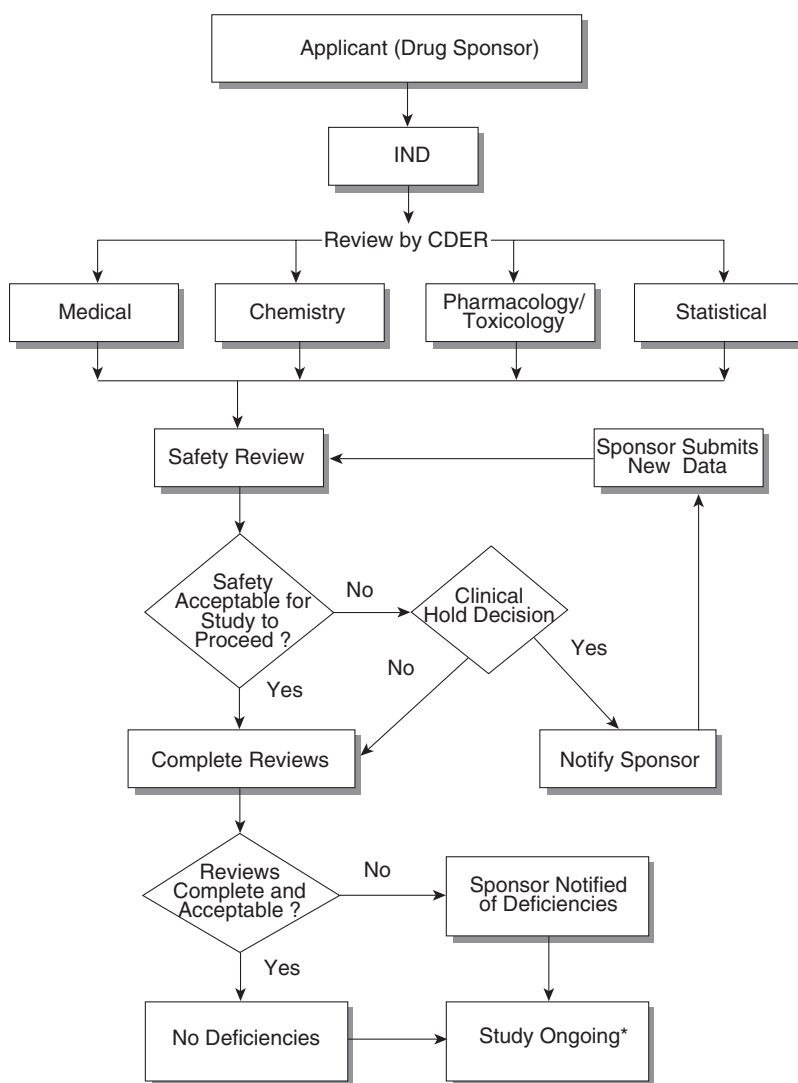
Direct FDA involvement in the new drug development process begins if data from the preclinical studies indicates that the new drug candidate is effective and reasonably safe to test in humans. In this case, the sponsor can submit an IND to formally notify the FDA of his or her intent to start clinical trials on human subjects. The IND falls in two categories and three types [13]. In this section we will cover the investigative IND, and in later sections we will cover the emergency use IND and the treatment IND. For more information on INDs, the reader is referred to the CDER web page titled Drug Applications ([www.fda.gov/Cder/regulatory/applications/ind\\_page\\_1.htm#Introduction](http://www.fda.gov/Cder/regulatory/applications/ind_page_1.htm#Introduction)).

An investigator IND may be submitted by a physician who both initiates and conducts an investigation and under whose immediate direction the investigational drug is administered or dispensed. A physician might submit a research IND to propose studying an unapproved drug, or an approved product for a new indication or in a new patient population. The IND application itself has several components but the information therein falls in three main areas:

1. Results of the preclinical animal pharmacological and toxicological studies together with any previous experience with the drug in humans. This information allows the FDA to decide if the compound is reasonably safe for initial testing in humans.
2. Information on the stability, composition, and manufacturing of the drug and the drug product. This information is provided to assure that the sponsor can supply consistent batches of the drug.
3. The proposed clinical studies protocols and the qualifications of the clinical investigators who will oversee the administration of the experimental compound to assess if people participating in the trial will be exposed to unreasonable risk and if the clinical investigators are qualified to fulfill their clinical trial duties.

Submitted with the IND also are commitments to obtain informed consents [14] from the research subjects and to obtain review of the study by an institutional review board (IRB) [15]. The informed consent document is given to the participants in a clinical trial and states clearly to the participants the purpose of the trial, how long they are expected to participate, what will happen in the study, and what the possible risks or discomforts are, possible benefits, other procedures or treat-

ments that might be available and advantageous to them, and that participation is voluntary and the participants can quit at any time should they so desire. An IRB is a committee composed of at least five medical and ethical experts designated by the institution where the clinical studies are to take place to review and approve clinical trials taking place within their jurisdiction to ensure that medical and scientific standards are maintained and to protect the rights of the human test subjects. The steps of the IND application process are summarized in Figure 1. After the submission of an IND, the FDA has 30 days to respond. During these 30 days, FDA experts review the IND and decide if the new drug is safe to be tested in human



\*While sponsor answers any deficiencies

**FIGURE 1** Steps an IND takes for review and approval. (Source: FDA.)

subjects. If the FDA approves the study and the IRB approves the proposed protocols, the clinical trials can begin.

### 9.1.2.2 Clinical Trials and New Drug Application (NDA)

In clinical trials, the effectiveness of the new drug in treating a disease or controlling a condition is compared to standard treatments or to no treatment at all (i.e., placebo). Side effects, toxicity, and metabolism are also studied. Because clinical trial participants represent only a small fraction of the target patient population and because drugs can work differently in different populations, it is important that the participants be representative of the wider general population by including people of various age groups, races, ethnic groups, and genders in the trials. Clinical trials usually are conducted in three main phases with a postmarketing phase. The three main phases of clinical trials differ in three main aspects: number and type of subjects required, purpose, and duration [16].

*Phase I* Phase I is the smallest of the three phases lasting up to 3 years and requiring 20–100 healthy volunteers. The main purpose of this phase is to determine dosing, how the drug is metabolized and excreted, and to identify the acute side effects.

*Phase II* Phase II trials are slightly larger than phase I, requiring 100–500 patients with the disease or medical condition that the new drug targets and may last from 2 to 4 years. The purpose of this phase is to collect information on the safety of the new drug and its efficacy. At the end of this phase, the drug sponsor meets with the FDA to discuss the results and how to proceed next. If the results indicate that the drug may be effective and the side effects are considered acceptable, the drug moves on to phase III.

*Phase III* Phase III is the largest of all phases requiring 1000–5000 patients and may take several years to finish. In this phase, the drug safety and effectiveness are further studied; and, if a standard treatment is available, the effectiveness of the new drug compared to that standard is also examined. When more and more participants are tested over longer periods of time, the less common side effects are more likely to be revealed. Phase III also establishes other aspects of the drug development process such as marketing claims, packaging, and storage conditions.

The role of the FDA in clinical trials is (1) to help protect the rights and welfare of the patients participating in the trial and (2) to verify the quality and the integrity of the data. To achieve these goals, the FDA's division of scientific investigation (DSI) conducts inspections of clinical investigators' study sites and reviews the records of the IRBs to make sure they are fulfilling their roles in patient protection. Also, the DSI seeks to determine whether the study was conducted according to the investigational plan, whether all adverse events were recorded, and whether the subjects met the inclusion/exclusion criteria outlined in the study protocol. At the conclusion of each inspection, FDA investigators prepare a report summarizing any deficiencies. In cases where they observe numerous or serious deviations, DSI classifies the inspection as "official action indicated" and sends a warning letter or

Notice of Initiation of Disqualification Proceedings and Opportunity to Explain (NIDPOE) to the clinical investigator specifying the found deviations [17]. The FDA usually has authority over clinical trials, and it can stop an ongoing trial or halt it if serious complications develop that puts the participants at high risk.

### 9.1.3 NDA AND BIOLOGICAL LICENSE APPLICATION (BLA)

After the conclusion of phase III of the clinical trials, the sponsor submits an NDA [18] to the FDA asking the agency to consider approving the new drug for marketing in the United States. The NDA contains all animal and human data and analyses of these data together with information on the chemistry, stability, and the proposed manufacturing of the new drug. After receiving an NDA, the FDA has 60 days to decide whether to file the NDA for review by the agency experts, or, if the NDA is incomplete, not to file it. The time the FDA spends in reviewing the NDA will be discussed in the timeline section of this chapter but outcome of the review process can be approved, approvable, or nonapprovable. “Approved” means the new drug product has met all the requirements and the sponsor can begin marketing the new drug. “Approvable” means the new drug product has some minor deficiencies that need to be addressed before approval. “Nonapprovable” means the FDA will not approve the new drug product the way it is submitted due to major deficiencies.

Biologics such as blood and its products, vaccines, antibodies, and the like are dealt with the same way. New biologics are subject to the same regulations and follow the same clinical testing as new drugs. The only difference is that to apply for marketing new biologics, sponsors need to submit a BLA not an NDA. BLAs are submitted to and reviewed by the CBER.

For many drugs approved for marketing, the FDA requires the sponsor to continue submission of clinical data to further validate their safety or effectiveness. In some cases, more studies are needed to find out more about a drug’s long-term risks, benefits, and optimal use or to test the drug in different populations of people such as children. These are the reasons behind postmarket surveillance.

### 9.1.4 TIMELINE

There is no doubt that drug development is a complicated, costly, and time-consuming process. The various steps a new drug endures from its synthesis to the time it makes it to the market as a drug vary in length from an average of 2 years in the research and development phase to an average of 7 years in clinical trials, as seen in Figure 2. It is hard to estimate how long before a newly synthesized chemical compound will make its way to the market as a drug product because each new drug has a different story and a unique path. Most potential drug candidates do not even make it to the clinical trials and are abandoned in the preclinical stage. In the research and development stage, a large number of chemical compounds are screened by researchers to identify only a few promising candidates called leads. Lead identification is the most time-consuming step in the drug discovery process because once leads are identified, structural modification can optimize

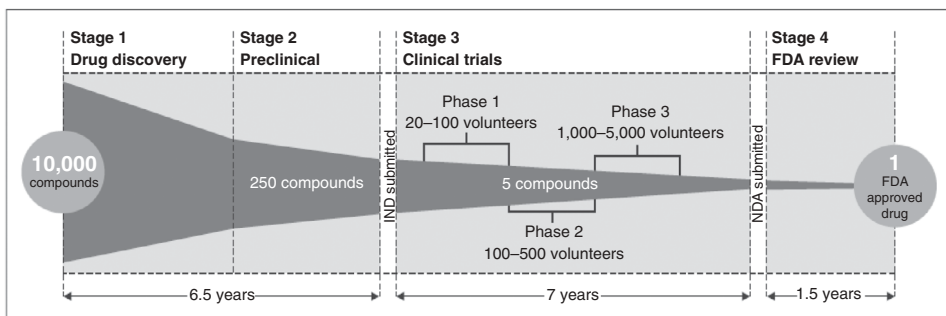
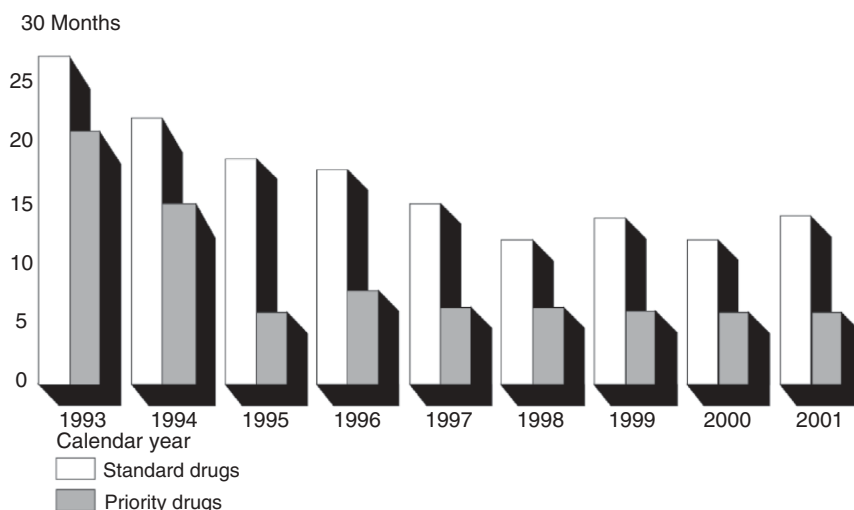


**FIGURE 2** Drug discovery, development and review process. (From Ref. 19.)

their activities or properties as drug candidates. Advances in high-throughput screening technology and computational drug design methods help accelerate this stage a little but, depending on the circumstances, this stage can take up to 5 years.

Once leads have been identified and optimized, toxicity and animal studies can begin. These are less time consuming than the previous step and may take up to 3 years. After the preclinical studies are completed, the new drug sponsor needs to submit an IND and get the FDA approval to begin the clinical studies. As stated above, the FDA has a period of 30 days to decide whether to hold the clinical trials or to give the green light. Clinical trials are the most time-consuming step in the process of new drug development and are the most critical step as well, especially phase II and phase III. Because of the diverse areas investigated and the comprehensive nature of these two phases, they take a long time to finish, sometimes up to 10 years. As the clinical trials evolve and more data become available, changes in the clinical protocols may become necessary. Since each change has to be approved by the IRB, the period of time the drug remains in clinical trials is further extended. All preclinical and clinical data must be included in the NDA submitted to the FDA for approval upon completion of the clinical trials. If the NDA is filed and in accordance with the Prescription Drug User Fee Act (PDUFA; [www.fda.gov/cder/pdufa/default.htm](http://www.fda.gov/cder/pdufa/default.htm)) the CDER should complete its initial review and act on at least 90% of all NDAs for standard drugs (those for which there are no perceived significant therapeutic benefits beyond those for available drugs) no later than 10 months after the applications are received and no later than 6 months for priority drugs (those that the FDA expects to provide significant therapeutic benefits beyond drugs already marketed) [19]. A 2002 general accounting office (GAO) report on the effect of user fees on FDA review times [20] has indicated that the median approval time for new drugs has dropped since the implementation of PDUFA in 1992. As shown in Figure 3, from 1993 to 2001, the median approval time for standard new drug applications has dropped from about 27 months to about 14 months and from about 21 months to about 6 months for priority new drugs. The reason for this drop in FDA approval times was attributed to the new resources that allowed the agency to recruit more reviewers and to upgrade its resources in information technology shortening the approval times significantly. Figure 2 summarizes the timeline and the different stages of the drug development process.





Source: FDA.

**FIGURE 3** Median approval times for standard and priority drug applications based on calendar year of approval, 1993–2001. (From Ref. 19.)

### 9.1.5 COST AND PROBABILITY OF SUCCESS

Much work has gone into determining how much the drug development process may cost from start to finish. Industry estimates on total research and development costs suggest an inflation-adjusted increase from almost \$16 billion annually in 1993 to almost \$40 billion in 2004. Out-of-pocket expenses for development of a single drug are estimated at \$403 million, with capitalized expenses nearing \$800 million (2000 dollars) [21]. Time and capital requirements grow at each clinical phase. Phase I studies have been estimated to cost \$30 million and take two to three years; Phase II studies may cost \$40 million and 2–4 years; phase III studies may cost \$86 million [21, 22]. Although size, duration, and design of the study may greatly vary, one estimate suggests that phase II clinical trials can cost anywhere from \$2000 to \$10,000 per subject [22]. Consumer advocates have questioned whether self-disclosure of drug development costs by industry may present an accurate picture of true costs, as a conflict of interest may be present [23]. Nonetheless, cost analysis of taking a drug to market is a ubiquitous part of the sponsor's decision-making process. Careful study design is essential in clinical studies to maximize data output while minimizing financial and health risks to the participants.

In addition to costs of drug development, a sponsor will also incur user fees when applying for FDA approval. The Prescription Drug User Fee Act (PDUFA) was instituted to provide resources for the FDA to review applications more quickly. Other legislation establishing user fees for medical devices and animal drugs has followed. The user fee encompasses an application fee for each drug, an establishment fee to cover the site(s) of drug production, and a product fee for each drug product in the application. Table 1 provides recent figures for user fees established by the PDUFA.

**TABLE 1 User Fees for Prescription Drug Applications in Selected Fiscal Years**

|                                                                    | 2006      | 2007      | 2008        |
|--------------------------------------------------------------------|-----------|-----------|-------------|
| Applications                                                       |           |           |             |
| Requiring clinical data                                            | \$767,411 | \$896,200 | \$1,178,000 |
| Not requiring clinical data or supplements requiring clinical data | 383,700   | 448,100   | 589,000     |
| Establishments                                                     | 264,000   | 313,100   | 392,700     |
| Products                                                           | 42,130    | 49,750    | 65,030      |

*Note:* Values have not been adjusted for inflation or capitalized in any way.

Determining whether or not a drug is likely to be approved is a challenging task. As noted in Figure 2, it is estimated that only one in five drugs that begins clinical trials is approved for market. At the NDA stage, the odds are much better. Of the NDAs submitted from 1993 to 2004, most (76%) were approved by the FDA, lesser amounts ongoing (17%) or withdrawn (7%) [19]. At each phase of clinical trials, good communication between the sponsor and the FDA team is vital. The FDA can suggest specific study designs to optimize the collection of useful and required data for its drug review. Ultimately, though, the drug or product must demonstrate an excellent safety and efficacy profile in order to ensure approval. The chance that a particular drug is approved for market depends on its performance in clinical trials.

### 9.1.6 FDA MEETINGS WITH DRUG SPONSORS

The FDA has a guidance document on the subject, Formal Meetings with Sponsors and Applicants for PDUFA Products, available at its website ([www.fda.gov/cder/guidance/2125fnl.htm](http://www.fda.gov/cder/guidance/2125fnl.htm)). Briefly, it describes the types of meetings, how and when to request a meeting, when to submit information and what information to submit prior to a meeting, procedures to conduct the meeting, and documentation of the meeting's focus and outcomes. Meetings can and should be scheduled throughout the process of investigating a drug or product, but the most critical meetings are the pre-IND meetings, end of phase II meetings, pre-NDA/BLA meetings, and the labeling meetings. At each of these checkpoints, the sponsor presents its clear scientific evidence of continuing (or discontinuing) the process, while the FDA provides feedback on the sponsor's plan and suggests improvements for a successful application.

The primary goal of the pre-IND meeting is "to introduce the drug to the FDA" [24]. If the sponsor is a small company or is not well known to the FDA, a second goal of this meeting is to present evidence that the company is qualified to perform the studies. The sponsor should present all scientific information about the drug, including any potential side effects identified in the nonclinical studies. The FDA team involved at this meeting will work with the sponsor to assess and reduce risk for the phase I study population. The best outcome from this meeting is that the FDA agrees with the sponsor that an IND can be submitted (which does not guarantee its approval).

A meeting at the end of phase II is almost universally advised. This meeting should be scheduled as soon as (1) phase II clinical data have established an effective dose and revealed pharmacokinetic and pharmacodynamic profiles that support

advancement to phase III and (2) the study design for phase III trials is complete. The phase III study design will be scrutinized thoroughly by the FDA prior to and at the meeting.

The pre-NDA/BLA meeting is recommended as a time to identify any potential pitfalls that might hinder the review of the submission. Depending on the type of drug involved, the FDA may use this meeting to discern the number and type of outside reviewers that may need to join an advisory committee. At this meeting, the FDA routinely confirms that the sponsor understands the NDA submission process and its timeline.

The labeling meeting is viewed as the last step in drug/product development prior to its approval. Although this step may seem small, the prescribing information included in product labeling can make the difference between a drug with minimal impact or one with a maximal impact on the market. A company with so much time, money, and effort in the drug development would certainly be heavily invested in securing a label that will allow the greatest impact on the market that the clinical data allow. Several rounds of negotiation on the final wording can occur. Once that hurdle is reached, the new product can be released onto the market.

The FDA can classify any meetings as type A, B, or C to prioritize the scheduling of the meeting. While sponsors can request a specific designation, the FDA is the authority on the matter. Type A meetings are the most urgent, requiring scheduling within 30 days of receiving the designation. Priority among type A meetings is given to those involving issues with a submitted NDA/BLA. Type B meetings must occur within 60 days of classification and type C within 75 days. The deadline for submission of supporting evidence is specific to the meeting type. Support documents are expected to be received by the FDA 2 weeks in advance of scheduled type A and C meetings. For type B meetings, documentation must be received 1 month in advance. It is vital for sponsors to be prepared to meet these deadlines when requesting a meeting with the FDA. For more information and advice on preparing for meetings with the FDA, see Grignolo's chapter titled "Meetings with the FDA" in *FDA Regulatory Affairs* [24].

### 9.1.7 REGULATION OF DRUGS AND BIOLOGICAL PRODUCTS BY FDA

The FDA is involved in regulation of drugs and biological products as they come to market and as they remain on the market. In addition to the regular INDs and NDAs mentioned earlier, the FDA regulates processes for approving generic drugs, over-the-counter (OTC) drugs, and follow-on drugs (modified chemical entities of approved first-in-class drugs). The FDA also has special rules and/or exemptions for expanded access and accelerated approval, orphan and pediatric drugs, and drugs for antiterrorism.

Generic drugs have the same active ingredient as an already approved, brand name drug (sometimes called the innovator drug). Applications for new generic drugs are examined in the FDA by CDER's Office of Generic Drugs. In order to be approved, generic drugs must be shown to be scientifically bioequivalent to the innovator drug. That is, the generic drug must have the same physical, pharmacokinetic, and toxicokinetic profiles as its brand-name counterpart, and it must produce the same pharmacologic effect for the same intended use. The differences between

a generic and a prescription drug are produced solely by changes in inert ingredients and/or changes in pill shape. Generic drugs are approved through an abbreviated new drug application (ANDA). The NDA is abbreviated because some or most of the clinical trials do not need to be repeated. If the drug or drug product is deemed to be bioequivalent to the innovator, that is, having the same bioavailability and pharmacokinetic profile, the same target effect, and the like, then it can be reasonably assumed to be as safe as the innovator drug.

There are a number of resources available to guide companies considering entry into the generic drug market. The FDA Office of Generic Drugs has a comprehensive website ([www.fda.gov/cder/ogd/](http://www.fda.gov/cder/ogd/)) dedicated to supporting new and worthy applications to approval for market. The FDA, together with industry feedback, has produced several guidance documents for preparation of NDAs and ANDAs. These guidance documents are neither law nor FDA rule, but rather advice from both sides of the application process on what to expect and how to present a well-prepared application. These guidance documents are fluid; they have been reviewed and updated as changes in the process or industry have warranted. A searchable database of guidance documents is available on the FDA website ([www.fda.gov/opacom/morechoices/industry/guidedc.htm](http://www.fda.gov/opacom/morechoices/industry/guidedc.htm)).

The guidance document on bioavailability and bioequivalence studies for oral drugs provides an excellent overview of the studies needed for ANDAs ([www.fda.gov/cder/guidance/3615fnl.htm](http://www.fda.gov/cder/guidance/3615fnl.htm)). It includes a general pharmacokinetic study design and data analysis. Potential generic drug producers can use the pharmacokinetic profile determined in clinical trials of the innovator drug as a benchmark for subsequent bioequivalence studies. The guidance provides several details on the types of studies to be completed, including pharmacokinetic, pharmacodynamic, and in vitro dissolution studies. Distinctions are made between immediate and modified release dosage forms.

The FDA also has a mechanism by which confidential, supporting information can be considered alongside the IND, NDA, ANDA, or export application. Detailed information involving many aspects from production facilities and processes to packaging materials can supplement an existing application under consideration using a confidential document known as the Drug Master File (DMF). The DMF is not required for submission of an IND, NDA, or ANDA, nor does the DMF replace these applications. The DMF may be particularly helpful in cases of patent-pending manufacturing processes, including some that may fall under Section 505(b)(2).

Section 505(b)(2) applications are used when some of the information within the application was not generated by or for the sponsor if the sponsor has no right of reference to the information. This application differs from an NDA in that the Section 505(b)(2) can be delayed to accommodate those with patent rights and/or exclusivity protections of a particular product. For example, an application to combine two previously approved drugs into a combination drug may necessitate the use of the 505(b)(2) application if the idea of that particular combination occurred outside the company. These types of applications have also been used when changes to existing drugs occur, such as a change from a prescription indication to an OTC indication or a change in the drug formulation. Further information on the subject is available in a guidance document titled Applications Covered by Section 505(b)(2) ([www.fda.gov/cder/guidance/2853dft.htm](http://www.fda.gov/cder/guidance/2853dft.htm)).

### 9.1.8 EXPANDED ACCESS AND ACCELERATED APPROVAL

The FDA has been making great strides in expanding the access to experimental therapies intended to treat life-threatening diseases and in accelerating the approval process of these promising treatments. The expanded access mechanisms allow severely ill patients with no response to approved therapies to have access to promising treatment drugs or devices. Expanded access can be granted through treatment IND protocols, parallel track protocols, or ordinary open-label studies that are a part of some normal NDAs. Generally, expanded access is only granted after the proposed drug or product has completed much of the clinical trial phase, when safety is established and efficacy is probable. To be approved for expanded access, four conditions must be met: (1) the drug or product must be intended to treat a serious or life-threatening disease, (2) no satisfactory alternative is available to treat that disease stage in that patient population, (3) the drug must either have an established IND protocol or all clinical trials must be completed, and (4) the sponsor must be pursuing marketing approval of the drug or product. Conversely, the parallel track mechanism allows patients with HIV/AIDS who have exhausted all other treatment options access to drugs in IND trials when the patients are not eligible to complete the trials. This mechanism is specific to HIV/AIDS-related products, and it has the potential to provide access earlier in the process than the treatment IND protocol.

In addition to expanded access mechanisms, the FDA supports new drug development in serious and life-threatening diseases through accelerated approval. A sponsor can apply for designation under the fast track drug development program. The drug or device must treat a serious aspect of a serious or life-threatening illness and address an unmet medical need.

Access to unapproved drugs may also be granted by filing a special exemption (also called compassionate exemption) or an emergency IND. Patients who are ineligible for a clinical trial may still be allowed access to the drug by filing as a special exemption. This requires investigator and sponsor approval, FDA consent, and an approved modification to the local IRB. With the permission of the drug supplier, a health care provider can alternatively file an emergency IND directly with the FDA to gain access to the drug or product. The health care provider must notify the IRB of his or her intent to use the drug; in a life-threatening situation in which specific criteria are met, according to 21 CFR 56.102(d), the drug may be administered prior to IRB approval.

### 9.1.9 ORPHAN DRUGS

The Office of Orphan Products Development is the subunit of the FDA that oversees, and sometimes funds, the clinical trials of orphan drugs. Enacted in 1983 and since revised, the Orphan Drug Act has provided tax and marketing incentives to entrepreneurs who want to study drug treatments or devices that are projected to be relevant only to small markets. These drugs and devices are considered “therapeutic orphans” if the intended target population is fewer than 200,000 individuals in the United States or in the case of preventative or diagnostic drugs, fewer than 200,000 individuals in the United States per year. When a drug or biologic receives

recognition as an orphan drug, this special status allows not only tax relief for costs related to development but also a waiver of the prescription drug user fee and a 7-year exclusivity of the drug on the market, allowing pharmaceutical companies time to recoup their expenses. Some consumer advocates propose that the exclusive license provides too great a benefit to the drug manufacturer and prevents drug access by prohibitively high end-user cost [25].

A sponsor can ask for designation as an orphan drug during or after the clinical trials. The sponsor submits documentation supporting the designation that the target disease is rare and that the drug will treat the disease. If orphan drug status is established, the investigators become eligible to compete for a research grant to defray the costs of a clinical trial. At the time of writing, the grant only provides funds for clinical trials of orphan drugs. This will facilitate approval of an unapproved drug or an unapproved use of a drug already on the market to treat a rare disease.

### 9.1.10 PEDIATRIC DRUGS

Pediatric patient populations experience growth and development that can interfere with drug absorption, distribution, metabolism, and excretion and can require special care in determining safe and effective dosing. The sponsor of a new drug or product intended to treat a disease or condition relevant to the pediatric population may request orphan drug designation, provided the drug is expected to treat 200,000 or less within this population within a given year. In addition to this designation, studies on pediatric populations for more common indications are regulated and required by the Pediatric Research Equity Act (PREA). PREA requires that all new NDAs and new BLAs for new chemical entities or new indication, dosage form, dosage strength, or route of administration contain an assessment of pediatric effectiveness unless a waiver or deferral is obtained. The FDA has posted a draft guidance document ([www.fda.gov/cder/guidance/6215dft.pdf](http://www.fda.gov/cder/guidance/6215dft.pdf)).

### 9.1.11 OTC DRUG PRODUCTS

Over-the-counter drug products comprise a special class of chemicals intended to treat a health condition that consumers may self-diagnose and self-medicate. In order to approve a drug for OTC use, the FDA must find that it is both safe and effective for the marketed use. OTC drugs are characterized by low health risk and high health benefit, low abuse and misuse potential, and adequate labeling for proper use.

OTC drugs can be marketed under two mechanisms: the NDA and the OTC drug monograph. The NDA process has been described above in detail. A change in existing OTC drug dosage form, dosage strength, or route of administration or the first time an OTC chemical entity comes to market requires new approval through the NDA. A drug previously available only by prescription must undergo NDA approval prior to introduction as an OTC drug, although sometimes this can fall under the guidelines of Section 505(b)(2). The sponsor of a prescription-to-OTC application must show that the drug is safe and effective for customer use without

the aid of a health care professional. Alternatively, an OTC drug can be marketed under an existing OTC drug monograph. The monograph regulates the active ingredients within an OTC product. Changes in inactive ingredients may not require further approval prior to marketing, provided that the active ingredient in the new product meets the standards of the monograph.

### **9.1.12 BEHIND-THE-COUNTER DRUGS**

At the time of writing, the FDA is considering a third class of drugs to join prescription and nonprescription drugs: the behind-the-counter drugs. This class would be available without a prescription but would require discussion with a pharmacist. It is unknown at this time how a new class of drugs might alter the drug approval process.

### **9.1.13 DRUGS FOR COUNTERTERRORISM**

In 2002, a measure was put into place to allow fast time-to-market for drugs or products intended to counteract the damaging effects of biological, chemical, radiological, and nuclear agents. The rule is officially titled Approval of Biological Products/New Drugs When Human Efficacy Studies Are Not Ethical or Feasible, but it is commonly called the animal rule. Because testing an antidote to these toxic substances in humans is unethical and unfeasible, the FDA must rely on well-characterized efficacy studies in animals and safety studies in humans and animals to determine whether to approve new drugs or products. Such determinations can only occur when the offending agent's mechanism of toxicity is understood, when the animal endpoints relate directly to human benefits, when the drug product's effect is determined in a species comparable to humans, and when the selection of an effective human dose is possible from the data. To date, two products have been approved by this animal rule: pyridostigmine bromide to combat nerve gas and hydroxycobalamin to treat cyanide poisoning.

### **9.1.14 GLOBALIZATION AND HARMONIZATION: FDA AND ICH**

The International Conference on Harmonisation (ICH) brings together drug regulatory bodies throughout the world to formulate international standards and to streamline national and international policies for establishing safety and efficacy of new and existing drugs through nonclinical studies and clinical trials. The FDA has a large number of guidance documents relating to harmonization; so many in fact that it has further subdivided each ICH topic into categories involving safety, efficacy, joint safety/efficacy, and quality. The common technical document (CTD) that the FDA uses as a format for all its electronic NDA, ANDA, BLA, and IND submissions was designed by the ICH. The FDA encourages sponsors to use the electronic CTD in hopes that it will increase the efficiency of global marketing approval ([www.fda.gov/cder/guidance/7087rev.pdf](http://www.fda.gov/cder/guidance/7087rev.pdf)). In addition to globalization of the drug approval process, the ICH with the FDA and other agencies is examining implications of drug



interactions, including the use of drugs developed outside one's country together with cultural remedies [26].

## REFERENCES

1. FDA (1999), *Food and Drug Administration: An Overview*, Publication No. BG99-2, U.S. Government Printing Office, Washington, DC.
2. Hiltz, P. J. (2003), *Protecting America's Health: The FDA, Business, and One Hundred Years of Regulation*, Alfred A. Knopf, New York.
3. Pisano, D. J. (2004), Overview of drug development and the FDA, in Pisano, D. J. and Mantus, D., Eds; *FDA Regulatory Affairs: A Guide for Prescription Drugs, Medical Devices, and Biologics*, CRC Press, Boca Raton, FL.
4. FDA (2002), A Guide to Resources on the History of the Food and Drug Administration; available at: <http://www.fda.gov/oc/history/resourceguide/default.htm>.
5. FDA (2003), *FDA's Sentinel of Public Health: Field Staff Safeguards High Standards*, Publication No. FS 01-7, U.S. Government Printing Office, Washington, DC.
6. FDA (2002), *FDA's Center on the Front Line of the Biomedical Frontier*, Publication No. FS 01-4, U.S. Government Printing Office, Washington, DC.
7. FDA (2002), *Better Health Care with Quality Medical Devices: FDA on the Cutting Edge of Device Technology*, Publication No. FS 01-5, U.S. Government Printing Office, Washington, DC.
8. FDA (2003), *Improving Public Health: Promoting Safe and Effective Drug Use*, Publication No. FS 01-3, U.S. Government Printing Office, Washington, DC.
9. FDA (2002), *Keeping the Nation's Food Supply Safe: FDA's Big Job Done Well*, Publication No. FS 01-2, U.S. Government Printing Office, Washington, DC.
10. Anon. (2007), CVM Introduction; available at: <http://www.fda.gov/cvm/aboutint.htm>; accessed July 14, 2007.
11. FDA (2007), NCTR's Mission; available at: <http://www.fda.gov/nctr/overview/mission.htm>; accessed July 14, 2007.
12. Code of Federal Regulations, 21 CFR Part 58.
13. FDA (2007), Investigational New Drug (IND) Application Process; available at: [http://www.fda.gov/Cder/regulatory/applications/ind\\_page\\_1.htm#Introduction](http://www.fda.gov/Cder/regulatory/applications/ind_page_1.htm#Introduction); accessed July 15, 2007.
14. Code of Federal Regulation, 21 CFR Part 50.
15. Code of Federal Regulation, 21 CFR Part 56.
16. Anon. (2006), Inside clinical trials: Testing medical products in people, FDA Consumer Mag., Publication No. FDA 06-1524G.
17. Anon. (2006), The FDA's drug review process: Ensuring drugs are safe and effective, FDA Consumer Mag., Publication No. FDA 06-1524G.
18. Code of Federal Regulations, 21 CFR Part 314.
19. GAO (2006), New Drug Development: Science, Business, Regulatory, and Intellectual Property Issues Cited as Hampering Drug Development Efforts, Report No. GAO-07-49.
20. GAO (2002), Effect of User Fees on Drug Approval Times, Withdrawals, and Other Agency Activities, Report No. GAO-02-958.
21. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22(2), 151–185.

22. Schacter, B. (2006), *The New Medicines: How Drugs Are Created, Approved, Marketed, and Sold*, Praeger, Westport, CT.
23. Angell, M. (2004), *The Truth about the Drug Companies: How They Deceive Us and What to Do About It*, Random House, New York.
24. Grignolo, A. (2004), Meeting with the FDA, in Pisano, D. J., Mantus, D., Eds., *FDA Regulatory Affairs: A Guide for Prescription Drugs, Medical Devices, and Biologics*, CRC Press, Boca Raton, FL.
25. Thamer, M., Brennan, N., and Semansky, R. (1998), A cross-national comparison of orphan drug policies: Implications for the U.S. Orphan Drug Act, *J. Health Polit. Policy Law*, 23(2), 265–290.
26. Huang, S. M., Temple, R., Throckmorton, D. C., and Lesko, L. J. (2007), Drug interaction studies: Study design, data analysis, and implications for dosing and labeling, *Clin. Pharmacol. Ther.*, 81(2), 298–304.

# 9.2

---

## *Phase I Clinical Trials*

**Elizabeth Norfleet and Shayne Cox Gad**

*Gad Consulting Services, Cary, North Carolina*

### **Contents**

|                                          |     |
|------------------------------------------|-----|
| 9.2.1 Overview                           | 245 |
| 9.2.2 Purpose and Objectives             | 246 |
| 9.2.3 Phase I Trial Design               | 247 |
| 9.2.4 Design Types                       | 247 |
| 9.2.5 Phase I Trial Characteristics      | 249 |
| 9.2.6 Critical Parameters to Measure     | 250 |
| 9.2.7 PK Parameters to Derive            | 251 |
| 9.2.8 Regulatory Requirements and Issues | 252 |
| References                               | 253 |
| Bibliography                             | 254 |

### **9.2.1 OVERVIEW**

Phase I trials [also referred to as FIH (first in human) or FIM (first in man)] are the earliest stage clinical trials of a new drug or device typically performed with just a few persons to determine the safety and pharmacokinetics of a new drug or biocompatibility of a new invasive medical device; for drugs, dosage or toxicity limits should be obtained. These rigorously controlled tests of a new drug or a new invasive medical device involve human subjects for such evaluation. In the United States, such trials are conducted with the concurrence of the Food and Drug Administration (FDA) or equivalent regulatory authority before proceeding to further clinical investigation. While it is generally the ideal to perform no more than two (a single-dose escalating and a multiple-dose escalating) or three (for oral drugs, a feed/fasted

is also usual) phase I studies before proceeding to phase II, it is not uncommon to perform as many as eight (formulation changes, etc.).

Phase I trials are conducted following prescribed preclinical work that establishes the new molecular entity (NME) to be safe and tolerable in animal and in vitro models. The animal models chosen for such evaluation should reflect the expected response in humans as closely as possible. After preclinical work has been successfully conducted, a complete and thorough copy of all preclinical data must be submitted to the FDA in the format of an investigational new drug (IND). Subsequent to the FDA's evaluation of the data, an initial phase I trial may then commence if the sponsor has not received a response from the FDA within 30 days of the IND submission. Similar pretrial review procedures exist for other countries and for medical devices. In the United States and the European Union (EU), review and approval by an institutional review board (IRB) or ethics committee (EC) prior to trial initiation is also required.

The primary objectives phase I trials seek to accomplish are to assess safety, tolerability, pharmacokinetics, and to determine the MTD (maximum tolerated dose). They typically are small trials ranging from 20 to 80 subjects and are relatively short, only lasting 6 months or less from initiation to trial completion. In general, phase I trials are performed in normal healthy volunteers (though the guinea pig zero effects may be operative [1]). However, in special cases such as drugs for life-threatening diseases such as cancer, AIDS (acquired immunodeficiency syndrome), amyotrophic lateral sclerosis (ALS), or where the drug is a combination of two already approved drugs, patients may constitute the initial trial population.

The importance of FIM trials is crucial to the progression of the new drug molecule. Identifying a safe dose and dosing regimen is vital, and in doing so the utmost precaution and care must be taken during this stage through strict monitoring.

### 9.2.2 PURPOSE AND OBJECTIVES

As stated above, the main objectives of phase I trials are to assess tolerability, monitoring the side effects in relation to increase of the doses, and confirm the maximum tolerated dose. Additionally, establishing a safety profile is paramount. These are three of the commanding goals, providing a firm basis for investigators to determine further testing and study design strategies. It is almost certainly the most critical "go/no-go" decisive point.

Other important aims are to obtain adequate information of the drug's basic pharmacokinetics (PK) such as absorption, distribution, metabolism, and excretion (ADME). Pharmacokinetics refers to the behavior of the drug in the body. Absorption is the process of a substance entering the body. Distribution is the dispersion or dissemination of substances throughout the fluids and tissues of the body. Metabolism is the transformation of the substances and its daughter metabolites. Excretion is the elimination of the substances from the body. This data gives rise to analyzing the pharmacological effects of the drug in the human body and developing a sufficient PK model to simulate exposure and response for intended dosing regimens, incorporating variability [2]. A good dose–response curve derived from PK analysis is greatly desired. Additionally, information is sought out to determine the drug's

bioavailability. Bioavailability is defined as the ability to deliver the drug in a usable form to the disease target. Phase I studies are also intended to determine whether the drug is best delivered orally, or by injection, or through the skin and by what regimen.

Furthermore, FIM trials study the drug's structure–activity relationship (SAR) with previously evaluated compounds and mechanism of action (MOA), which provide the basis to investigate biological phenomena or disease processes. However, data regarding efficacy [pharmacodynamics (PD)], though a highly desired aspect of the drug, is rarely gathered during phase I. Some studies will often shed enough early light on possible efficacy to support the decision on further continuance of the study.

### 9.2.3 PHASE I TRIAL DESIGN

More recently, it has become the common practice to divide phase I trials into two phases or separate stages, phase Ia and phase Ib. Phase Ia studies include the first dose in humans and are typically short-term single-dose studies to confirm safety before beginning a larger, more extensive trial. The studies on average are usually comprised of six cohorts, with about seven subjects per cohort. The main design usually entails escalating the dose with each new cohort while stringently monitoring for safety, with a final objective of establishing the MTD and a solid PK profile. Oftentimes a placebo is included to provide a more accredited safety evaluation.

Phase Ib studies are typically more comprehensive repeat-dose studies with the same goals being safety, tolerability, and PK on a repeat-dose level in order to assess the drug's therapeutic effect. Other frequent questions addressed in phase Ib studies include food effects and gender differences.

Phase I trials are also different from subsequent trials in that not only are they not conducted at a sponsor's facility (no clinical trials are), but they are almost always conducted by and at contract research organizations (CROs) that have this as their business.

### 9.2.4 DESIGN TYPES

When planning a clinical trial, the design and determination of the dose range depends on the MTD established in preclinical studies. The most common trial design is dose escalation/ascending. This study involves the gradual increase of drug dosages to determine the amount that delivers the best balance of high efficacy and acceptable side effects. It is basically viewed as a building block design. The primary objective is to determine MTD. The MTD is defined as the highest dose that is tolerated with no adverse side effects observed. If the dose is not tolerated, then the dose just below is the MTD. Typically, one may dose the subject, look for adverse events, and if none are present, increase dose and repeat. One may have a fixed dose increment or an even more risky approach is to bounce around until the MTD is established. One preferred way to determine AEs (adverse events) is by comparison to a placebo group. Conducting thorough dose-ranging and dose–response studies early in product development reduces the possibility of later failed phase II or III

studies. Of drugs tested in phase I, 50–70% are abandoned because of problems with safety or efficacy [3]. At the end of each phase of dosing, typically a data safety monitoring board (DSMB) meeting is held to decide if the drug is safe enough to continue, and if yes, typically one will want to recruit more patients. When analyzing a dose-escalating study the following should be included:

- Comparison between treatment groups (sometimes several doses polled).
- Comparison of active groups versus placebo.
- Comparison of before and after dosing (temporal factors).
- Evaluate AE incidence and complete a summarization.

In some instances, mainly for bioequivalence studies, a crossover design is used. The ICH E9 [4] guideline defines the crossover design as a study in which each subject is randomized to a sequence of two or more treatments and hence acts as his own control for treatment comparisons. This simple maneuver is attractive primarily because it reduces the number of subjects and usually the number of assessments needed to achieve a specific power, sometimes to a marked extent. In the simplest  $2 \times 2$  crossover design, each subject receives each of two treatments in randomized order in two successive treatment periods; often separated by a washout period. It is the type of experiment where experimental units are given several treatments in succession. The order of treatments ought to be set by some random procedure and each subject receives both treatments. In this type of trial design comparisons within subjects are performed.

Advantages of a crossover study are that it is useful for reversible effects (e.g., bioequivalence), fewer subjects are needed, and more precise comparisons are obtained.

If applicable, one can have more than two treatments in which an extended crossover design may be applied (which the FDA has been leaning more toward since 2000). An extended crossover design includes two treatments with four periods.

Crossover designs are not flawless, however, and contain a number of problems that can invalidate the concluding results. The chief difficulty concerns carryover, that is, the residual influence of treatments in subsequent treatment periods; therefore a washout period of at least 6 half-lives is desired. The main cause of carryover is that the treatment effects lingering in PK studies, because the washout is too short thus leading to overlap of dose effects.

When analyzing a crossover study the following should be taken into consideration:

$Y = \text{period} + \text{sequence} + \text{treatment} + \text{subject (sequence)} + \text{carryover} + \text{error}.$

Subject is usually considered a random effect.

In a  $2 \times 2$  crossover, sequence is confounded with carryover—if there is a significant carryover, use the first period only.

Comparisons are intrasubject (treatment comparisons).

When analyzing a bioequivalence (BE) crossover study the following should be taken into consideration:

BE is typically testing utilizing the crossover design.

Response is area under the curve (AUC) and the peak concentration of the drug ( $C_{\max}$ ).

Key endpoint is the ratio of treatment A AUC over treatment B AUC.

A and B are equivalent if 90% confidence interval for ratio lies entirely between (0.8, 1.25) 80% and 125%.

Perform analysis on log AUC and log  $C_{\max}$ .

## 9.2.5 PHASE I TRIAL CHARACTERISTICS

The initial (true FIM) administration of a new drug must be undertaken with some circumspection. Nonclinical (animal) safety studies give good predictions of clinical safety and expected toxicity almost all of the time, with application of current scaling factors for species difference providing confidence that initial doses will present minimal risk. Still, in normal volunteers there are exceptions—particularly highly humanized proteins for which animals lack the mechanisms for response, and therefore provide no assurance of safety (as with the TGN1412 drug trial).

A similar situation occurs when phase I trials are performed in patients (or, indeed, with the first trial in patients if earlier trials were done in normal volunteers). In the latter (patient) treatment case, one should carefully consider which organ, metabolic, or protective systems may already be compromised—and therefore for which one must be very attentive to even weak indication of adverse effect [as with the FIAU (fialuridine) clinical trial].

In either of these cases, or indeed in almost any true FIM dosing, it is generally wise to offset the dosing of the initial cohort—that is, separate dosing of individual cohort members by an appropriate period of time.

Both types of these designs share many similar characteristics indicative of phase I trials. Both serve to provide the information to accurately evaluate safety, tolerability, MTD, and PK. Additionally, they are relatively short and small trials, comprised of about 20–80 subjects usually healthy volunteers, with the exceptions for trials dealing with life-threatening diseases such as HIV/AIDS, cancer, ALS, and the like. The FDA explicitly draws the distinction between a healthy volunteer and a subject or patient. A healthy volunteer is defined as a healthy person who agrees to participate in a clinical trial for reasons other than medical and receives no direct health benefit from participating. Whereas a human subject is defined as an individual who is or becomes a participant in research, either as a recipient of the test article or as a control. A subject may be either a healthy human or a patient (21 CFR 50.3).

Whether it be a healthy volunteer or a patient, they must satisfy the subject/patient criteria outlined in the sponsor's protocol and must provide informed consent. Since phase I trials incorporate dosing a person with an investigational new drug, extremely close monitoring for safety and determination of a tolerated dose is crucial.

A placebo is almost always included as a standard component in phase I trials as well. The main advantage of using a placebo is to serve as a control and decrease bias in an attempt to make some sort of quantitative assessment of the drug's efficacy. A disadvantage to using a placebo is mainly seen in trials for life-threatening illnesses, and in some such cases a placebo is not required. A common approach is



to include a single placebo subject in each cohort. By the end of an ascending dose trial, enough placebo subjects have been accumulated to allow for an estimation of placebo (and nocebo) effects.

The issue of blinding in clinical trials is imperative for the prevention of bias and integrity of the data. The ICH E9 [4] definition of blinding is as follows:

Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from the influence that the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of endpoints, the handling of withdrawals, the exclusion of data from analysis, and so on. The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.

Phase I trials are usually double blind. ICH E9 [4] definition of double blind:

A double-blind trial is one in which neither the subject nor any of the investigator or sponsor staff involved in the treatment or clinical evaluation of the subjects are aware of the treatment received. This includes anyone determining subject eligibility, evaluating endpoints, or assessing compliance with the protocol. This level of blinding is maintained throughout the conduct of the trial, and only when the data are cleaned to an acceptable level of quality will appropriate personnel be unblinded.

## 9.2.6 CRITICAL PARAMETERS TO MEASURE

### Safety

#### A. Clinical Examinations

1. Physical
2. Vital signs (usually considered as part of the physical examination)
3. Height and weight (state of dress is usually specified, e.g., socks)
4. Neurological or other specialized clinical examinations

#### B. Clinical Laboratory Examinations

1. Hematology
2. Clinical chemistry
3. Urinalysis
4. Virology (viral cultures or viral serology)
5. Immunology or immunochemistry (e.g., immunoglobins, complement)
6. Serology
7. Microbiology (including bacteriology and mycology)
8. Parasitology (e.g., stool for ova and protozoa)
9. Pulmonary function tests (e.g., arterial blood gas)
10. Other biological tests (e.g., endocrine, toxicology screen)
11. Stool for occult blood (specify hemocult or Guaiac method)
12. Skin tests for immunologic competence
13. Medicine screen (usually in urine) for detection of illegal or nonprotocol-approved medicines
14. Bone marrow examination

15. Gonadal function (e.g., sperm count, sperm motility)
16. Genetics studies (e.g., evaluate chromosomal integrity)
17. Stool analysis using in vivo dialysis
- C. Probe for Adverse Reactions
- D. Psychological and Psychiatric Tests and Examinations
  1. Psychometric and performance examinations
  2. Behavioral rating scales
  3. Dependence liability
- E. Examinations Requiring Specialized Equipment (selected examples)
  1. Audiometry
  2. Electrocardiogram (EKG)
  3. Electroencephalogram (EEG)
  4. Electromyography (EMG)
  5. Stress test
  6. Endoscopy
  7. Computed tomography (CT) scans
  8. Ophthalmological examination
  9. Ultrasound
  10. X rays
  11. Others

### 9.2.7 PK PARAMETERS TO DERIVE

The purpose of human pharmacokinetic studies is to examine the rate of absorption, distribution, metabolism, and excretion of a drug. Findings from these studies describe how the drug travels through the body and where and how it is eliminated. PK data allows for the detection of drug levels in human blood/urine samples. After PK information is obtained, a dose–response curve should be plotted that describes the change in effect on the subject caused by differing levels of dose exposure to a stressor. Studying dose–response, and developing dose–response models, is central to determining the safe and hazardous levels and dosages for the drug. A good dose–response curve is obviously highly desired.

The dose–response curve (Fig. 9.2.1) defines the relationship between dose and response based on the following assumptions: (1) response increases as dose increases; (2) there is a threshold dose—a dose below which there is no effect. The quality of results depend on the placement of time points; therefore you want to see at least a 24-hour profile, with two or three time points for each component. However, sometimes limitations arise such as the ability to draw blood, the patience of the study population, and sometimes the study design or medical needs.

Another consideration taken when evaluating PK data is the food effect, which assesses how food affects the absorption of the drug; but this is only tested for with orally administered compounds. Other vital parameters to be drawn and evaluated are listed and defined as follows:

- $AUC_{0-\infty}$  represents the total amount of drug absorbed by the body, irrespective of the rate of absorption. This is useful when trying to determine whether two

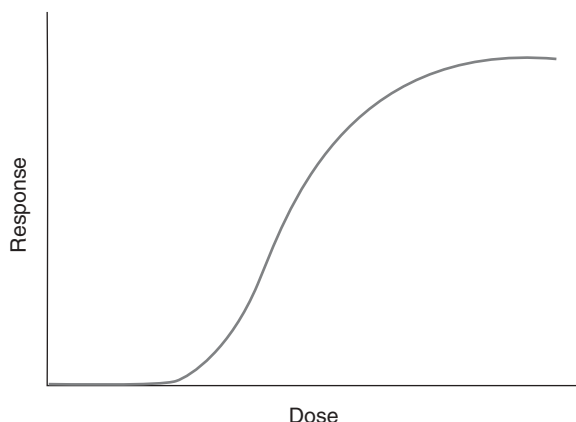


**FIGURE 1** Dose–response curve.

formulations of the same dose (e.g., a capsule and a tablet) release the same dose of drug to the body.

- $AUC_{0-T}$  represents the average concentration over a time interval,  $AUC/T$ .
- $C_{max}$  represents the peak concentration of the drug in plasma.
- $T_{max}$  represents the duration of time to reach the peak concentration of drug in plasma, beginning from administration of the drug.
- $T_{1/2}$ —half life, or the time it takes for half of the administered or absorbed dose to be cleared or metabolized.
- $V_0$ —central and peripheral. The theoretical volume in which the drug is homogeneously distributed and is basically dependent upon the lipid or water solubility of the drug and its particular affinity for given tissues or structures.
- Clearance—the volume of plasma that is completely cleared of drug per unit time.
- MRT (mean residence/residual time)—The average total time molecules of a given dose spend in the body. Thus, this can only be measured after instantaneous administration.

## 9.2.8 REGULATORY REQUIREMENTS AND ISSUES

In order to proceed into FIM or phase I trials, permission must be obtained from both the national drug regulatory authority in the country where the trial will take place [in the United States it is the Food and Drug Administration (FDA); in the European Union it is the European Medicines Agency (EMA)], and the national ethics committees, the institutional review board (IRB) in the United States. The IRB is an independent group officially authorized to approve, monitor, and review biomedical and behavioral research involving humans with the objective to protect the rights and welfare of the subjects. In the United States, the FDA and Health and Human Services (HHS) regulations have empowered IRBs to approve, require modifications to gain approval, or disapprove research. An IRB performs critical monitoring tasks for research conducted on human subjects that are scientific,

ethical, and regulatory. IRBs are required to have at least five members with varying backgrounds to promote complete and adequate review of research activities commonly conducted by the institution. The purpose of an IRB review is to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of humans participating as subjects in a research study. The review will cover materials such as the protocol, informed consent documents, and advertisements, with special attention paid to trials that may include vulnerable subjects, such as pregnant women, children, prisoners, the elderly, or persons with diminished comprehension.

The U.S. FDA requires an IND application that includes all the nonclinical and technical data to review. Typically, sponsors will request a pre-IND meeting with the FDA to discuss safety issues related to the proper identification, strength, quality, purity, or potency of the investigational drug, as well as to identify potential clinical hold issues. The pre-IND meeting should focus on the specific questions related to the planned clinical trials. The meeting should also include a discussion of various scientific and regulatory aspects of the drug as they relate to safety and/or potential clinical hold issues. The IND application contents fall primarily into three categories: animal pharmacology and toxicology studies, chemistry and manufacturing information, and clinical protocols and investigator information. Animal pharmacology and toxicology studies are comprised of preclinical data to allow an evaluation as to whether the NME is reasonably safe for initial testing in humans. Also included are any previous experience with the drug in humans (often foreign use). Chemistry and manufacturing information includes information pertaining to the chemical composition, manufacturing methods, stability, and controls used for manufacturing the drug substance and the drug product. The chemical stability and activity of the product must also have been tested. This information is reviewed to ensure that the company can adequately produce and supply consistent and active batches of the drug. Clinical protocols and investigator information consist of detailed protocols for proposed clinical studies to determine whether the initial-phase trials will expose subjects to unnecessary risks. Also, information on the qualifications of the clinical investigators who are to oversee the administration of the experimental compound is included in order to assess whether they are qualified to fulfill their clinical trial duties. An investigator's brochure (IB), which is a document intended to educate the trial investigators of the pertinent facts concerning the trial drug they need to know to conduct their study with the least hazard to the subjects is also submitted. Furthermore, commitments to obtain informed consent from the research subjects, to obtain review of the study by an institutional review board (IRB), and to adhere to the investigational new drug regulations are included. If the sponsor has not received a response from the FDA within 30 days, the trial may begin.

## REFERENCES

1. Helms, R., Ed. (2002), *Guinea Pig Zero: An Anthology of the Journal for Human Research Subjects*, Garret County Press, New Orleans.
2. Chien, J. Y., Friedrich, S., Heathman, M. A., et al. (2005), Pharmacokinetics/pharmacodynamics and the stages of drug development: Role of modeling and simulation, *AAPS J.*, 7(3), E544–549.

3. Lee, C.-J., Lee, L. H., Wu, C. L., et al. (2006), *Clinical Trials of Drugs and Biopharmaceuticals*, CRC Press, Boca Raton, FL.
4. ICH E9 (2005), *Statistical Principles for Clinical Trials*. International Conference on Harmonization.

## BIBLIOGRAPHY

- Gallin, J. I., and Ognibene, F. P., Eds. (2007), *Principles and Practices of Clinical Research*, 2nd ed., Academic Press, Burlington, MA.
- Green, S., Benedetti, J., and Crowley, J. (2002), *Clinical Trials in Oncology*, 2nd ed., Chapman & Hall/CRC Press, Boca Raton, FL.
- Machin, D., Day, S., and Green, S. (2004), *Textbook of Clinical Trials*, Wiley, Hoboken, NJ.
- O'Grady, J., and Joubert, P. (1997), *Handbook of Phase I/II Clinical Drug Trials*, CRC Press, Boca Raton, FL.
- O'Grady, J., and Linet, O. (1990), *Early Phase Drug Evaluation in Man*, CRC Press, Boca Raton, FL.
- Rang, H. P. (2005), *Drug Discovery and Development: Technology in Transition*, Churchill Livingstone Elsevier, Oxford, UK.
- Stone, J. (2006), *Conducting Clinical Research*, Mountainside MD Press, Cumberland, MD.
- U. S. Food and Drug Administration, Center for Biologics Evaluation and Research. (2001), Guidance for Industry IND Meetings for Human Drugs and Biologics Chemistry, Manufacturing, and Controls Information.

# 9.3

---

## *Phase II Clinical Trials*

**Say-Beng Tan<sup>1-4</sup> and David Machin<sup>3-5</sup>**

<sup>1</sup>*Singapore Clinical Research Institute, Singapore*

<sup>2</sup>*Duke–NUS Graduate Medical School*

<sup>3</sup>*Division of Clinical Trials and Epidemiological Sciences, National Cancer Centre, Singapore*

<sup>4</sup>*Clinical Trials and Epidemiology Research Unit, Singapore*

<sup>5</sup>*Children's Cancer and Leukaemia Group, University of Leicester, Leicester, United Kingdom*

### **Contents**

|         |                                                                   |     |
|---------|-------------------------------------------------------------------|-----|
| 9.3.1   | Overview                                                          | 256 |
| 9.3.1.1 | Phase II Trials                                                   | 256 |
| 9.3.2   | Planning a Phase II Trial                                         | 257 |
| 9.3.2.1 | Choice of Endpoint                                                | 257 |
| 9.3.2.2 | Eligibility                                                       | 258 |
| 9.3.2.3 | Choice of Design                                                  | 258 |
| 9.3.3   | Single-Stage Designs                                              | 260 |
| 9.3.3.1 | Fleming–A'Hern                                                    | 260 |
| 9.3.4   | Two-Stage Designs                                                 | 261 |
| 9.3.4.1 | Gehan                                                             | 261 |
| 9.3.4.2 | Simon Optimal and Minimax                                         | 262 |
| 9.3.5   | Phase II Trials with Survival Endpoints                           | 264 |
| 9.3.5.1 | Case and Morgan                                                   | 264 |
| 9.3.6   | Efficacy and Toxicity in Phase II Trials                          | 265 |
| 9.3.6.1 | Bryant and Day                                                    | 266 |
| 9.3.7   | Bayesian Approaches                                               | 267 |
| 9.3.7.1 | Motivation                                                        | 267 |
| 9.3.7.2 | Overview of Bayesian Approaches in the Context of Phase II Trials | 267 |
| 9.3.7.3 | Bayesian Single and Dual Threshold                                | 268 |
| 9.3.8   | Randomized Phase II Trials                                        | 271 |
| 9.3.8.1 | Simon, Wittes, and Ellenberg (SWE)                                | 271 |

|         |                             |     |
|---------|-----------------------------|-----|
| 9.3.9   | Trial Conduct and Reporting | 272 |
| 9.3.9.1 | Trial Conduct               | 272 |
| 9.3.9.2 | What to Report              | 273 |
| 9.3.10  | Concluding Remarks          | 274 |
|         | References                  | 275 |

### 9.3.1 OVERVIEW

New compounds are continually being developed with the intended objective that some of these will prove useful in the treatment of human diseases. Any new compound has to undergo a rigorous development process before it can be introduced for standard clinical use. New compounds are first tested for safety in laboratory and animal studies whose objectives are the characterization of the drug's pharmacology, toxicology, metabolism, and other properties.

The new compound is then tested in human subjects in phase I studies that, if successful, are the first of an eventual series in humans. These studies are designed to determine the metabolic and pharmacological actions of the drug, the side effects associated with increasing doses (to establish a safe dose range), and, if possible, to gain early evidence of activity. The focus is typically on determining the toxicity profile of the drug and on finding a potentially therapeutic effective dose.

After a suitable dose has been established in phase I trials, phase II trials are conducted with the main objective being to evaluate the drug's activity for patients with a particular disease or condition, as well as to determine the level of (short-term) side effects and possible risks associated with the use of the drug.

If sufficient preliminary evidence of effectiveness is found in these phase II trials, suggesting worthwhile potential therapeutic efficacy of the drug, the drug may be further evaluated in randomized phase III trials, where it will be compared with the current standard treatment in a larger number of patients.

#### 9.3.1.1 Phase II Trials

Although there is some variation in terminology and objectives depending on the disease or condition in question, phase II trials are usually single-armed studies with the objective being to investigate the antidisease activity of the new therapeutic regimen. Most of these trials evaluate at least one new treatment relative to a standard. So they are inherently comparative, even though the "standard" treatment information is usually historical and is not obtained prospectively as part of the phase II study itself.

In certain circumstances, there may be several candidate agents for phase II testing; the problem here is to select that which has the most potential to be effective and hence be tested in a phase III trial. This leads to randomized phase II trial designs, which seek to reduce the apparent variability in response rates observed in different studies of the same compound. Some factors that contribute to the variability include patient type, definition of response, interobserver variability in response evaluation, drug dosage and schedule, reporting procedures, and sample



size. Patients in randomized phase II studies are randomized to one of several experimental treatments, but the limited sample size of these trials does not provide sufficient statistical power to make reliable treatment comparisons.

In areas such as oncology, phase II trials often focus on both safety and efficacy of the new therapeutic regimen. Safety is usually assessed in terms of toxicity rates but is not always part of the formal design process. Efficacy is typically measured using tumor response, often the percentage decrease in tumor size compared with that before treatment commences, and so patients enrolled in phase II trials need to have measurable disease.

**Example: Phase II Trial—Carcinosarcoma of Female Genital Tract** Van Rijswijk et al. [1] conducted a phase II trial in 48 women with carcinosarcoma of the genital tract. Although the activity of the combination of cisplatin, doxorubicin, and ifosfamide was established (overall response rate 56%), they did not recommend this treatment combination but suggested that combinations “with more favorable toxicity profiles should be explored.”

## 9.3.2 PLANNING A PHASE II TRIAL

### 9.3.2.1 Choice of Endpoint

As already indicated, phase II trials seek to assess whether a new regimen is active enough to warrant a comparison of its efficacy with the standard treatment regimen in a phase III trial. Thus, appropriate endpoints need to be chosen to allow for such an assessment to be made. For example, in human immunodeficiency virus (HIV) research, suitable endpoints might include measures of viral load or immune function. In cardiovascular disease, we might look at blood pressure or lipid levels.

In certain situations, there is not necessarily an obvious measure to take. For example, in oncology, although one may regard tumor shrinkage as a desirable property of a cytotoxic drug when given to a patient, it is not immediately apparent how this is to be measured. Were every tumor of regular spherical shape, the direction in which it is measured is irrelevant. Furthermore the diameter, a single dimension, leads us immediately to the volume of the tumor. However, no real tumor will comply with this ideal geometrical configuration, and this has led to measures such as the product of the two largest (perpendicular) diameters to describe the tumor and then a reduction in this product to indicate response. Precisely what is the best measure to assess tumor shrinkage has been discussed by an international panel and reported in detail by Therasse et al. [2].

More generally, they offer guidelines to encourage more uniform reporting of outcomes particularly for clinical trials. Investigators of future trials may argue about the fine details, and no doubt in time these guidelines will need revision, but they would be foolish to ignore these recommendations when conducting and subsequently reporting their study.

If there are “justifiable” reasons why other criteria should be used, or the recommendations cannot be followed for whatever reason, then before the study commences, these should be reviewed by the investigating team. There is little point in

conducting a study using measures not acceptable to other groups, including referees for the clinical journals, as little note will then be taken of the results. The best option is to follow the guidelines for the primary endpoint, use the “local” measures for secondary reporting, and contrast the two in any discussion.

Sometimes it may be that the true endpoint of interest is difficult to assess for whatever reason. In this case, a surrogate may be sought. For example, when investigating the possibilities of a novel marker for prognosis, it may be tempting to use disease-free survival (DFS) as a “surrogate endpoint” for the overall survival (OS) time of patients with the cancer concerned. The reason being that for many cancers, relapse occurs well before death, and so the evaluation of the marker can occur earlier in time than would the case if OS was to be observed.

More generally, a surrogate endpoint is a biomarker (or other indicator) that is intended to substitute for a (often) clinical endpoint and predict its behavior. If a surrogate is to be used, then there is a real need to ensure that it is an appropriate surrogate for the (true) endpoint of concern.

### 9.3.2.2 Eligibility

Common to all phases of clinical trials is the necessity to define precisely who are eligible subjects. This definition may be relatively brief or complex depending on the substance under test.

At the very early stages of the process, it is particularly understandable that great care is taken in subject and particularly patient choice. In these situations, when relatively little is known about the compound, all the possible adverse eventualities have to be considered. This usually results in quite a restricted definition as to those that can be recruited.

Once the possibility of some activity (and hence potential efficacy) becomes indicated, then there is at least a prospect of therapeutic gain for the patient. In this case, the investigators may expand the horizon of eligible patients but simultaneously confine them to those in which a measurable response to the disease can be ascertained.

**Example: Eligibility for Phase II Trial—Gemcitabine in Nasopharyngeal Carcinoma** Foo et al. [3] specify that patients were to have histologically confirmed undifferentiated carcinoma arising from the nasopharynx, *bidimensionally measurable* disease not within any prior radiotherapy fields, be between 18 and 75 years, with an Eastern Cooperative Oncology Group (ECOG) performance status (PS) < 2. In addition there were seven clinical chemistry limits that had to be satisfied before inclusion was possible.

### 9.3.2.3 Choice of Design

There are a relatively large number of alternative designs for phase II trials. These include single-stage designs in which a predetermined number of patients are recruited, and two-stage designs in which patients are recruited in two stages and the move to stage 2 being consequential on the results observed in stage 1. Multi-stage designs have also been proposed, but the practicalities of having several

decision points have limited their use because of the inherent further delays involved with each extra stage. Most phase II trials are of a single-arm, noncomparative design. However, randomized phase II selection designs in which the objective is to select only one, the “best,” of several agents tested simultaneously, are strongly recommended in some situations.

Although the usual (single) endpoint for phase II studies typically involves some binary measure of activity, designs are available when activity is assessed by survival time, and when the dual endpoints of activity (response) and acceptable levels of toxicity are stipulated by the design.

With such a plethora of different options for phase II designs, it is clearly important that the investigators choose the design that is best for their purpose. In some cases the choice will be reasonably clear. For example, if one has several compounds to test at the same time, then the randomized selection design will be preferred to (say) a series of parallel single-arm studies. In other circumstances, the patient pool may be very limited and a key consideration will be the maximum numbers of patients that might have to be recruited. Features to guide investigators in their choice are summarized in Table 1.

In this chapter, we will briefly discuss a number of different phase II designs. However, a detailed discussion of the designs and sample size tables is beyond the scope of this chapter. Interested readers are referred to the *Sample Size Tables for Clinical Studies* [4]. The accompanying software also allows for the implementation of all the designs we discuss.

**TABLE 1 Comparative Properties of Alternative Phase II Designs**

| Single Stage       | No Stopping Rules            |                                                                                                                  |
|--------------------|------------------------------|------------------------------------------------------------------------------------------------------------------|
| Fleming–A’Hern     | Sample size fixed.           | Size determined at the design stage.                                                                             |
| Randomized         | Sample size fixed.           | Size determined at the design stage and depends on the number of compounds under test.                           |
| Two Stage          | Allow Early Termination      |                                                                                                                  |
| Gehan              | Maximum sample size unknown. | Final sample size depends on the number of responses in stage 1.                                                 |
| Simon—Optimal      | Maximum sample size fixed.   | Stage 1 sample size chosen to ensure inactive compound does not go to stage 2.                                   |
| Simon—Minimax      | Maximum sample size fixed.   | Designed for maximum sample size to be a minimum.                                                                |
| Tan—Machin         | Maximum sample size fixed.   | Stage 1 sample size chosen to ensure inactive compound does not go to stage 2.                                   |
| Two Stage          | Allow Early Termination      | Dual Endpoint                                                                                                    |
| Bryant—Day—Optimal | Maximum sample size fixed.   | Stage 1 sample size chosen to ensure inactive or too toxic compound does not go to stage 2.                      |
| Two Stage          | Allow Early Termination      | Survival Endpoint                                                                                                |
| Case—Morgan        | Maximum sample size fixed.   | Sample size chosen to minimize either expected duration of accrual or expected total study length for the trial. |

### 9.3.3 SINGLE-STAGE DESIGNS

In planning a phase II trial, there are several options to consider, including the number of stages within each design, alternative designs for essentially the same situation, as well as randomized designs. In this section, we look at the Fleming–A'Hern single-stage design. For this design, the endpoint is some measure of anti-disease activity and this translates into a measure of response.

A key advantage of such single-stage designs is that once the sample size has been determined, this translates directly into the number of patients that need to be recruited. This contrasts with, for example, two-stage designs in which the total numbers eventually recruited depends on the response rate in those recruited to stage 1.

In considering the design of a phase II trial of a new drug, the investigators will usually have some knowledge of the activity of other drugs for the same disease. The anticipated response to the new drug is therefore compared, at the planning stage, with the observed responses to other therapies. This may lead to the investigators prespecifying a response probability that, if the new drug does not achieve, results in no further investigation. They might also have some idea of a response probability that, if achieved or exceeded, would certainly imply that the new drug has activity worthy of further investigation, perhaps in a phase III randomized trial to determine efficacy. If a phase II trial either fails to identify efficacy or overestimates the potential efficacy, there will be adverse consequences for the next stage of the development process.

#### 9.3.3.1 Fleming–A'Hern

The Fleming [5] single-stage design for phase II trials recruits a predetermined number of patients to the study, and a decision about activity is obtained from the number of responses among these patients.

To use the design, the investigators first set the largest response proportion as  $\pi_0$ , which, if true, would clearly imply that the treatment does not warrant further investigation. They then judge what is the smallest response proportion,  $\pi_{\text{New}}$ , which would imply that the treatment certainly warrants further investigation. This means that the one-sided hypotheses to be tested in a phase II study are:  $H_0: \pi \leq \pi_0$  versus  $H_A: \pi \geq \pi_{\text{New}}$ , where  $\pi$  is the actual probability of response that is to be estimated at the close of the trial.

In addition to specifying  $\pi_0$  and  $\pi_{\text{New}}$ , it is necessary to specify  $\alpha$ , the probability of rejecting the hypothesis  $H_0: \pi \leq \pi_0$  when it is in fact true, together with  $\beta$ , the probability of rejecting the hypothesis  $H_A: \pi \geq \pi_{\text{New}}$  when that is true. Note that  $\alpha$  is often termed the *test size* or *significance level* and  $1 - \beta$  is referred to as the *power*.

With these inputs, Fleming [5] details the appropriate sample size to be recruited, along with the minimum number of responses that would need to be observed in order for the null hypothesis to be rejected.

However, A'Hern [6] repeated the calculations for sample size using exact binomial probabilities and, in general, these are greater than those of Fleming, although the two calculations are often in close agreement. As a consequence, his calculations should supersede those of Fleming.

To implement the design, the appropriate number of patients is recruited, and once all their responses are observed, the response rate (and corresponding confidence interval) is calculated. A decision with respect to efficacy is then made.

**Example: Fleming–A'Hern Design—Sequential Hormonal Therapy in Advanced and Metastatic Breast Cancer** Iaffaioli et al. [7] used A'Hern's design for two phase II studies of sequential hormonal therapy with first-line anastrozole (study 1) and second-line exemestane (study 2) in advanced and metastatic breast cancer. For study 1 they set  $\alpha = 0.05$ ,  $\beta = 0.1$ ,  $\pi_0 = 0.5$ , and  $\pi_{\text{New}} = 0.65$ . With these inputs, the design specifies a sample size of 93, with 55 being the minimum number of responses required for a conclusion of "efficacy." The study in the end recruited 100 patients, with 8 complete responses and 19 partial responses observed. These give an estimated response rate of 27% with 95% confidence interval (CI) 19.3–36.4% calculated using the method described by Newcombe and Altman [8]. This is much lower than the desired minimum of  $\pi_{\text{New}} = 65\%$ . For study 2, the investigators set  $\alpha = 0.05$ ,  $\beta = 0.1$ ,  $\pi_0 = 0.2$ , and  $\pi_{\text{New}} = 0.4$ , giving rise to a sample size of 47 with a minimum of 15 responses required. The trial eventually recruited 50 patients, with 1 complete response and 3 partial responses observed. These give an estimated response rate of 8% (95% CI 3.2–18.8%). Again this is much lower than the desired minimum of 40%. As a consequence, neither drug should be recommended for testing in a phase III trial.

### 9.3.4 TWO-STAGE DESIGNS

In many situations, investigators may be reluctant to embark on a single-stage phase II trial requiring a (relatively) large number of patients exposed to a new and uncertain therapy. In such circumstances, a more cautious approach may be to conduct such a study, but in a series of stages and review progress at the end of each stage. In two-stage designs, patients are recruited in two stages, and the move to stage 2 is consequential on the results observed in stage 1. The main advantage of such a design is that the trial may stop, after relatively few patients have been recruited, should the response rate appear to be (unacceptably) low. The disadvantage is that the final number of patients required is not known until after stage 1 is complete.

#### 9.3.4.1 Gehan

In the approach suggested by Gehan [9], a minimum requirement of efficacy,  $\pi_{\text{New}}$ , is set and patients are recruited in two stages. If no responses are observed in stage 1, patients are not recruited for stage 2. On the other hand, if one or more responses are observed, then the size of the recruitment to the second stage depends on their number.

To implement the design, the appropriate number of patients is recruited in stage 1, and, once all their responses are observed, a decision whether or not to proceed to stage 2 is taken. If stage 2 is implemented, then once recruitment is complete and all assessments made, the response rate (and corresponding CI) is calculated. A decision with respect to efficacy is then made. If stage 2 is not activated, the response

rate (and CI) can still be calculated for the stage 1 patients despite failure to demonstrate efficacy. This procedure applies to all the two-stage designs we will discuss.

**Example: Gehan Design—Dexverapamil and Epirubicin in Nonresponsive Breast Cancer** Lehnert et al. [10] used the Gehan design for a phase II trial of the combination dexverapamil and epirubicin in patients with breast cancer. For stage 1 they set  $\pi_0 = 0.2$  and  $\beta = 0.05$  obtaining a stage 1 sample size of 14. Of these 14 patients, 3 responses were observed, resulting in a further 9 patients needing to be recruited for stage 2. Finally a total of 4 (17.4%) responses was observed from the total of  $14 + 9 = 23$  patients with 95% CI for  $\pi$  from 7 to 37%.

### 9.3.4.2 Simon Optimal and Minimax

In the approach suggested by Simon [11], patients are recruited in two stages and there are two alternative designs.

One is *optimal* in that the expected sample size is minimized if the regimen has low activity. This implies that an important focus is to ensure that as few patients as possible receive what turns out to be an ineffective drug by not continuing to stage 2 in these circumstances. In this context, “expected” means, the average sample size that would turn out to have been used, had a whole series of studies been conducted with the same design parameters in situations where the true activity is the same.

The other, the *minimax* design, minimizes the maximum sample size for both stages combined, that is, the sum of patients required for stage 1 and stage 2, is chosen to minimize the maximum trial size within the parameter constraints as set by the design.

For either design, the designs imply that the one-sided hypotheses to be tested in a phase II study are  $H_0: \pi \leq \pi_0$  versus  $H_A: \pi \geq \pi_{\text{New}}$ , where  $\pi$  is the actual probability of response, and  $\pi_0, \pi_{\text{New}}$  are as defined before. It is also necessary to specify  $\alpha$  and  $\beta$  as for the Fleming–A’Hern design.

The trial then proceeds by recruiting  $n_{S1}$  patients in stage 1 from which  $r_{S1}$  responses are observed. Then a decision is made to recruit  $n_{S2}$  patients to stage 2 if  $r_{S1} > R_{S1}$ , where  $R_{S1}$  is the minimum number of responders required as indicated by the design. Otherwise the trial is closed at the end of stage 1. At the end of the second stage, the drug is rejected for further use if a predetermined total number of responses are not observed.

**Optimal versus Minimax** In determining which design to use, the minimax design may be more attractive than the optimal design when the difference in anticipated total sample size is small and the patient accrual rate is low. The optimal designs have smaller stage 1 than the minimax designs, and so this smaller stage 1 reduces the number of patients exposed to an inactive treatment if this turns out to be the case. In cases where the patient population is very heterogeneous, however, a very small stage 1 may not be desirable because the first patients entered into the study may not be entirely representative of the wider eligible population. In this case, a larger stage 1 may be preferred and the minimax design chosen.

**Example: Simon Minimax Design—Gemicitabine in Metastatic Nasopharyngeal Carcinoma** In a phase II trial of gemicitabine in previously untreated patients with metastatic nasopharyngeal carcinoma (NPC), Foo et al. [3] utilized the Simon minimax design. With  $\alpha = 0.05$ ,  $\beta = 0.2$ ,  $\pi_0 = 0.1$ , and  $\pi_{\text{New}} = 0.3$ , the design gives:

*Stage 1* Sample size of 15 patients: If responses less than 2, stop the trial and claim gemicitabine lacks efficacy.

*Stage 2* Overall sample size of 25 patients for both stages. Hence 10 more patients were to be recruited. If the total responses for the two stages combined is less than 6, stop the trial as soon as this is evident and claim gemicitabine lacks efficacy.

Once the phase II trial was conducted, the investigators observed 3 and 4 responses in stages 1 and 2, respectively, giving an estimated response rate of 7/25 or 28% (95% CI 14–48%).

**Example: Simon Optimal Design—Gemicitabine in Metastatic Nasopharyngeal Carcinoma** Suppose in the phase II trial of gemicitabine in metastatic NPC designed by Foo et al. [3] that the Simon optimal, rather than the minimax design, had been planned. With the same design values, we want to investigate what difference this makes to the patient numbers and responses required.

Again we have  $\alpha = 0.05$ ,  $\beta = 0.2$ ,  $\pi_0 = 0.1$ , and  $\pi_{\text{New}} = 0.3$ , but the use of the optimal design now gives the following results:

*Stage 1* Sample size of 10 patients: If responses less than 2, stop the trial and claim gemicitabine lacks efficacy.

*Stage 2* Overall sample size of 29 patients for both stages. Hence 19 more patients are to be recruited. If total responses for two stages combined is less than 6, stop the trial as soon as this is evident and claim gemicitabine lacks efficacy.

In this case, for the same design parameters, the optimal design has five fewer patients in stage 1 of the design, but four more patients if the trial goes on to complete stage 2, than the corresponding minimax design. The number of responses to be observed are the same, in each stage, for both designs, however.

**Example: Simon Minimax Design—Paclitaxel for Unresectable Hepatocellular Carcinoma** Chao et al. [12] state in their methods that a Simon [11] design was used in which if the response rate was  $\leq 3$  of 19 in the first stage, then the trial would be terminated. The authors set  $\alpha = 0.1$ ,  $\beta = 0.1$  but did not specify  $\pi_0$  or  $\pi_{\text{New}}$ . With a back calculation, it is possible to deduce that the minimax design was chosen with  $\pi_0 = 0.2$ ,  $\pi_{\text{New}} = 0.4$ , and a stage 1 sample size of 17. In this trial, 0 responses were observed in stage 1 and so stage 2 was not implemented. This implies that the response rate  $\pi$  is estimated by 0/17 or 0% with 95% CI of approximately 0–19%. Thus, even with an optimistic view of the true response rate as possibly close to 19%, this is far below the expectations of the investigators who set  $\pi_{\text{New}}$  as 40%.



### 9.3.5 PHASE II TRIALS WITH SURVIVAL ENDPOINTS

Although many phase II trials have disease response as a (binary) outcome, survival times or at least survival proportions at a fixed time are sometimes more relevant.

As already mentioned, a disadvantage of two-stage designs is that the final number of patients required is not known until after recruitment to stage 1 is complete and response in all these patients has been assessed. This poses a particular difficulty if the endpoint of concern is the time from initiation of treatment to some event (perhaps the death of the patient), which is expressed through the corresponding survival time. In this case there will be a variable, and possibly extended, period of observation necessary to make the requisite observations. However, estimates of survival at a prechosen fixed point in interval time (say 1-year poststart of treatment) can be estimated using the Kaplan–Meier (KM) technique, which takes into account censored observations. Censored survival time observations arise when a patient, although entered on the study and followed for a period of time, has not as yet experienced the “event” defined as the outcome for the trial. For survival itself “death” will be the event of concern, whereas if event-free survival was of concern, the event may be recurrence of the disease. Appropriate methods for survival time analysis are described by Machin et al. [13].

In this context, when considering the design of a phase II trial of a new drug, the investigators will usually have some knowledge of the activity of other drugs for the same disease. The anticipated survival rate of the new drug is therefore compared, at the planning stage, with that observed with other therapies. This may lead to the investigators prespecifying a survival probability that, if the new drug does not achieve, results in no further investigation. They might also have some idea of a survival probability that, if achieved or exceeded, would certainly imply that the new drug has activity worthy of further investigation, perhaps in a phase III randomized trial to determine efficacy.

#### 9.3.5.1 Case and Morgan

In the Case and Morgan [14] two-stage phase II trial designs, “survival” times are utilized in place of binary response variables. The “survival” times usually correspond to the interval between the registration of the patient into the study or the commencement of the phase II treatment, and the time at which the event of primary concern occurs, for example, recurrence of the disease, death, or either of these.

When considering the Case–Morgan designs, it is important to distinguish between *chronological* time—that is, the date on which the trial recruits its first patient, the date of the planned interim analysis, the date the trial closes recruitment, or the date all patient follow-up ends—from the time *interval* between start of therapy and the occurrence of the event. Trial conduct is concerned with chronological time while trial analysis is concerned with interval time. We denote the former by  $D$  and the latter by  $t$ .

The KM estimate at any follow-up time  $t$ , is denoted  $S(t)$ . Thus, for example, when  $t = 1$  year, the KM estimate at that time point is denoted  $S(1)$ . In general, a convenient time point, which we denote by  $T_{\text{Summary}}$ , is chosen by the investigators and the

corresponding  $S(T_{\text{Summary}})$  estimates the proportion of patients experiencing the event at that time point.

Typically, observing the event takes longer and is more variable in its time of occurrence than, for example, tumor response rate. This implies that any two-stage phase II design using such an endpoint may require a period between stage 1 and (the potential) stage 2. This time window is to allow sufficient events to accumulate for the stage 1 analysis so that a decision can be taken whether or not to continue to stage 2. The time window is added to the duration of stage 1, and its necessity may require suspending patient recruitment during this interval. Clearly, this will extend the total duration of the study. The Case–Morgan designs eliminate the need for this time window.

To implement the design, the investigators set for a particular interval time,  $t = T_{\text{Summary}}$ , the largest survival proportion as  $S_0(T_{\text{Summary}})$ , which, if true, would clearly imply that the treatment does not warrant further investigation. The investigators then judge what is the smallest survival proportion,  $S_{\text{New}}(T_{\text{Summary}})$ , that would imply the treatment warrants further investigation. This implies that the one-sided hypotheses to be tested in the study are:  $H_0: S(T_{\text{Summary}}) \leq S_0(T_{\text{Summary}})$  versus  $H_{\text{New}}: S(T_{\text{Summary}}) \geq S_{\text{New}}(T_{\text{Summary}})$ , where  $S(T_{\text{Summary}})$  is the actual probability of survival, which is to be estimated at the close of the trial. In addition to specifying  $S_0(T_{\text{Summary}})$  and  $S_{\text{New}}(T_{\text{Summary}})$ , it is necessary to specify  $\alpha$  and  $\beta$ .

With these inputs, there are then two variants of the Case–Morgan design, depending on whether we wish to minimize the expected duration of accrual (EDA) or the expected total study length (ETSL) for the trial. These are defined as follows:

$$\begin{aligned} \text{EDA} &= D_{\text{Stage1}} + (1 - P_{\text{Early}})D_{\text{Stage2}} \\ \text{ETSL} &= D_{\text{Stage1}} + (1 - P_{\text{Early}})(D_{\text{Stage2}} + T_{\text{Summary}}) \end{aligned}$$

where  $D_{\text{Stage1}}$  and  $D_{\text{Stage2}}$  are the durations of stage 1 and stage 2 of the trial, respectively, and  $P_{\text{Early}}$  is the probability of stopping at the end of stage 1.

**Example: Case and Morgan—Gemcitabine and External Beam Radiotherapy for Resectable Pancreatic Cancer** Case and Morgan [14] consider the design of a phase II trial of the effectiveness of adjuvant gemcitabine and radiotherapy in the treatment of patients with resectable pancreatic cancer. The outcome measure used was 1-year survival, and they planned to test the null hypothesis that 1-year survival is 35% against an alternative of 50%. Thus,  $T_{\text{Summary}} = 1$ ,  $S_0(1) = 0.35$ ,  $S_{\text{New}}(1) = 0.50$ . Further  $\beta = 0.1$  and  $\alpha = 0.1$ .

With these inputs, the ETSL design suggests that stage 1 recruits 54 patients and stage 2 83 patients. With the EDA design the corresponding sample sizes are 46 in stage 1 and 79 in stage 2, resulting in a total sample size of 125.

### 9.3.6 EFFICACY AND TOXICITY IN PHASE II TRIALS

In situations where the toxicity of an agent undergoing phase II testing is poorly understood, it may be desirable to incorporate toxicity considerations into the trial

design. We now discuss phase II trial designs for the situation in which both a minimum level of activity and a maximum level of (undesirable) toxicity are stipulated in the design. Such designs expand on the Simon two-stage designs discussed earlier.

### 9.3.6.1 Bryant and Day

Bryant and Day [15] point out that a common situation when considering phase I and phase II trials is that although the former primarily focuses on toxicity and the latter on efficacy, each in fact considers both. This provides the rationale for their phase II design that incorporates toxicity and activity considerations. Essentially, they combine a design for activity with a similar design for toxicity in which one is looking for both *acceptable* toxicity and *high* activity.

The design implies that two, *one-sided* hypotheses are to be tested. These are that the true response rate  $\pi_R$  is either  $\leq \pi_{R0}$ , the maximum response rate of no interest, or  $\geq \pi_{R\text{New}}$ , the minimum response rate of interest. Further the probability of incorrectly rejecting the hypothesis  $\pi_R \leq \pi_{R0}$  is set as  $\alpha_R$ . Similarly,  $\alpha_T$  is set for the hypothesis  $\pi_T \leq \pi_{T0}$  where  $\pi_T$  is the maximum nontoxicity rate of no interest. In addition, the hypothesis  $\pi_T \geq \pi_{T\text{New}}$  has to be set together with  $\beta$ , the probability of failing to recommend a treatment that is acceptable with respect to both activity and (non-)toxicity. (The terminology is a little clumsy here as it is more natural to talk in terms of “acceptable toxicity” rates rather than “acceptable nontoxicity” rates. Thus  $1 - \pi_{T0}$  is the highest rate of toxicity above which the drug is unacceptable. In contrast,  $1 - \pi_{T\text{New}}$  is the lower toxicity level below which the drug would be regarded as acceptable on this basis.)

In the Bryant and Day design, toxicity monitoring is incorporated into the Simon [11] design by requiring that the trial is terminated after stage 1 if there is an inadequate number of observed responses or an excessive number of observed toxicities. The treatment under investigation is recommended at the end of stage 2 only if there are both a sufficient number of responses and an acceptably small number of toxicities in total.

To implement the designs, the appropriate number of patients is recruited to stage 1, and once all their responses and toxicity experiences are observed, a decision whether or not to proceed to stage 2 is taken. If stage 2 is implemented, then once recruitment is complete and all assessments made, the response and toxicity rates, along with their corresponding CIs, are calculated. A decision with respect to efficacy and toxicity is then made. If stage 2 is not activated, the response rate and toxicity rates can still be calculated for the stage 1 patients despite either failure to demonstrate activity, too much toxicity, or both.

**Example: Bryant and Day Design—Ifosfamide and Vinorelbine in Ovarian Cancer** González-Martín et al. [16] used the Bryant and Day two-stage design with a cutoff point for the response rate of 10% and for severe toxicity of 25%. Severe toxicity was defined as grade 3 and 4 nonhematological toxicity, neutropenic fever, or grade 4 thrombocytopenia. They do not provide full details of how the sample size was determined, but their choice of design specified a stage 1 of 14 patients and stage 2 a further 20 patients. In the event, in these advanced platinum-resistant ovarian cancer patients, the combination of ifosfamide and vinorelbine was

evidently very toxic. Hence the trial was closed after 12 patients with an observed toxicity level above the 25% contemplated. In fact, this corresponds to a design with  $\alpha_R = \alpha_T = 0.1$ ,  $\beta = 0.2$ ;  $\pi_R^0 = 0.1$ ,  $\pi_{R\text{New}} = 0.3$ ;  $\pi_{T0} = 0.25$ ; and  $\pi_{T\text{New}} = 0.45$ . On this basis, the completed stage 1 trial of 14 patients proceeds to stage 2 if there are at least 2 responses *and* there are also no more than 2 patients with high toxicity. The stage 2 trial size is a further 20 patients, to a total of 34 for the whole trial, and sufficient efficacy with acceptable toxicity would be concluded if there were 6 or more responses observed *and* 10 or fewer with high toxicity.

### 9.3.7 BAYESIAN APPROACHES

#### 9.3.7.1 Motivation

For most of the designs discussed thus far, the final response rate is estimated by  $R/N$ , where  $R$  is the total number of responses observed from the total number of patients recruited  $N$  (whether obtained from a single- or two-stage design). This response rate, together with the corresponding 95% CI, typically provide the basic information for the investigators to decide if a subsequent phase III trial is warranted. However, even after the trial is completed, there often remains considerable uncertainty about the true value of  $\pi$ . For example, in the trial reported by Lehnert et al. [10] using the Gehan design, a 17% response rate was observed from 23 patients with the corresponding 95% CI for  $\pi$  from 7 to 37%. In the trial of previously treated patients with metastatic nasopharyngeal cancer, conducted by Foo et al. [3], a high response rate of 48% (95% CI of 33 to 63%) was reported. This result is consistent with both a true response rate as small as 33% and one as high as 63%, an almost twofold difference. The inevitable uncertainty arising from phase II trials with small sample sizes suggests that Bayesian approaches may be useful for phase II trials.

#### 9.3.7.2 Overview of Bayesian Approaches in the Context of Phase II Trials

The foundation of the Bayesian approach is Bayes' theorem, which can be expressed as

$$\text{post}(\pi|x) \propto \text{lik}(x|\pi)\text{prior}(\pi)$$

which involves combining the likelihood  $\text{lik}(x|\pi)$  with the prior distribution,  $\text{prior}(\pi)$ , to give the posterior distribution,  $\text{post}(\pi|x)$ . The  $\text{prior}(\pi)$  summarizes what we know about  $\pi$  before the trial commences, while  $\text{lik}(x|\pi)$  describes the data to be collected from the trial itself. Finally  $\text{post}(\pi|x)$  summarizes all we know about  $\theta$  once the trial is completed.

Many phase II trials involve endpoints that are binary (e.g., whether a response occurs or not). For such endpoints, the prior distribution may be assumed to be of the form

$$\text{prior}(\pi) \propto \pi^{a-1}(1-\pi)^{b-1}$$

This is a Beta distribution with parameters  $a$  and  $b$ , which can take any positive value. When  $a$  and  $b$  are integers, such a distribution corresponds to a prior belief equivalent to having observed  $a$  responses out of a hypothetical  $T = (a + b)$  patients. This is then similar to the situation modeled by the binomial likelihood distribution in which we have  $x$  as the number of responses from  $N$  patients.

Combining the above prior with a binomial likelihood results in a posterior distribution of the form

$$\text{post}(\pi|x) \propto \pi^{a+x-1}(1-\pi)^{b+N-x-1}$$

It can be seen that this too is a Beta distribution, but of the form Beta  $(a + x, b + N - x)$ .

As mentioned previously, the posterior distribution represents our overall belief at the close of the trial about the distribution of the population parameter  $\pi$ . Once we have obtained the posterior distribution, we can calculate the exact probabilities of  $\pi$  being in any region of interest or obtain summary statistics such as its mean value.

The prior distribution summarizes the information on  $\pi$  before the trial commences. The general way in which each of these are derived is as follows. The shape of a Beta distribution is dependent on the values of the parameters  $a$  and  $b$ , and each of the priors will have particular values associated with them. However, eliciting values for  $a$  and  $b$  is typically not an easy process. Instead, it is often much easier to obtain values for the mean ( $M$ ) and variance ( $V$ ) of the corresponding prior distribution. Once obtained, these values can then be used to obtain  $a$  and  $b$  by solving the simultaneous equations:

$$M = \frac{a}{a+b} \quad V = \frac{ab}{(a+b)^2(a+b+1)}$$

which give

$$a = \frac{M[M(1-M)-V]}{V} \quad b = \frac{(1-M)[M(1-M)-V]}{V}$$

More generally, prior distributions could be elicited either from relevant external data (see e.g., Tan et al. [17]) or from subjective clinical opinion [18] or a combination of the two.

For more detailed overviews of Bayesian approaches, the reader is referred to Berry and Stangl [19], Spiegelhalter et al. [20], and Tan [21]. In the particular situation of two phase II trials conducted “in parallel” using a two-stage design, Bayesian approaches allow for the information from both trials to be taken into account when making decisions regarding whether to proceed to stage 2 of each trial or not [22].

### 9.3.7.3 Bayesian Single and Dual Threshold

In the Tan–Machin (TM) two-stage single-threshold design (STD) [23, 24], the focus is to estimate, for example, the posterior probability that  $\pi > \pi_{\text{New}}$ , so that if this is

high, at the end of the phase II trial, the investigators can be reasonably confident in recommending the compound for testing in a phase III trial.

The investigator first sets the minimum interest response rate  $\pi_{\text{New}}$  and  $\pi_{\text{Prior}}$  the anticipated response rate of the drug being tested. However, in place of  $\alpha$  and  $\beta$ ,  $\lambda_1$  (the required threshold probability following stage 1 that  $\pi > \pi_{\text{New}}$ ) and  $\lambda_2$  ( $> \lambda_1$ ) (the required threshold probability after completion of stage 2 that  $\pi > \pi_{\text{New}}$ ) are specified. Further, once the first stage of the trial is completed, the estimated value of  $\lambda_1$ , that is,  $u_1$ , is computed and a decision made whether or not to proceed to stage 2. Should the trial continue to stage 2 then, on trial completion,  $u_2$  is computed. Note that the trial only goes into stage 2 if the estimate of  $\lambda_1$ , at the end of stage 1, exceeds the design value. Efficacy is claimed at the end of stage 2 only if the estimate of  $\lambda_2$ , obtained from all the data, exceeds the design value.

The design determines the sample sizes for the trial based on the following principle. Suppose that the trial was to be conducted and that  $X_1$  and  $X_2$  represent the resulting data obtained from stage 1 and stage 2, respectively. Now, suppose also that the (hypothetical) response proportion underlying  $X_1$  and  $X_2$  is just larger than the prespecified  $\pi_{\text{New}}$ , say  $\pi_{\text{New}} + \varepsilon$ , for some small  $\varepsilon > 0$ . We then want the smallest overall sample size,  $N_{\text{TM}}$ , that will enable the posterior probability at the end of the trial, denoted  $\Pr(\pi > \pi_{\text{New}} \mid X_1, X_2)$  or more briefly  $\Pr(\pi > \pi_{\text{New}})$ , to be at least  $\lambda_2$ . At the same time, we also want the smallest possible stage 1 sample size  $n_{\text{TM1}}$ , which is just large enough so that the posterior probability at the end of stage 1,  $\Pr(\pi > \pi_{\text{New}} \mid X_1)$  or more briefly  $\Pr(\pi > \pi_{\text{New}})$ , is at least  $\lambda_1$ . Tan and Machin [23] suggest planning values for  $(\lambda_1, \lambda_2)$  as (0.6, 0.7), (0.6, 0.8), or (0.7, 0.8) and also set a value of  $\varepsilon = 0.05$ .

Tan and Machin [23] also propose an alternative two-stage dual-threshold design (DTD). This design is identical to the STD except that the stage 1 sample size is determined not on the basis of the probability of exceeding  $\pi_{\text{New}}$  but on the probability that  $\pi$  will be less than the “no further interest” proportion,  $\pi_0$ . This represents the response rate below which the investigator would have no further interest in the new drug. Thus  $\pi_0$  functions as a lower threshold on the response rate, as opposed to the upper threshold represented by  $\pi_{\text{New}}$ . The rationale behind this aspect of the DTD is that we want our stage 1 sample size to be large enough so that, if the trial data really does suggest a response rate that is below  $\pi_0$ , we want the posterior probability of  $\pi$  being below  $\pi_0$ , to be at least  $\lambda_1$ . The design determines the smallest stage 1 sample size that satisfies this criterion. The trial only goes into stage 2 if the estimate of  $\lambda_1$  exceeds the design value and efficacy is claimed at the end of stage 2 only if the estimate of  $\lambda_2$  exceeds the design value.

The DTD requires the investigators to set  $\pi_{\text{Prior}}$  as the anticipated value of  $\pi$  for the drug being tested. A convenient choice may be  $(\pi_0 + \pi_{\text{New}})/2$ , but this is not a requirement. Further  $\lambda_1$  is set as the required threshold probability following stage 1, that  $\pi < \pi_0$ , while  $\lambda_2$  is the required threshold probability that, after completion of stage 2,  $\pi > \pi_{\text{New}}$ . (Note that unlike in the case of STD, it is no longer a requirement that  $\lambda_1 < \lambda_2$ .) Once stage 1 of the trial is completed, the estimated value of  $\lambda_1$ , that is  $l_1$ , is computed, and, should the trial continue to stage 2 then on its completion,  $u_2$  is computed. The latter is then used to help make the decision as to whether or not a phase III trial is suggested. As with the STD, Tan and Machin [23] suggest planning values for  $(\lambda_1, \lambda_2)$  as (0.6, 0.7), (0.6, 0.8), or (0.7, 0.8) and also set a value of  $\varepsilon = 0.05$ .

The original Tan–Machin [23] designs work on the basis of having a “vague” prior distribution. According to Mayo and Gajewski [25], this corresponded to having a prior sample size of 3. Furthermore, Tan and Machin imposed some practical constraints on the designs to encourage their adoption in practice. In particular, they constrained the total study size,  $N_{TM}$ , to be a minimum of 10 and a maximum of 90, with stage 1 size,  $n_{TM1}$ , having a minimum size of 5 and a maximum of  $N_{TM} - 5$ . For these and other reasons, Mayo and Gajewski [25] as well as Wang et al. [26] have suggested modifications to the original Tan–Machin design.

**Example: Tan–Machin STD Design—Gemcitabine in Metastatic Nasopharyngeal Cancer** Tan and Machin [23] reanalyzed the phase II trial of Foo et al. [3] for previously treated patients as if they had been designed using STD. First, they back-calculated from the two-stage Simon minimax design utilized, that this choice implied for their STD values of  $\lambda_1 = 0.728$  and  $\lambda_2 = 0.774$ , respectively. Using the actual trial data they then computed with the data at the close of stage 1,  $\Pr(\pi > \pi_{New}) = u_1 = 0.997$  (which is clearly greater than  $\lambda_1 = 0.728$ ). Further for the data at the close of stage 2, this probability was reestimated to be  $\Pr(\pi > \pi_{New}) = u_2 = 0.999$  (which is clearly greater than  $\lambda_2 = 0.774$ ). So *had* the STD been used, this reanalysis suggests that, at the end of stage 1, continuation to stage 2 would have been appropriate. Further information at the end of stage 2 very strongly recommended that gemcitabine can be considered to have sufficient activity for phase III evaluation.

**Example: Tan–Machin DTD Design—Gemcitabine in Metastatic Nasopharyngeal Cancer** Similarly, reanalyzing the chemonaïve study of Foo et al. [3], but now on the basis that a DTD was conducted, all the actual trial data gives an estimate of  $\Pr(\pi < \pi_0) = l_1 = 0.003$  or equivalently expressed by  $\Pr(\pi > \pi_0) = 1 - l_1 = 0.997$ . The estimate for  $\Pr(\pi > \pi_{New}) = u_2 = 0.445$ . Thus, we are very confident that  $\pi > \pi_0$  but not so sure that  $\pi > \pi_{New}$ . These together suggest that the response rate lies within the region of uncertainty,  $\pi_0 \leq \pi \leq \pi_{New}$ , as this has a reasonably high probability of  $0.997 - 0.445 = 0.552$  or 55%.

**Example: Tan–Machin DTD—Combination Therapy for Nasopharyngeal Cancer** A phase II trial using a triplet combination of paclitaxel, carboplatin, and gemcitabine in metastatic nasopharyngeal carcinoma was conducted at the National Cancer Centre, Singapore, by Leong et al. [27]. The trial was expected to yield a minimum interest response rate of 80% and a no further interest response of 60%. The anticipated response rate was assumed to be equal to the minimum interest response rate and the overall threshold probability at the start and end of the trial is set to be 0.65 and 0.7, respectively. The sample size of the trial was calculated using the DTD. With the no further interest response rate  $\pi_0 = 0.6$ , the minimum interest response rate  $\pi_{New} = 0.8$ , the anticipated response rate  $\pi_{Prior} = 0.8$ , the minimum desired threshold probability at the start of the trial  $\lambda_1 = 0.65$ , and the minimum desired threshold probability at the end of the trial  $\lambda_2 = 0.7$ , we obtain the following design:

*Stage 1* Sample size of 19 patients; if responses less than 15, stop the trial as soon as this becomes apparent and declare lack of efficacy. Otherwise complete stage 1 and commence stage 2.



*Stage 2* Overall sample size of 32 patients for both stage, hence 13 stage 2 patients to be recruited; if total responses for the two stages combined is less than 28, stop trial as soon as this becomes apparent and declare lack of efficacy.

### 9.3.8 RANDOMIZED PHASE II TRIALS

Most phase II trials are of a single-arm, noncomparative design. However, in certain circumstances there may be several compounds available for potential phase III testing in the same type of patients, but practicalities imply that only one of these can go forward for this subsequent assessment. Since there are several options, good practice dictates that the eligible patients should be randomized to the alternatives. This can be achieved by using a randomized phase II selection design in which the objective is to select only one, the “best,” of several agents tested simultaneously.

The randomized designs overcome the difficulties pointed out by Estey and Thall [28] when discussing single-arm trials where the actual differences between response rates associated with the treatments (*treatment effects*) are confounded, with differences between the trials (*trial effects*), as there is no randomization to treatment. Consequently, an apparent treatment effect may in reality only be a trial effect.

#### 9.3.8.1 Simon, Wittes, and Ellenberg (SWE)

The Simon, Wittes, and Ellenberg [29] design is a randomized (single-stage) phase II design that selects from several candidate drugs that with the highest level of activity. This approach chooses the observed best treatment for the phase III trial, however small the advantage over the others. The trial size is determined in such a way that if a treatment exists for which the underlying efficacy is superior to the others by a specified amount, then it will be selected with a high probability. Although details of the random allocation process are not outlined below, this is a vital part of the design implementation. Details are provided by, for example, Machin and Campbell [30].

When the difference in true response rates of the best and next best treatment is  $\delta$ , Simon et al. [29] allow for the computation of sample sizes depending on the desired probability of correct selection,  $P_{CS}$ , and the number of treatments being tested,  $g$ . The response rate of the worst treatment is denoted  $\pi_{\text{Worst}}$ .

To implement the design, the appropriate number of patients are recruited and randomized to the  $g$  groups. Once all their responses are observed, the response rates are calculated for each drug under test and that with the highest recommended for phase III testing.

**Example: Gemcitabine, Vinorelbine, or Docetaxel for Advanced Non-Small-Cell Lung Cancer** A randomized phase II trial of single-agent gemcitabine, vinorelbine, or docetaxel in the treatment of elderly and/or poor performance status patients with advanced non-small-cell lung cancer was conducted at the National Cancer Centre of Singapore [31]. The design was implemented with the probability of correctly selecting the best treatment assumed to be 90%. It was anticipated that the single-agent activity of each drug has a baseline response rate of approximately

20%. In order to detect a 15% superiority of the best treatment over the others, we wished to determine how many patients should be recruited per treatment for the trial. For the difference in response rate  $\delta = 0.15$ , smallest response rate  $\pi_{\text{Worst}} = 0.2$ , probability of correct selection  $P_{\text{CS}} = 0.90$ , and treatment groups,  $g = 3$ , the design gives a sample size of  $m = 44$  per treatment group. Thus the total number of patients to be recruited is given as  $N = 3 \times 44 = 132$ .

**Example: Non-Hodgkin's Lymphoma** Itoh et al. [32] describe a randomized two-group phase II trial comparing dose-escalated (DE) with biweekly dose-intensified (DI) CHOP in newly diagnosed patients with advanced-stage aggressive non-Hodgkin's lymphoma. Their design anticipated at least a 65% complete response (CR) rate in both groups. To achieve a 90% probability of selecting the better arm when the CR rate is 15% higher in one arm than the other, at least 30 patients would be required in each arm. In that event, they recruited 35 patients to each arm and observed response rates with DE and DI of 51 and 60%, respectively. Their follow-on study, a randomized phase III trial, compares DI CHOP with the standard CHOP regimen.

### 9.3.9 TRIAL CONDUCT AND REPORTING

In phase III trials, much emphasis has been placed on developing standards for the good conduct and reporting of clinical trials. Among the aspects looked at are issues relating to informed consent, registration of subjects, monitoring the trial, and common standards for the reporting of trials. It is well known that many trials go unreported, leading to publication bias being an important concern. This motivates proposals that all phase III trials need to have their protocols formally registered before a trial can even begin [33].

Unfortunately, the conduct and reporting of phase II trials often does not meet the high standards demanded of the phase III randomized controlled trial. All phases of the clinical trials process crucially affect the final conclusions made regarding the usefulness of new treatments. As such, phase II trials also need to be conducted and reported to the highest standards, and we advocate that the standards applied to phase III trials should also be extended to phase II wherever appropriate. Furthermore, there are some particular considerations that need to be taken into account for phase II trials, which we now discuss.

#### 9.3.9.1 Trial Conduct

Since most phase II trials are not comparative, in the taking of informed consent, only details of the procedures that are to be involved and any potential side effects and risks need to be explained. It would be important to explain to the patient that any therapeutic benefit hoped for, such as tumor shrinkage, may or may not transfer into benefit for the patient with respect to (say) increased survival or improved quality of life. In the case of randomized phase II trials of Simon et al. [29], then the usual considerations applicable to those for a phase III trial would apply.

Once consent has been taken and patients recruited into the trial, there is then a need to properly register and monitor them as for phase III trials. Moreover, unlike for phase III trials, continuous monitoring of patient responses may occur. This has particular implications should a two-stage design have been used. Typically, such designs would result in a delay in the recruitment process between stage 1 and stage 2, resulting in a longer trial. However with continuous monitoring of responses, stage 2 may be triggered before the formal recruitment to stage 1 is complete if there are already sufficient responses. Nevertheless, in such a situation, we recommend that a formal review of the stage 1 results should still be carried out.

### 9.3.9.2 What to Report

Considerable effort is required in order to conduct a clinical trial of whatever type and size, and this effort justifies reporting of the subsequent trial results with careful detail. However, there is a wide variation in the quality of the standard of reporting of clinical trials. In phase III trials, major strides in improving the quality have been made and pivotal to this has been the Consolidation of the Standards of Reporting Trials (CONSORT) statement described by Begg et al. [34] and amplified by Moher et al. [35]. CONSORT describes the essential items that should be reported in a trial publication in order to give assurance that the trial has been conducted to a high standard. This is an internationally agreed recommendation, adopted by many of the leading medical journals, although there are still some who do not appear to insist that their authors comply with the requirements.

Although the CONSORT statement primarily applies to phase III and not phase II trials, many of their principles can and should also be applied to these. In particular, Table 2 highlights some of the relevant items from CONSORT that should be applied. A diagram showing the flow of patients through the trial (Fig. 1) should also be given.

**TABLE 2 Selected Key Items Included in Phase II Clinical Trial Report**

|                                                       |                                                                                                                                                                                                                       |
|-------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Participants                                          | Eligibility criteria for participants and the setting and locations where the data were collected.                                                                                                                    |
| Intervention                                          | Precise details of the intervention intended, and how and when they were actually administered.                                                                                                                       |
| Objectives                                            | Specific objectives and hypotheses.                                                                                                                                                                                   |
| Outcomes                                              | Clearly defined primary and secondary outcome measures                                                                                                                                                                |
| Sample size                                           | How sample size was determined                                                                                                                                                                                        |
| Randomization and blinding (randomized phase II only) | Details of method used to generate the random allocation sequence—including details of strata and block size.<br>Method used to implement the random allocation—numbered containers, central telephone, or Web-based. |
| Statistical methods                                   | Description of the extent of the blinding in the trial—investigator, participant.<br>Statistical methods used for the primary outcome(s).                                                                             |
| Participant flow                                      | Flow of participants through each stage of the trial (see Figure 1).                                                                                                                                                  |
| Recruitment                                           | Dates defining the periods of recruitment and follow-up.                                                                                                                                                              |
| Follow-up                                             | As many patients as possible to be followed up. Dropouts should be reported by treatment group.                                                                                                                       |

Source: Adapted from Moher and co-workers [35].

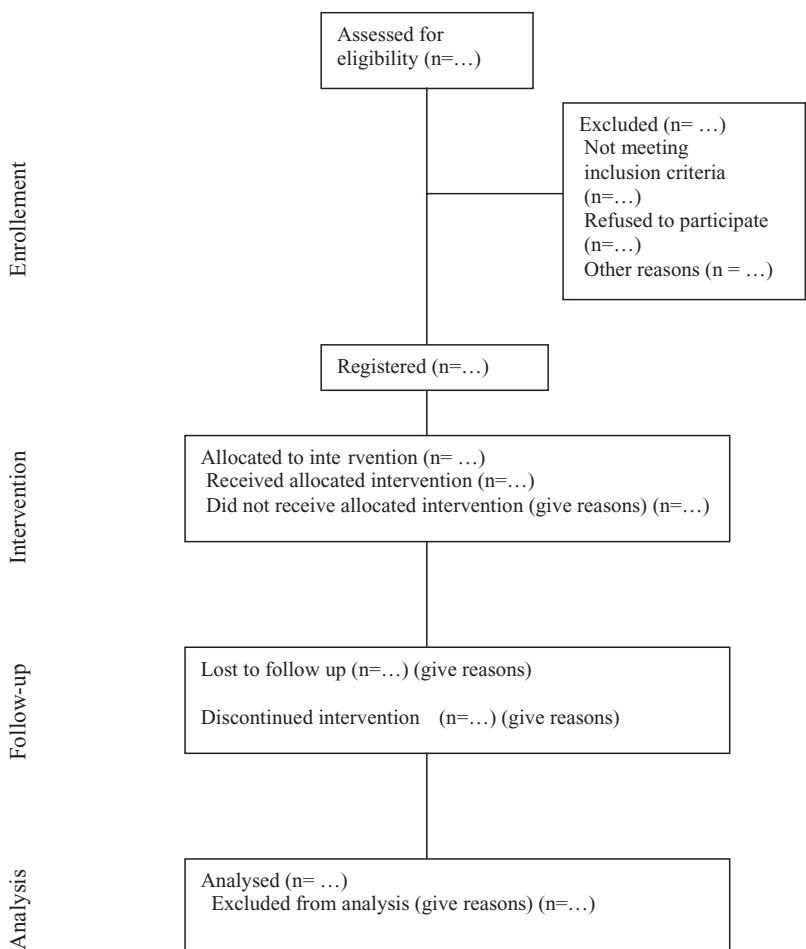


**FIGURE 1** Template of diagram showing flow of participants through a phase II trial.

**9.3.10 CONCLUDING REMARKS**

Although phase II studies are typically of modest size relative to phase III trials, the temptation to conduct these studies without due attention to detail should be resisted. In fact, these studies (imprecise though they may be) provide key information for the drug development process. It is therefore essential that they are carefully designed, painstakingly conducted, and meticulously reported in full. Table 3 summarizes the key design and conduct issues.

It is also important to again emphasize that all patients should be registered for the trial (and hence are in the trial database) and that the final report includes information on all these patients. This is particularly important if a review process of, for example, each objective response in a phase II trial may reveal that certain patients admitted to the trial were either not truly eligible, had not received the full treatment as specified by the protocol, or could not be evaluated for the endpoint. Perhaps it is unclear whether or not they had sufficient tumor shrinkage for a satis-

**TABLE 3 Design and Conduct Issues for Phase II Trials**


---

|                                                                                           |
|-------------------------------------------------------------------------------------------|
| Clearly define patient eligibility.                                                       |
| Clearly define the measures of response (and toxicity).                                   |
| Choose a single- or two-stage design.                                                     |
| Consider the importance of not proceeding to stage 2 if activity low.                     |
| Consider whether a CI or threshold probability approach is to be used for interpretation. |
| Consider the possibility of a randomized selection design.                                |
| Ensure that all patients are registered.                                                  |
| Ensure all evaluations are made.                                                          |
| Ensure the final report details information on all patients.                              |

---

factory response. It must be clear in the study protocol itself, and in the subsequent report of the study results, whether these “ineligible,” “noncompliant,” and “none-valuable” patients are or are not included in the reported response rates. This equally applies for any assessment of toxicity, whether or not toxicity is a formal endpoint for the design as it is in the Bryant–Day design of phase II.

It should also be emphasized that phase II trials should never be seen as an alternative to well-designed (large) randomized phase III trials. This is because the small sample sizes in phase II trials give rise to estimates with very wide confidence intervals (i.e., a high level of uncertainty). Hence any conclusions drawn from such trials cannot be confirmatory.

Finally, it should be noted that the results of separate single-arm phase II trials should also generally not be used for comparative purposes because of the potential confounding of treatment effects with trial effects. Consequently, an apparent treatment effect may in reality only be a trial effect. As for randomized phase II trials, their objective is to select the “best” of several agents tested for further testing in a phase III trial and not as alternatives to phase III trials.

## REFERENCES

1. van Rijswijk, R. E., Vermorken, J. B., Reed, N., Favalli, G., Mendiola, C., Zanaboni, F., Mangili, G., Vergote, I., Guastalla, J. P., ten Bokkel Huinink, W. W., Lacave, A. J., Bonnefoi, H., Tumulo, S., Rietbroek, R., Teodorovic, I., Coens, C., and Pecorelli, S. (2003), Cisplatin, doxorubicin and ifosfamide in carcinosarcoma of the female genital tract. A phase II study of the European Organization for Research and Treatment of Cancer Gynaecological Cancer Group (EORTC 55923), *Eur. J. Cancer*, 39, 481–487.
2. Therasse, P., Arbuck, S. G., Eisenhauer, E. A., Wanders, J., Kaplan, R. S., Rubinstein, L., Verweij, J., van Glabbeke, M., Van Oosterom, T., Christian, M. C., and Gwyther, S. G. (2000), New guidelines to evaluate the response to treatment in solid tumors, *J. Nat. Cancer Inst.*, 92, 205–216.
3. Foo, K.-F., Tan, E.-H., Leong, S.-S., Wee, J. T. S., Tan, T., Fong, K.-W., Koh, L., Tai, B.-C., Lian, L.-G., and Machin, D. (2002), Gemcitabine in metastatic nasopharyngeal carcinoma of the undifferentiated type, *Ann. Oncol.*, 13, 150–156.
4. Machin, D., Campbell, M. J., Tan, S. B., and Tan, S. H. (2009), *Sample Size Tables for Clinical Studies*, 3rd ed. Wiley-Blackwell, Chichester.
5. Fleming, T. R. (1982), One-sample multiple testing procedure for Phase II clinical trial, *Biometrics*, 38, 143–151.

6. A'Hern, R. P. (2001), Sample size tables for exact single stage Phase II designs, *Statist. Med.*, 20, 859–866.
7. Iaffaioli, R. V., Formato, R., Tortoriello, A., Del Prete, S., Caraglia, M., Pappagallo, G., Pisano, A., Fanelli, F., Ianniello, G., Cigolari, S., Pizza, C., Marano, O., Pezzella, G., Pedicini, T., Febbraro, A., Incoronato, P., Manzione, L., Ferrari, E., Marzano, N., Quattrin, S., Pisconti, S., Nasti, G., Giotta, G., Colucci, G., and Southern Italy Oncology Group (2005), Phase II study of sequential hormonal therapy with anastrozole/exemestane in advanced and metastatic breast cancer, *Br. J. Cancer*, 92, 1621–1625.
8. Newcombe, R. G., and Altman, D. G. (2000), Proportions and their differences, in Altman, D. G., Machin, D., Bryant, T. N., and Gardner, M. J., Eds., *Statistics with Confidence*, 2nd ed., British Medical Journal Books, London, pp. 45–56.
9. Gehan, E. A. (1961), The determination of the number of patients required in a preliminary and follow-up trial of a new chemotherapeutic agent, *J. Chronic Dis.*, 13, 346–353.
10. Lehnert, M., Mross, K., Schueller, J., Thuerlimann, B., Kroeger, N., and Kupper, H. (1998), Phase II trial of dexverapamil and epirubicin in patients with non-responsive metastatic breast cancer, *Br. J. Cancer*, 77, 1155–1163.
11. Simon, R. (1989), Optimal two-stage designs for phase II clinical trials, *Controlled Clin. Trials*, 10, 1–10.
12. Chao, Y., Chan, W.-K., Birkhofer, M. J., Hu, O. Y.-P., Wang, S.-S., Huang, Y.-S., Liu, M., Whang-Peng, J., Chi, K.-H., Lui, W.-Y., and Lee, S.-D. (1998), Phase, II., and pharmacokinetic study of paclitaxel therapy for unresectable hepatocellular carcinoma patients, *Br. J. Cancer*, 78, 34–39.
13. Machin, D., Cheung, Y.-B., and Parmar, M. K. B. (2006), *Survival Analysis: A Practical Approach*, 2nd ed., Wiley, Chichester.
14. Case, L. D., and Morgan, T. M. (2003), Design of Phase II cancer trials evaluating survival probabilities, *BMC Med. Res. Method.*, 3, 6.
15. Bryant, J., and Day, R. (1995), Incorporating toxicity considerations into the design of two-stage Phase II clinical trials, *Biometrics*, 51, 1372–1383.
16. González-Martín, A., Crespo, C., García-López, J. L., Pedraza, M., Garrido, P., Lastra, E., and Moyano, A. (2002), Ifosfamide and vinorelbine in advanced platinum-resistant ovarian cancer: Excessive toxicity with a potentially active regimen, *Gynecol. Oncol.*, 84, 368–373.
17. Tan, S. B., Dear, K. B. G., Bruzzi, P., and Machin, D. (2003), Strategy for randomized clinical trials in rare cancers, *BMJ*, 327, 47–49.
18. Tan, S. B., Chung, Y. F. A., Tai, B. C., Cheung, Y. B., and Machin, D. (2003), Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma, *Controlled Clin. Trials*, 24, 110–121.
19. Berry, D. A., and Stangl, D. K. (1996), *Bayesian Biostatistics*, Marcel Dekker, New York.
20. Spiegelhalter, D. J., Myles, J. P., Jones, D., and Abrams, K. R. (1999), An introduction to bayesian methods in health technology assessment, *BMJ*, 319, 508–512.
21. Tan, S. B. (2001), Introduction to Bayesian methods for medical research, (invited article), *Ann. Acad. Med. Singapore* 30, 444–446.
22. Tan, S. B., Machin, D., Tai, B. C., Foo, K. F., and Tan, E. H. (2002), A Bayesian re-assessment of two phase II trials of Gemcitabine in metastatic nasopharyngeal cancer, *Br. J. Cancer*, 86, 843–850.
23. Tan, S. B., and Machin, D. (2002), Bayesian two-stage designs for phase II clinical trials, *Stat. Med.*, 21, 1991–2012.

24. Tan, S. B., Wong, E. H., and Machin, D. (2004), Bayesian two-stage design for phase II clinical trials, in Chow, S. C. Ed., *Encyclopedia of Biopharmaceutical Statistics*, 2nd ed. (online), <http://www.dekker.com/servlet/product/DOI/101081EEBS120023507>, Marcel Dekker, New York.
25. Mayo, M. S., and Gajewski, B. J. (2004), Bayesian sample size calculations in phase II clinical trials using informative conjugate priors, *Controlled Clin. Trials*, 25, 157–167.
26. Wang, Y. G., Leung, D. H. Y., Li, M., and Tan, S. B. (2005), Bayesian designs with frequentist and Bayesian error rate considerations, *Statist. Methods Med. Res.*, 14, 445–456.
27. Leong, S. S., Tay, M. H., Toh, C. K., Tan, S. B., Thng, C. H., Foo, K. F., Wee J. T.S., Lim, D., See, H. T., Tan, T., Fong, K. W., and Tan, E. H. (2005), Paclitaxel, carboplatin and gemcitabine in metastatic nasopharyngeal carcinoma: A phase II trial using a triplet combination, *Cancer*, 103, 569–575.
28. Estey, E. H., and Thall, P. (2003), New designs for phase 2 clinical trials, *Blood*, 102, 442–448.
29. Simon, R., Wittes, R. E., and Ellenberg, S. S. (1985), Randomized phase II clinical trials. *Cancer Treat. Rep.*, 69, 1375–1381.
30. Machin, D., and Campbell, M. J. (2005), *Design of Studies for Medical Research*, Wiley, Chichester.
31. Leong, S. S. (2005), A randomized phase II trial of single agent gemcitabine, vinorelbine or docetaxel in patients with advanced non-small cell lung cancer who have poor performance and/or are elderly, *J. Thoracic Oncol.*, 2, 230–236. *Protocol SQLU01*.
32. Itoh, K., Ohtsu, T., Fukuda, H., Sasaki, Y., Ogura, M., Morishima, Y., Chou, T., Aikawa, K., Uike, N., Mizorogi, F., Ohno, T., Ikeda, S., Sai, T., Taniwaki, M., Kawano, F., Niimi, M., Hotta, T., Shimoyama, M., and Tobinai, K. (2002), Randomized phase II study of biweekly CHOP and dose-escalated CHOP with prophylactic use of lenograstim (glycosylated G-CSF) in aggressive non-Hodgkin's lymphoma: Japan Clinical Oncology Group Study 9505, *Ann. Oncol.*, 13, 1347–1355.
33. Dickerson, K., and Rennie, D. (2003), Registering clinical trials, *JAMA*, 290, 516–523.
34. Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schultz, K. F., Simel, D., and Stroup, D. F. (1996), Improving the quality of reporting randomized controlled trials: The CONSORT statement, *JAMA*, 276, 637–639.
35. Moher, D., Schultz, K. F., Altman, D. G., and the CONSORT Group (2001), The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials, *Lancet*, 357, 1191–1194.





# 9.4

---

## *Designing and Conducting Phase III Studies*

**Nabil Saba, John Kauh, and Dong M. Shin**

*Emory University School of Medicine, Winship Cancer Institute,  
Department of Hematology and Oncology, Atlanta, Georgia*

### **Contents**

|          |                                                                             |     |
|----------|-----------------------------------------------------------------------------|-----|
| 9.4.1    | Overview and Background                                                     | 280 |
| 9.4.2    | Drug Background and Information                                             | 281 |
| 9.4.3    | Objectives                                                                  | 282 |
| 9.4.4    | Primary and Secondary Endpoints                                             | 282 |
| 9.4.5    | Selection Criteria for Patient Population: Inclusion and Exclusion Criteria | 283 |
| 9.4.6    | Trial Design                                                                | 284 |
| 9.4.7    | Starting Dose and Dose Modification                                         | 285 |
| 9.4.8    | Evaluation of Drug Efficacy: World Health Organization (WHO) Criteria       | 287 |
| 9.4.9    | Toxicity Evaluation                                                         | 288 |
| 9.4.10   | Pathology                                                                   | 289 |
| 9.4.11   | Statistical Considerations                                                  | 290 |
| 9.4.11.1 | Hypothesis                                                                  | 290 |
| 9.4.11.2 | Sample Size                                                                 | 291 |
| 9.4.11.3 | Power                                                                       | 291 |
| 9.4.11.4 | <i>P</i> Values                                                             | 291 |
| 9.4.11.5 | Randomization                                                               | 291 |
| 9.4.11.6 | Stratification                                                              | 292 |
| 9.4.11.7 | Blinding                                                                    | 292 |
| 9.4.11.8 | Interim Analyses                                                            | 292 |
| 9.4.12   | Data-Monitoring Committee                                                   | 293 |
| 9.4.13   | Correlative Studies                                                         | 295 |

|                                         |     |
|-----------------------------------------|-----|
| 9.4.14 Adverse Event Reporting          | 295 |
| 9.4.15 Appendices (Usual Items)         | 297 |
| 9.4.16 Informed Consent Forms and HIPAA | 297 |
| 9.4.17 Closing                          | 300 |
| References                              | 301 |

## 9.4.1 OVERVIEW AND BACKGROUND

Phase III studies are defined as randomized controlled either single- or multicenter studies enrolling large numbers of patients with the purpose of rendering a definitive assessment of how effective the studied intervention is, compared with the current standard of care. The randomized clinical trial is characterized by two or more therapeutic treatment groups. These groups are sometimes referred to as *arms*, particularly in cancer trials. One treatment may be a placebo control in which a biologically inert substance (in a drug trial) is used. Note that while most phase III clinical trials compare two medications or devices, some trials compare three or four medications, doses of medications, or devices against each other. The protocol is the written operating manual of a trial, which ensures that researchers in different locations perform the study uniformly, allowing the data to be pooled and analyzed.

The U.S. National Institutes of Health, (NIH) organize clinical trials into five different types [1]:

*Treatment Trials* Test experimental treatments, new combinations of drugs, or new approaches to surgery or radiation therapy.

*Prevention Trials* Look for better ways to prevent disease in people who have never had the disease or to prevent a disease from returning. These approaches may include medicines, vitamins, vaccines, minerals, or lifestyle changes.

*Diagnostic Trials* Conducted to find better tests or procedures for diagnosing a particular disease or condition.

*Screening Trials* Test the best way to detect certain diseases or health conditions.

*Quality of Life Trials (or Supportive Care Trials)* Explore ways to improve comfort and quality of life for individuals with a chronic illness.

A randomized phase III controlled trial is the study design that can provide the most compelling evidence that the study treatment causes the expected effect on human health. Currently, some phase II and most phase III drug trials are designed as randomized, double blind, and placebo controlled. After preliminary evidence from phase I and II studies suggesting effectiveness of the drug has been obtained, phase III trials are designed to gather additional information to assess the overall benefit to risk relationship of the drug for possible use as a new standard of care.

Many NIH programs encourage or require the use of protocol templates, such as the ones available here at <http://ctep.cancer.gov/guidelines/templates.html>.

While following template guidelines can help guide authors, care should be exercised when using a template since not every part of a template will necessarily apply in a given study.

As is the case for all research protocols, writing a clinical protocol is a task that requires the collaboration and input of many individuals with various expertise to help reach the goal of a scientifically plausible and sound study. The writing of the protocol is of utmost importance and is a necessary condition on which the quality of the study itself depends.

#### 9.4.2 DRUG BACKGROUND AND INFORMATION

Phase III protocols, like phase I and II protocols, require a section detailing the scientific rationale for a protocol and the justification medically and scientifically of the hypothesis in question. This section should be organized in a logical and sequential manner.

First, the background section should focus on justifying the need for the study by elaborating first on where we are in the field. Here, a review of disease incidence and its impact on morbidity and mortality may help. Reviewing the results from similar prior studies or phase I and phase II studies testing the drug in question is also warranted. This will be a prerequisite to argue the need to perform the phase III trial and discuss what the trial will potentially add that would be important for patient care.

It is particularly important in phase III trials to indicate what the standard of care is in this patient population, so that if there is no standard of care an observation or placebo arm can be justified. In case there is a known standard, the control arm will usually be that standard, and the investigational drug will be used in the experimental arm(s).

If the drug in question is already established as an acceptable standard of care in the studied patient population, and the investigator wishes to perform a pharmacokinetic (PK) study with different dosing and administration schedules to be tested in several arms to determine the best dosing and schedule method, this design is that of a phase IV study.

It is important to learn about the drug and/or the intervention, and for the principal investigator (PI) to elaborate in the background section on the drug characteristics, the evidence derived from phase I and phase II trials and preclinical studies. Even though issues such as dosing, administration, and route of delivery will be expected to be detailed in the treatment plan section, it is helpful to introduce this information in the background section. Critical information will include the following: dose/dosing schedule, possible need for a vector, route of delivery, and method of preparation.

It is also advisable to discuss the possible benefits of performing the phase III trial in the background section, as this argument is closely linked to the findings of prior phase I or II studies. The benefits may not be restricted here to clinical outcome of disease but may also include ease of administration, convenience to patients, a better side effect profile, or a potential improvement in quality of life. Here, the authors may also allude to the patient population in question and elaborate briefly on the rationale behind the main inclusion and exclusion criteria as the opportunity is there to discuss these. This should be a prelude to the inclusion and exclusion criteria section in which a more detailed and complete description of these criteria is expected.

### 9.4.3 OBJECTIVES

In the initial development of a phase III protocol, it is essential for the investigator to try to answer several questions that will help in writing the protocol and define the objectives. Important questions to try to answer include: What is the expected outcome? What is the intervention? For how long will the intervention last? What patients or subjects is the intervention targeting? How many participants are needed? How can the potential benefit be optimized while minimizing potential harm?

The objectives should be stated clearly as they constitute the hypothesis in question. As in any trial, each objective needs to be discussed in the statistical section. The objective section should be concisely and simply written, often as a numbered or bulleted list. Here the authors should not include any arguments or justifications but rather be straightforward in describing the objectives. For a phase III trial the objectives usually are to compare two different approaches. For example, in a therapeutic trial the comparison usually involves two treatment plans, and the objective may seek to compare the survival of patients on each arm or the objective responses to the respective treatments. Other possible parameters may include the comparison of symptoms or quality of life parameters between arms or an assessment of the safety and tolerability of the treatment.

### 9.4.4 PRIMARY AND SECONDARY ENDPOINTS

It is important to remember that in a phase III study, as is the case in other designs, the primary endpoint will dictate the statistical method of analysis, the sample size, and stopping rules. Therefore, discussing these questions with a statistician while designing the study objectives cannot be overemphasized. For example, certain considerations such as the number of patients needed to reach statistical significance and the accrual potential in certain diseases may make the objectives difficult to achieve.

In a therapeutic trial, the primary endpoints must usually measure clinical outcome and will have the major impact on the rest of the study design as they will also influence the inclusion and exclusion criteria and the follow-up plan during or after the intervention. Secondary endpoints are usually easier to achieve and often revolve around the primary endpoints. They may include other parameters such as quality of life, assessing the ease of administration of agents in therapeutic trials, but in general they will not have as great an impact on the statistical section as the primary endpoints.

To better elucidate the choice of these endpoints, it is helpful for the author to include a section entitled Rationale for Selection of Endpoints. This will help the investigators review their decisions thoroughly and avoid major pitfalls that may result from a poor selection of endpoints. For example, choosing progression-free survival (PFS) as an endpoint for the trial may be more justifiable for certain types of malignancies such as sarcomas, which are generally incurable if being treated in the metastatic setting. Furthermore, prior interventions have not been shown to affect overall survival (OS) in this disease population—hence the rationale behind

choosing PFS as a primary clinical endpoint. The authors in this case may still be interested in looking at OS since this is a more questionable endpoint that is less likely to be reached in this disease, making it more suitable as a secondary endpoint. The authors may wish to add other endpoints as well that will be labeled exploratory endpoints and added to the secondary endpoints. These may include, for example, assessing changes in cancer-related symptoms.

#### **9.4.5 SELECTION CRITERIA FOR PATIENT POPULATION: INCLUSION AND EXCLUSION CRITERIA**

In any clinical trial, including phase III trials, selection criteria for patient enrollment must be well defined. The purpose of this process is to clearly define the subset of the general population to be investigated in the trial. When determining these criteria, it is important to consider the possible effects that the intervention may have on the subjects and any possible side effects already known or possibly anticipated by the investigator. Other important factors include the ability of the subjects to understand the nature of the intervention and be able to give a valid informed consent. In addition, for studies in which effectiveness is an endpoint, subjects will need to undergo exams that would determine the effectiveness (or lack thereof) of the intervention.

Exclusion criteria are designed to protect subjects with an expected high risk of side effects from the intervention or examinations required by the study, and to prevent the inclusion of biases in the study by including patients with serious comorbidities that could impact the outcome of a phase III trial.

The inclusion or exclusion criteria are, therefore, the medical or social criteria based on which a person may or may not be allowed to enter a clinical trial. These criteria are based on factors such as age, gender, type and stage of a disease, and other medical conditions. Careful attention should be paid while writing the inclusion and exclusion criteria since poorly written criteria have resulted in a number of ineligible and inevaluable patients being enrolled to a study, as well as the unnecessary exclusion of patients who could have been successfully enrolled.

Before proceeding with writing the eligibility criteria for a phase III trial, the investigator should be aware of certain facts: Criteria that are poorly written may dictate a poor rate of accrual to the study and may undermine its scientific validity and the ability to generalize its findings. Eligibility criteria present one of the most important obstacles to accrual to phase III clinical trials. It is imperative when writing these criteria to avoid confusion and to be simple and clear. Problems with being too selective in the inclusion and exclusion criteria include being unable to generalize the findings, limiting patient accrual, and possibly increasing the cost of the study. The investigator should aim to keep the number of criteria listed to a minimum, keeping only those necessary for the validity of the study and for the preservation of patient safety.

The study should enroll sufficient participants to be able to determine whether the endpoints of the study are met but should not enroll greater numbers than are needed to achieve statistical significance.

Factors such as the characteristics of the disease, the availability of alternative interventions, the availability of participants, the study endpoints, and the desired precision of the outcome may all influence the desired number of patients to enroll.

#### 9.4.6 TRIAL DESIGN

The study design section of a phase III protocol should include sufficient information for the participating site to develop a comprehensive clinical algorithm for enrolled patients on all arms of the study. It usually describes in a stepwise fashion all procedures required by the study.

This section may include the following: a description of the initial evaluation of patients, the treatment plan for both arms of the study, the need to use certain procedures, information on the agent(s) to be used including the investigational agent(s) or other standard agents used in the treatment plan, dose scheduling, and dose modification of all agents.

In phase III clinical trials where it is imperative to compare a new modality with the considered standard of care (or with placebo when there is no established standard of care), some fundamental principles must be followed:

1. The groups must be alike in all important aspects and differ only in the intervention each group receives.
2. The concept of randomization implies that each participant has the same chance of receiving any of the interventions specified in the study.
3. The randomization (allocation process) is carried out using a chance mechanism so that neither the participant nor the investigator will know *in advance* which will be assigned.
4. An effort should be made to avoid conscious or subconscious influences. In designing a clinical trial, a sponsor must decide on the target number of patients who will participate. The sponsor's goal is usually to obtain a statistically significant result showing a significant difference in outcome.

The number of patients required to give a statistically significant result depends on the chosen endpoints of the trial. The larger the sample size or number of participants in the trial, the greater the statistical power. However, in designing a clinical trial, this consideration must be balanced with the fact that more patients will lead to a more expensive trial. Because of their large size and relatively long duration, phase III trials are the most expensive, the most time-consuming and difficult trials to design and run, especially in therapeutic trials addressing the treatment of chronic medical conditions. In determining the sample size, it would help to have information on the nature of the condition being treated, the desired precision of the outcome, some knowledge of the effectiveness of the intervention, the usual outcome of the studied patient population with current standards of care, and the availability of alternative treatments. It is important to make sure that the objectives and study design portion of the statistical section are identical to those described in the objectives section.



It is also imperative to pay attention to the definitions of toxicities in the statistical section. These should match those in the safety and adverse events section.

The advantages of a phase III design over other study designs is that randomization tends to result in comparable groups, which validates the statistical analysis of the data and renders the study more meaningful in addressing the effectiveness of the intervention in terms of patient outcome compared with phase I or phase II designs.

Limitations of phase III studies may possibly include the inability to generalize results to all patients with the same condition or disease because of preselection criteria imposed, as participants may not represent the general study population. Investigators may be faced with challenges in recruiting patients as randomization may be hard to accept by some.

Certain parameters may render a phase I or II study to be more appropriate for a certain investigational agent than a phase III trial. Factors such as lack of effectiveness of the standard of care, therapies that have not been well investigated, or an expected dramatic response may be better investigated in a phase II study.

In summary, when writing the study design section of a phase III protocol, attention to the following items is required:

1. Make sure that a statement of the primary and secondary endpoints to be measured during the trial is included.
2. Describe the design of the phase III trial, for example, double blind, placebo controlled, parallel design. A schematic diagram of trial design, procedures, and stages would also help.
3. Describe the measures taken to minimize/avoid bias including, for example, randomization or blinding.
4. Include a description of the intervention for all arms of the trial including the control arm. This needs to encompass dosage and regimen of the investigational as well as noninvestigational drugs. A description of the dosage form, packaging, and labeling of the investigational and noninvestigational product(s) must be included.
5. The expected duration of subject participation needs to be specified, and a description of the sequence and duration of all trial periods, including follow-up, needs to be well stated.
6. Make sure to include a description of the stopping rules in the statistical section or discontinuation criteria for individual participants.
7. Take measures to assure accountability for the investigational product and the placebo if this applies.

#### **9.4.7 STARTING DOSE AND DOSE MODIFICATION**

In a phase III trial, it is important to reiterate that the drug will be administered only to randomized patients. Once the study drug has been administered to an enrolled patient, this will continue until the discontinuation criteria have been met.

These criteria should have been specified in the design section and may include intolerable side effects, a documentation of progression of disease, significant deviation from the protocol (a protocol violation), noncompliance, or a patient's decision to withdraw (these items do not need to be respecified in this section).

Under this section, it is important to specify the route and schedule of drug administration on each of the phase III study arms. It is also important to specify the time of administration if deemed important by the investigator(s). For oral medications, it should be specified whether the drug should be taken with water and with or without food. For oral medications, a mechanism for patients to report missed doses should be in place and should be specified and described in this section (e.g., a diary card).

All anticipated drug-related events should be described in the protocol as well as the investigator's brochure (IB). The authors should be clear in describing the severity of the potentially reported events, as well as the clinical guidelines they will be using to decide on the appropriate management of any of these events. A table detailing the management of study drug modifications for all the anticipated and possibly nonanticipated drug reactions should be included. This should include side effects expected of the study medication but also should offer information on the drugs used in control arms of the study (i.e., the conventional standard of care agents).

Additional sections should detail the management of specific side effects. For example, how to clinically manage mouth sores in addition to dose modifications. Information should include the method to be used for diagnosis of certain side effects. For example, "pneumonitis is to be diagnosed by bronchoscopy if suspected."

Guidelines should specify when to resume the administration of the drug, what conditions need to be met for the resumption of the drug administration, and what dose to use for each of these conditions.

Dose modification guidelines and interruption guidelines should also be included for unanticipated events such as emergencies unrelated to the drug administration. Here, it is important to specify when is it acceptable to continue the drug and keep patients on the trial, if deemed appropriate (e.g., "within two weeks from the time of interruption").

A section on formulation of the drug(s), packaging, and labelling should also be included. Here all chemical ingredients of the drug in question should be specified. Information on storage and stability should also be given (temperature and expected shelf life of the product). This is not needed for drugs used in the control arm or drugs that are considered to be used as part of the standard of care.

Dispensing of the drug and dosing compliance is usually the pharmacy's responsibility and this should be specified. The site must use the appropriate dispensing and log accountability provided by the sponsor. These logs are to be maintained by the study pharmacist. The principal investigator or a delegate is responsible for discarding all unused study drugs, and this should be clearly stated in the protocol. During the trial and after termination, patients are responsible for returning all unused supplies and any missing items should be investigated. This also applies to study drugs on phase III trials and not conventional approved drugs for treating the disease in question.

### 9.4.8 EVALUATION OF DRUG EFFICACY: WORLD HEALTH ORGANIZATION (WHO) CRITERIA

For phase III clinical trials, the investigator must determine the frequency of evaluation of response. This depends on the disease being investigated and the new agent being introduced. For uniformity and to reduce biases, responses to an intervention may need to be reviewed by a single source. For example, a central review where imaging studies are submitted from different participating centers may be designated to assess tumor response to a novel anticancer drug.

For measurements of response to an anticancer drug, measurable lesions are defined as those that can be accurately measured in at least two dimensions. Attention should be made to certain factors. For example, tumor lesions that are situated in a previously irradiated area should not, in general, be considered measurable.

Other potentially nonmeasurable lesions may include small lesions [longest diameter <20mm with conventional techniques or <10mm using spiral computed tomography (CT) scan], bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusions, lymphangitis, inflammatory breast disease, abdominal masses [not followed by CT or magnetic resonance imaging (MRI)], and cystic lesions.

In general, for assessing response, index lesions should be selected. This is done on the basis of their measurability in two dimensions and their suitability for accurate repeated measurements (by imaging techniques such as CT or MRI). A sum of the longest diameter (LD) and greatest perpendicular diameter of all index lesions will be calculated and reported as the baseline sum. The baseline sum will be used as a reference by which to characterize the objective tumor response. All other lesions (or sites of disease) including any measurable lesions over and above the index lesions should be identified as nonindex lesions and should also be recorded at baseline. Measurement of these lesions is not required, but the presence or absence of each should be noted throughout follow-up.

Evaluation of index lesions for phase I, II, or III clinical trials involve definitions of response, as set by the WHO or according to the RECIST criteria [2]. The WHO criteria follow.

#### **For Index lesions**

*Complete Response (CR)* Disappearance of all known disease, determined by two observations not less than 4 weeks apart.

*Partial Response (PR)* A > 50% decrease in the total tumor load of the lesions that have been measured to determine the effect of therapy not less than 4 weeks apart. The observations must be consecutive.

For single lesion, >50% decrease in tumor area (multiplication of longest diameter by the greatest perpendicular diameter).

For multiple lesions, a 50% decrease in the sum of the products of the perpendicular diameters of the multiple lesions. In addition there can be no appearance of new lesions or progression of any lesion.

*Stable Disease (SD)* A 50% decrease in total tumor area cannot be established nor has a 25% increase in the size of one or more measurable lesions been demonstrated.

*Progressive Disease (PD)* A >25% increase in the area of one or more measurable lesions or the appearance of new lesions.

### **For Nonindex Lesions**

*Complete Response (CR)* Complete disappearance of all known disease for at least 4 weeks.

*Partial Response (PR)* Estimated decrease in tumor area of >50% for at least 4 weeks.

*Stable Disease (SD)* No significant change for at least 4 weeks. This includes stable disease, estimated decrease of <50%, and lesions with estimated increase of <25%.

*Progressive Disease (PD)* Appearance of any new lesions not previously identified or an estimated increase of >25% in existent lesions.

**Duration of Overall Response** The duration of overall response is measured from the time measurement criteria are met for CR or PR (whichever is first recorded) until the first date that recurrent or progressive disease is objectively documented (taking as reference for progressive disease the smallest measurements recorded since the treatment started, best response scan).

The duration of overall CR is measured from the time measurement criteria are first met for CR until the first date that recurrent disease is objectively documented.

Duration of stable disease: Stable disease is measured from the start of the treatment until the criteria for progression are met, taking as reference the smallest measurements recorded since the treatment started.

**Placebo Effect** In a phase III clinical trial, a placebo may be used in one of the arms. A placebo is an inactive substance given to one group of participants, while the drug being tested is given to another group. A physical or emotional change, occurring after a substance is taken or administered, that is not the result of any special property of the substance is called a placebo effect, which may be beneficial, reflecting the expectations of the participant and, often, the expectations of the person giving the substance. The results obtained in the two groups are then compared to see if the investigational treatment is more effective in treating the condition than the placebo substance [1].

## **9.4.9 TOXICITY EVALUATION**

For phase I, II, or III clinical trials an adverse event (AE) is defined as any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product that does not necessarily have a causal relationship with the treatment. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medicinal (investigational) product, whether or not considered related to the medicinal (investigational) product. Pre-existing conditions that worsen during a study are to be reported as AEs.

The descriptions found in the revised National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 3.0 will be utilized for AE reporting. All appropriate treatment areas should have access to a copy of the CTCAE version 3.0. A copy of the CTCAE version 3.0 can be downloaded from the Cancer Therapy Evaluation Program (CTEP) web site (<http://ctep.cancer.gov>).

In addition to the definition of AEs, the safety (or adverse events) section should encompass a list of adverse events expected with the use of each study drug.

If the phase III trial involves blinding, a description of the unblinding process should be included. In addition, the human subject protection section should be included as is the case in all clinical trials and should encompass a discussion of minority representation, the inclusion or exclusion of subjects who are vulnerable because of mental capacity, language barriers or other factors, and a description of the methods for patient recruitment.

An informed consent for a phase III trial should include the following according to the Office of Human Subjects Research (OHSR) [3]:

1. Comprehension of the information provided.
2. Disclosure of relevant information to prospective research subjects.
3. Voluntary agreement of the subject, free from coercion.
4. Statement that the study involves research.
5. Purpose of the research and the length of the study.
6. Description of risks and benefits.
7. Discussion of alternative therapies.
8. Confidentiality policy.
9. Compensation for injury.
10. Contact for further questions/information.
11. Statement of voluntary participation.
12. The use of simple language and the avoidance of technical words, while being thorough and complete. The avoidance of phrases that may be coercive. The insight of a patient's advocate may be very helpful in writing the consent.
13. The principal investigator should personally review the consent form thoroughly.

#### 9.4.10 PATHOLOGY

As in other clinical trial designs, in a phase III trial, it is imperative especially with cancer-related therapies to be specific as to which pathologic subtype of cancer is to be included in the study. The inclusion and exclusion criteria should specify what pathologic subtypes are admissible to the studies and which are to be excluded. The lack of this information could result in jeopardizing the study by including patients who have heterogeneous diseases with different clinical behaviors.

If the study is multiinstitutional, the investigators may choose to have the pathology slides reviewed by a central pathology team for independent confirmation of the tissue diagnosis. This, as is the case with radiology central reviews, will minimize the introduction of subjective biases. Any discordance between the central pathol-

ogy assessment and the investigator site should be a subject for discussion between the sponsors and the principal investigator.

If the study calls for correlative examination of blood samples or archival tissues, the protocol should specify the mechanism by which the tissue is collected from the different institutions and consenting patients. It also helps to guide the participating institutions about the preferred tissue processing methods, for example, whether it is preferable that archival specimens be formalin fixed, or, if a block is not available, what is required, for example, a specified number of unstained slides to help determine a certain tissue marker. This entails a close coordination between the investigators, the central pathology labs, and the different pathology departments of participating centers.

It is strongly advisable for large phase III or multiinstitutional trials that a laboratory manual detailing the procedure of collection of material including archival material or whole blood specimens from participating patients, and shipping of the material be included as a separate document in the protocol.

#### 9.4.11 STATISTICAL CONSIDERATIONS

Given the complex nature of large clinical trials, a well-versed biostatistician is an integral part of the study design team and should be consulted in the initial planning stages. Although the study statistician will complete the statistical section of the trial protocol, investigators must have a basic understanding of the concepts involved.

##### 9.4.11.1 Hypothesis

As with any well-designed scientific experiment, phase III clinical trial investigators must begin with a hypothesis, defined as a supposition that appears to explain a group of phenomena and is advanced as a basis for further investigation [4]. The investigators must then design a trial to test the hypothesis.

In order to test a hypothesis, one must first state the null hypothesis ( $H_0$ ), which is defined as a statement that there is no difference between a standard and experimental treatment. The alternative hypothesis ( $H_1$ ) is a statement that there is a difference between a standard and experimental treatment [5]. If data from a clinical trial reject the null hypothesis, then the alternative hypothesis is proven true. There are three general rules when stating  $H_0$  and  $H_1$ :

1. The null hypothesis is the hypothesis of “no difference.”
2. Whatever is to be detected or supported should be the alternative hypothesis.
3. Statistical hypothesis are always set up in the hopes of being able to reject  $H_0$  and thereby accept  $H_1$ .

An example would be that a drug company has developed a new chemotherapy for pancreatic cancer, which it feels has a greater than 50% chance of shrinking tumors. If RR denotes response rates then:

$$H_0: RR \leq 50\%$$

$$H_1: RR > 50\%$$

Once a clinical trial is completed, the investigators must analyze the available data to make a determination on whether the data support  $H_0$  or  $H_1$ . Errors in reaching a conclusion from clinical trials can and do occur. When they occur, two types of errors that can be made are:

1. Type I error—rejecting the null hypothesis when it is actually true;  $\alpha$  is the probability of making a type I error.
2. Type II error—not rejecting the null hypothesis when it is false;  $\beta$  is the probability of making a type II error.

#### 9.4.11.2 Sample size

The number of patients to be enrolled to reach a statistically meaningful answer at the conclusion of a clinical trial is based upon the assumptions made during the planning stages of the trial.

#### 9.4.11.3 Power

Power is defined as the probability of rejecting the null hypothesis when it is indeed false or, conversely, accepting the alternative hypothesis is true when it really is true. In other words, the probability of obtaining a statistically significant result is known as the “power” of a trial. A general rule regarding power calculations of clinical trials is that as the number of planned subjects enrolled in a clinical trial increases, the power of the study increases as well. Conversely, the smaller the planned number of subjects accrued to a clinical trial, the bigger the difference between the standard and the experimental arms must be to demonstrate a statistically significant difference. Generally, the sample size of most phase III trials is adjusted to a power of either 0.80 or 0.90, assuming a difference between the standard and experimental arms is the smallest considered clinically meaningful.

#### 9.4.11.4 P Values

The  $P$  value is the probability of obtaining by chance a result at least as extreme as that observed, even when the null hypothesis is true and no real difference exists [4].

Alternatively, the  $P$  value can be thought of as the probability that the observed result is due to chance alone. Generally,  $P$  values  $<0.05$  are considered significant. The  $P$  value is calculated after the statistical test has been performed; if the  $P$  value is less than  $\alpha$ , the null hypothesis is rejected.

#### 9.4.11.5 Randomization

In clinical trials there are many sources of biases, ranging from physician judgments to patient self-selection, to referral patterns, all of which may skew data. Randomly



selecting trial subjects for either control versus experimental treatment is an effective way to minimize biases [6–10]. It is important to note that randomization does not completely remove all biases, but rather minimizes their effect on overall study results. Some investigators have argued that randomization is not absolutely necessary in all cases as matched historical or current controls can be identified. However, matching can only be performed with respect to known factors, and most investigators would agree there are many more unknown variables that cannot be accounted for in the study design. Thus without a randomization, a “statistically significant result” may be the result of a nonrandom difference in the distribution of unknown prognostic factors. Generally, for phase III clinical trials, which are large and costly to perform, randomization is a necessity to minimize the possibility of obtaining a nonmeaningful result.

There are many ways to randomize subjects, each with their respective advantages and disadvantages.

#### **9.4.11.6 Stratification**

In the case of known prognostic factors, stratifying subjects at the time of randomization may improve overall results as the study is more likely to be comparing two similar population vurses two populations that are skewed with an overabundance of a particular group. This can be accomplished by creating separate randomization lists for each stratum. For example, a clinical trial investigating a new agent in unresectable or metastatic pancreatic cancer with overall survival as the primary endpoint should prospectively stratify patients based upon locally advanced disease or metastatic disease as these two patient populations have known different survival. Otherwise, the study results may be uninterpretable due to a mixture of two patient populations with different expected survival estimates.

#### **9.4.11.7 Blinding**

Trials set so that participants do not know which intervention they are receiving are known as single-blinded trials. Those in which neither researchers nor participants know who is in the investigational or control group are called double-blinded trials. Double-blinded trials ensure that people assessing the outcome will not be influenced by knowing which intervention a participant is receiving and also that ancillary follow-up treatment will be the same.

#### **9.4.11.8 Interim Analyses**

Interim analyses are commonly performed in phase III clinical trials to evaluate current data and determine if a difference between the standard and experimental arms are sufficient enough to warrant premature disclosure of study results. Much controversy surrounds utilizing interim analysis data for stopping trials prematurely for efficacy. Supporters argue if a significant benefit of an experimental treatment is seen, why withhold treatment that may benefit patients; while detractors would argue that given more time the differences in the two treatments may vanish and stopping a trial would prevent fully evaluating the alternative hypothesis. In the setting of an interim analysis demonstrating significantly inferior results from an

experimental treatment, supporters would want to stop that trial from exposing trial subjects to an inferior experimental arm, while detractors would argue in favor of continuing the trial as the differences between the standard and experimental arms could potentially lessen or even reverse if the trial is permitted to run its full course.

In addition to ethical concerns, interim analyses also have potential biostatistical drawbacks. Fleming and colleagues reported that if interim statistical significance tests are performed every 3 months on a study planned for 3 years, the probability of reaching a statistically significant result prematurely can be as high as 26% [11]. This error is known as the type I error of the analysis plan. Therefore, in order for an interim analysis to be valid and useful, the analysis should be preplanned and not occur too frequently.

To minimize the potential negative impact of premature results influencing patients and trial investigators, phase III trials should have a data-monitoring committee (DMC) consisting of experienced research personnel [12, 13]. Due to potential conflict of interest issues with investigators and continuation of the trial, interim results are usually not shared with investigators unless the trial DMC authorizes release of the data. There are several different statistical designs for planning interim analysis [10, 14–16]. Haybittle and colleagues' method involves discounting interim differences unless the difference is statistically significant at the two-sided  $P < 0.0025$  level. If the interim differences are not significant at that level, the trial continues until its originally intended size. Other methods have a predetermined number of planned interim analyses and the  $P$  value that is used to determine statistical significance depends on the number of analyses performed during the trial.

#### 9.4.12 DATA-MONITORING COMMITTEE

Data-monitoring committees (also known as data safety monitoring boards) consist of a group of individuals with various areas of expertise charged with overseeing the conduct of a clinical trial in terms of safety and efficacy. Not all phase III trials are required to have a DMC, as in the case of low-risk trials. However, the majority of therapeutic trials will have a DMC. Depending on the sponsor and the trial, DMC membership and responsibilities will differ.

The U.S. Food and Drug Administration (FDA) defines a clinical trial DMC as:

A group of individuals with pertinent expertise that reviews on a regular basis accumulating data from one or more ongoing clinical trials. The DMC advises the sponsor regarding the continuing safety of trial subjects and those yet to be recruited to the trial, as well as the continuing validity and scientific merit of the trial. When a single DMC is responsible for monitoring multiple trials, the considerations for establishment and operation of the DMC are generally similar to those for a DMC monitoring a single trial, but the logistics may be more complex. [17]

Data-monitoring committees came into use in the 1960s. They were first utilized in large randomized multicenter trials sponsored by the federal government. The first DMC was formed as a result of an NIH external advisory group. This DMC was charged with monitoring the safety, efficacy, and trial conduct of several National

Heart Institute trials [18]. Current DMCs operate based upon the decades of experience gained since the first DMC [19]. Recently, the use of DMCs has grown in popularity. Due to issues related to conflict of interest and the use of mortality endpoints, integration of DMCs into industry-sponsored trials has been increasing.

Some governmental agencies that sponsor clinical trials may require DMCs for certain trials; however, the U.S. FDA does not currently require the use of DMCs except for research studies in emergency settings in which informed consent requirement is exempted. There are no absolute criteria for requiring a DMC, but generally accepted situations are a controlled trial of any size that will compare rates of mortality or major morbidity.

The composition of a DMC is critical; a poorly constituted DMC may put study subjects' lives at risk and also risk the scientific validity of a trial. Criteria commonly used in selecting members are:

1. Relevant experience in area of study
2. Experience on other DMCs
3. Absence of conflict of interest

Data-monitoring committees usually consist of clinicians with expertise in the area of interest and at least one biostatistician with experience with the statistical methods utilized in the trial being monitored. Depending on the trial, other potential members may include medical ethicists, toxicologists, epidemiologists, pharmacologists, or other individuals with potential interest or experience in the area of study. Ideally, members of a DMC should have no relationship to the study sponsors or investigator. In reality this is nearly impossible since members of a DMC are usually selected or nominated by the study sponsors or investigators.

In addition, DMCs should have an established charter describing the duties of the DMC prior to initiation of a study. Important components of the charter should include:

1. Meeting schedule and format (telephone versus face to face).
2. Meeting structure—ideally DMC meetings should be closed to all study sponsors and investigators, but the FDA does allow interaction between the DMC and study investigator/sponsors to clarify issues that may arise during a review.
3. Outline format of interim data to be presented to the DMC.
4. Study plan for interim analysis and stopping rules for futility for superior efficacy.

Although there are potential benefits to be derived from participation in clinical research, the investigational review boards (IRBs) and the NIH must ensure, to the greatest extent possible, the safety of study participants, that they do not incur undue risk and that the risks versus benefits are continually reassessed throughout the study period. Every clinical trial should have provision for data and safety monitoring. A variety of types of monitoring may be anticipated depending on the nature, size, and complexity of the clinical trial. In many cases, the principal investigator would be expected to perform the monitoring function [20].

### 9.4.13 CORRELATIVE STUDIES

Secondary endpoints can lead to meaningful trial results, in particular, laboratory or pathology correlatives are useful. Specimens can be obtained during the period of study enrollment and then banked for future use. Certainly, preplanned laboratory or pathological correlative studies can be fruitful, but saved specimens can prove to be invaluable for future study should new information become available or new scientific techniques are developed. The general process for evaluating the usefulness of correlative studies includes:

1. Identifying a clinical hypothesis
2. Identifying the laboratory parameters that may be influenced by the clinical treatment
3. Identifying the correlative hypothesis

An example of the utility of correlative studies is the International Adjuvant Lung Cancer Trial (IALT). This trial was designed to evaluate the effect of cisplatin-based adjuvant chemotherapy on survival after complete resection of non-small-cell lung cancer (NSCLC) [21]. Two years after the publication of the IALT results, Olaussen and colleagues hypothesized that expression of excision repair cross-complementation group 1 (ERCC1) protein could predict survival from cisplatin-based adjuvant chemotherapy in NSCLC as there is a significant amount of *in vitro* data supporting the rationale of platinum resistance to the expression of the ERCC1 messenger RNA (mRNA) in cell lines. In order to carry out their analysis, the investigators obtained 761 tumor specimens from IALT subjects and evaluated the expression of the ERCC1 protein. In the end, the investigators concluded, “patients with completely resected non-small-cell lung cancer and ERCC1-negative tumors appear to benefit from adjuvant cisplatin-based chemotherapy, whereas patients with ERCC1-positive tumors do not” [22].

An example of a prospectively designed study to bank tumor specimens is the Eastern Cooperative Oncology Group study E3200. The purpose of E3200 was to evaluate response, time to progression, and overall survival of patients with advanced colorectal cancer who have failed therapy with irinotecan and 5-fluorouracil when treated with oxaliplatin and fluorouracil with or without bevacizumab or bevacizumab alone. There were no preplanned correlative studies to be performed on the tumor specimens, but the investigators utilized the study as an opportunity to collect biologic samples prospectively for future use. One can imagine that these specimens could be analyzed at a later date to search for a biomarker of response as a wealth of clinical data for each patient was also prospectively collected during the clinical trial.

### 9.4.14 ADVERSE EVENT REPORTING

The timely reporting of adverse events that occur during the conduct of a trial is highly critical in order to maximize subject safety. Potential adverse events due to an experimental treatment under investigation must be quickly disseminated to all

study investigators so as to alert all study-related personnel to possible events and minimize future adverse events.

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) was convened to bring together experts from the pharmaceutical industry and the regulatory authorities of Europe, Japan, and the United States to discuss scientific and technical aspects of product registration. The overall purpose of this conference was to make recommendations to allow uniform interpretation and application of technical guidelines for product development across the world in order to minimize duplicate testing of new products. ICH defined the following terms to clarify terminology used for drug event reporting:

Adverse event (AE) is defined as “any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment” [23]. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory result), symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product.

Adverse drug reaction (ADR) is defined as “a response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of disease or for modification of physiological function” [23].

Serious adverse event (SAE) is defined as “a serious adverse event (experience) or reaction is any untoward medical occurrence that at any dose:

- Results in death
- Is life-threatening” [23]

Several mechanisms have been put in place by the FDA as well as other governing bodies such as the NCI. In the case of industry-sponsored studies, the sponsor is expected to collect all adverse event data and disseminate to all study sites to alert of potential adverse events.

MedWatch, is a FDA safety information and adverse event reporting program that serves both health care professionals and the medical product-using public. The FDA receives adverse drug reaction reports from manufacturers as required by regulation. Health care professionals and consumers can voluntarily send reports through the MedWatch program as well. These reports become part of a database. The ICH has issued guidance to facilitate the standardization of the data elements for the transmission of individual case safety reports regarding safety reporting (ICH E2B), which describes the content and format for the electronic submission of reports from manufacturers.

The FDA codes all reported adverse events using a standardized international terminology, MedDRA (the Medical Dictionary for Regulatory Activities), then uses this data to monitor for potential unforeseen toxicities. The data collected through the MEDWatch program is then incorporated into the Adverse Event Reporting System (AERS), which is a computerized information database designed to support the FDA's postmarketing safety surveillance program for all approved drug and therapeutic biologic products. The ultimate goal of AERS is to improve

the public health by providing the best available tools for storing and analyzing safety reports.

The CTEP branch of the National Cancer Institute utilizes a Web-based system for the electronic submission of expedited reports on protocols utilizing a CTEP-sponsored investigational new drug (IND) known as the Adverse Event Expedited Reporting System (AdEERS) to rapidly collect and disseminate clinical-trial-related adverse events [24].

#### 9.4.15 APPENDICES (USUAL ITEMS)

An appendix contains information that is nonessential to the understanding of the clinical trial protocol but may further clarify points without burdening the body of the protocol. Each appendix should be identified by a Roman numeral in sequence, for example, Appendix I, Appendix II, and so on. Each appendix should contain different material.

Appendices should include but are not limited to:

1. Informed consent form
2. Guidelines for study-related procedures not written in the body of the protocol such as:
  - a. Pathology submission guidelines and instructions
  - b. Correlative laboratory processing guidelines
3. Any additional instructions that are to be given to the study subjects

#### 9.4.16 INFORMED CONSENT FORMS AND HIPAA

For trials subject to the requirement of the U.S. FDA, which would be virtually all phase III clinical trials, the informed consent form (ICF) must meet the requirements of Title 21 of the Code of Federal Regulations, chapter 50 [25]. The Code of Federal Regulations (CFR) is the codification of the general and permanent rules published in the *Federal Register* by the executive departments and agencies of the federal government. It can be access at <http://www.gpoaccess.gov/cfr/about.html>. Chapter 50 covers the rules governing the protection of human subjects. Each ICF must also contain the six elements listed in 21 CFR 50.25(b) that are determined to be appropriate for a study. However, the IRB governing the study has the final authority for ensuring the adequacy of the ICF. In addition, the IRB may also impose additional requirements not listed in 21 CFR.

To improve the approval process of a multiinstitutional trial sample or draft consent, documents should be developed that each study site may alter to satisfy local IRB requirements. During the course of a study, many changes and amendments are likely to be made to the protocol as well as the ICF. In order to minimize confusion of multiple versions of ICFs, version numbers and IRB approval dates should be labeled on each version of the ICF approved.

**21 CFR 50.20: General Requirements for Informed Consent** Language in the ICF must be written in simple and easy to understand terms, as the vast majority of

study participants are not likely to possess medical knowledge sufficient to comprehend the risk and benefits involved in a modern-day clinical trial. 21 CFR 50.20 states the following:

Except as provided in §50.23, no investigator may involve a human being as a subject in research covered by these regulations unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative. An investigator shall seek such consent only under circumstances that provide the prospective subject or the representative sufficient opportunity to consider whether or not to participate and that minimize the possibility of coercion or undue influence. The information that is given to the subject or the representative shall be in language understandable to the subject or the representative. No informed consent, whether oral or written, may include any exculpatory language through which the subject or the representative is made to waive or appear to waive any of the subject's rights, or releases or appears to release the investigator, the sponsor, the institution, or its agents from liability for negligence.

**Non-English-Speaking Subjects** In the case of subjects who do not possess a sufficient grasp of the English language to critically evaluate an ICF, a translated ICF document in the native language of the potential study subject should be presented. While a translator may be used intermittently, translators should not be used routinely due to potential variations between individual translators. In addition, local IRBs may have additional rules regarding foreign language ICFs; as such, investigators should consult with the local IRB before proceeding to translate an ICF since this is an expensive procedure.

In the event of an illiterate subject, provided the subject is able to (1) retain the ability to understand the concepts of the study and evaluate the risk and benefit of being in the study when it is explained verbally and (2) is able to indicate approval or disapproval to study entry, they may be entered into the study. To further validate the consenting process of an illiterate individual, it is also suggested to have an impartial third-party witness as well as videotaping the entire consent process.

**Assent of Children** Guidelines for ICFs regarding children are not specifically addressed in the 21 CFR; instead the FDA defers judgment to the IRB overseeing the particular study. For children old enough to read and comprehend, many IRBs require signatures from both the parents and the child before study entry.

### **21 CFR 50.25: Elements of Informed Consent**

- (1) A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental.

The ICF must explicitly state that the purpose of a clinical trial is research as the relationship between a patient–physician is different than that between a trial participant–investigator.

- (2) A description of any reasonably foreseeable risks or discomforts to the subject.



The risks of the trial must be explicitly stated in the ICF. Many IRBs will also require quantification of risks, that is, the expected percentage of patients who are likely to experience a side effect such as nausea or diarrhea.

- (3) A description of any benefits to the subject or to others which may reasonably be expected from the research.

The potential benefits of a trial must be stated as a possible benefit and not be overstated as a likely event. Common phrases used in ICFs include “this trial may or not benefit you directly, but may help other people with the same medical problem as you.”

- (4) A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.

The ICF must offer alternative choices to the patient if available.

- (5) A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained and that notes the possibility that the Food and Drug Administration may inspect the records.

The ICF must contain a statement explaining the confidentiality of the subject’s medical records as they relate to the study and also state who may have access to the subject’s medical records such as the FDA, IRB, or study sponsors.

- (6) For research involving more than minimal risk, an explanation as to whether any compensation and an explanation as to whether any medical treatments are available if injury occurs and, if so, what they consist of, or where further information may be obtained.

The ICF must contain a statement regarding compensation if any will be provided. The ICF must also explain whether there is compensation available in case of injury but must not waive or appear to waive the rights of the subject. Acceptable language includes “no funds have been set aside for,” “[the cost] will be billed to you or your insurance.”

- (7) An explanation of whom to contact for answers to pertinent questions about the research and research subjects’ rights, and whom to contact in the event of a research-related injury to the subject.

The ICF must clearly state who should be contacted in the event of questions that arise regarding study participation. Generally, the contact person listed should be the principle investigator at a study site.

- (8) A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

The ICF must clearly state that participation in a study is voluntary. Common phrases used include “you may participate in this trial if you wish. Your decision to participate in this trial will not affect your care. The options to the trial are \_XXXX\_. You and your physician can discuss these options and you must decide what is best for you.”

The following are additional required elements of an ICF when applicable:

- (1) A statement that the particular treatment or procedure may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant) which are currently unforeseeable.
- (2) Anticipated circumstances under which the subject’s participation may be terminated by the investigator without regard to the subject’s consent.
- (3) Any additional costs to the subject that may result from participation in the research.
- (4) The consequences of a subject’s decision to withdraw from the research and procedures for orderly termination of participation by the subject.
- (5) A statement that significant new findings developed during the course of the research which may relate to the subject’s willingness to continue participation will be provided to the subject.
- (6) The approximate number of subjects involved in the study.

**Health Insurance Portability and Accountability Act (HIPAA)** In 1996, the Department of Health and Human Services (HHS) established national standards for electronic health care transactions and national identifiers for providers, health plans, and employers as a result of the administrative simplification provisions of the Health Insurance Portability and Accountability Act (HIPAA) [26]. These standards seek to prevent the misuse of confidential medical information. Penalties for breaking HIPAA regulations are severe (see below), as such investigators and IRBs must make sure to be HIPAA compliant. Penalties for HIPAA infractions are:

- (a) Be fined not more than \$50,000, imprisoned not more than 1 year, or both.
- (b) If the offense is committed under false pretenses, be fined not more than \$100,000, imprisoned not more than 5 years, or both.
- (c) If the offense is committed with intent to sell, transfer, or use individually identifiable health information for commercial advantage, personal gain, or malicious harm, be fined not more than \$250,000, imprisoned not more than 10 years, or both.

#### 9.4.17 CLOSING

Designing and implementing phase III clinical trials is a highly complex undertaking that requires good collaboration among all members of a well-trained multidisciplinary team. Involvement of the multidisciplinary team should be encouraged as early as possible in the design and conceptualization stage of protocol development. As with any scientific experiment, obstacles to the activation and implementation

of phase III trials will undoubtedly occur; however, well-written and designed trials will help to complete the study in a timely fashion.

## REFERENCES

1. Clinical trials.gov. (2007), Understanding Clinical Trials; available at <http://clinicaltrials.gov/ct2/info/understand>.
2. Therasse, P., Arbuck, S. G., Eisenhauer, E. A., et al. (2000), New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada, *J. Nat. Cancer Inst.*, 92(3), 205–216.
3. Office of Human Subjects Research (OHSR). (2006), Information sheet 6: Guidelines for writing informed consents; available at [ohsr.od.nih.gov/info/sheet6.html](http://ohsr.od.nih.gov/info/sheet6.html).
4. Anderson, D. M., ed. (2003), *Dorland's Illustrated Medical Dictionary*, Saunders, Philadelphia.
5. Milton, J. S. (1999), *Statistical Methods in the Biological and Health Sciences*, 3rd ed., McGraw-Hill, New York.
6. Byar, D. P., Simon, R. M., Friedewald, W. T., et al. (1976), Randomized clinical trials. Perspectives on some recent ideas, *N. Engl. J. Med.*, 295(2), 74–80.
7. Chalmers, T. C., Block, J. B., and Lee, S. (1972), Controlled studies in clinical cancer research, *N. Engl. J. Med.*, 287(2), 75–78.
8. Hill, A. B. (1951), The clinical trial, *Br. Med. Bull.*, 7(4), 278–282.
9. Peto, R., Pike, M. C., Armitage, P., et al. (1976), Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design, *Br. J. Cancer*, 34(6), 585–612.
10. Peto, R., Pike, M. C., Armitage, P., et al. (1977), Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples, *Br. J. Cancer*, 35(1), 1–39.
11. Fleming, T. R., Green, S. J., and Harrington, D. P. (1984), Considerations for monitoring and evaluating treatment effects in clinical trials, *Control. Clin. Trials*, 5(1), 55–66.
12. Fleming, T. R., Ellenberg, S., and DeMets, D. L. (2002), Monitoring clinical trials: issues and controversies regarding confidentiality, *Stat. Med.*, 21(19), 2843–2851.
13. Smith, M. A., Ungerleider, R. S., Korn, E. L., et al. (1997), Role of independent data-monitoring committees in randomized clinical trials sponsored by the National Cancer Institute, *J. Clin. Oncol.*, 15(7), 2736–2743.
14. Fleming, T. R., Harrington, D. P., and O'Brien, P. C. (1984), Designs for group sequential tests, *Control. Clin. Trials*, 5(4), 348–361.
15. O'Brien, P. C., and Fleming, T. R. (1979), A multiple testing procedure for clinical trials, *Biometrics*, 35(3), 549–556.
16. Pocock, S. J. (1982), Interim analyses for randomized clinical trials: The group sequential approach, *Biometrics*, 38(1), 153–162.
17. US Department of Health and Human Services. (2006), Guidance for Clinical Trial Sponsors—Establishment and Operation of Clinical Trial Data Monitoring committees; available at <http://www.fda.gov/cder/guidance/index.htm>.
18. Heart Special Project Committee (1988), Organization, Review and Administration of Cooperative Studies (Greenberg Report): A Report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967, *Control. Clin. Trials*, 137–148.

19. Ellenberg, S., Geller, N., Simon, R., et al. (1993), Proceedings of Practical Issues in Data Monitoring of Clinical Trials, Bethesda, Maryland, January 27–28 1992, *Stat. Med.*, 12, 415–616.
20. NIH Clinical Trials Committee. (1979), Clinical Trial Activity, *NIH Guide*, 8(8), 29.
21. Arriagada, R., Bergman, B., Dunant, A., et al. (2004), Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer, *N. Engl. J. Med.*, 350(4), 351–360.
22. Olaussen, K. A., Dunant, A., Fouret, P., et al. (2006), DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy, *N. Engl. J. Med.*, 355(10), 983–991.
23. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Expert Working Group. (1994), ICH Harmonised Tripartite Guideline: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting E2A; available at <http://www.ich.org/cache/compo/475-272-1.html#E2A>.
24. NCI Cancer Therapy Evaluation Program (CTEP). (2005), CTEP, NCI Guidelines: Adverse Event Reporting Requirements; available at [http://ctep.cancer.gov/protocolDevelopment/default.htm#adverse\\_events\\_adeers](http://ctep.cancer.gov/protocolDevelopment/default.htm#adverse_events_adeers).
25. US Food and Drug Administration (FDA). (2007), Code of Federal Regulations (CFR) Title 21, Chapter 50, Protection of Human Subjects; available at <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=50>.
26. US Department of Health and Human Services. (1996), Health Insurance Portability and Accountability Act (HIPAA); available at <http://www.cms.hhs.gov/HIPAAGenInfo>.

# 9.5

---

## *Phase IV: Postmarketing Trials*

**Karl Wegscheider**

*Department of Medical Biometry and Epidemiology, University Hospital  
Eppendorf, Hamburg, Germany*

### **Contents**

|          |                                                      |     |
|----------|------------------------------------------------------|-----|
| 9.5.1    | Introduction                                         | 304 |
| 9.5.1.1  | Life of a Drug                                       | 304 |
| 9.5.1.2  | Chapter Overview                                     | 305 |
| 9.5.1.3  | Commitment of Phase IV                               | 305 |
| 9.5.2    | Definition                                           | 306 |
| 9.5.3    | Objectives                                           | 306 |
| 9.5.3.1  | Starting Point of Phase IV                           | 306 |
| 9.5.3.2  | Open Issues                                          | 307 |
| 9.5.3.3  | Study Objectives in Phase IV                         | 307 |
| 9.5.4    | Approaches                                           | 310 |
| 9.5.4.1  | Controlled Randomized Trials                         | 310 |
| 9.5.4.2  | Cluster-Randomized Trials                            | 311 |
| 9.5.4.3  | Cross-Sectional Observational Studies                | 313 |
| 9.5.4.4  | Longitudinal Studies/Cohort Studies                  | 314 |
| 9.5.4.5  | Case-Control Studies                                 | 315 |
| 9.5.4.6  | Aggregate-Level Studies/Ecological Studies           | 316 |
| 9.5.4.7  | Medical Registers                                    | 317 |
| 9.5.4.8  | Surveillance                                         | 318 |
| 9.5.4.9  | Systematic Review, Meta-Analysis, and Metaregression | 318 |
| 9.5.4.10 | Health Economics/Health Technology Assessment (HTA)  | 319 |

|                                                              |     |
|--------------------------------------------------------------|-----|
| 9.5.5 Options                                                | 320 |
| 9.5.5.1 Data Sources                                         | 320 |
| 9.5.5.2 Experimental versus Observational Approach           | 321 |
| 9.5.5.3 Quality Assessment, Monitoring, and Auditing         | 321 |
| 9.5.5.4 Registration of Studies, Access to Data, and Results | 322 |
| 9.5.6 Impact of Phase IV Trials                              | 322 |
| References                                                   | 322 |

## 9.5.1 INTRODUCTION

### 9.5.1.1 Life of a Drug

The development of a drug has many aspects in common with the raising of a child. Both processes take place in typical phases that follow an unalterable order. The next stage is generally only attainable after the previous stage has been completed. Preclinical development may be seen as the embryonic stage, the first intake of the drug as the moment of birth. This moment bears special risks that cannot be foreseen even by ideal preclinical work. “Birth deformities” of the drug may become apparent only after a few human beings have been exposed. Fortunately, such observations are not frequent since the dosage will be kept low at the beginning and then increased only in small steps, analogous to the steady and rapid growth of a neonate. Phase I can be seen as the early childhood, where the drug is tested in healthy subjects in a protected environment under ideal conditions and the constant monitoring of the producers (parents) and specially trained experts charged with tending to the “education” and “socialization” of the product. Phase II is the first systematic contact with patients. For the first time, performance is demanded, and thus this phase parallels the time in school. From now on, failures have consequences. However, some kind of adaptation is allowed as far as the potential of the new drug is concerned. The ultimate indication and dosage are not yet defined and may be adapted to new experiences as the talents of the child are detected and subsequently supported and promoted. However, at some time (at the passage from phase II to phase III) the intended use of the drug has to be determined, and in phase III the drug has to prove its ability to attain the therapeutic goals for which it was presumed qualified. This point in time resembles the moment when a decision is made on the future career or field of specialization. The evaluation of the phase III trial(s) resembles the final examinations, and the licensing of a drug is the equivalent of the degree that entitles the holder to practice a particular profession.

Many people think that with the awarding of some kind of certificate for vocational or professional qualification and the first “real” employment, that children are grown-up and no longer require their parents’ care. But any parent who has had the pleasure of bringing a child this far along the path to maturity knows that is not really the case because the personal life as well as professional careers are still full of challenges, competition, and risks. Support is further required, but not only requested and given by the parents who usually have a personal interest in the well-being of their now grown-up child, but also by friends and family members.

Likewise, many people think that with the licensing of a pharmaceutical the research investment has or should come to an end, and that the marketed drug

should just be used as it was licensed in order to earn money for its producers. But as is the case for any grown-up child, a drug faces competition, has to adapt to new environments and changing demands, may fail in certain situations, and has to be defended against suspicions and accusations. What is more, information is needed and has to be provided on many aspects and has to be propagated to the public. Thus, continuing critical support is needed and will be given not only by the producers but by many other parties such as government, funding institutions, researchers, health insurances, patient organizations, and the media. Thus, there is a need and a demand for postmarketing trials of a drug, justifying the introduction of another phase, namely phase IV, for this kind of study, and this demand is met more and more frequently as we all become more and more aware of the intricate challenges with which a “grown-up”, that is, a licensed drug has to contend.

However, unlike the premarketing phases, in phase IV there is no standard approach, no definite order of study sequence, no unique methodology, nor any agreement on optimality. Instead, there is a still growing diversity of study types, organizational forms, funding models, regulatory guidelines and recommendations, and a rich and complex scientific as well as public debate on licensed drugs.

### 9.5.1.2 Chapter Overview

In this chapter, a broad overview is provided on the wealth of methodological approaches to phase IV trials. After definition and an overview of the objectives, each approach is presented separately, including the motivation behind it, the standard design, the challenges one has to meet during conduct, the standard evaluation procedure, the interpretations and common misinterpretations, and the limitations of the approach, shedding light on the place of the approach in the concert of phase IV studies. Phase IV studies on statins, on the use of antibiotics for coughs, and on the safety profile of oral contraceptives will serve as examples. The chapter ends with a summary of the options for phase IV trials and the decisions that have to be met in the planning stage of a phase IV trial by sponsors and investigators.

### 9.5.1.3 Commitment of Phase IV

The life of an individual human being inevitably ends with death. At first glance this might not seem to be so inevitable for a drug, but, although its life time is not biologically limited, it is useful to think of a drug's life as finite. A drug may be well known, have a large market share, and may even be identified with its substance class; nevertheless, it is always at risk of being topped by a better (i.e., more effective, more easily administered, more tolerable, or safer) drug, or by a cheaper drug, or by a drug that is simply better promoted or more fashionable. It may be frivolously accused of serious adverse reactions or it may be superfluous because the disease is conquered, for instance, by a new vaccine. Like human beings, a drug may die from an accident, that is, is withdrawn from the market because of safety concerns, but usually it gradually loses its dominance; it remains on the market but exists only in a niche (a kind of “retirement”) and disappears some day without many people noticing it, possibly because it is no longer economically viable. It may survive in another form, as a generic or as part of a combination, but the patent and brand name, its identifying characteristics cease to exist.



In summary, with a few exceptions, a drug can, and certainly will, die at some indefinite future date. Undoubtedly, the death will be a severe loss not only for the producers but also for the community since the deep knowledge of the drug gained by experience over an abundance of years of application collected in phase IV studies will at least in part die with the drug. After all, an established drug therapy provides stability and confidence, like an old friendship where you know the strengths and weaknesses of the friend and have learned to live with them. Does this mean that phase IV research is futile or pointless? Should we refrain from any post-marketing research, enjoy the licensing success, and avoid the unavoidable risks of any medical research? Quite the contrary is the case. If we do not study drugs in the market, we are in danger of being surprised by speculations that can be neither proven nor rebutted, and nevertheless can end the life of a drug by ruining its image. Lipobay and Vioxx [1] and Waxman [2] are prominent examples of this. What is more, we may miss opportunities for the use of a drug that may never come to light, and this may shorten its lifetime. A series of trials with different approaches is thus not only recommendable for medical and economic reasons but an ethical commitment for manufacturers, scientists, and society alike and thus not only be taken seriously but encouraged.

## 9.5.2 DEFINITION

A phase IV trial is a clinical study where the investigational therapy includes the use of a licensed drug in the approved indication. (Studies of a licensed drug in a new indication are usually referred to as phase IIIb trials.) The term *trial* is usually reserved for studies where the drug under investigation is administered for study purposes. In a strict sense, observational studies on routine use or medical registers are not covered by this definition. Nevertheless, the term *clinical trial* is sometimes used somewhat more loosely for clinical studies in general. In this chapter, we will reserve the term *trial* for interventional studies. It should, however, be noted that this chapter will cover other approaches as well in order to give a more complete account of the diversity of options covered by the concept *phase IV study*.

## 9.5.3 OBJECTIVES

### 9.5.3.1 Starting Point of Phase IV

At the end of phase III/beginning of phase IV, there is already a considerable body of knowledge about a drug. The proof of principle was given in phases I and II. Phase II/III added some knowledge on dosage and application. From phase III we know that the drug works in a variety of clinical settings and is able to improve clinical relevant endpoints.

On the other hand, the knowledge of the drug at the time of market introduction will always be limited in some sense due to the legal restrictions, economic considerations, and the primary targeting of approval. Up to phase III, the application of the drug has taken place under the conditions of controlled trials, that is, under

steady observation and quality control, in environments that were fully aware of the research motivation behind the application, and with strictly controlled inclusion and exclusion criteria defined in order to receive a homogenous study population to minimize the risk of a negative result. In many phase III trials, the criteria are so cautiously defined that only a small percentage of the future users of the drug can be included, resulting in study populations that are not representatively selected. At the time of licensing, we only have experience on the application of a drug in a very specific population, for a limited range of doses, under absolutely ideal circumstances, and in a number of patients too small to address less frequent complications and safety concerns that nevertheless may become very relevant if the exposure to the drug is extended for routine care of patients.

### 9.5.3.2 Open Issues

Correspondingly, optimization of use and safety concerns are the most prominent reasons for conducting postmarketing trials. Just recently, questions of the financial balance have joined the club and sometimes even dominate, starting as add-ons to clinical study projects and then becoming an independent research area. The more traditional form of studies aiming at economic aspects consists of marketing trials that address the propagation and use of a drug. These studies are frequently suspected of being undertaken predominantly for the purpose of propagation by merely performing the study and not for research purposes. This suspicion indicates that there may be a conflict of interest between different study objectives in phase IV. Conflicts of interest are a minor concern in phases I–III but are rather typical for phase IV studies where the scope of research questions is broader and more stakeholders with potentially different interests are participating in the research process.

### 9.5.3.3 Study Objectives in Phase IV

Table 1 gives an overview of common study objectives in phase IV, organized according to the four main research areas of phase IV. Optimization of use is required since many decisions made in phases I–III were taken for the purpose of gaining a license for the drug. Thus, only the most promising formulations and dosages are developed and tested in a targeted study population. What is more, the decisions in phases I–III were based on experiences from small samples that were not necessarily conclusive in nature. A further process of fine-tuning is therefore desired and should be based on a solid empirical base, big enough to allow even sophisticated considerations. The first six objectives belong to this category. Furthermore, the prelicensing trials were powered for overall evaluations. But patients' reactions to a drug may vary considerably, and thus an adaptation of the therapy to different subgroups of patients and several influences on drug efficacy are required and necessitate further studies, as the other objectives of this section demonstrate.

Safety issues are critical for the life of a drug. At the time of licensing, the finalized trials have had optimal power for the proof of efficacy. However, the power of the complete database is usually far too small to draw reliable conclusions about

**TABLE 1 Objectives of Phase IV Studies**

| <i>Optimization of Use</i>           |                                                                                                                                            |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| IMPROVING TREATMENT STRATEGY         |                                                                                                                                            |
| Opt-01                               | Optimization of application form                                                                                                           |
| Opt-02                               | Optimal dosage for different groups of patients, exploration of dose–response curve                                                        |
| Opt-03                               | Optimal duration of therapy                                                                                                                |
| Opt-04                               | Optimization of the circadian pattern                                                                                                      |
| Opt-05                               | Evaluation of further clinical endpoints                                                                                                   |
| Opt-06                               | Implementation of the drug in complex therapeutic strategies                                                                               |
| ADAPTATION TO SPECIAL PATIENT GROUPS |                                                                                                                                            |
| Opt-07                               | Specification of indication, improvement of diagnostics to establish the indication                                                        |
| Opt-08                               | Adaptation to age, gender, race, in particular, pediatric or geriatric use of the drug                                                     |
| Opt-09                               | Use in patients with comorbidity                                                                                                           |
| Opt-10                               | Assessment of individual variability of drug effects, metabolism                                                                           |
| EXPLORATION OF EFFECT MODIFIERS      |                                                                                                                                            |
| Opt-11                               | Identification of patient characteristics that increase or diminish drug effects, in particular, education, habits, lifestyle              |
| Opt-12                               | Efficacy in various clinical settings                                                                                                      |
| Opt-13                               | Effect modification due to environmental factors, in particular, food intake, sociodemographic factors                                     |
| Opt-14                               | Compliance studies                                                                                                                         |
| <i>Safety Issue</i>                  |                                                                                                                                            |
| Saf-01                               | Evaluation of the frequency of routine side effects/expected adverse effects                                                               |
| Saf-02                               | Detection and enumeration of unexpected or rare side effects and delayed effects                                                           |
| Saf-03                               | Cumulative effects of long-term use, accumulation effects in patients                                                                      |
| Saf-04                               | Interactions with concomitant therapies, in particular, drug interactions                                                                  |
| Saf-05                               | Permanent safety surveillance to track new developments that raise safety concerns                                                         |
| Saf-06                               | Addiction and abuse                                                                                                                        |
| Saf-07                               | Inadequate prescriptions                                                                                                                   |
| Saf-08                               | Risk/benefit ratio in different subgroups                                                                                                  |
| <i>Health Economy</i>                |                                                                                                                                            |
| Eco-01                               | Cost/benefit relation, cost effectiveness                                                                                                  |
| Eco-02                               | Head-to-head comparison to alternative therapeutic approaches, in particular, competing drugs                                              |
| Eco-03                               | Economic evaluation of complex therapeutic therapies that incorporate the use of the drug                                                  |
| <i>Marketing</i>                     |                                                                                                                                            |
| Mar-01                               | Accompanying and assisting the transition from clinical development to marketplace                                                         |
| Mar-02                               | Demonstration and implementation of the proper use of the drug                                                                             |
| Mar-03                               | Investigation and improvement of the compicance of physicians as well as of patients                                                       |
| Mar-04                               | Acceptance and performance of the drug in different markets, e.g., changes due to a switch from prescription to over-the counter marketing |
| Mar-05                               | Exploration of convenience of application, patient satisfaction, effects on quality of life                                                |

safety issues. This is not only due to the fact that serious adverse events are in most cases rare and may appear with some delay but also because the random fluctuation in the normal population, the treated disease itself and other medication and treatment will result in adverse events that may or may not be caused or aggravated by the drug under study. Extreme sample sizes are required to detect and validate adverse drug reactions caused by a drug in the presence of a natural background of events. Thus, other approaches in addition to randomized trials are required to address these issues, for purely economic reasons. Nevertheless, this kind of evaluation is mandatory since the public is extremely reluctant to accept even small safety risks for the sake of an average benefit in the patient population. The main reason for this reluctance presumably is the fact that a serious adverse event is always personal, while the benefit of a drug is statistical and anonymous. Thus, risk–benefit calculations do not dispense manufacturers from the commitment to meticulously search for each case when a drug might have caused harm. The objectives under the corresponding heading in Table 1 demonstrate how many different aspects of safety have to be addressed in one way or the other as soon as a drug is on the market.

Health economic studies traditionally investigate cost–benefit balances and the cost effectiveness of a drug. The limited availability of funds drives the increase in the number of such studies in phase IV. Economic studies usually require bigger sample sizes and other evaluation methods than clinical trials. Hence, economic evaluations accompanying phase III trials are usually not sufficient, and planned phase IV studies are required. Increasing competition in health markets let people ask for more direct head-to-head comparisons between therapies that help to define the place and potential function of a drug in the health care environment. While the two upper objectives in that section aim at a drug-related evaluation, more and more people take the perspective that the utility of drugs has to be judged according to their contribution to complex interventions in the health system. A statin, for instance, may be studied for its cost effectiveness (Eco-01) (see Table 1); it may be studied in direct comparison to another statin or a different lipid-lowering therapy (Eco-02), or as part of a multiway lipid-lowering program for a certain target population (Eco-03).

Marketing begins with the transition from clinical development to the marketplace. It is useful to accompany this process with a scientific evaluation. However, studies for market introduction are frequently understood as opportunities to promote the drug to physicians and patients as yet unfamiliar with the substance, making the study something akin to advertisement. The scientific aspect of the study may thus be compromised, the more so if the study is explicitly performed for the purpose of the demonstration and establishment of study drugs in the market. While these two study objectives are frequent in phase IV, they are not generally accepted by the scientific community as part of research. The scientific portion plays a greater role in compliance studies, in studies on the performance of the drug under different distribution regimes, and in the evaluation of patients' experiences and opinions when applying the drug. These three objectives are sometimes undervalued. In fact, availability, compliance, and patients' satisfaction may contribute considerably to the success of a theoretically useful drug in a certain health care setting and thus may be very important for all health system participants in order to actually reap the potential benefit of a new drug.

## 9.5.4 APPROACHES

The diversity of objectives in phase IV is reflected by a diversity of approaches. This section describes 10 approaches, defined by a characteristic methodology. Each of these approaches is useful for a certain subset of the objectives listed in Table 1. First, the motivation behind the approach is given. The design is described, followed by a summary of the typical challenges in performing such a trial. Then an accompanying example is provided. This is followed by an account of the standard evaluation. In the final sections, recommendations for interpretation of the results are offered and the limitations of the approach discussed.

### 9.5.4.1 Controlled Randomized Trials

**Motivation** Scientific conclusions are always based on comparisons. However, in everyday life as in studies, comparisons are always in danger of bias, that is, it may be that a difference is caused by other factors than the condition (in this case the drug) under study. Bias can never be definitely excluded, but it can be controlled by probability. For this purpose, a random allocation of the drug to the patient is required.

Correspondingly, the randomized controlled trial (RCT) is the gold standard for phase III trials. In phase IV, RCTs are the preferred approach for all objections related to the optimization of use, and for objectives Eco-02 and Mar-03 (see Table 1), whenever it is possible to perform such a trial. RCTs may be performed as superiority trials to prove some advantage of a drug over the comparator or as non-inferiority trials proving at least comparable efficacy of a drug to a competing active treatment.

**Design** The RCT is prospective. A study protocol has to be prepared that precisely defines the study population and describes the inclusion procedure. After inclusion, patients have to be randomized, preferably by means of an external randomization procedure. Further treatment, observation, and follow-up should be identical for all groups under comparison. A sample size calculation has to be performed in advance for the study to be adequately powered for the primary endpoint.

**Challenges** There are potential biases that may take place after randomization. To avoid differential reporting, blinding of patients, treating physicians, and readers is recommended. Differential drop-out or missing values may compromise the study and should be avoided as much as possible. Major challenges in conducting such trials are recruitment problems and limited compliance.

### Examples

1. A phase IV head-to-head comparison between two statins was performed to test the assumption that 40 mg pravastatin daily is at least equivalent to more aggressive lipid reduction by 80 mg atorvastatin daily after acute coronary syndrome [3]. The study demonstrated superiority of the more aggressive therapy.

2. To assess the efficacy of pravastatin in the elderly, a double-blind randomized trial versus placebo (PROSPER) was performed in patients aged 70–82 with a history of or risk factors for vascular disease [4]. In the conclusion, the extension of the treatment strategy from middle-aged to elderly people is recommended.

**Evaluation** In the simplest case, the evaluation of a superiority trial consists of two-group comparisons by Fisher's exact tests or chi-square tests for binary outcomes and by two-sample *t* tests or nonparametric tests for continuous outcomes. An adjustment for baseline differences between patients may increase the power of the trial but is not mandatory. The evaluation of noninferiority trials is more complex. In any case, the presentation of the results using point estimates and confidence limits is recommended.

**Interpretation** If the study conduct was regular, cause–effect conclusions can be drawn for the primary endpoint and for the significant secondary endpoints if the study is positive. Results that are not significant are harder to interpret due to the fact that failure to prove an assumed effect may follow from lack of power or low quality of the study.

The intergroup comparisons of a randomized trial may be considered unbiased with a high level of confidence. This is even true if baseline differences between groups are observed. Randomization is expected to evenly distribute chances (success probabilities) between groups, not each of the baseline characteristics. Contrary to a widespread notion, the correctness of randomization cannot be checked empirically but only by checking of the randomization procedure.

Subgroups should be prespecified and never be analyzed separately but using interaction tests. Usually, the trial is not adequately powered for subgroup considerations.

**Limitations** The ability to draw generalized conclusions about the results crucially depends on the representativeness of the study population and may be limited due to a restrictive patient selection. RCTs are optimized for intergroup comparisons but not for extrapolation to routine health care in general. Analyses that were not prespecified and any kind of intragroup comparisons are not protected by randomization and may be biased to a relevant extent, although they are derived within the framework of an RCT.

#### 9.5.4.2 Cluster-Randomized Trials

**Motivation** Drug interventions usually address the individual patient. However, several patients may be cared for by one and the same health care institution or treated by the same physician. The drug intervention itself may even be provided within the framework of a program that includes several physicians and institutions and thus groups of patients. If so, patients within one institution or program tend to be more similar in their results than patients from different programs. This kind of effect is called a *cluster effect*. If cluster effects are present, it may be recommended not to randomize patients to drugs but to randomize institutions to drug interventions on the cluster level.

**Design** In a cluster-randomized trial, clusters (e.g., general practitioners, hospitals) are randomized to the study drugs, and all patients within a cluster receive the same drug according to the same schedule. Patients should be representative of their cluster, for example, by sequential inclusion or random selection. Patients within a cluster are allowed to communicate and exchange experiences.

In phase IV, cluster-randomized trials are typically applied when the introduction and propagation of a new drug is under study or when drugs are incorporated in prevention programs or general health programs.

### **Challenges**

1. A sufficient number of clusters (30 or above) have to participate. The required total number of patients frequently is a multitude (3–5 times) of the required sample size in trials with randomization of individual patients, depending on the extent of the cluster effect.
2. Institutions in the intervention group are confronted with a significant amount of effort and extra work. Institutions in the control group have only few advantages from participating in the study. Thus, recruitment problems are frequent. Special incentives for control group institutions may be helpful.
3. Clusters should not cooperate.
4. Blinding is frequently difficult or impossible.

### **Examples**

1. To study the clinical efficacy and the cost effectiveness of a rosuvastatin-based compliance program, general practitioners were randomized to take part with their patients in the program or not [5].
2. In a cluster-randomized trial, Coenen et al. [6] study the effect of a tailored peer intervention to implement a practice guideline for acute cough for optimizing antibiotic prescribing.

**Evaluation** In the presence of cluster effects, classical statistical evaluation methods are no longer valid, resulting in too many significances. Correct, but not very efficient, is an evaluation where cluster averages are calculated and compared between groups as if they came from individual patients. The best way of analyzing clustered data is the use of a multilevel model that allows the distinction between cluster-level effects and patient-level effects. Since relevant baseline differences are more frequent than in patient-based RCTs, due to the usually small number of clusters, adjustment for baseline is recommended.

**Interpretation** Cluster-randomized trials are less balanced and thus more vulnerable to biases than individually randomized trials. On the other hand, cluster-randomized trials are better capable to cope with real-world situations and thus may be more externally valid.

**Limitations** Recruitment problems, baseline imbalances, and differential drop-outs are frequent and may compromise the between-group comparisons or the



generalizability of the trial. Despite successful randomization, it may be difficult to maintain a comparable study conduct in all random groups due to external influences on institutions.

#### 9.5.4.3 Cross-Sectional Observational Studies

**Motivation** Hypotheses concerning effect modification or suspicions of adverse drug reactions are frequently gained from samples where drug exposure data, patient characteristics, and environmental factors are collected from patients at one point in time. These studies are quick and easy to perform and evaluate since only correlations have to be analyzed, but they do not provide evidence on drug efficacy and are just hypotheses generating. In phase IV, these studies provide first impressions on potential factors for optimization of use or on potential safety risks.

**Design** Planned observational study. Selection of a representative sample from a target population, collection of data from available sources, and one-time application of questionnaires.

**Challenges** Completeness of sample and records.

#### Examples

1. By means of a questionnaire, Coenen et al. [7] study the reasons for prescribing antibiotics in general practices. They explore the answers by factor analysis, identifying three latent dimensions, one of them being “nonmedical reasons.”
2. In an observational study nested in an RCT, Little et al. [8] studied the influence of patient pressure, perceived patient pressure, and perceived medical need on the investigations, referral, and prescribing decisions of general practitioners by logistic regression.

**Evaluation** Correlation and regression analyses, data-mining techniques. Frequently, a reduction of dimensionality is helpful, for example, by factor analysis or structural equation modeling.

**Interpretation** The study results in the demonstration of associations, that is, of coincidences that may or may not be substantial.

**Limitation** The numerical extent of associations depends heavily on the other variables in the model, and the interpretation is rarely unique. Associations thus depend on other associations and cannot be generalized. Even if the associations are stable over different models and times (if cross-sectional analyses are repeated), the associations do not provide evidence on cause–effect relationships. Likewise, missing associations do not prove the absence of cause–effect relations. Moreover, cross-sectional studies are not able to demonstrate individual courses or any changes with time. Thus, these kinds of studies can only be hypotheses generating and do not provide direct evidence and require additional studies with higher levels of evidence.

#### 9.5.4.4 Longitudinal Studies/Cohort Studies

**Motivation** The duration of phase III trials is usually limited to the minimum required to demonstrate clinical efficacy. Long-term efficacy and safety remain to be demonstrated in phase IV. While it is preferable to study long-term effects in RCTs, it is frequently not possible to maintain a randomization regime over many years, for economic, practical, or ethical reasons, the more so as the investigational drug has proven to be advantageous in phase III. A prospective study of drug effects can then be performed by building a well-defined cohort of patients who are routinely treated by the drug under investigation and by following the cohort for the period of drug intake. This cohort can then be compared to other cohorts, treated differently, from the same study or cohorts from other studies (“historical cohorts”).

In phase IV, cohort studies are commonly applied to evaluate long-term tolerance and safety but also to study the associations between changes in intermediate or surrogate parameters and clinical endpoints and to learn about long-term morbidity and prognosis and their determinants.

**Design** Cohort studies are planned epidemiological, purely observational studies. Groups of patients are recruited as for RCTs, but there is no random allocation of study drugs to patients. Instead, patients qualify for the study by the treatment (“exposure”) they receive in clinical routine. Data collection and follow-up are predefined and are performed with identical procedures for all patients, that is, additional follow-up examinations may be performed for study purposes.

**Challenges** Completeness of follow-up is a daunting task, in particular if the drug intervention is lifelong.

#### **Examples**

1. Simes et al. [9] use the cohorts of the LIPID trial on pravastatin to study the relationship between lipid levels and clinical outcomes on long-term intervention.
2. In order to study the safety profile of rosuvastatin beyond the analyses within phase II and phase III trials, Shepherd et al. [10] formed new cohorts from 27 phase II and phase III trials with different comparators, according to drug and dose ranges. Frequencies of adverse events per year of drug intake were calculated for each of these cohorts.

**Evaluation** While in RCTs adjustment for baseline is only optional to increase the power of the study, in cohort studies it is mandatory to minimize the selection bias resulting from lack of randomization. Repeated measurements require special statistical model building, called longitudinal data analysis.

**Interpretation** Treatment effects are usually quantified by regression coefficients. Effects may be time dependent and effect modifiers may be present. The interpretation of results of longitudinal data analysis may thus be involved and requires professional support.

**Limitations** Even with extended baseline adjustment or inclusion of time-dependent covariates, there is still a danger of hidden bias that cannot be controlled by probabilistic calculations. The evidence level of cohort studies is thus generally considered lower than that of RCTs. On the other hand, actual patient populations are usually better represented by cohorts than by RCT populations.

Cohort studies usually take a long time and thus are not appropriate to raise or to give quick answers to safety concerns. However, in the long run they may contribute substantially to the precise specification of risk and to the exploration of mechanisms.

#### 9.5.4.5 Case–Control Studies

**Motivation** Safety concerns, however substantiated, cannot wait for resolution by prospective long-term, follow-up studies such as RCTs or cohort studies. Retrospective studies are an alternative that can strengthen or mitigate the suspicion of a side effect in a shorter time interval. Case–control studies (directly comparing cases with and appropriate controls without the symptom suspected being drug related) and similar study types (such as case cohort studies where exposure and nonexposure time intervals are compared within the same patients) are widespread retrospective study arrangements.

In phase IV, case–control studies are widely used for the analysis of specified safety concerns (Saf-02, Saf-03, Saf-04, Saf-06, Saf-07) or for the study of the influence of environmental factors and lifestyle on drug effects (Opt-11, Opt-13); see Table 1.

**Design** Cases are systematically detected based on patients' records. Controls matched for other potential factors (e.g., age, sex, primary disease) are selected from a comparable population without symptoms. Drug exposure is compared between cases and controls.

**Challenges** The validity of the study decisively depends on the use of a comparable control group. Since there is usually no “natural” control group, the choice has to be justified, or several control groups should be used. Further, a bias-free selection of cases and controls has to be guaranteed. Since the exposure (and possibly some baseline characteristic that have to be determined for adjustment) has to be investigated for past periods of time, lack of documentation may compromise the study.

**Example** Suissa et al. [11] report the results of a case–control study comparing the risks of venous thromboembolism between oral contraceptives of the first and third generations. The data stimulated a fierce debate for years on the existence of an elevated risk in third-generation brands, several new studies, regulatory activities, and even litigation due to the lack of data from trials with a higher level of evidence.

**Evaluation** The standard statistical analysis requires the fit of a conditional (i.e., stratified) logistic regression model to the data.

**Interpretation** Drug effects are determined by adjusted odds ratios calculated from the final model that are directly comparable to odds ratios from prospective studies.

**Limitations** Case-control studies are vulnerable to any type of conscious or subconscious selection. They are discussed predominantly with respect to the appropriateness of the choice of controls. Results heavily depend on the choice of model. Thus, results are rarely unanimously accepted. Evidence is considered to be moderate and only supportive to other findings.

#### 9.5.4.6 Aggregate-Level Studies/Ecological Studies

**Motivation** Relevant data may not be available on an individual level but only for communities, states, or countries (e.g., prescription rates, frequencies of side effects, environmental exposure, costs for certain medical interventions). In these cases, analyses can only be performed on an aggregate level. In phase IV, this situation frequently occurs in the exploration of environmental effect modifiers, health economic studies, and studies on safety.

**Design** Collection of data from different aggregate data collections or linkage with individual data bases.

**Challenges** Linkage of different data sources requires the existence of comparable identifiers. Different databases may use different definitions of items that do not match. Selection criteria may vary. Data quality may vary.

**Example** Juurlink et al. [12] use aggregate-level prescription claims data and combine them with hospital records in order to analyze changes in the rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study (RALES). For this purpose they use time-series analysis methods.

**Evaluation** Statistical evaluation has to take the clusters and the resulting correlation structure into account. This can be done by using time-series analysis, Generalized Estimating Equations (GEE) models, or multilevel models, by meta-analyses or metaregressions, among others.

**Interpretation** Aggregate-level studies are purely observational. The interpretation follows the rules for patient-based cross-sectional studies; see above. Additionally, the different data levels have to be taken into account when interpretations are given.

**Limitations** In addition to the limitations of other cross-sectional studies, aggregate-level studies are vulnerable to the so-called ecological fallacy. It may occur that an association observed between clusters of individuals cannot be confirmed on an individual level or may even reveal an inverse association. Likewise, associations present on an individual level may not show up in the aggregate level analysis. In particular, trend parallelism may be spurious since any simultaneous trends will be correlated even if there is no causal relation, just by random

coincidence. Thus, while aggregate-level analysis are popular with politicians, media, and the public due to its handiness, they may be and frequently are extremely misleading, the more so as the control of other potential influences is usually poor.

#### 9.5.4.7 Medical Registers

**Motivation** Due to the rapid development of electronic documentation in medicine, routine data are available for most patients and clinical conditions. The quality of these data is improving since it is being used more frequently for the management and financing of health institutions. However, documentation standards still differ. Nevertheless, it has in the meantime become possible to routinely collect such data for certain groups of patients from comprehensive databases by defining interfaces and by regular data exchange, with a justifiable amount of effort. Such databases, called medical registers or pharmaco-economic databases, allow quick answers to questions arising on the use and the safety of several drugs at the same time. Registers are almost exclusively phase IV, useful for almost all of the objectives of Table 1.

**Design** Medical registers and pharmacoepidemiological databases try to give an image of a defined target population that is as complete as possible. Cross-sectional as well as longitudinal data are collected on an individual level. There will usually be no interventions, neither for experimental nor for observational purposes.

Registers require a protocol where the procedures are defined. However, other than in planned epidemiological studies, no research hypotheses are formulated in advance but will be defined according to the current state of debate, while further data are collected. If complete and high-quality data are available from routine databases, register population can even be defined retrospectively if organized by inception.

**Challenges** The validity of a register depends on its completeness and data quality.

**Example** Graham et al. [13] retrospectively define drug-specific inception cohorts of statin and fibrate users to study the incidence of hospitalized rhabdomyolysis.

**Evaluation** Registers and pharmacoepidemiological databases include cross-sectional and longitudinal elements and are thus similar to epidemiological studies, but broader and more exploratory than those approaches. The evaluation requires complex statistical modeling. Simple univariate or bivariate statistics usually are misleading.

**Interpretation** Registers are predestinated for the determination of rates and frequencies under changing conditions. They allow the study of associations but face the same problems as epidemiological studies.

**Limitations** Registers are purely observational. While their findings are most representative, they do not provide evidence for cause–effect relationships.

#### 9.5.4.8 Surveillance

**Motivation** Once a substantiated safety concern has been established, regular surveillance becomes mandatory. Surveillance is a typical phase IV safety activity.

**Design** Spontaneous reporting systems are the earliest type of surveillance. Other ways of organization are regular survey, sentinels (a selection of institutions that report the rates of interest on a regular base), and specific pharmacoepidemiological databases/registers on drug utilization. The use of Internet platforms is still experimental.

**Challenges** Traditionally, underreporting by physicians is the greatest challenge of surveillance. Considerable effort is required to improve participation. Independence of the responsible body, transparency, a careful practice of publication, and open access can serve to convince physicians of the usefulness of their contribution.

**Examples** Several reporting systems and routine databases are run by manufacturers, agencies, and pharmacological academic institutions.

**Evaluation** Prespecified safety concerns can be addressed with advanced statistical standard methodology. Extraction of unanticipated signals within huge databases is a continuing challenge. Data-mining techniques and in particular graphical tools are used, but no general methodological standards exist.

**Interpretation** The appraisal of surveillance data should take into account both aspects of statistical significance and those of medical relevance.

**Limitations** The perception of the data is usually not free from conflicts of interest. At present, no generally accepted rules exist as to when or which actions should be taken.

#### 9.5.4.9 Systematic Review, Meta-Analysis, and Metaregression

**Motivation** With the extended use of a drug after approval, the number of studies and the body of evidence increases rapidly. Information has to be summarized not only to facilitate an overview but also to realize the advantages in statistical power and differentiation that can be gained by combining data from different studies. Reviews and meta-analyses that summarize the results are recommended for all objectives in phase IV.

**Design** Reviews can be performed on a nonformal base as systematic reviews with a qualitative summary, or on a formal base with an established statistical meta-analysis technique with quantitative summaries. The formal analysis can be performed on the patient level if it is possible to combine the individual data of several studies into one database, or on the study level if only summary measures are available by using techniques of formal meta-analysis or metaregression. Every review and meta-analysis requires the prespecification of a protocol.

**Challenges** Completeness of studies/research results is difficult to attain due to limited access to studies not published in English and in established journals. Literature searches may also be incomplete due to the reluctance of journals to publish negative results (resulting in publication bias). Reporting standards may differ between studies to an extent that prevents pooling of results. Direct access to source data may be refused.

### **Examples**

1. The Control Treatment Trialists' Collaborators combined the data of 90,056 participants from 14 randomized statin trials to further evaluate the efficacy and safety of cholesterol-lowering treatment [14].
2. The Cochrane collaboration publishes systematic reviews and aggregate-level meta-analyses to several medical fields, provided by working groups of experts.

**Evaluation** Systematic reviews should follow the rules of the Cochrane collaboration for literature search and appraisal of the quality of the collected studies. For aggregate-level meta-analysis, an established statistical methodology exists. Random effects models are generally preferred to demonstrate potential heterogeneity between studies. Individual data-based meta-analyses should be analyzed by statistical modeling similar to the model building in epidemiological studies, but including the factor "study" to demonstrate heterogeneity.

There is an open debate as to what extent reviews and meta-analyses should rely on high-quality studies with a high level of evidence (RCTs) only or should include information of nonrandomized studies.

**Interpretation** Interpretation is usually based on overall summary measures, but heterogeneity should be considered as well.

**Limitations** Even if only high-quality RCTs are used, this does not automatically translate into a high-quality meta-analysis. Selection biases and an insufficient amount of data may limit interpretation. Aggregate-level meta-analyses and metaregressions may be compromised by the ecologic fallacy, as with individual aggregate-level studies, despite randomization of the contributing studies.

#### **9.5.4.10 Health Economics/Health Technology Assessment (HTA)**

**Motivation** Costs of a useful drug intervention may be too high to be financed in certain health care settings, or advantages of drug interventions may be judged negligible because the prices for this advantage may be found to be too high. Thus, health economic evaluations or the study of changes to be expected after the introduction of a new health technology are requested to a steadily growing extent.

**Design** Costs of interventions can be evaluated directly as part of a clinical study or determined separately and then combined with clinical data. HTA uses several data sources and combines them in probabilistic models based on Bayesian decision trees.



**Challenges** Costs originate at several levels, directly in connection with health care (direct costs) and indirectly due to the consequences of being ill or disabled. A complete evaluation of costs up to the regarded level requires the use of many sources of information. For cost-effectiveness calculations or HTA, a one-dimensional measure of efficacy is required. If several aspects of efficacy are to be taken into regard, a summary measure combining information on different scales has to be used. Quality-adjusted life-years [15] is such a concept that incorporates quality of life into the efficacy measure, but this kind of measure is not unchallenged.

**Example** Hirsch et al. [16] study the cost effectiveness of rosuvastatin as compared to other statins by retrospective drug cost evaluation for the patients in a head-to-head RCT comparing rosuvastatin to atorvastatin.

**Evaluation** Costs can be compared between study groups by classical statistical methods, but care is required since cost distributions are usually extremely skewed and variable. Baseline adjustment is thus recommended. Cost effectiveness is mostly measured by incremental cost-effectiveness ratios (ICER) that reveal additional variability since denominator and numerator are subject to random variation. Since costs may vary in different settings and time periods, it is good practice to add sensitivity analyses that demonstrate the sensitivity of the results to price changes or clinical improvements.

Health technology assessment predominantly uses Markov chain Monte Carlo (MCMC) methods for the calculation of probabilities along complex decision trees [15].

**Interpretation** Health economic analyses result in imprecise estimations of costs for improvements. HTA also gives an impression of the future distribution of outcomes in patient populations.

**Limitations** The results do not allow automatic decisions on the implementation of new technologies since the definition of boundaries is a question of values and political priorities. The selection of summary measures inevitably requires the definition of subjective weights that may not be accepted by everyone.

## 9.5.5 OPTIONS

Before performing a phase IV study, authorities, sponsors, and investigators have to make some decisions crucial to the impact of the results once the study is finished. The following represent some remarks on the decision criteria.

### 9.5.5.1 Data Sources

Studies can be performed using data from literature and completed studies or collecting own data. Actually, these two options are less an alternative than usually thought. A study should always start with a review of the present knowledge, including the knowledge to be expected from ongoing studies. If information is judged to be insufficient, existing databases that have not yet been evaluated should be

checked for accessibility, data quality, and potential impact of an analysis. New studies should be planned in a way that they optimally contribute to future reviews and meta-analyses. The development of strategic plans incorporating series of parallel studies, studies with different methodology, and performance of studies conditional on the results of previous studies are as recommended in phase IV as they were in phase III. However, it has to be remembered that in phase IV, other than before approval, the manufacturer is not the only potential sponsor of studies. Since the drug is licensed, several players can perform studies as long as they have a sufficient budget at their disposal. There may be conflicts of interest between different parties. Joint sponsorships resulting in linked databases of different origin may help to regulate such conflicts of interest in advance and certainly are interesting options for future phase IV activities.

### **9.5.5.2 Experimental versus Observational Approach**

Among methodologists there is a continuing debate whether phase IV studies on safety should be predominantly RCTs or observational studies [17]. As an example, the Cochrane Institute and many supporters of evidence-based medicine prefer RCTs and meta-analyses, attributing them the highest level of evidence since they feel that bias and spurious cause–effect relationships are to be feared most. Pharmacoepidemiologists refer to the lack of representativeness of many RCTs and argue that RCTs clarify safety concerns with an unacceptable time lag insufficient to guarantee patients' safety. The debate was intensified by recent withdrawals or black-box warnings of approved drugs due to new data on previously discussed potential side effects [1]. However, the debate should not be allowed to obscure the fact that the two approaches are not exclusive alternatives. In fact, experimental as well as observational studies contribute unique information on drugs. The question should thus rather be whether the amount of unbiased information and representative information is sufficient when a funding decision has to be met. What is more, both approaches are able to validate each other, the more so if the different study types are linked directly; an epidemiologic study may well be nested within an RCT; an RCT can more easily be performed if patients are recruited within a registry framework. Phase IV gives even more freedom than the premarketing phases to combine different activities into common projects.

### **9.5.5.3 Quality Assessment, Monitoring, and Auditing**

At present, the standards of phase III are not universally applied in phase IV, as far as data quality and control are concerned. The absence of regulatory requirements and the lower level of evidence on drug efficacy of some types of phase IV studies may have contributed to this situation. However, from a methodological as well as from an economical or political point of view, there is no reason why in a pharmacoepidemiological database on drug safety, for instance, a poor data quality should be more acceptable than in an RCT for optimization of use. The collection and evaluation of spurious data is per se not reasonable and not profitable and should be avoided, the more so as spurious findings, for example, on safety, can have a similar impact on public opinion to that of correct data can have.

#### 9.5.5.4 Registration of Studies, Access to Data, and Results

In phase III, trials are performed to gain knowledge about a drug and to convince agencies of the approvability of a new drug. Phase IV trials always have to address a broader audience. Not only agencies, but also physicians, patients, the scientific community, and the public have to be convinced as far as the virtue of a drug is concerned. For this purpose, it is essential that the audience trusts the results and has confidence in the proper performance of a study. Important milestones on the way to trustworthiness are the registration of all study activities to avoid the impression that unwanted results are covered up and open access to the data and the results once a study is finished.

#### 9.5.6 IMPACT OF PHASE IV TRIALS

Up to now phase IV trials and studies are frequently understood as voluntary exercises that may or may not be performed and do not need coordination or comprehensive development plans, special quality investments, transparency, or scientific evaluation. However, this attitude contrasts with the enormous impact that phase IV trials or the lack of these trials can have on the reception and use of a drug. If extended RCTs for the optimization of use are lacking, no safety studies or registries exist, economic consequences are not clear, and the relationship to competing drugs and other therapies is open. When all available data are locked in secret databases and reserved for the management of the manufacturing company, the position of a drug on the market is extremely vulnerable, and even random observations and rumors cannot be challenged due to a lack of data for scientific authorities who are able to clarify concerns. The performance of phase IV thus is in everybody's best interest.

The limited realization of phase IV studies may in part be due to the fact that the postapproval period is only sparsely regulated. However, there are more and more voices asking for a better regulation. Pharmacovigilance in particular is thought to be organized insufficiently and ineffectively. Rules are not clear, the predictability of legal decisions is low, and the twists of public attention and opinion are not predictable either. The only way to overcome the resulting risks is the production of solid data and transparent analyses that establish generally accepted facts upon which future work can rely.

## REFERENCES

1. Psaty, B. M., and Furberg, C. D. (2005), COX-2 inhibitors—lessons in drug safety, *N. Engl. J. Med.*, 352, 1133–1135.
2. Waxmann, H. A. (2005), The lessons of Vioxx—drug safety and sales, *N. Engl. J. Med.*, 352, 2576–2578.
3. Cannon, C. P., Braunwald, E., McCabe, C. H., Rader, D. J., Rouleau, J. L., Belder, R., Joyal, S. V., Hill, K. A., Pfeffer, M. A., and Skene, A. M. (2004), Intensive versus moderate lipid lowering with statins after acute coronary syndromes, *N. Engl. J. Med.*, 350, 1495–1504.
4. Shepherd, J., Blauw, G. J., Murphy, M. B., Bollen, E. L. E. M., Buckley, B. M., Cobbe, S. M., Ford, I., Gaw, A., Hyland, M., Jukema, J. W., Kamper, A. M., Macfarlane, P. W.,

- Meinders, A. E., Norrie, J., Packard, C. J., Perry, I. J., Stott, D. J., Sweeney, B. J., Twomey, C., Westendorp, R. G. J., and the PROSPER Study Group (PROspective Study of Pravastatin in the Elderly at Risk). (2002), Pravastatin in elderly individuals at risk of vascular disease (PROSPER): A randomised controlled trial, *Lancet*, 360, 1623–1630.
5. Willich, S. N., Müller-Nordhorn, J., Sonntag, F., Völler, H., Meyer-Sabellek, W., Wegscheider, K., Windler, E., and Katus, H. (2004), Economic evaluation of a compliance-enhancing intervention in patients with hypercholesterolemia: Design and baseline results of the Open Label Primary Care Study: Rosuvastatin Based Compliance Initiatives to Achievements of LDL Goals (ORBITAL) study, *Am. Heart J.*, 148(6), 1060–1067.
6. Coenen, S., van Royen, P., Michiels, B., and Denekens, J. (2004), Optimizing antibiotic prescribing for acute cough in general practice: A cluster-randomized controlled trial, *J. Antimicrobial. Chemother.*, 54, 661–672.
7. Coenen, S., Michiels, B., van Royen, P., Van der Auwera, J.-C., and Denekens, J. (2002), Antibiotics for coughing in general practice: A questionnaire study to quantify and condense the reasons for prescribing, *BMC Family Practice*, 3, 16–25.
8. Little, P., Dorward, M., Warner, G., Stephens, K., Senior, J., and Moore, M. (2004), Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study, *BMJ*, 328, 444–447.
9. Simes R. J., Marschner, I. C., Hunt, D., Colquhoun, D., Sullivan, D., Stewart, R. A. H., Hague, W., Keech, A., Thompson, P., White, H., Shaw, J., and Tonkin, A. (2002), Relationship between Lipid Levels and Clinical Outcomes in the Long-Term Intervention with Pravastatin in Ischemic Disease (LIPID) Trial, *Circulation*, 105, 1162–1169.
10. Shepherd, J., Hunninghake, D. B., Stein, E. A., Kastelein, J. J. P., Harris, S., Pears, J., and Hutchinson, H. G. (2004), Safety of rosuvastatin, *Am. J. Cardiol.*, 94, 882–888.
11. Suissa, S., Blais, L., Spitzer, W. O., Cusson, J., Lewis, M., and Heinemann, L. (1997), First-time use of newer oral contraceptives and the risk of venous thromboembolism, *Contraception*, 56, 141–146.
12. Juurlink, D. N., Mamdani, M. M., Lee, D. S., Kopp, A., Austin, P. C., Laupacis, A., and Redelmeier, D. A. (2004), Rates of hyperkalemia after publication of the randomized aldactone evaluation study, *N. Engl. J. Med.*, 351, 543–551.
13. Graham, D. J., Staffa, J. A., Shatin, D., Andrade, S. E., Schech, S. D., Grenade, L. L., Gurwitz, J. H., Chan, K. A., Goodman, M. J., and Platt, R. (2004), Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs, *JAMA*, 292, 2585–2590.
14. Cholesterol Treatment Trialists (CTT) Collaborators (2005), Efficacy and safety of cholesterol-lowering treatment: Prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins, *Lancet*, 366, 1267–1278.
15. Friedland, D. J., Go, A. S., Davoren, J. B., Shlipak, M. G., Bent, S. W., Subak, L. L., and Mendelson, T. (1998), *Evidence-Based Medicine—A Framework for Clinical Practice*, Prentice Hall, London.
16. Hirsch, M., O'Donnell, J. C., and Jones, P. (2005), Rosuvastatin is cost-effective in treating patients to low-density lipoprotein-cholesterol goals compared with atorvastatin, pravastatin and simvastatin: analysis of the STELLAR trial, *Eur. J. Cardiovasc. Prevention Rehab.*, 12(1), 18–28.
17. Okie S. (2005), Safety in numbers—Monitoring risk in approved drugs, *N. Engl. J. Med.*, 352, 1173–1176.



# 9.6

---

## *Phase IV and Postmarketing Clinical Trials*

**Ali Miraj Khan**

### **Contents**

|          |                                                                                              |     |
|----------|----------------------------------------------------------------------------------------------|-----|
| 9.6.1    | Introduction                                                                                 | 326 |
| 9.6.2    | Phase IV Clinical Studies                                                                    | 327 |
| 9.6.3    | Postmarketing Surveillance                                                                   | 328 |
| 9.6.4    | Adverse Drug Reactions                                                                       | 328 |
| 9.6.4.1  | Methods for Monitoring Adverse Drug Reactions                                                | 329 |
| 9.6.5    | Pharmacovigilance and Postmarketing Surveillance                                             | 330 |
| 9.6.6    | PMS: Linkage System and Data Bank                                                            | 331 |
| 9.6.7    | Postmarketing Surveillance: FDA Perspective                                                  | 332 |
| 9.6.8    | Drug Regulation and the FDA                                                                  | 332 |
| 9.6.8.1  | FDA: Globalization and Recent Activities in Drug Safety                                      | 333 |
| 9.6.9    | Postmarketing Surveillance: Australia Perspective                                            | 333 |
| 9.6.10   | Confirmation and Quantification in Denominator-Based Systems                                 | 334 |
| 9.6.11   | United Kingdom Prescription Event-Monitoring (PEM) Program                                   | 335 |
| 9.6.12   | Adverse Drug Reactions and Resource-Poor Countries                                           | 335 |
| 9.6.12.1 | Pharmacoeconomic Impact of Drug Toxicity in Developing Countries                             | 336 |
| 9.6.12.2 | Basic Requirements for Effective Postmarketing Surveillance Program in Resource-Poor Country | 336 |
| 9.6.13   | WHO and Other Authorities for PMS                                                            | 337 |
| 9.6.13.1 | Safety Surveillance of Antiretroviral Drug and RaPID Program                                 | 338 |
| 9.6.13.2 | Baseline Program in Hospitals of Resource-Poor Countries                                     | 338 |

|           |                                                                                                   |     |
|-----------|---------------------------------------------------------------------------------------------------|-----|
| 9.6.13.3  | Possible Approach to Develop Drug Safety Monitoring Program for Hospital in Resource-Poor Country | 339 |
| 9.6.13.4  | Inference for Future Buildup in Poor Countries                                                    | 339 |
| 9.6.13.5  | Data Mining in Postmarketing Surveillance                                                         | 340 |
| 9.6.13.6  | Spontaneous Reporting Databases                                                                   | 340 |
| 9.6.13.7  | Prescription Event-Monitoring Databases                                                           | 340 |
| 9.6.13.8  | Linked Administrative Databases                                                                   | 340 |
| 9.6.13.9  | Electronic Medical Records                                                                        | 340 |
| 9.6.13.10 | Other Databases                                                                                   | 341 |
| 9.6.13.11 | Data Preprocessing                                                                                | 341 |
| 9.6.13.12 | Application of Data-Mining Techniques in PMS                                                      | 341 |
| 9.6.14    | Discussion                                                                                        | 342 |
| 9.6.15    | Future Perspective                                                                                | 342 |
| 9.6.16    | Conclusion                                                                                        | 344 |
|           | Bibliography                                                                                      | 344 |

## 9.6.1 INTRODUCTION

Pharmaceutical products are developed and marketed to promote human health and protect people from various ailments. The identification or development of drugs involves a lot of time and resources, and to bring drugs to the marketing stage involves assessment of different aspects of the product in a very scientific and credible manner. Yet, the assessment of the products does not end just at marketing; assessment of drug effectiveness and, particularly, their safety continues, constituting phase IV clinical trials.

Typically, development of drugs involves several stages of experimentation in animals followed by trials in humans before large-scale phase IV clinical assessments. Animal experimentation precedes trials in humans and are intended to provide information on the acute toxicity, organ damage, dose dependence, absorption and bioavailability, the pharmacokinetics and dynamics, the metabolism, mutagenicity, teratogenicity, species specificity, and lethal dose. After animal experimentations providing encouragement for further development of the drug, clinical trials are planned that typically have three stages: phase I, phase II, and phase III conducted sequentially.

The phase I clinical studies involve a small number of healthy volunteers that allows investigators to understand the effects an investigational drug has in humans, with the goal to understand what happens to the investigational drug in the body from the time it is administered until its elimination from the body, on how human bodies react to the new drug from a safety and tolerability point of view. Study participants are monitored for the occurrence and severity of any side effects they may experience.

Phase II clinical studies are designed to evaluate the safety and efficacy of an investigational drug in patients with a specific disease or condition. Phase II clinical trials are typically conducted in a group of patients who are usually selected based upon their being at the same stage of a disease. The patients are given various doses of the drug and closely monitored to compare the effects and to determine the safest dosing regimen. In many instances, multiple phase II clinical studies are conducted to test the drug in a variety of patient populations with different indications.



Phase III clinical studies are designed to confirm the safety and efficacy of an investigational drug, and the dosage regimen chosen, in large numbers of patients with a specific disease or condition. These studies, as in the earlier phases, may involve one or more “treatment arms,” which allow the investigational drug to be compared to other available treatments or to be tested for effectiveness in combination with other therapies. The safety and efficacy of the new drug is compared to that of the currently accepted standard treatment. Information obtained from phase III studies is used to determine how the drug is best prescribed to patients in the future. The complete information available on the new drug is submitted to health authorities. But, for safety reasons, these clinical trials of drugs are designed to exclude women of childbearing age, pregnant women, and/or women who become pregnant during the study. In some cases the male partners of these women are also excluded. Also children and very aged persons are frequently excluded from these clinical studies.

The purpose of the next phase, that is, phase IV clinical study, is to increase knowledge surrounding use of a new drug by increasing the number of humans that provides information on its effectiveness and more importantly adverse events when rolled out onto the market. “There are no therapeutic roses without thorns.” Thus, the general use of a particular drug depends on availability of alternatives, the determination of the frequency and severity of adverse effects associated with the new drug, and finally on risk–benefit and cost–benefit assessments. Approval for marketing of a drug is usually granted based on adequate efficacy and safety, as demonstrated through well-designed phase I to phase III clinical trials.

It is, however, important to remember the limitations of these phases, which are conducted on a smaller number of humans and, thus, do not represent real-life situations. They generally represent only a limited population with specific characteristics and thus underrepresent the greater patient populations and provide short-term outcomes. A number of issues are monitored after marketing of drugs to determine their performance as they are prescribed in increasing numbers: Will the drug provide equal or greater benefit to people compared to other available drugs on the market? Will this be more cost effective? Will its safety profile be equal to or better than the alternatives? Will it provide better compliance?

### 9.6.2 PHASE IV CLINICAL STUDIES

Phase IV clinical trials involve the surveillance of the safety and adverse reactions of drugs in the market. With progression of medical services, the need for phase IV clinical trials is definitely greater now. Phase IV clinical trials encompass the detection, assessment, understanding, and prevention of drug-related problems. Both regulatory authorities and the sponsoring company may require phase IV studies for competitive purpose to find a new market for the drug. The safety surveillance is designed to detect any rare or long-term adverse effects of drugs over a larger patient population and for a longer time period than was possible during the phase I–III clinical trials. Withdrawal of a drug from the market or restricted to certain uses may be recommended if harmful effects are discovered by phase IV clinical trials.

Pharmacoepidemiology is the primary scientific discipline now engaged in postmarketing drug safety surveillance. The most common epidemiologic study designs used in phase IV studies are randomized trials, cohort studies, and case-control studies. Regardless of their design, all studies involve human subjects and these studies need to be conducted by qualified investigators according to a written protocol duly reviewed and approved by an institutional review and ethics board. Promotional activities conducted under the guise of postmarketing studies are never acceptable. Changing regulatory environment, increasing concerns about the safety of new drugs, and varied uses for large-scale, real-world data on marketed drugs' safety and efficacy are primary drivers of the growth seen in the phase IV study environment today.

Postmarketing study is an important element of commercialization that enables companies to expand existing markets, enter new markets, develop and deliver messages that directly compare their products with the competition, and secure a better position in competitive markets.

### 9.6.3 POSTMARKETING SURVEILLANCE

Management of patients using drugs on a long-term as well as a short-term basis raise several complex issues such as toxicity, interaction with other drugs, as well as efficacy and effectiveness. In addition, age, sex, nutritional status, disease duration, comorbidity, ability to tolerate adverse events, and socioeconomic status are important considerations that influence the outcome of individual patients taking drugs. Randomized clinical trials alone are not sufficient to address all of these issues. Multiple therapies with attendant adverse effects are also common. Thus, postmarketing surveillance is important to accumulate information of adverse effects, influence of personal attributes mentioned above, identification of risk factors, quantitative measurement of safety and long-term safety/toxicity, study of potential risk groups (children/elderly, pregnant women, etc.), and identification of new indications and unexpected beneficial effects. It, additionally, provides useful information on the patterns of use and characteristics of drug users, inappropriate use of drugs, for example, addiction, impact on compliance, the possibility of medication errors and intoxication, quality of life during drug use, and utility and cost assessments.

### 9.6.4 ADVERSE DRUG REACTIONS

Drugs are intended to improve and make life comfortable for humans. But, they definitely are not without risk and may cause lesser or greater harm to many people. The unintended effects of drugs are harmful to the recipients and generally are referred to as adverse drug reactions (ADRs). Classical examples of a few ADRs are provided in Table 1.

Adverse drug reactions might account for over 10% of all hospital admissions in some countries. For example, a prospective study conducted in a European country reviewed the cause of hospital admissions of 18,820 patients over a 6-month period and noted 6.5% of admissions to be related to an ADR. The projected annual cost of such admissions to the National Health Service of that country was very high.

**TABLE 1 Adverse Drug Reactions**

| Drug                  | Adverse Drug Reaction    |
|-----------------------|--------------------------|
| Phenyl butazone       | Aplastic anemia          |
| Chloramphenicol       | Pancytopenia             |
| Ethambutol            | Optic atrophy            |
| Erythromycin estolate | Hepatitis                |
| Aspirin               | GI hemorrhage            |
| Steroid               | Hyperglycemia            |
| Oral contraceptives   | Thromboembolism          |
| Clindamycin           | Pseudomembranous colitis |
| Reserpine             | Depression               |
| Co-trimoxazole        | Agranulocytosis          |

According to another survey published in the *Journal of American Pharmacists' Association* in 2001, the cost of identified drug-related morbidities and deaths in a North American country was much higher than expected. Costs associated with drug-related problems have been estimated to be greater since 1995. From developing countries, such information is not readily available, but it is believed that the incidence of adverse events is greater with considerable cost.

#### 9.6.4.1 Methods for Monitoring Adverse Drug Reactions

There are several methods that can be employed for evaluation and monitoring of adverse drug reactions. Selection of an appropriate method depends on factors such as program objectives and setup and the availability of human resources. A few common methods to evaluate ADRs include:

1. *Case Reports* The publication of single case reports, or a series of cases, on adverse drug events in the medical journals is an important way of compiling unreported and serious reactions.
2. *Anecdotal Reporting* Individual medical practitioners through anecdotal reports provide the majority of first reports of adverse drug reactions. They describe the occurrence of an event in association with use of a particular drug. Because of the anecdotal nature, such events need to be verified; however, confirmation may not be possible.
3. *Spontaneous Reporting* It is the principal continuing method for monitoring safety of marketed drugs. Along with the use of other methods, ADR programs are still based on spontaneous reporting system in many countries, both developing and developed. Clinicians are encouraged, and in some countries physicians are specially instructed, to report any or all reactions associated with medication. Usually, attention is focused on newly marketed drugs. Reporting of ADRs of a serious nature has also been made mandatory in many developed countries. This system is simple to operate and usually does not interfere with clinical practice to any great extent. Absence of an adequate control group is the main drawback of spontaneous reporting and, consequently, the difficulty of knowing the relative risk of treated patients. Despite limitations, voluntary spontaneous reporting systems remain the most

cost-effective approach for postmarket identification of new ADRs. Here the purpose is to register new problems or the occurrence of an increased incidence of previously reported problems. However, underreporting is one of the major limitations in the spontaneous reporting system. Despite regulations, health care professionals report only a variable proportion of adverse drug events with temporal association with a drug.

4. *Intensive Event Recording* In certain medical setups a group of people are given the responsibility to scrutinize a defined population specifically to detect adverse reactions associated with specific drugs. However, because of shorter periods of observation of individual patients and a relatively small number of people targeted in such studies, this method has not been very effective in the detection of drug toxicity.
5. *Cohort Studies (Prospective)* This is a very useful epidemiological method in which a cohort with a particular attribute (e.g., recipients of a drug) is followed prospectively and compared for some outcome (e.g., toxicity of the drug) with another cohort not possessing the attribute. Good control of data, accurate checking of outcome, good estimation of risk, study of many different outcomes, etc., are advantages of this method. Disadvantages are high cost and long wait.
6. *Case-Control Studies (Retrospective)* This is a good epidemiological study design. In this type of study the investigator selects the case group and the control group on the basis of outcome—i.e., having drug-related toxicity of interest versus not having drug-related toxicity of interest—and compares the groups in terms of their frequency of past exposure to possible risk factors. Case-control studies are relatively inexpensive, can be done quickly, and are a frequently used type of epidemiological study that can be carried out by a small team. Its retrospective nature limits the conclusions.
7. *Case-Cohort Studies* The case-cohort study is a recently developed useful modification of the case-control study. It is a hybrid epidemiological design. This design allows direct estimation of the risk ratio from a fixed cohort.
8. *Meta-Analysis* This is a quantitative analysis of two or more independent studies (already carried out) to determine an overall effect and to describe reasons for variation in study results. Meta-analysis aims to establish associations between drugs and adverse events, to estimate the frequency of ADRs, and to identify subgroups at increased risk for ADRs.
9. *Use of Population Statistics* Birth defect registers and cancer registers can be used if a drug-induced event is very frequent. If such assessment raises suspicion, then further observational studies are initiated.

### 9.6.5 PHARMACOVIGILANCE AND POSTMARKETING SURVEILLANCE

Pharmacovigilance is a relatively new and emerging science that is expected in the future to serve the purpose of postmarketing surveillance (PMS), but now it is not a well-established academic activity. Although a few universities offer specific courses in pharmacovigilance, the current curricula for training programs of profes-

sions in clinical medicine, clinical pharmacy, clinical pharmacology, and medical biology are not geared to provide all of the skills needed in this discipline. One of the reasons could be that pharmacovigilance involves a wide range of subjects, such as pharmacology, epidemiology, clinical medicine, data management, and drug legislation and communication, and they do not easily fit within the area of competence of any of the existing academic departments. It is anticipated that pharmacovigilance will be taught as an academic subject at the university level in the near future. Currently, the International Society of Pharmacovigilance (ISoP) provides specialized ad hoc training courses in pharmacovigilance. Each country might establish such an organization to provide a stronger base in effective and credible training services for relevant individuals/group of individuals.

### 9.6.6 PMS: LINKAGE SYSTEM AND DATA BANK

Many countries and related organizations involved in health care activities either already have or have the capacity to establish comprehensive databases using modern technologies at reasonable expenditure. Existence of such databases would immensely help establish a comprehensive and cost-efficient system for pharmaco-epidemiology. With the multicenter nature of such databases, postmarketing surveillance programs have new possibilities. A wide spectrum of data is covered by automated databases of general practice and, therefore, at the population level it provides opportunities for a closer look at the impact of medications. However, special methods need to be established for determining the abuse potential of non-prescription drugs, such as many analgesics, in order to determine the extent of abuse and the temporal changes over periods since marketing a new drug. For nearly last two decades, automated medical record linkage systems have been utilized to evaluate postmarketing drug safety. Recently, for adequate assessment, networks of databases have been employed to identify problems that require large or diverse populations. Despite the potentials, establishment of such an effective system might not be easy. Because it involves merging of data from different systems, their reliability, variations in coding conventions for drug formularies, and the need to protect the confidentiality of medical records are likely to make the task difficult.

The European Active Surveillance Study (EURAS) is an ideal example of record linkage for postmarketing surveillance. This is a prospective, controlled, multinational, postmarketing, nonintervention cohort study of users of oral contraceptives. The study aims to detect the incidence of rare serious adverse side effects associated with the use of new and old oral contraceptives, particularly to determine incidence of thromboembolic events. This surveillance is under routine medical practice in European countries. The study is based on postal questionnaires and semiannual, active follow-up with validation of reported events by the physicians treating those women under surveillance. To ensure low loss to follow-up rates, a multifaceted, four-level follow-up procedure has been established.

Before the introduction of automated databases, studying large cohorts of drug users was extremely expensive and required major administrative and financial efforts to assemble and ensure complete and accurate collection of the required follow-up information. The computerized information system, initiated in the late

1970s, facilitated all these efforts and allowed highly efficient research. In the United Kingdom, the general practitioner (GP) traditionally acts as a gatekeeper to services within the National Health Service (NHS). A comprehensive record of written prescriptions, outpatient diagnoses, and referral letters to hospitals resides in the individual patient's record that GPs maintained in their offices. Since the early 1990s, the database has been known as the General Practice Research Database (GPRD), belonging to the UK Department of Health, and currently it is maintained by the Medicine Control Agency (MCA). Truly population-based data is collected by GPRD, which includes both outpatient and inpatient clinical information with a size that makes follow-up possible on large cohorts of users of specific drugs.

### **9.6.7 POSTMARKETING SURVEILLANCE: FDA PERSPECTIVE**

Drug review process of the U.S. Food and Drug Administration (FDA) is recognized worldwide as a gold standard for the approval of drugs. To improve the standard, there have been significant additions during the last several decades in response to advances in medical science. Recently, drugs are approved by the FDA after comprehensive assessment, including safety evaluations. Major changes have taken place in the evaluation of new drugs, including complete understanding of their metabolism, interactions with other drugs, and potential differences of effectiveness or safety in people of different age groups, sex, and races. Focusing on the potential problems with the greatest care and importance, internal guidelines are in place to describe systematic approaches for assessment of safety through comprehensive review. Approval of drugs are granted by the FDA after a sponsor demonstrates risk–benefit assessment favoring use in specific population(s) and that the drug meets the standard for accepted safety and efficacy. It is conceivable that premarketing studies will never elucidate all the information about the effectiveness or all the risks from use of a new drug. The FDA recognizes that it is not possible to assess or anticipate all possible effects of a drug from the clinical trials that precede approval. So, the FDA has created a postmarket drug safety program designed to collect and assess adverse events following the drug's approval and marketing with an aim to detect unexpected serious adverse events and to take required action(s). The FDA collects information from adverse event reports filed by drug manufacturers, postmarketing clinical trials, and spontaneous reporting of adverse events by practicing physicians, pharmacists, and consumers.

### **9.6.8 DRUG REGULATION AND THE FDA**

In the United States, both the prescription and over-the-counter drugs are essential, and their numbers are increasing. Newly marketed drugs are saving lives, reducing suffering, and improving the quality of life of millions, as do older drugs. Despite rising costs of health care, innovative newer medicines are necessary to improve life expectancy. Also, drugs are used to prevent or slow the progress of many diseases, thus avoiding costly invasive treatments, hospitalizations, and nursing home stays. However, there are undesirable consequences by widespread use of drugs. Inherent

side effects, misuse, overuse, underuse, medication mix-ups, and intentional abuse are increasing. So, appropriate monitoring and addressing of all these problem are required. Thus, the responsibilities for oversight of the entire life cycle of drugs has been undertaken by the FDA. It includes premarket testing and development through drug approval, postmarket surveillance, and risk management. The objective of the FDA is to ensure that safe and effective new drugs are available as quickly as possible and that drugs already on the market remain safe and maintain the highest quality. The FDA maintains its efforts to translate new scientific advances into benefits for patients and to take advantage of new developments in monitoring the performance of marketed drugs. The FDA and its Center for Drug Evaluation and Research (CDER) dedicate at least half of their premarket review activities to evaluating the safety of investigational drugs and safety of clinical trial participants. The FDA's Office of Regulatory Affairs (ORA) support the premarket drug review process by monitoring clinical trials and their results.

#### 9.6.8.1 FDA: Globalization and Recent Activities in Drug Safety

At every stage of the drug life cycle, the FDA is working to improve the mechanism that supports drug safety and risk management. With the evolution of science over drug development, related activities have been expanded, and the focus is now on biomarkers and pharmacogenomics, which are likely to help maximize the safe and effective use of drugs in individual patients. More scientific methods and new tools are now in the works to target a specific drug for specific patients to maximize benefits and minimize associated risks. One of the current important objectives is to prevent adverse events by predicting drug safety problems before they can cause an insult. The aim is to combine new understandings of disease and its origins at the molecular level, including adverse events resulting from drugs, with the emerging knowledge of genetic and biologic characteristics of individual patients that determine how they will react to drug treatment. If successful, these innovative and risk-assessment-based drug development processes will improve FDA's efficiency for early disclosure of safety-related problems. Moreover, the FDA is now exploring, testing, and developing new methods of data mining, signal detection, and analysis of electronic data related to global health care.

The FDA is now very aware of the globalization trends. To respond to this global challenge in a proactive manner, the FDA has developed a variety of solutions. The FDA has entered into a number of cooperative relationships with foreign regulators to formalize arrangements for advancement of drug quality. The FDA is continuing work to encourage greater transparency and to ensure that patients and health care providers have access to emerging new safety information.

#### 9.6.9 POSTMARKETING SURVEILLANCE: AUSTRALIA PERSPECTIVE

In Australia, guidelines have been applied in industry-sponsored postmarketing surveillance studies on drug safety, and most are observational studies that used cohort design. The guidelines, however, apply for other types of studies, including case-control studies and in intensified monitoring and various forms of recorded release. The term *postmarketing surveillance (PMS) study* need not mean a mere



passive follow-up, but actually it includes scientifically rigorous studies of products that are approved for registration and marketing, and its purpose is to collect reliable information on drug safety. However, they are not clinical trials of registered products or studies designed primarily for marketing purposes despite their scientific validity. PMS studies are generally initiated by the producers but might also be requests from other parties. Their aim is very specific—drug safety question or testing a hypothesis that often results from voluntary reporting. The guidelines for postmarketing surveillance studies have been developed to design and conduct scientifically rigorous and ethical studies to maximize validity of the results and increase potentials for acquiring useful newer information on the product. The ultimate purpose of a PMS study is to find an answer to a specific medical question that is new, significant, and unlikely to be answered by existing voluntary reporting systems.

In company-sponsored PMS studies that are typically observational rather than experimental, the decision to enroll a patient needs to be made after making an independent clinical decision to prescribe a specific drug concerned a particular patient. Prior to commencement, study proposals are sent to the Australian Drug Regulatory and Advisory Council with clear statements with regards to its aims and objectives, the question of clinical significance under investigation, proposed methodology including data collection and analyses procedures, and identification of company officially designated as responsible for the study. The notification of adverse reactions occurring during the PMS study should be done according to the pharmacovigilance guidelines.

The sponsor company is responsible to provide a timely report on the study outcomes. The product to be studied must have been approved for registration in Australia, and only patients having an approved indication for use of the product can be enrolled in PMS studies. Company-sponsored PMS studies should not be a disguised marketing or promotional exercise. Any payment offered to the medical professional must be commensurate with the work involved. The guidelines describe scientific approaches that the pharmaceutical companies are encouraged to follow in designing and conducting the postmarketing surveillance studies.

#### **9.6.10 CONFIRMATION AND QUANTIFICATION IN DENOMINATOR-BASED SYSTEMS**

A number of epidemiological approaches have been developed for postmarketing surveillance and quantification. Some of them have control groups to make relative risk estimates. Population-based claims databases are large databases that can link prescription-dispensing information (e.g., drug, dose, duration, and date) to claims data for medical interventions via unique patient identifiers. These systems can generate event rates (numerators) for the denominator of exposed persons and person-years of exposure and allow comparison of rates observed in unexposed patients. Here, the limitations of the systems are: ability to record the events of interest and the reliability. The availability of medical records in some of these systems allows validation of event reports and more in-depth analysis of adverse events, making formal nested case-control studies possible. However, in the event of very rare events these databases, despite being large, might still be relatively small for acquiring adequate numbers of exposures to the drugs of interest.

### 9.6.11 UNITED KINGDOM PRESCRIPTION EVENT-MONITORING (PEM) PROGRAM

In the UK, all prescriptions for newly marketed drugs are entered into an automated database after retrieval from the British Prescription Pricing Authority. Prescription data provided to the Drug Safety Research Unit (DSRU) include the name of the patient and the prescribing general practitioner (GP). Subsequently, each prescribing GP received the questionnaires (“green cards”) 6 months after the first prescription for each of the patients. Questionnaires request adverse events during and after treatment, reason(s) for discontinuation, if any, patient age, indication for treatment, starting and stopping dates of treatment, and dosage. They are tailored for individual drugs and additional questions might also be incorporated.

Exposure-specific registry is another cohort technique employed that attempts to identify and follow patients who have been exposed to a drug but have not yet had an adverse outcome (e.g., hypertension treatment registry created before development of side effects). These registries may also be constructed on diseased populations (e.g., patients with gluten-induced enteropathy). Availability of denominator is a major advantage in these systems. Registries may provide also background rates for unexposed or differently exposed patients. Registries also might facilitate more complete follow-up of incident cases. These resources have also limitations common to other databases. Rates are provided for reported events without any control group. Limitations also include the imprecise nature of data linkage of drugs to events, and the temporal relationships between drug and events sometimes are not well established. In most claims data, unless an event leads to hospitalization, the event is not captured. Sometimes, very indirect methods are adopted, e.g., some may capture office visit events and often use of a drug is taken as an indicator of an event. In addition, claims interpreted as representing adverse drug reaction might not actually reflect the actual clinical event or an outcome of interest (e.g., not all claims of “convulsions” in such databases are actually episodes of epilepsy, and similarly, syncope, vasovagal attacks, etc., are often found events). These are limitations that prevent right quantification of the incidence of drug-related adverse reactions.

### 9.6.12 ADVERSE DRUG REACTIONS AND RESOURCE-POOR COUNTRIES

The type A reactions include normal and augmented response to drugs and are dose related; reactions are usually predictable and preventable on the basis of known pharmacology of drugs. The incidence of type A reactions is greater and usually associated with higher morbidity. It may be managed to overcome or prevent such reactions by reducing the dose, e.g., bradycardia associated with  $\beta$ -adrenoreceptor blockers and bleeding in association with anticoagulants.

The type B reactions are usually unpredictable and are often caused by immunological and pharmacogenetic mechanisms. The type B reactions are not dose-dependent, are relatively rare, often not predictable and preventable, and sometimes cause serious morbidities and death. Examples of this kind of reactions include malignant hyperthermia caused by anesthetics and many immunological reactions. It is difficult to classify other adverse drug reactions.

In resource-poor countries, frequent use of complementary medicines might interact with conventional allopathic drugs. Thus it may predispose to adverse drug

reactions, factors that need to be considered in their interpretation. A majority of drugs used in resource-poor countries are usually imported from other countries. The ingredients used in drugs may vary and the genetic makeup of the population might also vary, thus may predispose to adverse drug reactions. These factors are not usually verified among populations of resource-poor countries before approval of these drugs. Therefore, the risk of occurrence of drug-related toxicities in such populations could be very high and, in fact, remains unknown.

Due to geographic variations, poor socioeconomic condition, high cost of modern medicines, and nonavailability of physicians in most of the rural areas, the people of developing countries face greater difficulties in accessing standard health care. Drug retail shops frequently serve as the first point of contact of population for seeking health care, thus predisposing the poor people to be more vulnerable to self-medication, a potential contributor to adverse effects of drugs. Moreover, poly-pharmacy is a very common practice in resource-poor countries, which increases the incidence of adverse drug reactions. It has been demonstrated that drug toxicity sharply rises with the number of drugs consumed.

Age is another very important factor for development of adverse drug reaction—the very old and the very young are more susceptible to adverse effects of drugs due to compromised physiological reserve. The aging population is more likely to have cardiovascular, pulmonary, musculo-skeletal, and metabolic disorders that necessitate several medications; and all these predispose them to increased incidence and increased severity of drug-related toxicities. Patients with impaired kidney and liver functions and those with intercurrent illness are at greater risk for adverse drug reactions. In poor countries these are not meticulously considered by the doctors while prescribing drugs because of limitations of laboratory tests. Race genetic polymorphism and hereditary factors are known to affect the bioavailability of drugs and thus are common important contributing factors for adverse drug reactions in poor regions. Also, in poor countries, the effect of drugs may vary from place to place, depending on climatic diversity or different cultural and social backgrounds.

#### **9.6.12.1 Pharmacoeconomic Impact of Drug Toxicity in Developing Countries**

The economic impact of adverse drug events is less explored and less reported in developing countries. Adverse drug events prolong hospital stay, result in new admissions, and increase treatment cost leading to increased burden on the health care systems in resource-poor countries. It has been observed that the total cost of drug-related morbidity and mortality is exceedingly high, which exceeds the cost for treatment of the primary illness. Therefore, ADR-related admissions in developing countries with limited health care budget stretched that even further.

#### **9.6.12.2 Basic Requirements for Effective Postmarketing Surveillance Program in Resource-Poor Country**

The minimal requirement for the establishment of a working program for postmarketing surveillance of drugs requires infrastructure, funding, expertise, and a trained staff. For resource-poor countries, there may be shortcomings in all of these areas.

However, some useful activities can be established with limited financial resources if sincere efforts are undertaken. It is possible to initiate a program for concurrent surveillance during drug therapy, based on reporting of suspected adverse drug reactions by pharmacists, physicians, nurses, and patients at an affordable cost. Alternatively, a prospective surveillance before drug therapy for patients with a high risk for adverse drug reactions may be initiated.

Doctors or drug prescribers should be adequately notified regarding the drug to be monitored for adverse reactions. Information regarding suspected adverse drug reactions should be reported to the pharmacy for complete collection of data and their analyses, including patients' name, their medical and medication histories, a description of suspected ADR, temporal relationship of the event with drug administration, any remedial treatment required, and sequel and outcome. The cause(s) of each suspected adverse drug reaction need to be evaluated based on the patient's medical and medication histories, the circumstances of the adverse event, the result of de-challenge and re-challenge (if any), consideration of alternative etiologies, and a literature review. Description of each of the suspected adverse drug reactions and their outcomes needs documentation in patients' medical records.

Serious or unexpected adverse drug reactions should be reported to the drug regulatory/health authority and the drug's manufacturer. All adverse drug reaction reports should be reviewed and evaluated by the authorized pharmacy and therapeutics committee. Adverse drug reaction report information should be disseminated to health care professionals for educational purposes. In all of these processes patient confidentiality must be preserved. An organization's ongoing quality assurance activities should incorporate findings from adverse drug reaction monitoring and reporting programs.

### 9.6.13 WHO AND OTHER AUTHORITIES FOR PMS

The WHO established its program as a pilot project in 1968 initially in 10 and later in many other countries. The WHO promotes pharmacovigilance at the country level working through the WHO Collaborating Centre for International Drug Monitoring (Uppsala Monitoring Centre). This pharmacovigilance program has following objectives: greater patient care and drug safety, special emphasis on prevention of undesired drug-related toxicities, better public health safety in relation to the use of drugs, provision of reliable information for more rational use of drugs, improved assessment of risk–benefit profile to encourage safer and more effective use of medicines, the resolution of apparent conflicting interests of public health, and ensuring the welfare of individual patients.

Today worldwide many other regulatory authorities are involved in monitoring the safety of drugs. Other authorities include the United Kingdom's Medicine Control Agency (MCA) and the Committee on Safety of Medicine (CSM), the Vaccines Adverse Event Reporting System (VAERS), co-administered by the Department of Health and Human Services (DHHS), the FDA, and the Centers for Disease Control and Prevention (CDC) of the U.S., the Adverse Drug Reactions Advisory Committee (ADRAC) of Australia, and MedWatch Program of the USFDA-Canadian Adverse Drug Reaction Monitoring Program.

### 9.6.13.1 Safety Surveillance of Antiretroviral Drug and RaPID Program

In some resource-poor countries, many children have started their lives with antiretroviral therapy, but these countries have no effective and functioning pharmacovigilance system to monitor appropriate use of these drugs and collection of drug safety data. Long-time, adequate financial efforts and expertise are required to establish new pharmacovigilance programs or to strengthen the existing ones. WHO and the Uppsala Monitoring Centre (UMC) have been conducting training seminars for more than 15 years in those countries. Despite these efforts, UMC estimates that excepting a few, these countries have limited capacity; they collect little adverse drug reaction information and it is not sufficient for analysis of formulation of policies to increase patient safety. Drugs exhibiting toxicity of concern need to be removed from use or withdrawn from the market. And those are to be replaced by other suitable drugs.

The most significant benefit of the RaPID program is that it can establish a country- and disease-specific drug safety program very quickly within a reasonable period and concurrently help build institutional capacity at the country level. The RaPID program can help national drug authorities in capturing ADR data and their inclusion into UMC/WHO's Vigiflow software, and analyze them by a team of RaPID program global experts with technical expertise in HIV/AIDS, TB, and malaria and drug-related toxicities. It is important to mention that all activities of the RaPID program are coordinated and jointly developed with the national drug authority. The RaPID program is being implemented in joint technical collaboration with leading pharmacovigilance institutions from the developed and the developing countries including the Uppsala Monitoring Centre in Sweden, the national regulatory authority of Switzerland (Swissmedic), and other organizations, which provide activity-specific expertise, for example, WHO (various departments), Ecumenical Pharmaceutical Network, and the Indian Institute of Health Management Research (a WHO collaborating center). The RaPID program team has significant expertise in designing and implementing pharmacovigilance programs. With WHO (Geneva) support, the RaPID program is currently conducting an evaluation in several countries of Africa to assess their capacity for collection of pharmacovigilance data and their analyses to formulate appropriate policy related to drug use. Currently, the RaPID program team is in discussion with several countries (Nigeria, Sri Lanka, Uganda, India) and nongovernmental organizations (NGOs) (Medicines Sans Frontier, PSI, etc.) to explore collaborations and develop strategies for conducting pharmacovigilance.

### 9.6.13.2 Baseline Program in Hospitals of Resource-Poor Countries

Recently, a few hospitals in developing countries have taken initiatives and initiated spontaneous reporting programs at the facility level. ADR reporting forms are placed in different wards and out-patient departments. The doctors, nurses, and pharmacists report the adverse drug reaction, if any, to the pharmacovigilance cell, newly developed units of the hospitals. Based on the report, the cell provides relevant information to the drug regulatory authority of the country. Since its inception, the program looks very promising. Plans are under way to upscale existing spontaneous reporting programs to fully developed pharmacovigilance programs of these hospitals in near the future—its ultimate goal.

### 9.6.13.3 Possible Approach to Develop Drug Safety Monitoring Program for Hospital in Resource-Poor Country

Through awareness programs all health care providers need to be encouraged to report all suspected adverse drug reactions using a reporting form, which should be readily available all the time in all patient care areas of the health facilities including wards, out-patient departments (OPDs), and pharmacy departments. After filling out the forms, the health care provider should send them to the drug safety monitoring unit of the hospital. Data collected from these forms can be utilized to evaluate the reactions in detail. Several adverse drug reaction reporting forms are available, such as “yellow card” and “blue card” systems, and one developed by MedWatch. Generic forms can be modified to meet the local needs. The adverse drug reactions and related information of the particular drug(s) should be provided to the concerned authority. The determination of the likelihood of a causal relation between a suspected drug and an adverse drug reaction is an essential step in identifying drug toxicity. Based on the causality assessment, an adverse drug reaction can be classified as one of the following categories: certain, probable, possible, unlikely, conditional/unclassified, and not assessable/not classifiable. There are several causality assessment scales such as the WHO causality assessment scale, Naranjos scale, Karch and Lasagnas scale, Schumock and Thornton scale, European ABO system, and the Bayesian neutral scale. Any one these scales can be used to conduct causality assessments. The preventability and predictability of the adverse drug reactions can be assessed by using the specified scales. Thus, incidence of particular adverse drug reaction can be detected against the total number of drugs used in the hospital during the particular period. All data captured by the program should be entered into its database, which will be data for that particular population group(s). The relevant drug reactions data and their analyses should be periodically disseminated through adverse drug reaction bulletins, journals, and newsletters.

### 9.6.13.4 Inference for Future Buildup in Poor Countries

In resource-poor countries, it should be the responsibility of all health care workers to develop strategies for reporting, monitoring, and preventing adverse drug reactions. The drug regulatory authorities in poor countries should take the necessary steps to promote safe use of drugs in their country. Initially, the program might begin at a single hospital for consolidation and training and later might be extended to other hospitals as the program matures. Generally, people also need to be educated on the correct use of drugs and on reporting of undesirable experiences in association with drug use. The medical curriculum should include and emphasize rational use of drugs and aspects of adverse drug reactions. A formal training program for the primary health care professionals on adverse drug reactions should also be established. Postmarketing surveillance studies must be made mandatory after marketing any new drug in the country. A well-developed adverse drug reactions monitoring system can contribute a lot to the health care system if the formed teams/programs are supported by highly dedicated and well-experienced staff and good data management systems. Appropriate addressing of these issues would help establish an effective program of pharmacovigilance in resource-poor situations.



### 9.6.13.5 Data Mining in Postmarketing Surveillance

An increasing interest has developed in data mining or knowledge discovery initiatives using robust databases with development of large electronic health data storage systems in recent years. Web mining and information science can simultaneously make their application in postmarketing surveillance of drugs. Data mining processes are often performed on existing databases, for the purposes of postmarketing surveillance considering high cost for collection of data. It is difficult to determine the requirement of the necessary size of the dataset. It depends on quality of data, the background frequency of the event(s), and the strength of the association of the event with the drug; however, for even moderately rare events large databases are required.

### 9.6.13.6 Spontaneous Reporting Databases

The spontaneous reporting of adverse drug reactions by health care providers to health authorities or drug companies is an important procedure in postmarketing drug surveillance. However, spontaneous reporting systems are vulnerable to inconsistent reporting. The accuracy of the data may remain uncertain. Yet, spontaneous reporting databases contain large amounts of data. These databases could, therefore, be mined to obtain details of ADRs.

### 9.6.13.7 Prescription Event-Monitoring Databases

Prescription event monitoring (PEM) system detects adverse drug reactions by mobilizing data of high quality from family physicians on a select group of patients exposed to a specific (new) drug, for a limited period of time. Lack of an adequate control group is an important limitation of PEM database mining.

### 9.6.13.8 Linked Administrative Databases

Large linked health administrative databases, such as Medicaid in the United States, and the Ontario provincial databases contain data on millions of subjects. These might also be used as a source for data mining. The Saskatchewan-linked administrative health care utilization database and the Tayside Medicines Monitoring (MEMO) are examples of linked medical health databases. Both of them have been utilized to identify risks of benzodiazepine therapy.

### 9.6.13.9 Electronic Medical Records

Electronic medical records (EMRs) contain large numbers of data: e.g., use of tobacco products, use of nonprescription drugs, symptoms and signs of diseases, laboratory data, and social circumstances on a smaller numbers of patients. All these might be used for data mining. The availability of large numbers of variables and their combined use makes possible the generation of hypotheses and the exploration of new diagnoses or adverse events of drugs. EMR has been used by some to investigate detection of known adverse drug reactions.



#### 9.6.13.10 Other Databases

Specialist and clinical trials databases and overdose or toxicology databases might provide valuable information. Signals for liver-related ADRs have been reported from a biochemistry laboratory database at a higher rate than that reported by physicians. Poison information centers also record details of ADRs and thus might be a good contributor of pharmacovigilance activity.

#### 9.6.13.11 Data Preprocessing

Data preprocessing involves sampling and verification of data quality to ensure that data are clean. Medical data might contain erroneous data, for example, putting an age of 120 years instead of 20 years and documentation of hysterectomy in males. Data can also be used in creating new variables of interest for analysis, for example, it might be possible to estimate socioeconomic status from the postal code.

#### 9.6.13.12 Application of Data-Mining Techniques in PMS

Data mining involves a number of statistical techniques, including cluster analysis, link analysis, deviation detection, and disproportionality assessment, which can be used to determine the presence and assessment of the strength of adverse drug reaction signals. It is also a systematic process whereby large databases are searched. Combinations of variables that occur at a higher frequency than expected are mined. Research has been conducted specifically on the application of data mining techniques in postmarketed drug-related toxicity surveillance. Development of this tool is promising to optimize limited resources and improve the level of consistency in safety reviews of approved marketed drugs. An improvement in the current understanding of adverse drug reaction patterns is expected through the use of data-mining techniques. A successful data-mining tool will allow for earlier detection of safety signals and drug–drug interactions with drugs already in the market. Recently, there has been potential use of data mining and knowledge discovery in databases for detection of adverse drug events in pharmacovigilance. Knowledge discovery in databases (KDD) is a technical process that may be employed to explore potential adverse drug reactions more effectively. KDD involves selection of data variables and databases, data preprocessing, data mining, and data interpretation and utilization. Recently, the spontaneous reporting database of WHO Uppsala Monitoring and the UK yellow card scheme use data-mining methods. These data-mining methods used in pharmacovigilance are those of disproportionality, such as the proportional reporting ratio and information component, which has been used to analyze the drug-related adverse events. Examples are: association of captopril and other angiotensin-converting enzymes with cough, the association of pericarditis with practolol but not with other  $\beta$ -blockers, and the association of terfenadine with heart rate and rhythm disorders; these all could be identified by mining the database. In view of the importance of adverse drug reaction and the development of massive data storage systems and availability of powerful computers, the use of data-mining techniques in knowledge discovery in medical databases is now of increasing importance in the process of pharmacovigilance. They are likely to be able to detect signals earlier.

### 9.6.14 DISCUSSION

Postmarketing surveillance programs as well as clinical phase IV studies both are definitely essential to ensure the safety of drugs. These studies disclose beneficial and detrimental effects of drugs, including the existence of causal effects, risk factors, economic consequences, and characterization of drug use in clinical practice. Drug-related diseases impart a considerable burden to health care systems globally, which has been demonstrated well in the Western countries. The situations in the low-income countries are not very clear. There, published reports on the incidence and cost of drug-related problems are few. It is a reasonable assumption that less rigorous control of drug quality, easy availability, and indiscriminate usage create problems associated with drug therapy. A majority of drug-related problems are preventable as demonstrated by studies conducted in developed countries.

Appropriate instructions should be available on the effective use of drugs, and people should be instructed to adhere to them. Then prevention or reduction of drug-related toxicity would be possible for thousands of patients in developing countries. There, suffering could be alleviated as well as many lives could be saved, while millions of dollars would be saved annually by the society. The absence of awareness on the magnitude of drug-related problems in many societies is common. Adverse effects of drugs often manifest gradually and the symptomatology often mimic those of common illness, and usually they manifest at home and thus remain largely undetected and unrecorded. The awareness of drug-related problems is inadequate among politicians, the general public, patients, and sometimes even in professionals to trigger and stimulate a wide discussion on its prevention. So, qualified people are required in the health care system to support industry with genuine interest in disclosing these problems to everyone. An elaborative system of rules and regulations is necessary to govern the drug surveillance process. In postmarketing drug surveillance, pharmaceutical manufacturers, drug regulatory agencies, and the end users (physicians and patients) all need to play essential roles to minimize the risks.

### 9.6.15 FUTURE PERSPECTIVE

The future of postmarketing surveillance will need the proper development of research methods that would be capable to cope with the specific problems posed by the newly marketed drugs and would be able to accurately suggest adjustments for reducing severity of drug-related adverse events. Phase IV clinical studies and postmarketing surveillance, including pharmacovigilance, is now a very growing science that would offer great scope for minimizing harm to patients taking drugs. It will also reduce costs of health care systems. From small beginnings, with the appropriate technological knowledge and skills, postmarketing surveillance and phase IV studies can make important contributions to the health of people in any country. A successful postmarketing drug surveillance program needs capacity building, for example, training of professionals, optimization of computerized databases for research purposes, their joint use with traditional epidemiological methods, and so forth. The impact of new vaccines and new drugs on well-defined populations

needs assessment, which will be a critical task and from public health perspectives its impact is important. The pharmaceutical industry and its business environment should change immensely in the future. Growing competition from generic drugs, competitive product market, and limitation of resource and development costs are future problems for pharmaceutical executives, and they should plan a balanced program in this regard. The objective should be to divert resources for promotion of only those drugs that will be unlikely to be rejected by the regulatory agencies or shall not be removed from the market due to safety problems. A pharmaceutical company will need to use appropriate technology to ensure production of a safe and effective drug to ensure a favorable risk–benefit ratio. Thus, to implement a proactive postmarketing surveillance and risk management strategy, drug companies should always adapt to changing regulations to ensure compliance for eventual approval of their drug products. The pharmaceutical industries must adhere to international regulations in addition to local rules and should be regulated by a global enterprise.

Therefore, drug companies should report and communicate drug-related safety issues regularly to every country where they market their products. The adoption of advanced pharmacovigilance tools will greatly help drug companies to generate and submit reports to specific countries or regions based on their regulations. Here, specifically, signal detection and data management tools, management technology solutions, and adverse-event reporting systems will be specially helpful. Effective pharmacovigilance is needed for future successful drug safety surveillance for which radical changes are required in signal detection, use of data mining and management process, and use of these technologies in phase IV clinical trials. Thus, completion of analysis of clinical data would be quick as compared to the past. Time saving shall reduce in development costs of drug and would allow early arrival of new drugs on the market. Simultaneously, termination of future clinical trials would be possible on early detection of safety issues and, therefore, patients would be protected. Thus company resources would be saved that might be spent on modernization of their industry-based technology and development of efficient manpower.

The European Union's postmarketing surveillance system has undergone major changes over the last few years through a stronger drug regulatory and pharmacovigilance system. The EU PMS systems have the power to mandate phase IV studies and impose financial penalties on companies that fail to comply with regulations and their obligations. In recent years, the FDA has adopted many changes and many are currently under review process to have a more proactive approach to pharmacovigilance in the near future. There has been a noticeable shift forwarded by regulatory authorities, in terms of formulation of policies on risk management, and also by pharmaceutical companies in terms of incorporation of pharmacogenomics in their development programs. The purpose of these initiatives are to better predict safety issues and minimize risks as quickly as possible.

In the future, companies need to ensure a close link between premarketing and postmarketing pharmacovigilance activities to increase transparency of their activities and improve communication of information about drugs to begin risk management planning at the inception. Many companies have now agreed to voluntarily register clinical trials and will post all clinical trial data of marketed drugs, whether favorable or unfavorable, on publicly accessible websites.

### 9.6.16 CONCLUSION

Despite its four-decade history, postmarketing surveillance remains a dynamic clinical and scientific discipline. It continues to play a crucial role in meeting the challenges posed by the ever increasing range of drugs, all of which carry an unpredictable potential for harm. It is essential that previously unknown adverse effects are reported, analyzed, and their significance communicated effectively to the people, and it should be ensured that they have the knowledge to interpret that information. For all drugs, there is a trade-off between the benefits and the potential for harm. The harm can be minimized by ensuring that drugs are of good quality, safety, and efficacy and are used rationally. At the same time, expectations and concerns of the patient should be taken into consideration when therapeutic decisions are taken. The practitioners should ensure that risks in drug use are anticipated and will be managed if they occur. This will create a sense of trust among patients taking drugs. So, improvement of communication between the health professionals and the public is essential. Also, updating health professionals to understand the efficacy/risk of medicines that they prescribe is necessary for this purpose. As adverse drug reactions are a significant public health problem in every health care system, postmarketing drug surveillance systems need to be improved continuously. Simultaneously, it is a grand challenge with a fundamental problem of great social and economic importance, and the solution needs significant increases in scientific knowledge and technical capabilities.

### BIBLIOGRAPHY

- Amery, W. K. (1999), Signal generation from spontaneous adverse event reports, *Pharmacoepidemiol. Drug Saf.*, 8, 147–150.
- Anon. (2002), MedDRA compared with WHO-ART, *Drug Saf.*, 25, 445–452.
- Anon. (2001), Details of methods: Explanation of data mining methods, *BMJ*, 322, 1207–1209.
- Anon. (1975), *Requirements for Adverse Reaction Reporting*, World Health Organization, Geneva, Switzerland.
- ASHP (1989), ASHP guidelines on adverse drug reaction monitoring and reporting, *AMJ Hosp. Pharm.*, 46, 336–337.
- Bagheri, H., Michel, F., Lapeyre-Mestre, M., Lagier, E., Cambus, J. P., Valdiguié, P., and Montastruc, J. L. (2000), Detection and incidence of drug-induced liver injuries in hospital: A prospective analysis from laboratory signals, *Br. J. Clin. Pharmacol.*, 50, 479–484.
- Barbone, F., McMahon, A. D., Davey, P. G., Morris, A. D., Reid, I. C., McDevitt, D. G., and MacDonald, T. M. (1998), Association of road-traffic accidents with benzodiazepine use, *Lancet*, 352, 1331–1336.
- Bate, A., Lindquist, M., Edwards, I. R., and Orre, R. (2002), A data mining approach for signal detection and analysis, *Drug Saf.*, 25, 393–397.
- Bate, A., Lindquist, M., Orre, R., Edwards, I. R., and Meyboom, R. H. (2002), Data-mining analyses of pharmacovigilance signals in relation to relevant comparison drugs, *Eur. J. Clin. Pharmacol.*, 58, 483–490.

- Bates, D. W., Cullen, D. J., Laird, N., Petersen, L. A., Small, S. D., Servi, D., Laffel, G., Sweitzer, B. J., Shea, B. F., and Hallisey, R. (1995), Incidence of adverse drug events and potential adverse drug events. Implications for prevention, *JAMA*, 274, 29–34.
- Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., Burdick, E., Hickey, M., Kleeefield, S., Shea, B., Vander Vliet, M., and Seger, D. L. (1998), Effect of computerized physician order entry and a team intervention on prevention of serious medication errors, *JAMA*, 280, 1311–1316.
- Beard, K. (1992), Adverse reactions as a cause of hospital admission in the aged, *Drugs Aging*, 2, 356–367.
- Biriell, C., and Edwards, R. (1997), Reasons for reporting adverse drug reactions—some thoughts based on an international review, *Pharmacoepidemiol. Drug Saf.*, 6, 21–26.
- Bombardier, C., Laine, L., Reicin, A., Shapiro, D., Burgos-Vargas, R., Davis, B., Day, R., Ferraz, M. B., Hawkey, C. J., Hochberg, M. C., Kvien, T. K., and Schnitzer, T. J. (2000), Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group, *N. Engl. J. Med.*, 343, 1520–1528.
- Brewer, T., and Colditz, G. A. (1999), Postmarketing surveillance and adverse drug reactions: current perspectives and future needs, *JAMA*, 281, 824–829.
- Brown, E. G., Wood, L., and Wood, S. (1999), The medical dictionary for regulatory activities (MedDRA), *Drug Saf.*, 20, 109–117.
- Bytzer, P., and Hallas, J. (2000), Drug-induced symptoms of functional dyspepsia and nausea. A symmetry analysis of one million prescriptions, *Aliment Pharmacol. Ther.*, 14, 1479–1484.
- Cerrito, P. (2001), Application of data mining for examining polypharmacy and adverse effects in cardiology patients, *Cardiovasc. Toxicol.*, 1, 177–179.
- Chakrabarti, S. (2002), *Mining the Web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann, San Francisco.
- Chowdhury, G. G. (1999), Template mining for information extraction from digital documents, *Library Trends*, 48, 181–207.
- Ciorciaro, C., Hartmann, K., and Kuhn, M. (1998), Differences in the relative incidence of adverse drug reactions in relation to age. An evaluation of the spontaneous reporting system of SANZ, *J. Suisse Med.*, 128, 254–258.
- Classen, D. C., Pestotnik, S. L., Evans, R. S., Lloyd, J. F., and Burke, J. P. (1997), Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality, *JAMA*, 277, 301–306.
- Colin-Jones, D. G., Langman, M. J., Lawson, D. H., and Vessey, M. P. (1985), Postmarketing surveillance of the safety of cimetidine: Mortality during second, third, and fourth years of follow up, *BMJ (Clin. Res. Ed.)*, 291(6502), 1084–1088.
- Crombie, I. K., Brown, S. V., and Hamley, J. G. (1984), Postmarketing drug surveillance by record linkage in Tayside. *J. Epidemiol. Community. Health*, 38(3), 226–231.
- Dunn, N., Freemantle, S., and Mann, R. (1999), Nicorandil and diabetes: A nested case–control study to examine a signal generated by prescription-event monitoring, *Eur. J. Clin. Pharmacol.*, 55, 159–162.
- Edwards, I. R. (2000), The accelerating need for pharmacovigilance, *J. Roy. Coll. Physicians Lond.*, 34, 48–51.
- Edwards, I. R., and Biriell, C. (1994), Harmonisation in pharmacovigilance, *Drug Saf.*, 10, 93–102.
- Egberts, A. C., Meyboom, R. H., and van Puijenbroek, E. P. (2002), Use of measures of disproportionality in pharmacovigilance: Three Dutch examples, *Drug Saf.*, 25, 453–458.

- Eisenhauer, L. A. (2002), Adverse drug reactions: A concern for clinicians and patients, *Clin. Excellence Nurse Practitioners*, 6, 3–7.
- Evans, S. J., Waller, P. C., and Davis, S. (2001), Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports, *Pharmacoepidemiol. Drug Saf.*, 10, 483–486.
- Faich, G. A. (1986), Adverse-drug-reaction monitoring, *N. Engl. J. Med.*, 314, 1589–1592.
- Fattinger, K., Roos, M., Vergees, P., Holenctein, C., Kind, B., Masche, U., Stocker, D. N., Braunschweig, S., Kullak-Ublick, G. A., Galeazzi, R. L., Follath, F., Gasser, T., and Meier, P. J. (2000), Epidemiology of drug exposure and adverse drug reactions in two Swiss departments of internal medicine, *Br. J. Clin. Pharmacol.*, 49, 158–167.
- Fisher, S., Bryant, S. G., Solovitz, B. L., and Kluge, R. M. (1987), Patient-initiated postmarketing surveillance: A validation study, *J. Clin. Pharmacol.*, 27, 843–854.
- Friedman, M. A., Woodcock, J., Lumpkin, M. M., Shuren, J. E., Hass, A. E., and Thompson, L. J. (1999), The safety of newly approved medicines: Do recent market removals mean there is a problem? *JAMA*, 281, 1728–1734.
- Heeley, E., Wilton, L. V., and Shakir, S. A. (2002), Automated signal generation in prescription-event monitoring, *Drug Saf.*, 25, 423–432.
- Honig, P. K., Woosley, R. L., Zamani, K., Conner, D. P., and Cantilena, L. R., Jr. (1992), Changes in the pharmacokinetics and electrocardiographic pharmacodynamics of terfenadine with concomitant administration of erythromycin, *Clin. Pharmacol. Ther.*, 52, 231–238.
- Honigman, B., Lee, J., Rothschild, J., Light, P., Pulling, R. M., Yu, T., and Bates, D. W. (2001), Using computerized data to identify adverse drug events in outpatients, *J. Am. Med. Assoc.*, 8, 254–266.
- Inman, W. H. (1981), Post marketing surveillance of adverse drug reactions in general practice. II: Prescription-event monitoring at the University of Southampton, *BMJ*, 282, 1216–1217.
- Ioannidis, J. P., and Lau, J. (2001), Completeness of safety reporting in randomized trials: An evaluation of 7 medical areas, *JAMA*, 285, 437–443.
- Jick, H., Madsen, S., Nudelman, P. M., Perera, D. R., and Stergachis, A. (1984), Post-marketing follow-up at Group Health Cooperative of Puget Sound, *Pharmacother.*, 4, 99–100.
- Johnson, J. A., and Bootman, J. L. (1996), Drug-related morbidity and mortality, *Arch. Intern. Med.*, 155, 1949–1956.
- Juntti-Patinen, L., and Neuvonen, P. J. (2002), Drug-related deaths in a university central hospital, *Eur. J. Clin. Pharmacol.*, 58, 479–482.
- Karch, F. E., and Lasagna, L. (1975), Adverse drug reaction—a critical review, *JAMA*, 234, 1236–1241.
- Kaufman, D. W., Rosenberg, L., and Mitchell, A. A. (2001), Signal generation and clarification: Use of case-control data, *Pharmacoepidemiol. Drug Saf.*, 10, 197–203.
- Kwansnik, B. H. (1999), The role of classification in knowledge representation and discovery, *Library Trends*, 48, 22–47.
- Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998), Incidence of adverse drug reactions in hospitalized patients—a meta-analysis of prospective studies, *JAMA*, 279, 1200.
- Leape, L. L., Cullen, D. J., Clapp, M. D., Burdick, E., Demonaco, H. J., Erickson, J. I., and Bates, D. W. (1999), Pharmacist participation on physician rounds and adverse drug events in the intensive care unit, *JAMA*, 282, 267–270.



- Lindquist, M., Stahl, M., Bate, A., Edwards, I. R., and Meyboom, R. H. (2000), A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database, *Drug Saf.*, 23, 533–542.
- Mann, R. D. (1998), Prescription-event monitoring—recent progress and future horizons, *Br. J. Clin. Pharmacol.*, 46, 195–201.
- Mey, C., Hentschel, H., Hippus, M., and Balogh, A. (2002), Documentation and evaluation of adverse drug reactions (ADR)—contribution from a poison information center, *Int. J. Clin. Pharmacol. Ther.*, 40, 102–107.
- Meyboom, R. H., Lindquist, M., Egberts, A. C., and Edwards, I. R. (2002), Signal selection and follow-up in pharmacovigilance, *Drug Saf.*, 25, 459–465.
- Mukherjee, D., Nissen, S. E., and Topol, E. J. (2001), Risk of cardiovascular events associated with selective COX-2 inhibitors, *JAMA*, 286, 954–959.
- Murphy, B. M., and Frigo, L. C. (1993), Development implementation and results of a successful multidisciplinary adverse drug reaction reporting program in a university teaching hospital, *Hosp. Pharm.*, 28, 1199–1240.
- Niu, M. T., Erwin, D. E., and Braun, M. M. (2001), Data mining in the US Vaccine Adverse Event Reporting System (VAERS): Early detection of intussusception and other events after rotavirus vaccination, *Vaccine*, 19, 4627–4634.
- Noan, B. A., and Brushwood, D. B. (2000), Adverse drug reactions in elderly patients: alternative approaches to post-surveillance, *J. Health Law*, 33, 383–454.
- Olsson, S. (2001), The need for pharmacovigilance, in Gupta, S. K., Ed., *Pharmacology and Therapeutics in the New Millennium*, Narosa, New Delhi, pp. 502–508.
- Patel, P., and Zed, P. J. (2002), Drug-related visits to the emergency department: How big is the problem? *Pharmacotherapy*, 22, 915–923.
- Perner, P., and Petrou, M. (1999), *Machine Learning and Data Mining in Pattern Recognition*, Berlin: Springer, Berlin.
- Pirmohamed, M., Breckenridge, A. M., Kitteringham, N. R., and Park, B. K. (1998), Adverse drug reactions, *BMJ*, 316, 1295–1298.
- Ramesh, M., Pandit, J., and Parthasarathi, G. (2003), Adverse drug reactions in a south Indian hospital—their severity and cost involved, *Pharmacoepidemiol. Drug Saf.*, 12, 687–692.
- Rawlins, M. D., and Thompson, J. W. (1977), Pathogenesis of adverse drug reactions, in Davies, D. M., Ed., *Textbook of Adverse Drug Reactions*, Oxford University Press, Oxford.
- Rawson, N. S., and Rawson, M. J. (1999), Acute adverse event signalling scheme using the Saskatchewan Administrative health care utilization datafiles: Results for two benzodiazepines, *Can. J. Clin. Pharmacol.*, 6, 159–166.
- Shakir, S. A., and Layton, D. (2002), Causal association in pharmacovigilance and pharmacoepidemiology: Thoughts on the application of the Austin Bradford-Hill criteria, *Drug Saf.*, 25, 467–471.
- Smith, D. L. (1993), The effect of patient non-compliance on healthcare costs, *Med. Interface*, 6, 74–76, 78, 84.
- Strom, B. L. (2000), How should one perform pharmacoepidemiology studies? Choosing among the available alternatives, in Strom, B. L., Ed., *Pharmacoepidemiology*, Vol. 3, Chichester: Wiley, Chichester, England, pp. 401–413.
- Strom, B. L., Carson, J. L., Morse, M. L., and LeRoy, A. A. (1985), The computerized on-line Medicaid pharmaceutical analysis and surveillance system: A new resource for postmarketing drug surveillance, *Clin. Pharmacol. Ther.*, 38, 359–364.



- Szarfman, A., Machado, S. G., and O'Neill, R. T. (2002), Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database, *Drug Saf.*, 25, 381–392.
- van Puijenbroek, E. P., Bate, A., Leufkens, H. G., Lindquist, M., Orre, R., and Egberts, A. C. (2002), A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions, *Pharmacoepidemiol. Drug Saf.*, 11, 3–10.
- van Puijenbroek, E. P., Egberts, A. C., Meyboom, R. H., and Leufkens, H. G. (1999), Signalling possible drug–drug interactions in a spontaneous reporting system: Delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole, *Br. J. Clin. Pharmacol.*, 47, 689–693.
- van Puijenbroek, E. P., van Grootheest, K., Diemont, W. L., Leufkens, H. G., and Egberts, A. C. (2001), Determinants of signal selection in a spontaneous reporting system for adverse drug reactions, *Br. J. Clin. Pharmacol.*, 52, 579–586.

# 9.7

---

## *Regulatory Approval*

**Fred Henry<sup>1</sup> and Weichung J. Shih<sup>2</sup>**

<sup>1</sup>*Drug Development and Regulatory Affairs, Taisho Pharmaceuticals R&D Inc.,  
Morristown, New Jersey*

<sup>2</sup>*Department of Biostatistics, School of Public Health,  
University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey*

### **Contents**

|         |                                                                   |     |
|---------|-------------------------------------------------------------------|-----|
| 9.7.1   | FDA Approval                                                      | 350 |
| 9.7.1.1 | Introduction                                                      | 350 |
| 9.7.1.2 | Regulations and Guidance Documents                                | 350 |
| 9.7.1.3 | Consultation with the FDA                                         | 352 |
| 9.7.1.4 | Investigational New Drug Applications                             | 354 |
| 9.7.1.5 | Clinical Hold: Issues and Resolution                              | 357 |
| 9.7.1.6 | IND Sponsor Reporting Requirements                                | 358 |
| 9.7.1.7 | Marketing Approval Requirements                                   | 359 |
| 9.7.1.8 | Summary                                                           | 361 |
| 9.7.2   | Fast-Track Approval                                               | 362 |
| 9.7.2.1 | Introduction                                                      | 362 |
| 9.7.2.2 | Fast Track Is More Than “Priority Review”                         | 362 |
| 9.7.2.3 | Request for Designation of Drug as Fast-Track Product             | 362 |
| 9.7.2.4 | Limitations                                                       | 363 |
| 9.7.2.5 | Accelerated Approval Issues (I): Clinical and Surrogate Endpoints | 363 |
| 9.7.2.6 | Accelerated Approval Issues (II): Postapproval Study              | 365 |
| 9.7.2.7 | Accelerated Approval Issues (III): Control of Type I Error Rate   | 365 |
| 9.7.2.8 | Examples of FTDDP                                                 | 367 |
|         | References                                                        | 371 |

## 9.7.1 FDA APPROVAL

### 9.7.1.1 Introduction

The Food and Drug Administration (FDA) is part of the Department of Health and Human Services (DHHS) in the U.S. federal government and is the regulatory agency responsible for marketing approval of new pharmaceutical products. Within the FDA are the Center for Drug Evaluation and Research (CDER), whose activities are the focus of this chapter, along with the Center for Biologics Evaluation and Research (which regulates vaccines and blood products), Center for Devices and Radiological Health, Center for Food Safety and Applied Nutrition, and the Center for Veterinary Medicine. The FDA has authority to write regulations relevant to interstate commerce associated with pharmaceuticals (as well as food, cosmetics, and veterinary medicines), which may then be enacted into law. Regulations may be enforceable by the FDA's compliance office. Guidelines created by the FDA, on the other hand, are meant to provide general direction to drug sponsors on many issues, without legal obligation. However, deviation from guidelines should be discussed with the agency in order to maintain a cooperative and common understanding of a drug's development.

It is important to keep in mind the FDA's concerns during drug development in order to anticipate and address them proactively during all phases of clinical development. The FDA's primary concern is always the protection of the health and rights of subjects involved in human drug research. In addition, during phases II and III the agency becomes concerned about the adequacy and quality of the scientific evaluation of the drug's safety and efficacy as part of a new drug application (NDA). This chapter will summarize the activities necessary to achieve and maintain regulatory clearance for the conduct of clinical trials and also describes some general strategies for the cooperative design of clinical programs to support marketing approval.

### 9.7.1.2 Regulations and Guidance Documents

Clinical trial conduct is governed by national laws that are supplemented by international, regional, and national guidance documents. Regulations regarding the conduct of clinical trials and the protection of human subjects involved in clinical trials in the United States, as well as many other regulations regarding food and drugs, are described in Title 21 of the Code of Federal Regulations (21 CFR) [1]. Clinical trials in the United States must conform to the regulations for investigational new drug (IND) applications, which are covered in 21 CFR Part 312 ("Investigational New Drug Applications"). Regulations regarding the content and format of NDAs are described in 21 CFR Part 314 ("Applications for FDA Approval to Market a New Drug"). Application of these regulations to the design and conduct of clinical trials will be addressed in greater detail throughout this chapter. Other important 21 CFR regulations for the clinical development of pharmaceuticals include Part 50, which covers the topic of "Protection of Human Subjects," and Part 56, "Institutional Review Boards."

In addition, the FDA has developed many guidelines for drug sponsors that provide direction regarding the application of the 21 CFR regulations, as well as

more specific information on the expectations that the FDA has for appropriate clinical study design and conduct. A key concept in the regulatory requirements for marketing approval is “substantial evidence” of efficacy from clinical trials, based on the Kefauver–Harris Amendments to the 1962 Food, Drug and Cosmetic Act, which was interpreted to mean that approval for the marketing of a new drug in the United States would be based on evidence from adequate and well-controlled investigations. This requirement would in turn imply that clinical trials had to be of sufficient size to have adequate statistical power to test important clinical hypotheses and had to meet the standard of being adequate and well controlled. To be adequate and well controlled a study has to have:

- A clear statement of objectives
- A design that permits a valid comparison with a control and provides a quantitative assessment of the drug effect
- A method of selection of subjects that ensures that they have the disease under study
- A method of patient assignment that minimizes bias
- A well-defined and reliable method of assessment of responses (outcome measures)

Several regulatory guidance documents from the FDA consider the reporting, evaluation, and interpretation of clinical trial evidence. These guidance documents were issued over a period of nearly 20 years. The first document is the Guideline for the Format and Content of the Clinical and Statistical Section of New Drug Applications [2]. This document addresses how the results of clinical studies are to be reported to the FDA. The second is the Guidance for Industry Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products [3]. The agency participates in the International Conference on Harmonisation (ICH), which is a group of representatives from government and pharmaceutical industry groups in Japan, Europe, and the United States, formed to facilitate the development of mutually agreed upon guidelines for the chemical, nonclinical, and clinical development of new drugs. The guidance documents developed in ICH are then considered for adoption by national authorities in the three represented areas. The ICH guidances are powerful in that they support the design of clinical studies that are accepted internationally, increasing the efficiency of global drug development. For instance, the format issues of the NDA content and format guidelines mentioned above have been largely superseded by the ICH M4: Common Technical Document for the Registration of Pharmaceuticals for Human Use [4], which provides a structure for marketing applications that can be submitted in most all countries.

Of the ICH guidances relevant to the conduct of trials and adopted by the FDA, those most applicable to the conduct of clinical trials and support for drug registration are:

- E1B      The Extent of Population Exposure to Assess Clinical Safety: For drugs intended for long-term treatment of non-life-threatening conditions

- E2A Clinical Safety Data Management: Definitions and standards for expedited reporting
- E3 Structure and Content of Clinical Study Reports
- E4 Dose Response Information to Support Drug Registration
- E6 Good Clinical Practice: Consolidated Guideline
- E7 Studies in Support of Special Populations Geriatrics
- E8 General Considerations for Clinical Trials
- E9 Guideline on Statistical Principles for Clinical Trials
- E10 Choice of Control Group and Related Issues in Clinical Trials
- M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals

Throughout this chapter, references will be made to several specific FDA and ICH guidances; however, they are all conveniently included on the FDA's website at <http://www.fda.gov/cder/guidance/>. Special attention should be paid to the ICH E9, as it also pertains to the aspect of the "substantial evidence" requirement. The FDA's current thinking on the matter of substantial evidence, meshing the views from these guidelines, can be found in a recent paper [5].

Once an organization has determined that it is interested in conducting a clinical trial in the United States, it is appropriate to begin plans for design of the protocol and its submission with supporting documentation to the FDA. The first clinical trial of a new investigational agent will require compilation and submission of an IND [6]. This submission includes information on the chemistry and manufacturing, non-clinical pharmacology, pharmacokinetics, and toxicology, as well as a summary of any previous human experience with the product. Thus, it is prudent to organize an internal team of individuals and external consultants as necessary who will contribute to the summarization of available information on the product and can address issues that may arise with regard to the IND submission. The formation of such a team is essential to establish a communication framework for discussing development issues, including regulatory submissions, meetings with authorities, responses to FDA questions, and so forth. Drug sponsors often do not adequately address cross-functional development issues, such as the implications of findings in animal studies to the clinical setting. A team appropriate to the size and structure of each organization can help to avoid neglect of such issues before they become regulatory roadblocks.

During the preparation of an IND application, questions may arise as to the contents of the submission, protocol design, or other development issues that would benefit from the FDA's input. In order to assist sponsors to best develop new products, the FDA has implemented a guidance on arranging meetings, Formal Meetings with Sponsors and Applicants of PDUFA Products [7], which is described in the next section. The first such opportunity in a product's development for a meeting with the FDA is referred to as the pre-IND meeting.

### 9.7.1.3 Consultation with the FDA

Formal meetings between drug sponsors and the FDA regarding the development of new drug products are possible in order to promote mutual understanding of issues relating to matters that arise during a research program (21 CFR 312.47).

Depending on the stage of development, these issues may involve chemistry/manufacturing, nonclinical or clinical research pertaining to the design of a development program, interpretation of study results and/or implications of research on other members of a therapeutic class, or special issues regarding a novel compound. The FDA has identified three basic types of meetings according to the timing and intention of the meeting.

- *Type A Meetings* Immediately necessary to discuss clinical holds, protocol assessments, or other development issues that are considered to be critical to the continued development of a product. Such meetings are normally scheduled within 30 days of a meeting request. An information package that includes the questions and background information for the agency to review, prior to the meeting, and that provide information necessary for their review must be submitted at least 2 weeks prior to the meeting (contents of the information package provided to the agency prior to formal meetings is described in the meetings guidance).
- *Type B Meetings* These meetings include discussions prior to submission of an initial IND (pre-IND meeting), certain meetings after completion of phase I, end of phase II, and meetings to discuss the content and format of new drug applications (pre-NDA meeting). Type B meetings are scheduled within 60 days of the agency's receipt of the meeting request and require submission of a briefing document at least 4 weeks prior to the meeting date.
- *Type C Meetings* Other meetings regarding the development and review of drug products. Type C meetings are scheduled within 75 days of meeting request receipt and briefing documentation is required at least 2 weeks prior to the meeting.

In addition to these sponsor meetings that the FDA allows, special protocol review by the FDA is also possible. A written request is necessary for each meeting desired by a drug sponsor. This request must include a brief summary of the product, type of meeting requested, an agenda and list of objectives for the meeting, along with administrative information regarding the scheduling of the meeting as listed in the meetings guidance. The reviewing division normally responds to meeting requests within 14 days of receipt of the request, at which time a mutually agreeable time and date are set for the meeting.

**Conducting and Documenting Formal FDA Meetings** After an introduction of the meeting participants and a brief discussion of the background and objectives of the meeting, the agenda of the meeting is typically driven by the list of questions and/or issues that the sponsor has submitted as part of the meeting request and briefing documentation. Agency representatives provide responses to each question or issues raised by the sponsor, and discussion takes place if necessary to clarify the agency's position. Many divisions at the FDA are now providing to sponsors prior to meetings a copy of the agency's internal meeting minutes where they discuss their responses before meeting face to face with the sponsor. Though consideration of the planned duration of the meeting is prudent, sponsors should take advantage of the opportunity of these meetings to ensure understanding of the agency's position and resolve any misunderstandings during the meeting discussion. A summary of

the proceedings normally forms the conclusion of the meeting, and the FDA prepares a written summary of the meeting as the official minutes, which is shared with the sponsor (usually within 30 days of the meeting). Any disputes regarding the minutes of these meetings may be addressed as per 21 CFR 10.75 (Administrative Practices and Procedures), 312.48, and 314.103.

#### 9.7.1.4 Investigational New Drug Applications

Key to the successful clinical development of a new product is completion and submission of an IND. The FDA's Guidance for Industry on the Content and Format of Investigational New Drug Applications (IND) for Phase I Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-derived Products [6], along with 21 CFR 312.22 and 312.23, provides a clear framework for the content and format of such submissions. While the required components of the IND are standardized by the Guidance and 21 CFR, the FDA has allowed some flexibility in the presentation of data in the application. Also, the agency has begun to accept electronic submissions of INDs, reducing the burdens of archiving and navigating paper-based submissions.

Whatever the therapeutic class, status of development, chemical qualities, or other product characteristics, sponsors of an IND will increase the likelihood of a positive outcome of the FDA's review if they view the application from the perspective of the agency staff who will be considering the documentation. Mindfulness of the agency's concerns for the safety and rights of subjects, and in later stages of development the quality of research design regarding establishment of a product's safety and efficacy, should be maintained throughout the preparation of the application. A logical format will also aid in the review and avoids confusion among the team that will be assembled to consider the application at the agency. Following the FDA's guidance on the content and format is a good start in this direction, but cross-functional critiquing the application by the sponsor throughout the compilation process will help to ensure that the technical information is presented effectively.

***Content and Format*** As per the regulations in 21 CFR 312.23, an IND must include the following components to be considered complete for a new drug product being proposed for clinical study.

***Cover Sheet (Form FDA-1571)*** Following a cover letter in standard business style addressed to the Center for Drug Evaluation and Research, a completed Form FDA-1571 (available on the FDA website) is required. This form provides a brief identification of the contents of the submission, along with administrative information regarding the product sponsor's responsible personnel and their contact information.

***Table of Contents*** A detailed description of the components of the application and their location within the submission.

***Introductory Statement and General Investigational Plan*** This brief section of the IND provides an opportunity to introduce the agency staff to the product, including



intentions for its development, potential indications for use, any regulatory history (within and outside the United States), and any exceptional aspects of the IND that might be important for the review team.

**Investigator's Brochure** The investigator's brochure should follow the ICH guidance on the standards for the content and format of this document.

**Protocol(s)** Phase I study protocols need not be of the same level of detail as those for phase II or III trials; however, they should minimally include an outline of the investigation including number of subjects, description of safety exclusions, doses and duration of planned treatments, and safety monitoring. Since the objectives of phase I trials are normally to assess the safety, tolerability, pharmacokinetics, and perhaps some pharmacodynamic effects of the new drug, plans for assessing each of these aspects of the product's activities should be described in a rational manner.

Phases II and III protocols are required to provide a detailed description of all aspects of the trial. In addition to the assessment of safety, efficacy measurements need to be described thoroughly, remembering that the concerns of the FDA expand to include assessment of the trial's scientific merit toward eventual submission of study results in support of a marketing application.

**Chemistry, Manufacturing, and Controls (CMC) Information** Detailed information regarding the active ingredient (the drug substance) and the drug product (the formulation, such as tablet, solution, etc.) are required as described in 21 CFR 312.23(a)(7) and the Guidance on IND Content and Format for each and all of the products being investigated in the study (including placebo and active comparators). For the drug substance, this information must include:

- A description of the drug substance, including its physical, chemical, and/or biological characteristics
- The name and address of its manufacturer
- The general method of preparation of the drug substance
- The acceptable limits and analytical methods used to assure the identity, strength, quality, and purity of the drug substance
- Information to support the stability of the drug substance during the toxicologic studies and the proposed clinical study(ies)

With regards to the drug product information, similarly detailed information is required, including:

- A list of all components, which may include reasonable alternatives for inactive compounds, used in the manufacture of the investigational drug product, including both those components intended to appear in the drug product and those that may not appear but that are used in the manufacturing process
- Where applicable, the quantitative composition of the investigational new drug product, including any reasonable variations that may be expected during the investigational stage

- The name and address of the drug product manufacturer
- A brief, general description of the method of manufacturing and packaging procedures as appropriate for the product
- The acceptable limits and analytical methods used to assure the identity, strength, quality, and purity of the drug product
- Information to support the stability of the drug substance during the toxicologic studies and the proposed clinical study(ies)
- A brief general description of the composition, manufacture, and control of any placebo to be used in the proposed clinical trial(s)
- A copy of all labels and labeling to be provided to each investigator
- A claim for categorical exclusion from or submission of an environmental assessment

The information presented in the CMC section of the IND should be prepared by or at least with the close involvement of professionals within the organization who have expertise in this area of development and experience in preparing CMC documentation for regulatory submission. An FDA guidance specific to requests for information from the FDA relevant to CMC issues at the pre-IND stage is available to assist with questions that may arise during the preparation of this section of the IND.

*Pharmacology and Toxicology Information [21 CFR 312.23(a)(8)]* A summary of the pharmacology (effect of the drug on animals) and information known about the absorption, distribution, metabolism, and excretion (pharmacokinetics) of the drug should be provided. Individual reports may be provided; however, a summary is often sufficient. The sponsor should consider the stage of development (and the FDA concerns at that stage of development) and what if any implications the results of these studies may have on the assessment of safety.

Because of the safety concerns at the FDA in reviewing IND documentation and regulating clinical trial conduct, much more detailed information is required from the results of toxicology studies compared to other nonclinical studies. Short- and long-term trials in animals need to be presented in a logical manner, paying particular attention to the implications that the results and methods used in the study have for the design of the clinical trial and the characteristics of the subjects involved in the study. ICH has issued several guidances regarding the design and reporting of toxicology studies to support clinical trials and marketing applications that the sponsor should consider. In addition, the ICH M3 guidance Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals [8] must be considered to assure that the duration of treatment in the proposed clinical study is supported by the duration and types of toxicology studies reported in the IND. Detailed listings of data from each toxicology study used to support the conduct of the clinical study must be presented in the IND.

*Previous Human Experience* A summary of any and all previous investigational and if available marketed use of the product should be provided. Individual study reports are not normally required.

**IND Review Process** Once an IND is submitted to the FDA, it is routed to the appropriate reviewing division for assembly of a review team of representatives from each of the functional areas (CMC, nonclinical pharmacology and toxicology, clinical) and initiation of the review process. The review clock starts at the date of receipt of the application at the FDA, and if no comments are received by the sponsor within 30 days, the application is considered to be cleared. No formal approval letter is typically provided by the FDA; however, if issues are identified that delay or halt the clinical trial, the agency will issue a clinical hold to the sponsor (see following section).

**Information Amendments and New Protocols (21 CFR 312.30)** New protocols and amendments to previously submitted protocols must be submitted to the FDA prior to enactment at the clinical site, in addition to obtaining necessary institutional review board approval. There is no need to await the FDA review; however, the agency may provide comments to a sponsor, which may impact study conduct.

The IND sponsors are also responsible for submitting new technical information (toxicology study reports, chemistry or manufacturing changes, etc.) to the FDA. These “information amendments” should be presented in a format that facilitates the agency’s review of the submitted information, whether or not comment from the agency is requested by the sponsor.

#### 9.7.1.5 Clinical Hold: Issues and Resolution

At any time during the review of an initial IND submission or during the conduct of a clinical program, the FDA may order a drug sponsor to delay or suspend clinical study conduct (21 CFR 312.42). Such “clinical holds” may be limited in nature to a specific protocol or stage of development (“partial clinical hold”) or may impact all clinical investigation of a product (“complete clinical hold”). There are many reasons that the agency may take such an action, including:

- Risk to the health of subjects
- Unqualified investigators
- Investigator brochure deficiencies
- Insufficient information in IND to assess subject risks
- Phase II or III study plan or individual study deficient in design

Unless there is immediate and serious risk to patients, the FDA will attempt to discuss and resolve any clinical hold issues prior to imposing the clinical hold. If necessary, the clinical hold is communicated to the sponsor and immediate action is required as described in the clinical hold letter.

Sponsors must submit an official Response to Clinical Hold in order to address the concerns of the agency. Within 30 days, the FDA will provide a response that either removes the hold or describes any outstanding deficiencies. Resumption of the clinical study on hold may not occur until official notification by FDA of its removal.

### 9.7.1.6 IND Sponsor Reporting Requirements

**Safety** Any serious and unexpected drug-related adverse event that occurs during clinical research of a product must be submitted by the drug sponsor (or designee) to the FDA, as well as all investigators participating in trials of the product. Submission of such events must occur within 15 calendar days of the sponsor's awareness of the event. If the event is fatal or life-threatening, the report to the FDA must be submitted within 7 days. Serious adverse events are any untoward medical occurrence that at any dose:

1. Results in death.
2. Is life-threatening.
3. Requires inpatient hospitalization or prolongation of hospitalization.
4. Results in persistent or significant disability/incapacity.
5. Is a congenital anomaly/birth defect.
6. Any other significant medical event that requires intervention to prevent the occurrence of one of the above.

A complete explanation of sponsor responsibilities for submission of serious adverse event reports is found in 21 CFR 312.32.

**Annual Reports** Each year, within 60 days of the anniversary date of the IND becoming effective, an annual report summarizing accumulated safety and other updates to the IND file must be submitted to the FDA. This IND annual report includes:

- A brief summary of the status of each study in progress or completed during the previous year
- A summary of frequent and serious adverse events
- A summary of all safety reports submitted in the previous year
- A list of subjects who died or dropped out of clinical studies due to adverse events
- A summary of new understandings that have been achieved regarding the drug's activity
- A list of nonclinical studies completed or in progress during the previous year and summary of significant findings
- A summary of chemistry and manufacturing changes
- An investigation plan for the upcoming year
- An updated investigator brochure (if changed in the previous year)
- Any unreported phase I protocol amendments
- Any significant foreign marketing developments (such as approvals or withdrawals)

Due to the need to integrate information from several functional groups (chemistry/manufacturing, nonclinical, clinical) into IND annual reports, a central point

within the sponsor's organization is typically identified to coordinate the compilation of this document. This is typically an activity of the regulatory affairs department of a drug sponsor; however, it may take place within a document management center or be prepared by an external vendor, depending on the size and organization of the drug sponsor. IND annual reports are submitted administratively to the IND file at the FDA. Once a product is marketed, NDA annual reports are also required; however, the IND report must continue to be submitted each year after product commercialization.

#### 9.7.1.7 Marketing Approval Requirements

**Pivotal Trials** As discussed earlier in this chapter, discussion and cooperation between a drug sponsor and the FDA is strongly encouraged by the agency to maximize the adequacy of the NDA submission to provide substantial evidence in support of the marketing application, including “adequate and well-controlled” studies that provide the information necessary for the agency to assess the safety and efficacy of the product. These “pivotal” trials will have the statistical power to reach a conclusion regarding the predetermined efficacy endpoint(s) and also contribute significantly to the safety database to allow a well-informed benefit–risk assessment by the agency. Additional studies that examine the drug's effects in special populations (elderly, hepatically impaired, etc.), are also required to elucidate the product's activity and to assist in the evaluation of a drug's safety in a less controlled environment during marketed use.

**NDA Format and Content** The overall format and content of a new drug application as described in the ICH M4 guidance includes the following administrative, chemistry/manufacturing, nonclinical, and clinical information.

#### Module 1: Administrative Information and Prescribing Information

- 1.1 Table of Contents of the Submission Including Module 1
- 1.2 Documents Specific to Each Region (e.g., application forms, prescribing information)

#### Module 2: Common Technical Document Summaries

- 2.1 CTD Table of Contents
- 2.2 CTD Introduction
- 2.3 Quality Overall Summary
- 2.4 Nonclinical Overview
- 2.5 Clinical Overview
- 2.6 Nonclinical Written and Tabulated Summary
  - Pharmacology
  - Pharmacokinetics
  - Toxicology
- 2.7 Clinical Summary
  - Biopharmaceutics and Associated Analytical Methods
  - Clinical Pharmacology Studies
  - Clinical Efficacy
  - Clinical Safety
  - Synopses of Individual Studies

**Module 3: Quality**

- 3.1 Module 3 Table of Contents
- 3.2 Body of Data
- 3.3 Literature References

**Module 4: Nonclinical Study Reports**

- 4.1 Module 4 Table of Contents
- 4.2 Study Reports
- 4.3 Literature References

**Module 5: Clinical Study Reports**

- 5.1 Module 5 Table of Contents
- 5.2 Tabular Listing of All Clinical Studies
- 5.3 Clinical Study Reports
- 5.4 Literature References

The important clinical sections of the application are modules 2.5 (Clinical Overview) and 2.7 (Clinical Summary). An outline of these sections follows:

**2.5 Clinical Overview**

- 2.5.1 Product Development Rationale
- 2.5.2 Overview of Biopharmaceutics
- 2.5.3 Overview of Clinical Pharmacology
- 2.5.4 Overview of Efficacy
- 2.5.5 Overview of Safety
- 2.5.6 Benefits and Risks Conclusions
- 2.5.7 References

**2.7 Clinical Summary**

- 2.7.1 Summary of Biopharmaceutic Studies and Associated Analytical Methods
  - 2.7.1.1 Background and Overview
  - 2.7.1.2 Summary of Results of Individual Studies
  - 2.7.1.3 Comparison and Analyses of Results across Studies
- 2.7.2 Summary of Clinical Pharmacology Studies
  - 2.7.2.1 Background and Overview
  - 2.7.2.2 Summary of Results of Individual Studies
  - 2.7.2.3 Comparison and Analyses of Results across Studies
  - 2.7.2.4 Special Studies
- 2.7.3 Summary of Clinical Efficacy
  - 2.7.3.1 Background and Overview
  - 2.7.3.2 Summary of Results of Individual Studies
  - 2.7.3.3 Comparison and Analyses of Results across Studies
  - 2.7.3.4 Analysis of Clinical Information Relevant to Dosing Recommendations
  - 2.7.3.5 Persistence of Efficacy and/or Tolerance Effects
- 2.7.4 Summary of Clinical Safety
  - 2.7.4.1 Exposure to the Drug
  - 2.7.4.2 Adverse Events

- 2.7.4.3 Clinical Laboratory Evaluations
- 2.7.4.4 Vital Signs, Physical Findings, and Other Observations Related to Safety
- 2.7.4.5 Safety in Special Groups and Situations
- 2.7.4.6 Postmarketing Data
- 2.7.5 References
- 2.7.6 Synopses of Individual Studies

The contents of each of these sections is described in detail within the M4 guidance; however, sponsors should be mindful again that this is a guidance that should be adapted to each drug product in cooperation with the agency during pre-NDA discussions as per the FDA/sponsor meeting guidance.

***FDA Review of Marketing Applications*** Once the NDA is submitted to the FDA, an initial review is conducted to consider if it contains sufficient information to complete their review. The agency then informs the sponsor of the application's acceptance and the review begins. Questions are usually raised by the agency and discussed with the sponsor during the review process, and labeling negotiations take place after the technical review is complete. Prior to completion of the FDA review, public discussion may take place in which one of the FDA advisory committees is assembled to consider specific questions for which the agency requests external advice. Advisory committees have no regulatory power but rather serve as consultants to the agency, which provides the final, formal approval action. Current standard timelines for completion of marketing application review will lead to an FDA action—approval, nonapproval, or approvable (upon completion of specific additional requirements)—within approximately 10 months.

#### **9.7.1.8 Summary**

In this first section of our chapter on regulatory approval, we have identified the basic requirements that need to be addressed in order to initiate clinical trials in the United States and the obligations of drug trial sponsors during the ongoing conduct of a clinical research program. The objective of investigational clinical studies is to achieve marketing approval of a new product and support the commercial positioning of the drug. Toward this end, we have provided a roadmap for discussions with the FDA and the preparation of essential drug applications. Though these concepts and tactics are common to most drug development projects, a customized strategy must be prepared for each new drug, which reflects its intended use and intrinsic characteristics. The FDA has developed guidance and regulations to assist sponsors in addressing most common and several special circumstances, including orphan drugs (which address diseases with small patient populations), “accelerated” or “priority” review, and accelerated development of products for life-threatening diseases not addressed with available therapies. Each of these categories of new drug products requires thorough discussion with the FDA during development in order to agree upon registration requirements. Accelerated development is considered in greater detail in the following section.



## 9.7.2 FAST-TRACK APPROVAL

### 9.7.2.1 Introduction

The Food and Drug Administration Modernization Act of 1997 has an intriguing section (No. 112) entitled Expediting Study and Approval of Fast Track Drugs (the act) [9]. In 1998, the FDA issued a Guidance for Industry: The Fast Track Drug Development Programs (the FTDD programs) [10] to meet the requirement of the act. The purpose of FTDD programs is to “facilitate the development and expedite the review of new drugs that are intended to treat serious or life-threatening conditions and that demonstrate the potential to address unmet medical needs” [9]. The act states that an application for approval of a fast track product may be approved if it is determined that “the product has an effect on a clinical endpoint or on a surrogate endpoint that is reasonably likely to predict clinical benefit” [9].

### 9.7.2.2 Fast Track Is More Than “Priority Review”

The FDA’s Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) have long-standing policies that describe criteria for review priority classification of marketing applications. Products regulated by CBER are eligible for priority review if they provide a significant improvement in the safety or effectiveness of the treatment, diagnosis, or prevention of a serious or life-threatening disease. Products regulated by CDER are eligible for priority review if they provide a significant improvement compared to marketed products in the treatment, diagnosis, or prevention of a disease; eligibility is not limited to drugs for a serious or life-threatening disease. A fast track product would ordinarily meet either the center’s criteria for priority review. Note, however, that for an NDA or BLA (biologic license application) manufacturers need not seek fast-track designation to be eligible for priority review.

### 9.7.2.3 Request for Designation of Drug as Fast-Track Product

Since *fast-track* programs are designed to facilitate the development and expedite the review of new drugs or biologics that are intended to treat serious or life-threatening conditions and that demonstrate the potential to address unmet medical needs, they emphasize the critical nature of close, early communication between the FDA and sponsors. Procedures such as pre-IND and end-of-phase I meetings are convened as ways to improve the efficiency of preclinical and clinical development. These meetings can help FDA and manufacturers reach early agreement on the design of the major clinical efficacy studies that will be needed to support approval. But the manufacturer of a new drug may request the FDA to designate the drug as a fast-track product concurrently with, or at any time after, submission of an application for the investigation of the drug. Within 60 calendar days after the receipt of a request, the FDA shall determine whether the drug that is the subject of the request meets the criteria. If the FDA finds that the drug meets the criteria (i.e., “serious or life-threatening and unmet medical condition” and “valid surrogate endpoint”), it shall designate the drug as a fast-track product and shall take appropriate actions to expedite the development and review of the application for approval

of such product. Under certain circumstances, the FDA may consider reviewing portions of a marketing application in advance of the complete NDA or BLA. Fast-track products may also be eligible to participate in the FDA's Continuous Marketing Applications Pilot 1 or Pilot 2 programs [11]. To facilitate FDA review, a submission for fast-track designation should contain all discussion and supporting documentation necessary to permit a reviewer to assess whether the criteria for fast-track designation are met without having to refer to information located elsewhere, yet should also not be voluminous. The amount of discussion and supporting documentation that shows that the criteria are met will vary. For example, little explanation or supporting documentation may be needed to establish that studying the drug in the treatment of a fatal condition with no approved treatment would qualify if the endpoint were mortality. More extensive explanation and supporting documentation would likely be submitted to show that for a nonfatal condition, serious or life-threatening aspects of the condition will be studied. Where acceptable therapy for the condition already exists, still more extensive discussion and supporting documentation would probably be submitted to establish that the new therapy has the potential to fill a medical need not met by existing therapy.

#### **9.7.2.4 Limitations**

Approval of a fast track product may be subject to the following requirements. The manufacturer needs to: (a) conduct appropriate postapproval studies to validate the surrogate endpoint or otherwise confirm the effect on the clinical endpoint and (b) submit copies of all promotional materials related to the fast track product during the preapproval review period and, following approval and for such period thereafter as the FDA determines to be appropriate, at least 30 days prior to dissemination of the materials. However, the FDA also may withdraw approval of a fast track product using expedited procedures if (a) the manufacturer fails to conduct any required postapproval study of the fast track drug with due diligence; (b) a postapproval study of the fast track product fails to verify clinical benefit of the product; (c) other evidence demonstrates that the fast track product is not safe or effective under the conditions of use; or (d) the manufacturer disseminates false or misleading promotional materials with respect to the product. Therefore, the fast track approval is only a conditional approval for the drug. Nevertheless, the prospect of early availability on the market of the drug has great implications for patients as well as for the drug manufacturer.

#### **9.7.2.5 Accelerated Approval Issues (I): Clinical and Surrogate Endpoints**

Since 1998 many health products have reached patients who suffered from HIV (human immunodeficiency virus), cancer, hepatitis C, and some other diseases sooner by using the fast-track act and the FTDD programs. In the mean time several scientific issues have also surfaced when following the FTDD procedures. The surrogate or intermediate endpoints and their potential in predicting clinical benefit is the major issue [12–15]. Manufacturers whose products are in fast track drug development programs may seek traditional approval based on data demonstrating an effect on clinically meaningful endpoints or well-established surrogate endpoints. Alternatively, they may seek approval under the accelerated approval regulations.

If a manufacturer seeks approval of a product in a fast track drug development program based on evidence of an effect on a less than well-established surrogate endpoint, the FDA may grant accelerated approval based on a determination that the effect on the surrogate endpoint is reasonably likely to predict clinical benefit (21 CFR 314.510 and 601.41). Drug approval under the accelerated approval regulations may also be based on demonstrated clinical effects that are not the desired ultimate benefit but are reasonably likely to predict such benefit (e.g., improved exercise tolerance in refractory heart failure might be considered reasonably likely to predict ultimate benefit) (21 CFR 314.510 and 601.41).

Section 506(b) essentially codifies in statute the FDA's accelerated approval regulations. A surrogate endpoint was defined in the preamble to the accelerated approval rule (57 FR 13234–35, April 15, 1992) as “a laboratory or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and that is expected to predict the effect of the therapy.” Although some surrogate endpoints are recognized as well-established and have long been a basis for approval (e.g., change in blood pressure or cholesterol), the accelerated approval rule allows reliance in specific circumstances on a “surrogate endpoint that, while ‘reasonably likely’ to predict clinical benefit, is not so well-established as the surrogates ordinarily used as bases of approval in the past” (57 FR 58942–44, December 11, 1992). To meet the statutory standard for approval, which requires the submission of “substantial evidence” to demonstrate effectiveness, “there must be evidence from adequate and well-controlled studies showing that the drug will have [its claimed] effect ... That effect will, in this case, be an effect on a surrogate endpoint.” (57 FR 58943–44).

With respect to approval based on clinical endpoints other than survival or irreversible morbidity, the preamble to the final accelerated approval rule pointed out that such approval would usually be considered (like other approvals based on a clinical finding) under traditional procedures (i.e., not under accelerated approval). Approval based on clinical endpoints other than survival or irreversible morbidity would “be considered under the accelerated approval regulations only when it is essential to determine effects on survival or irreversible morbidity in order to confirm the favorable risk/benefit judgment that led to approval” (57 FR 58946). The following examples illustrate types of clinical endpoints that could be a basis for approval with a requirement for further studies under the provisions of the modernization act and the accelerated approval rule:

- Clinical endpoints measuring short-term benefit in a chronic condition where short-term benefit per se does not outweigh risk and where durability of benefit is uncertain but expected.
- Clinical endpoints measuring lesser symptoms or signs of a serious disease (e.g., weight loss, appearance) when the resulting benefits do not per se outweigh risks but are expected to lead to a favorable effect on ultimate outcome, which would outweigh risks.
- Clinical endpoints measuring substantial benefits otherwise suitable for ordinary approval but where there exists a significant but limited concern that the

treatment may adversely affect ultimate outcome. Where such concerns are minimal, ordinary approval would be used. Where the concerns are substantial, data regarding ultimate outcome would be required preapproval. Between these extremes, accelerated approval may be considered.

#### **9.7.2.6 Accelerated Approval Issues (II): Postapproval Study**

The required postapproval study of the fast track drug has a broadened definition. That is, the postapproval study does not have to be a separate study but can be a proper continuation/extension of the same study upon which the early submission for accelerated approval is based. This use of the same study for both accelerated and final approval has been a common practice in HIV. An obvious advantage of such a proper continuation of the same study, instead of initiating a new study, is that much time and resource can be saved. It is also conceivable that in certain situations of severe diseases (such as pancreatic cancer) continuation/extension of the same trial, often very large in sample size, is the only feasible way to study the treatment effect under investigation. This broadened definition may also fit in what has been discussed in the literature [16] as “seamless phases II and III,” except that the “phases” may not be labeled as such in this case. However, there are design and conduct considerations to continue a study properly so as to ensure that the final submission of the product satisfies the usual requirements of the traditional approval. For example, the protocol has to design ways to avoid excessive “crossovers” in the continuation to maintain the integrity of the trial for the later definitive endpoint.

#### **9.7.2.7 Accelerated Approval Issues (III): Control of Type I Error Rate**

Controlling the probability of type I error is an important issue from the regulatory and consumer standpoints. It should be recognized that the tentative, conditional approval of a fast track drug requires control of a different kind of type I error rate and that the level of control also needs to be considered. This stems from the fact that often an intermediate (surrogate) endpoint is used at the fast track review and the final approval is based on the confirmation of the treatment effect on the clinical (primary) endpoint.

Since the objective and endpoint often change between two possible approvals according to the design, this raises some interesting questions: Do we view the fast track approval submission as an interim analysis of the whole study or view the final analysis as a “poststudy extension” (i.e., the fast track submission is the main study)? If the fast track approval submission is an interim analysis, how does this type of interim analysis differ from the usual ones? What is the implication of the type I error rate when the first accelerated approval is only a conditional one? How should the type I error rate(s) be calculated and controlled? These issues have been discussed in Shih et al. [17].

The practice of continuation or extension of a study is not uncommon in clinical trials. But the current situation has a unique characteristic in that the continuation is to investigate a more important clinical endpoint (alias the primary endpoint), which requires a longer time to manifest than the surrogate or intermediate endpoint. When the study is to continue for a different mission, it is natural to view the

latter part as an extension. If the latter part is an extension, the first part for the expediting approval must be the main part. Therefore, the type I error rate for the surrogate or intermediate endpoint at the accelerated approval submission (the main study) and that for the primary endpoint at the final submission (the “internal” extension study) may be controlled separately, each at the conventional level.

Another popular view of this situation is that this is a two-stage design with one interim analysis. After all, the first part is an “early” submission. However, it should be noticed that there are differences in the current situation from the usual interim analyses in the group sequential design setting, at least from the outset.

In the usual group sequential design setting, study monitoring is based on a single endpoint. Here, in this situation two endpoints are involved in the sequential monitoring process. Furthermore, the two endpoints are “ordered” differently by their relative emphasis and expected outcomes at the two stages. The usual sense of overall (or experiment-wise) type I error rate means the probability of falsely rejecting any true hypothesis. However, this concept is not useful for the situation where one hypothesis ultimately dominates the other at a later time point with more data.

Shih et al. [17] showed that the final approval type I error rate only involves the primary clinical endpoint. The next question is whether the tentative, conditional approval for an early release of the drug on the market should be protected by controlling another false-positive rate prior to the confirmation. If affirmative, how much? There is no definitive answer to the above question since the answer may well be dependent on how much risk the consumers (patients) need to be protected from using a yet-to-be confirmed drug. That in turn depends on how restrictive the early use of the drug would be. It is reasonable to suggest that, if the time gap between the conditional early approval and the confirmatory final approval is short and that the use of the drug is limited by the label during this gap time, the allowance for the tentative, conditional approval type I error rate could be greater (i.e., more relaxed) than the situation of a wider use of the drug during a longer gap time. In any case, this is clearly another kind of type I error rate, which should not be confused with the one for the final approval, and Shih et al. [17] showed that these two error rates should not be added directly. Specifically, denote

$\alpha_F$  = final approval type I error rate

$\alpha_E$  = early submission type I error rate

Since the early submission can also be the final submission, the above two error rates overlap. Subtracting this overlap,

$\alpha_A$  = tentative, conditional accelerated approval type I error rate during the gap time

Shih et al. [17] suggested considering the following scenarios:

1. Control  $\alpha_F$  at 0.05 level.
2. Control  $\alpha_F$  at 0.05 level and, in addition, control  $\alpha_E$  also at 0.05 level.

3. Control  $\alpha_F$  at 0.05 level and, in addition, control  $\alpha_A$  at 0.01 level.
4. Control  $\alpha_F + \alpha_A$  at 0.05 level, for example,  $\alpha_F = 0.04$  and  $\alpha_A = 0.01$ .

Scenario 1 views that the ultimate approval is the final conclusion of the drug's efficacy, hence only  $\alpha_F$  needs to be controlled at the conventional 5% level. Scenario 2 views that the continuation of the trial is an extension study. Each section of the submission process is controlled at the 5% level. Scenario 3 controls the final submission at the conventional 5% and the tentative conditional type I error rate at 1% (implying to impose a stricter risk protection for the accelerated approval). Scenario 4 is the most conservative way of controlling the accelerated and final approvals together at the 5% rate.

In the approaches described above, the clinical and the surrogate endpoints may be controlled at different significance levels at the first stage. This is an advantage in terms of design flexibility, but not necessarily in terms of statistical power. Another approach is to test the endpoints in an ordered fashion at the interim stage so that we may use the primary and secondary nature of the endpoints and gain some power over scenario 4. This is described briefly as follows. At stage 1, we will test the surrogate endpoint first, then the clinical endpoint, since the former is considered more hopeful to show a beneficial effect in terms of statistical power. If both endpoints are "significant" at the first stage, we will stop the trial. Otherwise, we will continue. At the second stage, we reverse the order, test the clinical endpoint, then the surrogate endpoint, since the clinical endpoint is more important in the clinical sense. Compared to the previous approaches, the surrogate and the clinical endpoints are tested at about the same significance level at the first stage (e.g., 0.01) and the second stage (e.g., 0.04). This is a result of the ordering of the test and is set according to the relative importance of the endpoints at the two stages. The strategy may be viewed as an application of Holms generalized weighted procedure [18].

Other scenarios will require different statistical strategy for the type I error rates protection. For example, it might be reasonable for some cases to require early termination of the trial if the secondary endpoint is negatively significant while the primary endpoint is not at the interim stage. Conditional power of the clinical endpoint based on the result of the surrogate endpoint may also be evaluated.

#### 9.7.2.8 Examples of FTDDP

Table 1 lists the fast-track products approved by CDER since 1998 through March 31, 2005.

As discussed previously, one of the crucial questions in the fast-track approval program is how consumers should be protected from a yet-to-be confirmed but promising drug. It would be interesting to research the statistics of the time for drug manufacturers to fulfill the final confirmation submission after receiving their fast track approvals and the number of failures and successes of the confirmation submission for the list in Table 1. This would give us some idea as to how much protection there should be in comparison with the usual 5% type I error rate required for a regular new drug application.

**TABLE 1 Fast-Track Products Approved by CDER (1998–March 31, 2005)**

| NDA or BLA #       | Proprietary Name<br>(Established or Proper Name)                                                 | Applicant              | Approval Date          | Use                                                                                                                                                                                 |
|--------------------|--------------------------------------------------------------------------------------------------|------------------------|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| BL103964<br>5039.0 | Pegasys (peginterferon<br>alfa-2a)                                                               | Hoffmann–La<br>Roche   | 2/25/2005              | Alone or in combination with ribavirin (Copegus) for treatment<br>of chronic hepatitis C in patients co-infected with HIV, who<br>have clinically stable HIV disease                |
| N21841             | Lamivudine 150 mg;<br>zidovudine 300-mg tablets<br>co-packaged with<br>nevirapine 200-mg tablets | Aspen Pharmacare       | 1/24/2005              | Treatment of HIV                                                                                                                                                                    |
| N21660             | Abraxane (paclitaxel protein-<br>bound particles for<br>injectable suspension)                   | American<br>BioScience | 1/7/2005               | Treatment of breast cancer after failure of combination<br>chemotherapy for metastatic disease or relapse within 6<br>months of adjuvant chemotherapy                               |
| N21673             | Clolar (clofarabine)                                                                             | Genzyme                | 12/28/2004             | Treatment of pediatric patients 1–21 years old with relapsed or<br>refractory acute lymphoblastic leukemia after at least two<br>prior regimens                                     |
| N21756             | Macugen (pegaptanib<br>sodium)                                                                   | Eyeteck Pharms         | 12/17/2004             | Treatment of exudative (wet) age-related macular degeneration                                                                                                                       |
| BL125103           | Kepivance (palifermin)                                                                           | Amgen                  | 12/15/2004             | Decrease the incidence and duration of severe oral mucositis in<br>patients with hematologic malignancies receiving myelotoxic<br>therapy requiring hematopoietic stem cell support |
| N20635<br>SE1-035  | Levaquin (levofloxacin)                                                                          | Ortho McNeil           | 11/24/2004             | Treatment of inhalational anthrax (postexposure)                                                                                                                                    |
| N20634<br>SE1-035  | Levaquin (levofloxacin)                                                                          | Ortho McNeil           | 11/24/2004             | Treatment of inhalational anthrax (postexposure)                                                                                                                                    |
| N21743             | Tarceva (erlotinib)                                                                              | OSI Pharms             | 11/18/2004             | Treatment of locally advanced or metastatic non-small-cell lung<br>cancer after failure of at least one prior chemotherapy<br>regimen                                               |
| N21677<br>N50794   | Alimta (pemetrexed)<br>Vidaza (azacitidine)                                                      | Lilly<br>Pharmion      | 8/19/2004<br>5/19/2004 | Treatment of non-small-cell lung cancer<br>Treatment of patients with certain subtypes of myelodysplastic<br>syndrome                                                               |
| N21264             | Apokyn (apomorphine)                                                                             | Bertek                 | 4/20/2004              | Treatment of acute, intermittent hypomobility, “off” episodes<br>(“end-of-dose wearing off” and unpredictable “on/off”<br>episodes) associated with advanced Parkinson’s disease    |
| BL125085           | Avastin (bevacizumab)                                                                            | Genentech              | 2/26/2004              | For the first-line treatment of patients with metastatic carcinoma<br>of the colon and rectum (in combination with intravenous<br>5-fluorouracil-based chemotherapy)                |



|                              |                                                                                                     |                                                    |                                    |                                                                                                                                                                                                                                                                                                                                |
|------------------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| BL125084                     | Erbitux (cetuximab)                                                                                 | ImClone Systems                                    | 2/12/2004                          | Treatment of EGFR-expressing, metastatic colorectal carcinoma in patients who are refractory to irinotecan-based chemotherapy (in combination with irinotecan); treatment of EGFR-expressing, metastatic colorectal carcinoma in patients who are intolerant to irinotecan-based chemotherapy (administered as a single agent) |
| N21462<br>N21548             | Alimta (pemetrexed disodium)<br>Lexiva (fosamprenavir calcium)                                      | Lilly<br>GlaxoSmithKline                           | 2/4/2004<br>10/20/2003             | Treatment of malignant pleural mesothelioma<br>Treatment of HIV                                                                                                                                                                                                                                                                |
| N21348<br>N21500<br>BL125011 | Zavesca (miglustat)<br>Emtriva (emtricitabine)<br>Bexxar (tositumomab and iodine I 131 tositumomab) | Actelion Pharms<br>Gilead<br>Corixa Corporation    | 7/31/2003<br>7/2/2003<br>6/27/2003 | Treatment of type I Gaucher disease<br>Treatment of HIV<br>Treatment of patients with CD20 positive, follicular, non-Hodgkin's lymphoma, with and without transformation, whose disease is refractory to Rituximab and has relapsed following chemotherapy                                                                     |
| N21602<br>N21399<br>BL125058 | Velcade (bortezomib)<br>Iressa (gefitinib)<br>Aldurazyme (laronidase)                               | Millennium Pharms<br>AstraZeneca<br>Biomarin Pharm | 5/13/2003<br>5/5/2003<br>4/30/2003 | Treatment of relapsed/refractory multiple myeloma<br>Treatment of non-small-cell lung cancer<br>Treatment of patients with Hurler and Hurler-Scheie forms of mucopolysacchar- idosis I (MPS I) and for patients with the Scheie form who have moderate to severe symptoms                                                      |
| BL103979                     | Fabrazyme (agalsidase beta)                                                                         | Genzyme Corporation                                | 4/24/2003                          | Use in patients with Fabry disease to reduce globotriaosylceramide (GL-3) deposition in capillary endothelium of the kidney and certain other cell types                                                                                                                                                                       |
| N21106                       | Somavert (pegvisomant)                                                                              | Pharmacia and Upjohn                               | 3/25/2003                          | Treatment of acromegaly                                                                                                                                                                                                                                                                                                        |
| N21481                       | Fuzeon (enfuvirtide)                                                                                | Hoffmann-La Roche                                  | 3/13/2003                          | Treatment of HIV                                                                                                                                                                                                                                                                                                               |
| BL103964 5000                | Pegasys (peginterferon alfa-2a)                                                                     | Hoffmann-La Roche                                  | 12/3/2002                          | PEGASYS, peginterferon alfa-2a, alone or in combination with COPEGUS, for the treatment of adults with chronic hepatitis C virus infection who have compensated liver disease and have not been previously treated with interferon alpha                                                                                       |
| N20541<br>SE1-010            | Arimidex (anastrozole)                                                                              | AstraZeneca                                        | 9/5/2002                           | Adjuvant treatment of early breast cancer in postmenopausal women                                                                                                                                                                                                                                                              |
| N21492<br>BL125019           | Eloxatin (oxaliplatin)<br>Zevalin (ibritumomab tiuxetan)                                            | Sanofi- Synthelabo<br>IDEC Pharms                  | 8/9/2002<br>2/19/2002              | Treatment of colorectal cancer<br>Treatment of patients with relapsed or refractory low-grade, follicular, or transformed B-cell non-Hodgkin's lymphoma (not acc. app.); including patients with Rituximab (Rituxan) refractory follicular non-Hodgkin's lymphoma                                                              |
| N21232                       | Orfadin (nitisinone)                                                                                | Swedish Orphan                                     | 1/18/2002                          | Treatment of hereditary tyrosinemia type I                                                                                                                                                                                                                                                                                     |

TABLE 1 *Continued*

| NDA or BLA #                    | Proprietary Name<br>(Established or Proper Name)                         | Applicant                             | Approval Date                       | Use                                                                                                                                                                                                |
|---------------------------------|--------------------------------------------------------------------------|---------------------------------------|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| N21356                          | Viread (tenofovir disoproxil fumarate)                                   | Gilead                                | 10/26/2001                          | Treatment of HIV                                                                                                                                                                                   |
| N21335<br>BL103948              | Gleevec (imatinib mesylate)<br>Campath (alemtuzumab)                     | Novartis Pharms<br>ILEX Pharms        | 5/10/2001<br>5/7/2001               | Treatment of chronic myeloid leukemia<br>Treatment of B-cell chronic lymphocytic leukemia (B-CLL) in patients who have been treated with alkylating agents and who have failed fludarabine therapy |
| N21227                          | Candidas (caspofungin acetate)                                           | Merck                                 | 1/26/2001                           | Treatment of aspergillus infections for patients who are refractory to or intolerant of other therapies                                                                                            |
| BL103772<br>1007.0              | Remicade (infliximab)                                                    | Centocor                              | 12/29/2000                          | Expand the indication to include the inhibition of progression of structural damage in patients with rheumatoid arthritis who have had an inadequate response to methotrexate                      |
| N21248<br>N21226                | Trisenox (arsenic trioxide)<br>Kaletra (lopinavir/ritonavir)<br>Capsules | Cell Therapeutics<br>Abbott Labs      | 9/25/2000<br>9/15/2000              | Treatment of acute promyelocytic leukemia<br>Treatment of HIV                                                                                                                                      |
| N21251                          | Kaletra (lopinavir/ritonavir)<br>Oral Solution                           | Abbott Labs                           | 9/15/2000                           | Treatment of HIV                                                                                                                                                                                   |
| BL103836<br>1001.0              | Actimmune (interferon gamma-1b)                                          | InterMune                             | 2/10/2000                           | Delaying time to disease progression in patients with severe, malignant osteopetrosis                                                                                                              |
| N20449<br>SE1-011               | Taxotere (docetaxel)                                                     | Aventis Pharms                        | 12/23/1999                          | For the treatment of locally advanced or metastatic non-small-cell lung cancer                                                                                                                     |
| N21007                          | Agenerase (amprenavir)<br>Capsules                                       | Glaxo Wellcome                        | 4/15/1999                           | Treatment of HIV                                                                                                                                                                                   |
| N21039                          | Agenerase (amprenavir) Oral<br>Solution                                  | Glaxo Wellcome                        | 4/15/1999                           | Treatment of HIV                                                                                                                                                                                   |
| N20978                          | Ziagen (abacavir sulfate) Oral<br>Solution                               | Glaxo Wellcome                        | 12/17/1998                          | Treatment of HIV                                                                                                                                                                                   |
| N20977                          | Ziagen (abacavir sulfate)<br>Tablets                                     | Glaxo Wellcome                        | 12/17/1998                          | Treatment of HIV                                                                                                                                                                                   |
| BL103795<br>BL103792<br>N020972 | Enbrel (etanercept)<br>Herceptin (trastuzumab)<br>Sustiva (efavirenz)    | Immunex<br>Genentech<br>DuPont Pharms | 10/2/1998<br>9/25/1998<br>9/17/1998 | Treatment of rheumatoid arthritis<br>Treatment of metastatic breast cancer<br>Treatment of HIV                                                                                                     |

## REFERENCES

1. Title 21 Code of Federal Regulations—Food and Drugs; available at: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm>.
2. U.S. Department of Health and Human Services (1988), Guidelines for the Format and Content of the Clinical and Statistical Sections of New Drug Applications. U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Rockville, MD.
3. U.S. Department of Health and Human Services (1998), Guidance for Industry Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration.
4. International Conference on Harmonisation Guideline M4; Common Technical Document for the Registration of Pharmaceuticals for Human Use.
5. Anello, C., O'Neill, R. T., and Dubey, S. (2005), Multicentre trials: A US regulatory perspective, *Statist. Methods Med. Res.*, 14, 303–318.
6. U.S. Department of Health and Human Services (1995), Guidance for Industry on the Content and Format of Investigational New Drug Applications (IND) for Phase I Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-derived Products, U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Rockville, MD.
7. U.S. Department of Health and Human Services (2000), Formal Meetings with Sponsors and Applicants of PDUFA Products, U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Rockville, MD.
8. International Conference on Harmonization Guideline M3; Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals.
9. U.S. Department of Health and Human Services (1997), Section 112 of the Food and Drug Administration Modernization Act of 1997: Expediting Study and Approval of Fast Track Drugs; available at: <http://www.fda.gov/cder/guidance/index.htm>.
10. U.S. Department of Health and Human Services (1998), Guidance for Industry: Fast Track Drug Development programs—Designation, Development, and Application Review; available at: <http://www.fda.gov/cder/guidance/index.htm>.
11. U.S. Department of Health and Human Services (2004), Guidance for Industry Fast Track Drug Development Programs—Designation, Development, and Application Review Procedural Revision 1. U.S. Department of Health and Human Services, Public Health Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER).
12. Ellenberg, S., and Hamilton, J. M. (1989), Surrogate endpoints in clinical trials: Cancer, *Statist. Med.*, 8, 405–414.
13. Fleming, T. R., Prentice, R. L., Pepe, M. S., and Glidden, D. (1994), Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research, *Statist. Med.*, 13, 955–968.
14. Fleming, T. R. (1994), Surrogate markers in AIDS and cancer trials (with discussions), *Statist. Med.*, 13, 1423–1435.
15. Neaton, J. D., Wentworth, D. N., Phame, F., Hoagan, C., Abrams, D. I., and Deyton, L. (1994), Considerations in choice of a clinical endpoint for AIDS clinical Trials, *Statist. Med.*, 13, 2107–2125.
16. Chi, G., Hung, H. M. J., and O'Neill, R. O. (2001), Some comments on “Adaptive trials and Bayesian Statistics in Drug Development” by Donald A. Berry, *Biopharmaceut. Rep.*, 9, 7–9.

17. Shih, W. J., Ouyang, P., Quan, H., Lin, Y., Michiels B., and Bijmens, L. (2003), Controlling type I error rate for fast track drug development programmes, *Statist. Med.*, 22, 665–675.
18. Holm, S. (1976), A simple sequentially rejective multiple test procedure, *Scan. J. Statisto.*, 65–70.

# 9.8

---

## *New Paradigm for Analyzing Adverse Drug Events\**

**Ana Szarfman, Jonathan G. Levine, and Joseph M. Tinning**

*Food and Drug Administration, CDER, Silver Spring, Maryland*

### **Contents**

|         |                                                                                                 |     |
|---------|-------------------------------------------------------------------------------------------------|-----|
| 9.8.1   | Introduction                                                                                    | 374 |
| 9.8.2   | Current Paradigms of Analysis                                                                   | 374 |
| 9.8.3   | Why We Need a Paradigm Change                                                                   | 375 |
| 9.8.4   | A New Paradigm: Informatics                                                                     | 376 |
| 9.8.4.1 | Data Standards and Interoperable Systems                                                        | 377 |
| 9.8.4.2 | High-Quality Data                                                                               | 378 |
| 9.8.4.3 | Restructuring Capabilities                                                                      | 383 |
| 9.8.4.4 | Data Analysis                                                                                   | 385 |
| 9.8.4.5 | Reproducibility                                                                                 | 389 |
| 9.8.4.6 | Maintenance                                                                                     | 389 |
| 9.8.5   | Examples of Practical Computer-Intensive Tools for Systematically Assessing<br>Drug Safety Data | 389 |
| 9.8.5.1 | Background                                                                                      | 389 |
| 9.8.5.2 | Analysis of Premarketing Data with WebSDM                                                       | 390 |
| 9.8.5.3 | Analysis of Postmarketing Data with MGPS and HBLR                                               | 390 |
| 9.8.5.4 | Other Data Resources                                                                            | 394 |
| 9.8.5.5 | Validation of New Methods                                                                       | 394 |
| 9.8.6   | Conclusions                                                                                     | 394 |
|         | References                                                                                      | 395 |

\*The opinions expressed in this chapter are solely those of the authors and do not necessarily reflect those of the United States Food and Drug Administration. This chapter originally appeared in *Computer Applications in Pharmaceutical Research and Development*, edited by Sean Ekins, Copyright © 2006 John Wiley & Sons, ISBN: 978-0-471-73779-7.

*Clinical Trials Handbook*, Edited by Shayne Cox Gad  
Copyright © 2009 John Wiley & Sons, Inc.

### 9.8.1 INTRODUCTION

Only a generation ago, 107 people in 15 states died within a few weeks after a new drug was placed on the market. Many of the victims were children. One victim was the best friend of the doctor who had prescribed the drug for him. The S.E. Massengill Company, which marketed the drug, had been looking for a solvent to dissolve sulfanilamide, a new antibiotic. A company chemist chose diethylene glycol, a chemical normally used as antifreeze. Diethylene glycol was effective in dissolving sulfanilamide but caused renal failure in the unsuspecting patients. The company owner shirked responsibility, stating, “We have been supplying a legitimate professional demand and not once could have foreseen the unlooked-for results.” And how could they have foreseen the results? One doesn’t find what one doesn’t look for. The drug was completely legal. At that time, in 1937, the U.S. Food and Drug Administration (FDA) did not require drug products to be tested for safety [1].

The passage of the Food, Drug, and Cosmetic Act in 1938 greatly helped to improve drug safety. However, challenges in accessing and understanding drug safety data in real time persisted. Computer technology was decades away in 1938. The FDA did not develop a computerized repository of postmarketing voluntary adverse event reports until 1968. FDA scientists analyzed adverse events with paper, pen, typewriter, and perhaps a mechanical calculator. Even in the 1980s, FDA scientists had little more at their disposal to review adverse event data than a typewriter or dedicated word processor. In the 1990s, personal computers and software programs made it possible to eliminate many of the paper processes involved in adverse event analysis. However, there were still no uniform data standards and interoperable systems in place, hindering efforts to truly analyze adverse events in a systematic, computerized way. To rectify this situation, the FDA and the pharmaceutical industry began constructing a computerized repository of premarketing and postmarketing clinical trial data that would enable more efficient data analysis and decision making.

### 9.8.2 CURRENT PARADIGMS OF ANALYSIS

Challenges in systematically analyzing drug safety data continue to the present day. The typical product of a traditional analytical method is a static, paper report. Such reports usually consist of a vast number of discrete, personal, ad hoc processes that cannot typically be used to perform subsequent comprehensive *reproducible* analyses as soon as additional analyses are needed. Having an analytical method that can be subsequently audited and *reproduced* is absolutely critical because essential drug safety decisions are made from these analyses. Unfortunately, many drug safety organizations still verify the accuracy of the data they analyze by manual, ad hoc methods when comparing the data stored in a database with the primary medical records. Auditing is done by a second-party review, again by manual ad hoc methods.

Current computerized analyses of adverse events still typically consist of a vast number of discrete, often personal, ad hoc processes that mimic paper and pencil methods. Some commercial-off-the-shelf (COTS) software tools (e.g., Adobe Acrobat, Microsoft Word, Excel) do have the capability to search for specific terms

in electronic documents/case reports and do have navigational tools with hyperlinks and full-text indexing that enable researchers to create their own hyperlinks. Some other COTS software tools (e.g., SAS, Excel, Access, JMP) allow importation of electronic case report tabulations (ECRT) for more detailed analysis.

However, many of these tools, while enabling markedly faster and more detailed analysis than paper-based methods, still mimic static, one-by-one “paperlike” reports with no real-time auditing capability. Moreover, these COTS do not have integrated data analysis and automated data screening capabilities and are not optimized for systematic analyses. Furthermore, the ad hoc analyses that these COTS produce lack interactive, automatic auditing *reproducible* functions. Thus these tools are often used to produce the same dense, unwieldy paper tables of counts and percentages that were created manually before personal computers became ubiquitous.

Humans should use computers to do functional work for them in the most efficient manner possible. However, we must not delude ourselves into thinking that the mere use of a computer to analyze adverse events will magically analyze these events in a systematic, efficient way. Computers do not *automatically* produce coherent, auditable results that can be subsequently reproduced with ease. Computers must be *actively* programmed through an iterative process involving tight communication between analysts and software developers until these processes are totally functional.

### 9.8.3 WHY WE NEED A PARADIGM CHANGE

A new paradigm for computer-assisted analysis of adverse drug events is sorely needed. Critics of the current paradigm emphasize the need for more transparency in the data review process, claiming that there is the potential for suppressing negative results or for hiding safety issues in both interventional and observational studies. It is often necessary to reanalyze the data in light of new information about a particular adverse event or class of events for a drug or class of drugs months—or even years—after the initial analysis. In many such situations, it is critical to have ready access to all preclinical, clinical trial, and postmarketing data. The findings from these updated analyses are potentially so influential that they can impact drug therapy recommendations for decades. However, ready access to all of the actual data and results is still lacking in many situations.

To properly assess drug safety, we must be able to systematically tap the information captured in the massive amounts of medical data collected in both premarketing and postmarketing settings. In addition, as stated above, we must also be able to *reproduce* findings in different repositories of medical data in an auditable way. However, two major issues in studying drug safety confront us.

The first issue lies in the whole realm of the human disease process itself. Many adverse drug events mimic diseases and vice versa. Is an “adverse event” really an adverse event, or is it merely a natural occurrence of a disease process that is entirely independent of drug exposure? The science of drug safety is often complicated by the lack of objective markers of drug toxicity that can systematically separate a disease process from an adverse drug event process [2]. Clinical trials, often viewed as the gold standard to assess efficacy, are simply too limited in scope to answer safety questions in a systematic way.



The second issue involves the whole process of data collection, transformation, and presentation. At the study level, important information needed to assess the safety of a new drug is often presented in idiosyncratic ways. For example, concomitant medications are often not translated into standard drug names, and there are often subtle errors in coding of events (which we discuss further). This lack of standards hinders the creation of an integrated safety database. Later, the data that may come from numerous preclinical, clinical, and postmarketing studies are often not collected with a common data standard and are not systematically integrated into a single cumulative database before analysis. Various personnel working in different organizations or in different sections of the same organization may perform analytical tasks with nonstandardized and nonintegrated data. Even when the premarketing adverse event data of a new drug *are* incorporated into an integrated summary of clinical trial safety data, the *totality* of the safety data from *all* pre- and postmarketing research are usually *not* integrated into a coherent, analyzable database that can be used for a comprehensive, real-time analysis.

It is therefore easy to see why this current drug safety paradigm, with its lack of standards in data collection and analysis, hinders the analysis of adverse events. Without data standards in place, it is difficult to build practical, reusable tools for systematic safety analysis. With no standard tools, truly standardized analyses cannot occur. Reviewers may forget their initial analytical processes if they are not using standardized data and tools. Comprehensive reproducibility and auditability, therefore, become nearly impossible. In practice, the same data sets and analytical processes cannot be easily reused, even by the same reviewers who produced the original data sets and analyses. Not using standardized tools slows the real-time systematic analysis and reanalysis of the data because of the large number of restructuring steps needed to perform these analyses and reanalysis. These obstacles restrict the ability of analysts and senior decision makers to gain a full understanding of the entire data and the results in a timely manner. The end result is that analysts cannot routinely access the *computerized* preclinical and clinical data that supported the marketing approval of a drug, a new indication, or a new dosage schedule.

#### 9.8.4 A NEW PARADIGM: INFORMATICS

We need to transition from quasi-computerized methods, in which the different elements of the analytical process are treated as discrete, “paper report” tasks, to a comprehensive *informatics approach*, in which the entire data collection and analysis is considered as a single reusable, extensible, auditable, and reproducible system. Informatics can be defined as the science of “storing, manipulating, analyzing, and visualizing information using computer systems”[3].

To analyze adverse drug events from an informatics framework, six major infrastructure elements must be in place:

- *Data Standards and Interoperable Systems* When interoperability is in place, standard, automated software tools for systematically analyzing the data can be constructed.

- *High-Quality Data* Miscoding, duplicate records, missing data items, and other data problems must be kept to a minimum.
- *Restructuring Capabilities* There must be the capability of restructuring the data for various types of analyses in a transparent way.
- *Systematic Analysis* The data must be systematically analyzed to gain insight regarding the associations between various treatments and medical conditions. This knowledge can assist in causality assessments.
- *Reproducible Capabilities* The data and analyses must be electronically accessible in real time, easily reanalyzable, and easily reproduced, even years after the adverse event data were collected.
- *Maintenance* The database and software that comprise the data systems must be maintained.

#### 9.8.4.1 Data Standards and Interoperable Systems

The foundation of any efficient computer-assisted data analysis system is the creation and use of data standards. Data standards consist of standard data file names for each predefined file, standard data elements in each data file, standardized names for each data element, and standard definitions for each data element.

Data standards are exceedingly important because they allow for the use of “interoperable systems.” The concept of interoperable systems can be illustrated by the situation that prevailed in the American railroad industry during the nineteenth century. At that time, there were many small, local railroad companies throughout the country. Each company, however, utilized different standards of rail gauges, that is, the distance between the two parallel rails on a track. The Baltimore and Ohio Railroad used a 4-foot, 8.5-inch gauge; railroads in the South used a 5-foot, 0-inch gauge; the Erie and Lackawanna Railroad used a 6-foot, 0-inch gauge, and so forth. The onset of the Civil War exposed the absurdity of the lack of rail gauge standards when such inconsistencies made it nearly impossible to accomplish very fundamental tasks, such as materiel transport [4]. At the 2005 FDA Science Forum, Secretary of Health and Human Services Michael O. Leavitt recounted the story of rail gauge inconsistencies to emphasize the need for interoperability to advance health information technology [5]. It becomes obvious, therefore, why standardized, unique codes for all data values across all stages of drug development and during the entire postmarketing period are so essential. Comprehensive data standards allow for the creation and systematic implementation of reusable software for analyzing medical records and adverse event data. These reusable tools enable the creation of standard analysis tools that can be shared among safety analysts for enhanced communication and learning and refined as existing analytical techniques are evaluated and extended.

All data fields (patient identification, date, sex, drug names, narratives, etc.) require rules regarding the data, because some characters, such as tabs and hard return characters, can interfere with attempts to store, read, and/or reorganize the data. For example, if the data are being migrated from Oracle into other data analysis software and data fields in Oracle contain tabs and hard return characters, the data from these fields may be split into several columns and rows upon migration into the other software. This fact is extremely important because information that

is misplaced in a data file may not be readily apparent to the user without the use of systematic approaches for analysis. This missing information may have profound and unpredictable influences at all levels of drug safety analysis, including studies using preapproval and postapproval pharmacoepidemiological data.

Ideally, the data from the whole drug development program, including accumulating postmarketing data, would be integrated into a single database. This integration would simplify the process of answering important, but previously *unforeseeable* questions (remember the sulfanilamide example). For instance, is a new dose increase as safe as the previously approved dosages? Is it enough to compare the two highest dosages with each other? A better answer could be obtained if the entire integrated clinical trial data across all dosages are studied. However, as mentioned above, drug safety data from multiple sources are not usually integrated into a single database. Compounding this problem is that many people tasked with managing data are focused on preparing a small portion of the data for a specific purpose and are not trying to create a single database that integrates all of the data. Integration of the entire safety data for a drug would enable better and faster communication among decision makers in various organizations (industry, regulatory agencies, etc.)

9.8.4.2 High-Quality Data

It is essential to have high-quality data in place for interoperable systems to function efficiently. Standard data structures can only be used to full advantage if they are combined with standard terminology for values populating a data element. Yet there are many potential pitfalls in data collection and configuration for analysis. Some of the more common pitfalls are discussed here, but this list is by no means comprehensive.

**Errors and Inconsistencies in Patient Identification** Whatever system is used to identify patients in a database, it is essential that a single, unique identifier be used for each patient. This identifier must be consistent throughout the database. Errors and inconsistencies in patient identification can significantly interfere with adverse event analysis. Examples include situations in which hyphens, commas, additional zeros, or other characters are not used consistently for identifying patients. In some cases, the *same* patient in a study may have *different* patient identification numbers listed in various tables [5] (Table 1). In such situations, correctly merging patient data from multiple sources/tables becomes essentially impossible because some of the data for the same patient either will be treated as missing or will appear as data for the wrong patient. In one new drug application (NDA) submitted to the FDA, a sponsor built a “unique patient identifier” for some tables by concatenating the study identification number plus the study site identification number plus the

TABLE 1 Example of Different Types of Unique Patient Identifiers for the Same Patients in a Clinical Trial

| Unique Patient Identifier in Data Tables | Unique Patient Identifier in Narrative Table              |
|------------------------------------------|-----------------------------------------------------------|
| 8023007                                  | Patient 08-023-007 is a 48-year-old Caucasian male, etc.  |
| 8031011                                  | Patient 08-031-011 is a 61-year-old Hispanic female, etc. |

patient within-site identification number, whereas in other tables the numbers were concatenated in a different sequence. In another NDA, a sponsor separated the concatenated identifiers by hyphens in some tables but not in others. In yet another NDA, a sponsor concatenated numeric with character identifiers (the latter with leading zeroes) in some tables, but not in others. In our experience, such problems are easier to detect when systematic approaches for analysis are in place.

***Errors and Inconsistencies in Categorical Variables*** For categorical variables such as sex, race, diagnosis, and the like it is essential that the values used to designate the different categories be precisely defined and consistently used. Precise variable value definitions and their consistent use greatly simplify the analysis at hand and also future analyses. For example, a human being can readily tell that “M,” “m,” “Male,” “male,” “hombre,” “homme,” and so forth all refer to the same sex. A computer, on the other hand, cannot readily make this distinction (unless specifically programmed to do so) and will therefore treat these items as different values for sex. In one FDA submission, a sponsor coded “male” by using the code “1” in some studies and the code “2” in other studies, resulting in the finding of “pregnancies in men.” In another submission, a similar problem resulted in males developing “female breast carcinomas.”

***Errors and Inconsistencies in Formatting Dates*** It is also necessary to use standardized formats to record dates. For example, Oracle stores a date as the number of seconds since January 1, 4712 BC, and then uses various functions to display the dates in more human-friendly formats. There are many different and personal ways for recording dates; one of the authors (AS) has noted at least 25 different ways in a single clinical trial submitted for FDA review! Should February 1, 2007 be recorded as 1 February 2007, 1 Feb 2007, 1 Feb 07, or 02/01/07? This problem still exists when data for numerical dates are extracted into software programs such as Excel that do not force the user to select a unique format for dates.

***Errors and Inconsistencies in Adverse Event Coding*** Adverse events are also subject to errors and inconsistencies by coders and data entry personnel. Many of these inconsistencies become very important when adverse events are analyzed by automated software.

Adverse events need to be coded consistently with respect to letter case. Problems can occur when there is discordant coding using all capital letters, all lowercase letters, or combinations thereof, as computer software will interpret these capitalization variations as different events. Letter case sensitivity can be important when two or more words are used to describe an adverse event. For example, some databases utilizing the Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary employ a coding system in which only the first letter of the first word of an adverse event is capitalized (e.g., “Atrioventricular block complete”). Failing to adhere to uniform letter case conventions across the data can result in severe errors in data analysis.

*Proper interpretation and coding* of events are also extremely important so that drug safety data can be appropriately analyzed. However, investigators and coders vary widely with respect to their health care training to properly interpret and code these events. Individual investigators may choose different terms to code the same

adverse event. Subjectivity in coding may be due to personal preferences of coders in the selection of terms versus the granularity of coding dictionaries that offer the coder a considerable array of terms to classify the reported adverse event. For example, kidney stones may be coded with a number of terms such as “kidney stones,” “nephrolithiasis,” “renal calculi,” and “renal calculus not otherwise specified.” There can also be variations in spelling of the exact same event code, such as the British spelling of “gynaecomastia” versus the American “gynecomastia.”

Coding dictionaries also change over time (sometimes even before a study is completed) because of revisions in coding terminology. In one NDA submitted to the FDA, the data were coded with different versions of the same coding dictionary without properly integrating the terms into a single coding version. Such a scenario may lead to partitioning events into too many terms and therefore mischaracterization of adverse events. Again, this misclassification, although not readily apparent, may have a profound impact on the results of the analysis performed.

It is especially difficult to code events when they occur as a constellation of signs, symptoms, and laboratory findings because a coder may inadvertently minimize or exaggerate the severity of an adverse event, depending on the selection of terms to code that event. The use of only one or two seemingly benign terms such as “muscle cramps” and “pain in limb” to describe rhabdomyolysis would not provide a comprehensive picture of the event. On the other hand, categorizing an isolated case of elevated transaminases (with no other associated laboratory findings) as acute liver failure would be inappropriate. Potential signals of adverse events may be obscured or distorted depending on how the events are grouped. Splitting events into multiple terms or grouping unrelated terms may erroneously underestimate the magnitude of a signal.

**Errors and Inconsistencies in Drug Names** Drug naming conventions are also exceedingly important. Although it may be expected that multiple names of the same medications would occur in postmarketing safety databases, this problem is also often seen in *premarketing* data. In one NDA submitted to the FDA, the data contained 900 different names for 150 unique concomitant drugs. Another NDA recorded 34,000 drug names for 2000 concomitant drugs mentioned in the study because contractors in different countries had submitted different names for the same medications.

**Errors and Inconsistencies in Numerical Data** It is also important to understand that *character* and *numeric* data are not interchangeable. Errors can occur when there is not a clear understanding between coding with characters versus coding with numbers. A variable may be inadvertently entered with character data in one study and numerical data in another study. For example, studies measuring the effects of a 30-mg dose of a drug may utilize a “30-mg” code in some studies and a “30” code in other studies. The “30-mg” code is treated as character data, but the “30” code is treated as numerical data. Efforts to successfully combine and analyze these studies may be hindered. The “30 mg” and “30” will be treated as two different doses if the numerical variable is treated as a character. If the combined variable is treated as a number, the data for the character variable will be treated as missing in some types of further analyses (e.g., regression analyses).

**Missing Information** Both premarketing and postmarketing collections of data are perpetually plagued by missing information. In premarketing and postmarketing clinical trials; patients can be lost to follow-up because of:

- Undocumented beginning and/or end of a medication or an adverse event (see Fig. 1)
- Undocumented death
- Undocumented serious adverse events
- Undocumented nonserious adverse events

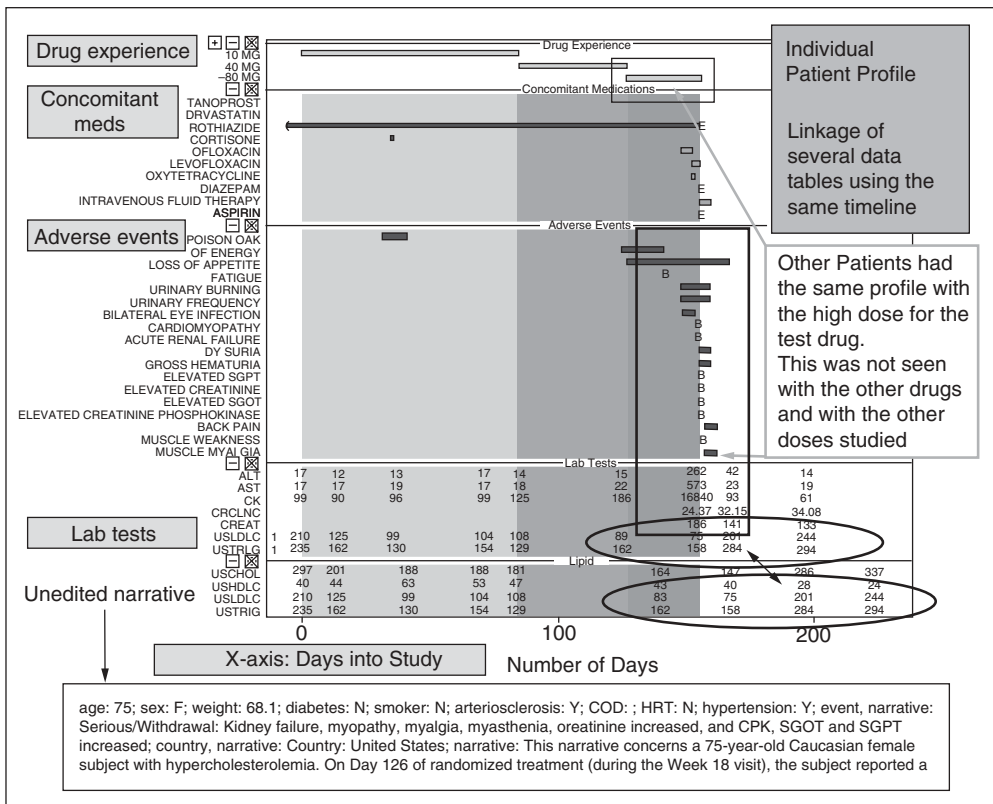


**FIGURE 1** An example of new drug application (NDA) data graphically displayed for an individual patient in a dose escalation clinical trial. This graph displays and links drug exposure information, clinical adverse events, concomitant medications, clinical laboratory values, demographic information, and narratives. The graph is divided into four major sections: The x axis for all four sections depicts the same time line, and the y axis the label of each row for each section. The top section displays drug exposure data for the test drug used in various doses (coded in a gray scale). The second section displays exposure to concomitant medications over time. The third section displays adverse events over time. The bottom sections display when laboratory tests were conducted and the results. Note in the areas highlighted by framed squares that clinical and laboratory adverse events were associated with the high dose of the test drug. Other patients had the same profile with the high dose for the test drug. This was not seen with the other drugs and with the other doses studied. Note that the beginning of an adverse event is displayed (B) but not the end for many adverse events. Note that the end of a concomitant medication is displayed (E) but not the beginning for some medications. Observe in the areas highlighted by framed ovals the discrepancies in the timing of the same laboratory results in different tables, making it difficult to assess whether these values occurred before or after an adverse event or a concomitant drug.



- Failure to assess adverse events that occur more than 30 days after the last dose of a drug
- Failure to assess adverse events occurring outside a scheduled time window
- Loss of interest of a patient or unwillingness to continue in a study
- Geographical moves by patients
- Loss of insurance coverage (in a postmarketing case-control study)
- Missing records due to technical errors during data migration (see below)

Even when patients remain in a study, information regarding adverse events may not be complete. The event itself may not be coded, even when the narrative includes relevant laboratory or other data (e.g., renal failure may not be coded, even though the narrative mentions abnormal blood urea nitrogen and creatinine levels and the need for dialysis). Moreover, adverse event data (such as laboratory tests, radiological reports, biopsy reports, and hospital records) collected outside a scheduled time window, outside of the study center, or after the study is completed *might not* be included in the final database for the study. This issue is important because unexpected adverse events do not occur at known prespecified time windows. “Tight windows” of adverse event data collection can also be a problem with drugs that have a very long half-life or the potential for causing a delayed, but serious condition.

The problem of missing data is even more pervasive in postmarketing drug safety databases. Most of these systems rely on voluntary reporting for which there are no well-defined protocols. Additionally, there are significant challenges in interpreting such data because of the wide variability of reporting sources (physicians, pharmaceutical companies, patients, attorneys, etc., as well as submissions from different countries). With electronic medical records, the presence of important relevant, but unreachable, data elements may not become apparent without the use of systematic approaches for analysis. This problem may be compounded in some epidemiological studies that ignore the presence of missing information in their analyses.

**Duplicate Information** Both premarketing and postmarketing databases may contain duplicate reports of adverse events. In postmarketing databases of voluntary reports, duplicate information on the same adverse event case may be submitted by several sources. For example, submission of information on the same event may be duplicated by the treating physician, the dispensing pharmacist, the nurse, the patient’s attorney, and/or the patient himself. Submission of information by drug companies may compound the problem, even though they are attempting to comply with regulatory requirements mandating submission of adverse events. Multiple drug companies may submit information on the same case, using their own unique patient identifiers, especially when the adverse event is associated with drugs manufactured by several different companies. There may also be a series of follow-up reports for the same case as additional information becomes available. However, updated patient identifiers for the same patient may not be linked to the original patient identifier. Duplicated information on the same event may also come from several teams working for the same drug company. Despite these problems, removal of duplicate reports is absolutely essential (though challenging) because superfluous information may result in false-positive signals of adverse events and wasted analysis time.



In postmarketing electronic longitudinal medical records, redundant and potentially contradictory information may come from several sources (e.g., reports from medical residents, their supervisors, the attending physicians). In these cases time stamping of each event may help to delineate the important sequences in understanding the adverse events.

**Other Inconsistencies** Inconsistencies in drug safety data due to difficulties in standardization also include a subject's primary diagnosis, differential diagnoses, relevant medical history, relevant physical exam findings, pertinent information from hospital records, and follow-up information (all of which may be subjective). There may also be a lack of consistency in the narrative summaries for individual patients and the data supporting the narratives. Indeed, it is difficult to clearly describe in the narratives the sequence of adverse events, medication, and laboratory results by using case report forms or line listings as source information. In complex patient records, case report forms and line listings may generate uneven temporal sequences of adverse events, concomitant medications, dosages, and so forth that cannot readily be comprehended. Such data need to be visualized by tools capable of displaying the complex, interrelated information on a common time line (Fig. 1).

Additional examples of variability in data collection (which, in turn, affects data interpretation) include questionnaires and physical exam forms. Questionnaires often utilize open-ended questions that allow great variability in the type and extent of adverse event information gathered. Physical exam forms—even when designed in a checklist format—may elicit variable collection of adverse event data; what is a serious event to one clinician may not be serious to another.

#### 9.8.4.3 Restructuring Capabilities

**Reconfiguration of Data** Drug safety data from different sources are often pooled or combined in databases. Reasons for combining data vary. In the case of premarketing studies, data from different sites are routinely combined because one site may not be able to recruit enough patients for a study. Data from different studies are often combined to increase sample size and therefore statistical power for detecting an uncommon adverse event.

In postmarketing safety surveillance databases, data from different countries or from different sources (physicians, patients, drug companies) may be combined in the same database in an attempt to obtain as much information about approved drugs as possible. Pooling or combining data can allow explorations of drug toxicity among various subgroups. Having a large database allows studying possible drug–drug, drug–disease, and drug–demographic associations.

When reconfiguring the data, several issues must be borne in mind. Combining data from different data sources can obscure potentially meaningful signals of adverse drug events [6]. For example, combining data for the term “colitis” with “ischemic colitis” may obscure the presence of ischemic colitis. Also, the criteria for reporting and coding an adverse event may differ among various data sources (e.g., countries with disparate regulatory requirements and different coding dictionaries). Reference ranges for normal values may vary, depending on the reporting source. When reference ranges vary, changes from baseline grouped

by treatment assignment may provide useful information. Patient populations may also vary in different study arms or in different countries where studies are conducted. Different populations may tolerate drugs differently or have varying levels of drug sensitivity. Study design may vary among sites, especially in terms of how outliers are handled and follow-up information is obtained. The selection of analytes and biomarkers of toxicity to measure may vary depending on the reporting source and when the data were collected (criteria for toxicity and case definition may be site specific and can change over time). Duration of drug exposure (as well as drug dose) may vary among studies. Because so many factors can influence the results, the process of transforming and combining data from different sources should be documented in a way that is easy for subsequent investigators to understand.

**Reconfiguration of Databases** Not only must data from different sources be preprocessed (“cleaned”), reconfigured, and validated before analysis, but entire databases must also sometimes be reconfigured and validated. This is especially the case if the database has evolved and has been maintained over a long period of time.

A good example of an evolving database is the FDA’s Adverse Event Reporting System (AERS) database containing reports of spontaneously submitted adverse events. AERS has undergone several configurations since its inception in 1968. This database was known as the Spontaneous Reporting System (SRS) from 1968 to 1997. Adverse events were coded into SRS with the COSTART (Coding Symbols for Thesaurus of Adverse Reaction Terms) dictionary. Only 1200 event codes were present in the COSTART dictionary. COSTART was replaced by the much more granular MedDRA (Medical Dictionary for Regulatory Activities) system of coding in 1997. MedDRA contains over 15,000 preferred term event codes, of which 10,000 are currently in use in the database. When SRS was modified to build AERS, adverse events coded with COSTART terms were mapped to MedDRA terms. Moreover, drug names in AERS have been and are still collected in free text form. There is substantial variation among reporting sources regarding the manner in which drug names are ultimately listed in adverse event reports. Drugs may be listed by their generic or trade names, with numerous and creative variations in spelling, abbreviations, spacing, and punctuation (see Section 9.8.4.2 on high-quality data). Thus what is now termed the AERS database is really a data set containing data that have undergone several organizational changes during more than three decades of data collection. This mapping, recoding, organizational reconfiguration, and validation of the database has been necessary to provide a uniform format for data analysis, yet this entire process has, understandably, been labor intensive and challenging.

With electronic medical records, multiple clinical records for the same patient may be treated as belonging to different patients during anonymization and migration of electronic medical records, tainting analytical conclusions. This problem may be difficult to untangle once the anonymized data migration takes place.

Sound analytical assessments require that analysts understand the manner in which the data were collected, reconfigured, migrated, and combined. These processes should be documented in a transparent way so that future investigators can readily understand the anonymization and migration in real time.

#### 9.8.4.4 Data Analysis

Current practices require that all the data collected be “cleaned,” reconfigured, and standardized in order to perform analytical and integrative tasks. These processes are complex, time consuming, and error prone—especially when there are many different personal standards in place. This process of “data cleaning” [7], reconfiguration, standardization, and integration must be done because the data are typically collected by several different contract research organizations, each with its own independent personal data collection standards. Because of personal standards in data cleaning and reconfiguration, many investigators end up analyzing only a small portion of the safety data, resulting in missed rare but serious adverse events or risk factors. If systematic data cleaning and reconfiguration are not done initially, then even seasoned investigators will still waste time constructing a new integrated database prior to analysis.

**Size of the Database** Database size is important in assessing drug safety in both premarketing and postmarketing settings. During clinical trials in the premarketing period, the number of subjects in a drug safety database often depends on the intended use of a product. For products intended for long-term treatment of non-life-threatening conditions, subjects studied may number in the thousands. For products intended for *short-term* treatment of *rare or life-threatening* conditions for which there are few effective treatment options available, a “smaller” number of subjects are studied, but there can be great subjectivity in defining the word “small” in such cases. This subjectivity is in part a reflection of the wide spectrum of disease severity for which such products might be indicated. For products intended for *chronic* treatment of *life-threatening* conditions, the number of subjects would need to be greater, but again, there is great potential for subjectivity.

Larger databases can help with risk–benefit decisions, but how can we achieve consensus on the exact size of the number of subjects needed for the database? A clinical trial database that contains limited information on a small number of subjects will likely lack the statistical power needed to detect differences in adverse events between control and treatment groups. A researcher may specify criteria for the minimum differences of adverse event rates between the two groups in an attempt to identify important safety signals, but ultimately these criteria are arbitrary. Even when a study enrolls a large number of subjects and records a large volume of data in a database, it is difficult to adequately identify all potential risks associated with a product. Some risks will only become apparent once a product is approved, that is, when hundreds of thousands or even millions in the general population are exposed. Adverse events in the postmarketing period are often collected in a voluntary manner through the use of spontaneous reporting systems such as FDA’s previously described AERS database or drug registries (e.g., clozapine). Yet extracting safety information from these databases—even if they are large—can still be challenging because background rates for various events can be difficult to obtain systematically. What is the background rate for headache, rash, decreased appetite, appendicitis, and fatigue—events that frequently occur in the population independent of drug therapy? Moreover, how does one assess the background rate of an event in a prespecified, but non-drug-exposed population compared to the rate of the same adverse event in a similar, but drug-exposed population? For example,

how does one assess the risk of confusion in elderly diabetic patients due to the effects of a drug prescribed for diabetes from baseline rates of confusion in elderly diabetic patients who are not receiving the same diabetes drug and concomitant drugs?

**High-Dimensional Aspect of Data** The high-dimensional aspect of data collected in a study can make the analysis of these data challenging. Even a simple clinical trial may have recorded dozens of measurements for each patient. More sophisticated studies may have hundreds of measurements per patient. Complex tests measuring the physiology of a specific system, such as pulmonary function tests or echocardiography, may be impossible to standardize over many different treatment groups or over a period of time because of technological changes and interpretation of findings. Pathological specimens (e.g., biopsies) may also be difficult to classify in a systematic and objective way. Other high-dimensional data include pharmacokinetic and pharmacodynamic information in phase III clinical trials, both of which are crucial in anticipating potential safety problems, especially in patients with impaired hepatic and renal function or in patients taking many concomitant medications. The challenge then lies in how to analyze such high-dimensional data. Unfortunately, we often do not have systematic methods for reducing the dimensionality of the data to find the subset of variables for which the treatments differ and the key statistics that describe the differences without losing important interdependent information.

**Variations Among Subjects** In the premarketing phase of assessing drug safety, it is important to have a study population that is not only representative of the target population but also sufficiently diverse in terms of demographics. This diversity will bolster the generalizability of the safety analysis. Diversity in the study population can be enhanced by including both males and females and also patients in different age, racial, body weight, and risk factor groups. Yet this same variability that is so necessary in increasing generalizability also presents challenges in analyzing data. For example, there may be a great deal of variability in renal and hepatic function among elderly patients. Plasma levels of a drug can also vary greatly if patients are given the same dose regardless of body weight, body surface area, or renal and hepatic function. There may also be great differences in sensitivity to the effect of a particular drug among individual patients. These issues may also taint the analysis of postmarketing safety data, including electronic medical records. Computerized safety analysis systems (discussed below) can aid in studying the effect of these variations by automatically generating stratified analyses to adjust for the impact of these patient attributes on adverse events.

**Temporal Relationships of Adverse Events** The temporal relationship between duration of product exposure and development of an adverse event is important in assessing causality. But how can data on temporal relationships be systematically summarized in a database containing thousands or even hundreds of thousands of subjects? Temporal relationships cannot be clearly elicited if only frequencies of adverse events between treatment and control groups are compared. There can be many disparities in the subjects' time of exposure or time at risk. Toxic manifestations of drugs may not occur until several months or even years after the initial exposure to the drug. How do we systematically assess delayed toxicity of a previ-

ously prescribed drug from the effect of a newly prescribed drug? Such a scenario occurred with reported cases of pancreatitis associated with valproic acid therapy, in which some cases appeared several years after therapy [2].

Sometimes an adverse event that occurs with a different frequency in the treatment group than in the control group is also qualitatively different in the two groups. For example, suppose a rare but serious vascular event occurs more frequently and earlier in the treatment group than the control group and is more likely to lead to discontinuation in the treatment group than in the control group. We need to describe that the event occurs more frequently and earlier in the treatment group and that it is more likely to cause discontinuation of treatment. This information is difficult to convey with tabular data, but often becomes clear when the data are presented in graphical form [8, 9, Figure 2].

**Effect of Concomitant Medication** Assessing drug–drug interactions is absolutely critical in evaluating the safety profile of a drug. Interactions can occur when one drug affects the absorption, distribution, metabolism, or excretion of another drug or drugs, producing additive or antagonistic effects on the other drugs. Various foods and dietary or herbal supplements (e.g., St. John’s wort) can also interact with drugs. Yet how do we systematically assess adverse events from one drug as opposed to adverse events from concomitant drugs or supplements?

**Effect of Preexisting Disease** The adverse event profile of a drug can be confounded because of the effects of underlying disease for which the drug may or may

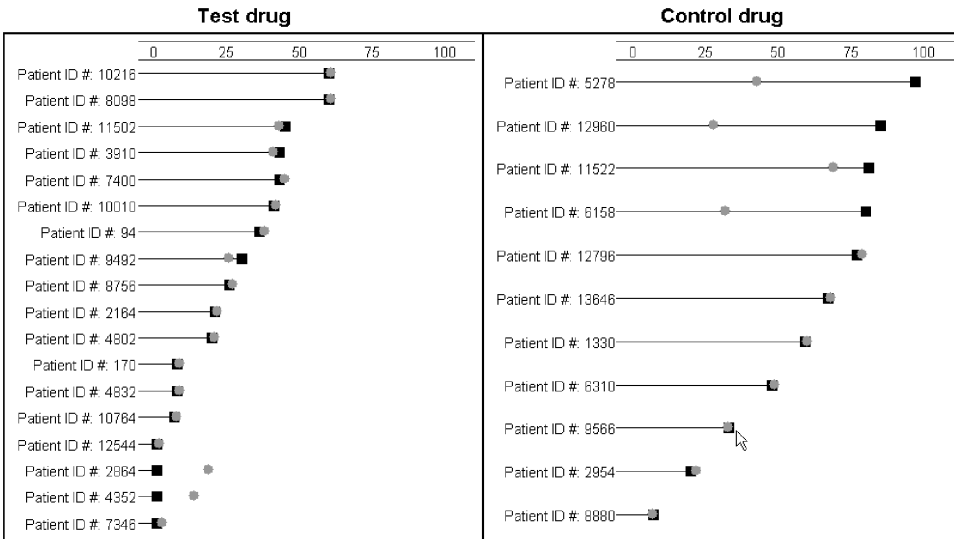


**FIGURE 2** A display that summarizes the duration of treatment (black squares) and the timing of serious vascular events (gray circles) for the subset of patients who withdrew from treatment because of an adverse event. Each line represents a single patient’s experience over time in days for the test (left panel) and the control drug (right panel). Patients are sorted by decreased duration of treatment. In this 1:1 randomized clinical trial there were 18 withdrawals due to a severe vascular adverse event with the test drug. This is in contrast with the control drug, with 11 withdrawals. Withdrawals with the test drug occurred sooner than with the control drug.

not be prescribed. Comorbidity can affect a drug's potential for inducing an adverse event. However, it is often difficult to separate the influence of preexisting disease when assessing the potential toxicity of a drug. How do we systematically separate the effects of a drug from a disease, the progression of that disease, or multiple diseases syndromes—each with its own varying rate of progression?

Preexisting disease, such as renal or hepatic disease, can especially influence the metabolism and excretion of certain drugs. In clinical trials, and in electronic longitudinal medical records, it is important to have sufficient variability in disease states and concomitant diseases among subjects in both the study and control groups. Investigators need to consider whether the adverse events that occur are due to abnormalities in the distribution, metabolism, and excretion of drugs as a result of underlying disease. These analyses could be systematically facilitated by having standardized ways of measuring blood (and in some cases, tissue) levels of drugs and their metabolites.

***Lack of Objective Markers of Drug Toxicity*** Some products have well-established, valid biomarkers that can be measured to track certain safety concerns. For example, a dose–response association with proteinuria, creatine phosphokinase, and urine myoglobin levels can be monitored to assess the safety of HMG-CoA reductase inhibitors (Fig. 1). However, there are often no specific markers (or pathognomonic clinical findings) of toxicity for many drugs. For most drugs under investigation—and most marketed drugs—practical tests to measure toxic drug or metabolite levels are not widely available [2]. Additionally, relying on product labeling of a drug in a similar class as a clue to investigate the toxicity of a drug is faulty, as labeling can be influenced by a number of factors including litigation and publicity.

***The Overwhelming Volume of Data to Be Analyzed*** Any given drug safety researcher—whether a statistician, clinician, epidemiologist, or safety evaluator—can only analyze so much nonstandardized data in a given time period. Premarketing databases contain very detailed data on thousands or perhaps tens of thousands of subjects. In some cases, premarketing data may have been collected over several decades for a wide array of indications. Postmarketing drug safety databases may contain millions of adverse event reports. The Composite Health Care System II database maintained by the U.S. Department of Defense contains over 9 million medical records. The FDA's AERS database contains over 2.5 million adverse event reports collected since 1968. As mentioned above, AERS utilizes the MedDRA classification system for its adverse event coding system. For these reports, approximately 10,000 MedDRA-preferred terms have been coded for 4000 generic drugs in the database. Thus over 40 million drug–event combinations are theoretically possible. This situation, compounded with the fact that the FDA currently receives over 1000 new reports of adverse events each day, exemplifies the daunting task that safety evaluators face when analyzing postmarketing adverse event data. With such a large volume of reports to review each day, exploring signals based on clinical judgment in combination with threshold reporting frequencies may not always be optimal or even practical. Such an approach makes it difficult to contextualize information regarding adverse events. How do we systematically access such huge databases to select appropriate variables for analyses? Without adequate drug expo-



sure data and baseline rates of disease processes (which may be erroneously attributed to “adverse drug events”) in specific populations at risk, how do we determine whether 20 cases of a specific drug–event combination is disproportionately frequent to merit further investigation?

**Subjective Analysis Strategies** There are also too many *subjective* analysis strategies in place. What is convincing to one analyst is not convincing to another. For example, what should be done when there are outliers in the data, that is, measured values of findings (such as laboratory values) or events that deviate substantially from the reference range? If outliers are ignored, important safety findings may not be identified; on other hand, outliers may represent errors in data collection. This was the case for one patient in an NDA submission. The patient had a serum creatinine value of 13 mg/mL (over 10 times the normal value, by any reference!) but a normal serum blood urea nitrogen (BUN) value.

**Limited Knowledge of Exposure and Reporting Rates in Postmarketing Data** Unlike clinical trials and electronic medical records in clinical practice, postmarketing voluntarily reported data contain limited information about the total number of patients exposed and the duration of exposure. This problem is compounded by the fact that adverse events are often underreported [2, 9].

#### 9.8.4.5 Reproducibility

Traditional analytical methods make extensive use of computers, but typically these methods still require constant restructuring of the data and multiple analytical tools. This endless restructuring wastes time and productivity and also makes the analytical processes difficult to document, audit, and reproduce in real time. This situation also makes it difficult to reconstruct and update analyses in real time when *new* adverse event data become available or when *new* questions need to be asked. The application of comprehensive data standards allows the use of integrated, reusable software for analyzing adverse event data. This integration facilitates the reproducibility of the results.

#### 9.8.4.6 Maintenance

Any computer database system will require maintenance. This maintenance includes such things as actively identifying and correcting data errors, ensuring that data can still be used with upgraded software and that this software can be used with upgraded hardware and data. Maintenance also includes actively testing and identifying computer bugs and adding new features and enhanced functions to the software.

### 9.8.5 EXAMPLES OF PRACTICAL COMPUTER-INTENSIVE TOOLS FOR SYSTEMATICALLY ASSESSING DRUG SAFETY DATA

#### 9.8.5.1 Background

Although 500,000 individuals were enrolled in clinical trials that were submitted to the FDA during 1990–1995 [10], the lack of a repository of clinical trial data, standardized data, and interoperable systems precludes us from efficiently tapping and



reanalyzing these data. This missed opportunity underscores the need for standardization and interoperable systems, as discussed above (see Section 9.8.4.1 on data standards and interoperable systems).

Drug safety reviewers spend a great amount of time learning the peculiarities of the data structure format and the variable names used with each NDA (and NDA supplement) submitted for marketing approval. As described above, for some NDAs the data from several studies must be incorporated into an integrated summary of safety data set and validated before performing a safety analysis; if every study uses a different data structure, this is an arduous task. To rectify this situation, the FDA is using the Clinical Data Interchange Standards Consortium (CDISC) Submission Data Tabulation Model (SDTM) format adopted as a standard by the FDA in July 2004. The implementation of such data standards allows for development of standard and comprehensive analytical tools that can automatically generate standardized and comprehensive analyses that can be used across numerous NDAs.

### 9.8.5.2 Analysis of Premarketing Data with WebSDM

An example of an analytical tool that utilizes CDISC data standards is WebSDM (Web Submission Data Manager) by Lincoln Technologies. The FDA has recently implemented WebSDM to analyze two NDAs in real time [11]. Although the original NDA data were submitted with nonstandard formats and the data had to be transformed into the Study Data Tabulation Model format before being loaded into WebSDM, the review process for these 2 NDAs was more efficient than with standard methods. In this case the data was transformed to demonstrate the concept that the use of standardized data simplifies the analytical process. WebSDM saves time because it first ensures that the data submitted to the FDA complies with CDISC format and then uses standard methods that enable automation of the analytical processes. Typically, FDA reviewers receive different data formats for each NDA; however, WebSDM eliminates the need for reviewers and supervisors to learn where the variables for analyses for each NDA are located (Fig. 3) and the different data formats for each NDA. For example, assessing potential liver injury by analyzing increases in serum alanine aminotransferase (ALT) and total serum bilirubin (TBILI) is done in one step instead of multiple cumbersome steps. It was also easier for the reviewers to fulfill the requirements of the NDA Review Template in a recent FDA guidance document for reviewers [12]. WebSDM also eliminates the need to reconfigure the data and the analytical tools for each new NDA analysis. WebSDM allows reviewers to use tailor-made, reusable tables and graphs of patient data in any NDA or supplement. The Sector Map graphical tool (with interactive drill-down capabilities) visualizes clinical trial data by highlighting higher-than-expected associations of adverse events compared to control groups. These features greatly simplify interpretation of the data. (See example with postmarketing data in Fig. 4.) These advanced graphical and analytical features are designed to simplify the interactive analysis of the clinical trial data.

### 9.8.5.3 Analysis of Postmarketing Data with MGPS and HBLR

For drugs already on the market, the FDA is utilizing the multi-item gamma poisson shrinker (MGPS) statistical algorithm [6, 9, 13] to systematically and simultaneously

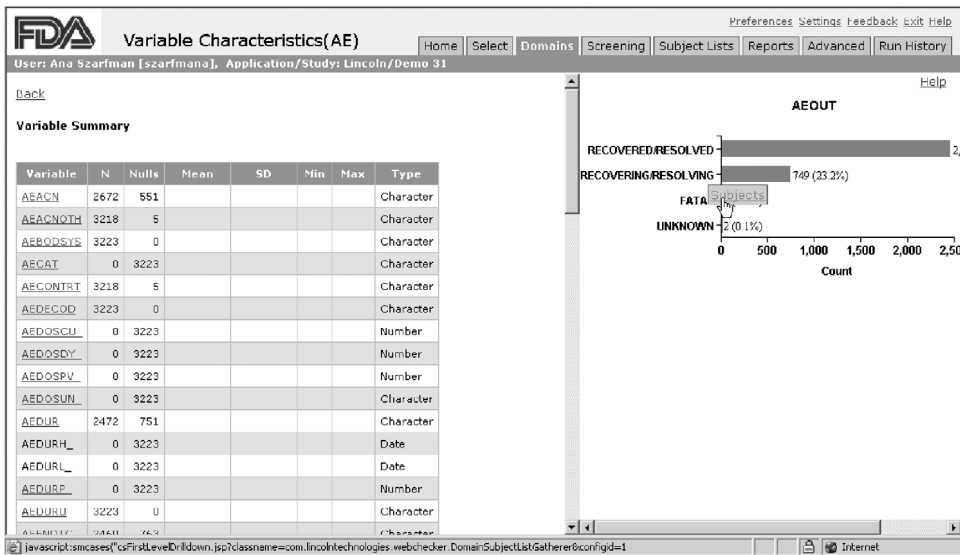
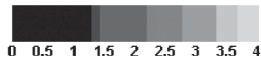


**FIGURE 3** Once the data are transformed into CDISC standards and integrated with a drug safety analysis system, the data can be easily analyzed. In this figure we show a sample screen from WebSDM a drug safety analysis system being evaluated by the FDA. This screen allows the user to view different attributes of the variables in a user-specified data set. When a variable is selected, a graphical display of the data is produced on the right-hand side of the window. The user can then select to visualize a graphical display of the individual patient profiles under the variable in a different window.

detect signals of higher-than-expected drug–adverse event associations in its post-marketing drug safety databases. To identify these signals, MGPS employs a disproportionality analysis of drugs and events, combined with Bayesian shrinkage. MGPS uses the independence model as the basis for computing the drug–event expected counts. MGPS includes a Maentel–Haenszel style approach for adjusting the expected counts for potential strata heterogeneity. When applied to the FDA’s AERS database, the MGPS program systematically stratifies the data by over 1000 categories (9 categories for age, 3 for sex, and 38 for year of report) to help adjust for background differences in relative reporting rates by these variables. The FDA thus far has focused its analytical efforts on the AERS database, but MGPS can be applied to any large drug safety database. The British Medicines and Healthcare Products Regulatory Agency (MHRA) has recently begun using MGPS as part of its pharmacovigilance program. MGPS also incorporates advanced graphical tools for analysis, including the Sector Map described previously (Fig. 4).

The FDA is also exploring another statistical algorithm, the hierarchical bayesian logistic regression (HBLR), to study signals that may be influenced by polypharmacy. HBLR corrects for confounding induced by concomitant medications (each with its own potentially strong adverse event associations) throughout the database [14, 15]. HBLR uses a prior distribution (estimated from the data) to improve the modeling of the joint associations for up to hundreds of drugs with a logistic regression response variable. The HBLR adjusts for both “signal absorption” and “signal masking.” Signal absorption is a phenomenon whereby an “innocent bystander” drug is falsely signaled as being associated with a particular adverse event simply

Cerivastatin



| Rank | SOC   | Term (PT)                 | EBGM   | AERS cases |
|------|-------|---------------------------|--------|------------|
| 1    | Renal | Myoglobinuria             | 15.416 | 104        |
| 2    | Nerv  | Myasthenic syndrome       | 12.449 | 1021       |
| 3    | Musc  | Myositis                  | 12.198 | 2670       |
| 4    | Musc  | Rhabdomyolysis            | 11.501 | 12024      |
| 5    | Genrl | Organ failure             | 10.121 | 685        |
| 6    | Renal | Chromaturia               | 9.437  | 3034       |
| 7    | Renal | Renal tubular necrosis    | 9.249  | 2547       |
| 8    | Inj&P | Muscle injury             | 9.025  | 998        |
| 9    | Inv   | Myoglobin blood increased | 8.807  | 527        |
| 10   | Inj&P | Polytraumatism            | 8.497  | 1105       |

NOTES: Additional restrictions for graph:

- Color controlled by: EBGM.
- Size controlled by: relative importance.
- Maximum intensity at signal score of 4.0.
- Omit rare terms used fewer than 100.0 times
- List 10 top scores.
- Show score indexes.
- Group by HLT
- Group by HLT
- Group by SOC
- Lowest level displayed: PT

Print  
Close

because of its frequent coprescription with another drug that is associated with that same event. Signal masking occurs in a database when there is failure to detect a weak signal for a particular drug because of the presence of other strong signals for the same adverse event, usually because of the homogeneity of drugs in the database [16]. Our initial experience indicates that HBLR may be a useful adjunct to MGPS in postmarketing safety assessments, especially in polytherapy regimens [15]. Dr. William DuMouchel has presented an overview of future empirical Bayes

**FIGURE 4** Sector map display of the MGPS data-mining profile for a drug, using a dictionary of medical terms. This display shows the safety profile of cerivastatin, a drug withdrawn from the U.S. market in August 2001 because of reports of fatal rhabdomyolysis, renal failure, and other organ failure [24]. The sector map shows strong signals for several serious muscle events including rhabdomyolysis within the Musc System Organ Class (SOC) and for renal failure within the Renal SOC. The strong renal failure signals with this drug were unexpected. In addition, there were huge differences between cerivastatin and other statins regarding the magnitude of the renal failure signals.

A sector map for data mining results is a visual presentation of data for a particular drug across all System Organ Classes (SOCs).

Each SOC is represented by a large tile in the sector map. Smaller tiles within each SOC tile represent Preferred Terms (PTs). A sector map graph is available for two-dimensional results of MGPS data mining runs.

PTs are ranked in descending order of EBGM values, a value that indicates the reporting association between a drug and an event (9). A color or a gray scale key (in this article) indicates relative ranking. The user can select to list the ranked PTs below the sector map.

Note: The primary path of a PT in the MedDRA hierarchy is used to determine where it appears in the sector map. If a PT is not in the event hierarchy associated with the configuration, it appears in a SOC tile named "Unknown".

The list of ranked PTs below the sector map includes the following information:

*Column Description*

*Rank:* Ranking of the term (combined with the drug) according to values of the EBGM values SOC containing the term.

*Term:* (PT) Specific PT.

*EBGM or other statistics used to configure the sector map*

If the Notes checkbox is checked, a Notes section provides information about the selection criteria and display options used for the graph.

- Color or shades of gray (shades of gray in this chapter), size, position in space, grouping, and ranking of tiles provide a "big picture" overview of the adverse event profile of a drug.
- Color or shades of gray: Light gray in this figure corresponds to stronger signals (high EBGM values).
- Size: A large tile (with a white border) defines each SOC (System Organ Class) in the MedDRA dictionary.
- Box size for each PT (preferred adverse event term): The box size is based on the number of serious cases of the term across all drugs in the AERS database. Thus, the box size of each PT is stable over different displays of different drugs.
- Position in space of each SOC and PT: SOC and PTs are always represented in the same area of the sector map. The position of each SOC and PT is stable over displays of different drugs.
- Grouping of PTs: PTs are grouped by high level term (HLT), high-level group (HLGT), and SOC.
- Ranking of PTs: PTs are ranked in descending order of EBGM values for each drug. EBGM: Signal Score. AERS cases: number of cases for the term in the AERS database. The PT "renal failure acute" is ranked 43rd with cerivastatin (see the top right rectangle within the Renal SOC, not displayed in the ranking for the top 10 drugs in the list below the graph) and "renal tubular necrosis" is ranked 7th with this drug. The PT "renal failure acute" has a larger box size than "renal tubular necrosis" because it has a much larger number of serious cases across all drugs in the AERS database.

methods for estimation of adverse event rates in clinical trials and active surveillance [17].

#### **9.8.5.4 Other Data Resources**

The number and size of databases containing drug safety information is growing rapidly, with some databases already containing millions of records [18, 19]. Analyzing several databases can help strengthen or refute a putative safety problem based on the results of the primary database analysis. Databases that can be analyzed include those maintained by various countries (e.g., the British General Practice Research Database or GPRD); various organizations (e.g., the World Health Organization and health maintenance organizations); various agencies (e.g., the U.S. Department of Defense, Department of Veterans Affairs, and Centers for Medicare and Medicaid Services Affairs), and others. Adapting new standard computer-intensive analytical tools to analyze data converted into standardized format will allow different experts to review each other's selection criteria and results so that conclusions can be more objectively studied and understood.

#### **9.8.5.5 Validation of New Methods**

Validation of new methods for analyzing drug safety data is challenging. There is no gold standard tool that can provide complete information about the whole spectrum of toxicity for a given drug and the magnitude and extent of this toxicity in specific subpopulations [9]. These facts, coupled with the discordant manner in which medical data are collected, make it very difficult to systematically analyze drug safety data in real time and to cross-reference multiple collections of medical data and results in a systematic way. The application of advanced computer methods offers a tremendous opportunity to analyze large databases in a timely and consistent manner and to learn about drug safety in a systematic way. These efforts will assist in creating gold-standard positive and negative signal definitions and methods given the data analyzed [20]. With these gold standards in place, we will be better able to further advance the art as well as the science of systematic drug safety assessment across databases.

### **9.8.6 CONCLUSIONS**

Even though great progress has been made since the passage of the Food, Drug, and Cosmetic Act in 1938, pharmaceutically related adverse events are still, unfortunately, responsible for a tremendous burden of pain and suffering in the United States [21], and old problems still persist around the world [22, 23]. Adverse events are also responsible for tremendous financial costs to taxpayers, insurance policyholders, insurance companies, and pharmaceutical companies. With such public health and economic costs in mind, we should carefully consider the strengths of new pharmacovigilance approaches while clearly acknowledging their limitations as well.

New tools that exploit the power of modern computer technology provide innovative approaches to help identify and investigate potential drug safety problems

in a systematic way. These new computer methods can assist in pharmacovigilance efforts because the results of an analysis from a particular drug safety database can be compared to the results from other, independent databases (e.g., clinical trial, health maintenance organization, or military medical databases). With interoperable standards in place, statisticians, epidemiologists and other analysts will be in a great position to improve drug safety and communication between all parties involved (consumers, clinicians, regulators, industry representatives, and legislators). By using the same software and data we can validate each other's selection criteria, results, and interpretation. As a result, researchers and policy makers will be better equipped to understand the limitations and biases of the data, leading to more objective decisions regarding drug safety.

## REFERENCES

1. <http://www.fda.gov/oc/history/elixir.html>; accessed April 2009.
2. Szarfman, A., Tonning, J. M., and Doraiswamy, P. M. (2004), Pharmacovigilance in the 21st century: new systematic tools for an old problem, *Pharmacotherapy*, 24, 1099–1104.
3. [http://www.janssen-cilag.com/gldisplay.jhtml?itemname=glossary&product=none#gl\\_informatics](http://www.janssen-cilag.com/gldisplay.jhtml?itemname=glossary&product=none#gl_informatics); accessed April 2009.
4. <http://www.trains.com/Content/Dynamic/Articles/000/000/003/011gsqfq.asp>; accessed on March 1, 2006.
5. 2005 FDA Science Forum: Advancing Public Health through Innovative Science. Washington, DC, April 27, 2005.
6. Datamining with applications in Genomics, Clinical Trials, and Post-marketing Drug Risk. Schering—Plough Workshop of May 31–June 1, 2001. Harvard School of Public Health. Section III: Data examples; available at: <http://www.hsph.harvard.edu/biostatistics/events/schering-plough/old/agenda2000-01.html> Video Clips. Statistical issues in drug safety monitoring. Schering—Plough Workshop of June 2–3, 2005. Harvard School of Public Health. Section III: Pharmacovigilance and Datamining for Drug Safety Monitoring; available at: <http://www.biostat.harvard.edu/events/schering-plough/old/agenda2004-05.html>; accessed on April 2009. Szarfman, A., Levine, J. G., Tonning, J. M., Use of Advanced Computer Methods to Simplify the Analysis of Complex Clinical Drug Safety Data; [http://webapps.sph.harvard.edu/content/Sch-Plo6205III\\_Unspecified\\_2005-06-02\\_04-03-PM.htm](http://webapps.sph.harvard.edu/content/Sch-Plo6205III_Unspecified_2005-06-02_04-03-PM.htm) (3rd presentation in the Video Clip); Accessed April 2009.
7. Barnett, S. T., and James, J. A. (1995), Measuring the clinical development process, *Appl. Clin. Trials*, 4, 44–52.
8. Szarfman, A., Talarico, L., and Levine, J. G. (1997), Analysis and risk assessment of hematological data from clinical trials: toxicology of the hematopoietic system, in Sipes, I. G., McQueen, C. A., Gandolfi, A. J., Eds., *Comprehensive Toxicology*, Vol. 4, Elsevier Science, New York; pp. 363–379.
9. Szarfman, A., Machado, S. G., and O'Neill, R. T. (2002), Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the U.S. FDA's spontaneous reports database, *Drug Saf.*, 25, 381–392.
10. Evelyn, B., Toigo, T., Banks, D., et al. (2001), *Women's Participation in Clinical Trials and Gender-Related Labeling: A Review of New Molecular Entities Approved 1995–1999*, June, Office of Special Health Issues, Office of International and Constituent Relations, Office of the Commissioner, U.S. Food and Drug Administration; available at: [http://www.fda.gov/cder/reports/womens\\_health/women\\_clin\\_trials.htm](http://www.fda.gov/cder/reports/womens_health/women_clin_trials.htm); accessed April 2009.



11. Cooper, C. K., Levine, J. G., Tonning, J. M., et al. (2005), Use of standards-based data and tools to improve the efficiency of the NDA Safety Review. 2005 FDA Science Forum. Abstract and Poster H-03; available at: [http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract\\_id=339&backto=author](http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract_id=339&backto=author); accessed April 2009.
12. Reviewer Guidance Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review. U.S. Department of Health 13a., and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER) February 2005; available at: <http://www.fda.gov/cder/guidance/3580fnl.htm>.
13. DuMouchel, W., and Pregibon, D. (2001), Empirical Bayes screening for multi-item associations, in *Seventh ACM SigKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, San Francisco.
14. DuMouchel, W. Manuscript in preparation.
15. Szarfman, A., DuMouchel, W., Fram, D., et al. (2005), Lactic acidosis: Unraveling the individual toxicities of drugs used in HIV and diabetes polytherapy by hierarchical Bayesian logistic regression data mining (abstract), 11th Annual FDA Science Forum, April 27–28, 2005; available at: [http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract\\_id=483&backto=author](http://www.accessdata.fda.gov/scripts/oc/scienceforum/sf2005/Search/preview.cfm?abstract_id=483&backto=author); accessed April 2009.
16. Gould, L. (2003), *Pharmacoepidemiol Drug Saf.*, 12, 559–574.
17. Workshop and Video Clip. Statistical issues in drug safety monitoring. Schering—Plough Workshop of June 2–3, 2005. Harvard School of Public Health. Section III: Pharmacovigilance and Datamining for Drug Safety Monitoring; available at: <http://www.biostat.harvard.edu/events/schering-plough/old/agenda2004-05.html>; accessed on April 2009. DuMouchel W. “Empirical Bayes Methods for Estimation of Adverse Event Rates in Clinical Trials and Active Surveillance”; available at: [http://webapps.sph.harvard.edu/content/Sch-Plough2005III\\_Unspecified\\_2005-06-02\\_04-03-PM.htm](http://webapps.sph.harvard.edu/content/Sch-Plough2005III_Unspecified_2005-06-02_04-03-PM.htm) (2nd presentation in the Video Clip); Accessed April 2009.
18. <http://www.gprd.com/home/>; accessed April 2009.
19. <http://www.nttc.edu/resources/funding/dod/sbir2003/osd032.htm>; accessed March 2006.
20. Levine, J. G., Tonning, J. M., and Szarfman, A. (2006), Reply: The evaluation of data mining methods for the simultaneous and systematic detection of safety signals in large databases: lessons to be learned, *Br. J. Clin. Pharmacol.*, 61, 105–113.
21. Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998), Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies, *JAMA*, 279(15), 1200–1205.
22. Centers for Disease Control and Prevention (CDC). (1996), Fatalities associated with ingestion of diethylene glycol-contaminated glycerin used to manufacture acetaminophen syrup—Haiti, November 1995–June 1996, *MMWR Morb. Mortal. Wkly. Rep.*, 45, 649–650.
23. Ferrari, L. A., and Giannuzzi, L. (2005), Clinical parameters, postmortem analysis and estimation of lethal dose in victims of a massive intoxication with diethylene glycol. *Forensic Sci. Int.*, 153, 45–51.
24. FDA Talk Paper (2001), Bayer Voluntarily Withdraws Baycol, FDA talk paper no T01-34, Aug 8.



# 10.1

---

## *Clinical Trials in Interventional Cardiology: Focus on XIENCE V Drug-Eluting Stent*

**J. Doostzadeh, S. Bezenek, W.-F. Cheong, P. Sood, L. Schwartz, and K. Sudhir**

*Clinical Science Department, Abbott Vascular Inc., Santa Clara, California*

### **Contents**

|          |                                                         |     |
|----------|---------------------------------------------------------|-----|
| 10.1.1   | Interventional Cardiology: History and Current Practice | 398 |
| 10.1.2   | Coronary Stent Design                                   | 399 |
| 10.1.2.1 | Bare-Metal Stents                                       | 402 |
| 10.1.2.2 | Drug-Eluting Stents                                     | 403 |
| 10.1.3   | Appropriate Control Groups for PCI Clinical Trials      | 410 |
| 10.1.3.1 | Coronary Artery Bypass Graft Surgery                    | 410 |
| 10.1.3.2 | Drug Therapy                                            | 411 |
| 10.1.3.3 | Plain Old Balloon Angioplasty                           | 412 |
| 10.1.3.4 | Stenting                                                | 413 |
| 10.1.4   | PCI Study Design Endpoints                              | 414 |
| 10.1.4.1 | Clinical Endpoints                                      | 414 |
| 10.1.4.2 | Angiographic Endpoints                                  | 415 |
| 10.1.5   | Safety                                                  | 415 |
| 10.1.6   | Postmarket Registries and Off-Label Stent Use           | 417 |
| 10.1.6.1 | Real-World Issues                                       | 417 |
| 10.1.6.2 | Characterizing Off-Label Stent Use                      | 418 |
| 10.1.7   | Peripheral Stents                                       | 419 |

|                           |     |
|---------------------------|-----|
| 10.1.8 Future Technology  | 422 |
| 10.1.8.1 Future Stents    | 422 |
| 10.1.8.2 Cardiac Tissue   | 423 |
| 10.1.8.3 Genetic Advances | 423 |
| 10.1.9 Conclusions        | 424 |
| References                | 425 |

### 10.1.1 INTERVENTIONAL CARDIOLOGY: HISTORY AND CURRENT PRACTICE

Coronary artery disease (CAD) affects approximately 13 million individuals in the United States and is considered the number-one killer among both men and women. It is estimated that more than 3 million individuals worldwide suffer myocardial infarctions (MI) annually because of CAD.

Remarkable changes have taken place in interventional cardiology since 1977, when Gruentzig and co-workers [1] performed a novel technique using an expandable balloon to compress obstructive lesions and reduce myocardial ischemia without a surgical procedure. Though the practice of using balloon catheters proved quite successful in the short term, long-term prognosis was less promising because of restenosis, which occurred in  $\geq 30\%$  of patients [2, 3]. This prompted the development of new techniques to remove plaque and the development of mechanical adjuncts, or stents, in order to maintain lumen patency following balloon angioplasty. Limitations in maintaining arterial patency have driven further development of interventional devices such as atherectomy catheters, stents, and laser therapy [4]. Coronary stenting after percutaneous transluminal coronary angioplasty (PTCA) was found in clinical trials to be a promising approach to reducing restenosis compared with angioplasty alone [5, 6]. Bare-metal stents (BMS) were the first type of stents used in percutaneous coronary intervention (PCI). Bare-metal stents reduced angiographic and clinical restenosis rates in de novo lesions compared to PTCA alone and decreased emergency coronary artery bypass graft (CABG) surgery. Stenting became quite prevalent in Europe in the 1980s and was increasingly used as the first line of PCI in the treatment of CAD. Methods developed in the 1990s to reduce or eliminate side effects of stenting included the administration of anti-coagulants, intravascular ultrasound, and brachytherapy. Systemic anticoagulation proved very effective and greatly reduced thrombus-associated complications. Brachytherapy was also effective at preventing in-stent restenosis, but its unfavorable side effects such as inhibition of healing around the stent and increased risk of cancer made it unsuitable for widespread use. The shortcomings of brachytherapy prompted further development of technology to maintain long-term arterial patency. This led to the emergence of stents designed to deliver drugs locally in order to inhibit neointimal hyperplasia without the serious effects of radiation or systemic drug administration. These coated, or drug-eluting, stents (DES) used various drugs encapsulated in different polymeric and nonpolymeric formulations. Drug-eluting stents were found to greatly reduce restenosis and maintain arterial patency. These advantages, combined with a lower incidence of side effects and a lower cost compared to surgical interventions such as bypass grafts, makes DES a great option for treatment of CAD [7].

Since the invention of stent devices, many clinical trials with various endpoints, sample size, and complexity have been designed. Device trials are designed to establish the safety and efficacy of the device by achieving a statistically significant difference in defined endpoints. The processes for receiving U.S. FDA approval for BMS and DES have required the submission of a pre-market approval (PMA) application; FDA requirements are evolving as stent technology has advanced. As shown in the Table 1, in clinical trials for BMS approved by the FDA, the sample size, total study duration, and clinical follow-up were lower compared to contemporary trials with DES. The primary objectives required by the FDA in the design of first-generation DES were surrogate endpoints, namely, the angiographic and intravascular ultrasound measures of the lumen diameter and the volume of plaque within the lumen. As a continuous and dichotomous variable that can demonstrate a statistically significant difference in small sample sizes, late luminal loss is a favored endpoint; in addition, its value in predicting target lesion revascularization makes it an excellent surrogate. However, the FDA has currently emphasized the need to demonstrate long-term safety of coronary stents and is shifting the focus of trials from surrogate endpoints [late loss (LL), percent diameter stenosis (%DS), and neointimal hyperplasia (NIH)] to clinical endpoints [cardiac death, MI, stent thrombosis (ST), major adverse cardiac events (MACE), and target vessel failure (TVF)]. Because of this shift, the clinical trial sample sizes for evaluating current DES are becoming larger than previously designed clinical trials. Current clinical trial designs for DES are mostly randomized, active-controlled, single-blind, parallel, two-arm, multicenter trials with clinical endpoints at 1 year and usually compare (newer) DES to (older) DES. The clinical follow-up should be designed for long-term safety (5 years) with angiographic, intravascular ultrasound (IVUS) and clinical endpoints. In addition to the complexity of procedural testing and follow-up duration, prespecified statistical hypothesis testing should be proposed for primary and secondary endpoints.

Another additional FDA requirement for the DES is evidence of continued safety in postmarket trials for the rare events such as ST. The rationale for this decision originated following the FDA warning notification of low-frequency event rates of ST involving the CYPHER stent in 2003 [8]. Current postmarket surveillance plans evaluate a variety of clinical endpoints, including Academic Research Consortium (ARC)-defined ST, as endpoints.

In conclusion, comparison of trial designs between BMS and DES reflect the increasing expectations for PMA approval by the FDA. DES clinical trials have increased in complexity in terms of trial design, study endpoints, hypothesis testing, statistical analysis, randomization, and sample size over the past 4 years. As coronary interventional therapy continues to evolve, efforts are made between the sponsor and the FDA to work collaboratively to generate evidence of efficacy and reasonable assurance of safety.

### 10.1.2 CORONARY STENT DESIGN

Stent design affects many properties of the stent, such as elastic recoil and rigidity—two undesirable stent characteristics [9, 10]. Previous studies have shown that differences in stent design and geometry can have a profound influence on late lumen loss [11] and neointimal proliferation [12], which in turn affect restenosis

**TABLE 1 Clinical Trial Summary for Bare-Metal and Drug-Eluting Stents**

| Trial Name                                | Design <sup>a</sup> | Trial Type   | Subjects         | Duration | Primary Endpoints                                                                                                                                 |
|-------------------------------------------|---------------------|--------------|------------------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Selected Bare-Metal Stent Trials</i>   |                     |              |                  |          |                                                                                                                                                   |
| DUET                                      | SA, OL, MC          | Registry—PMA | 320 <sup>b</sup> | 6 months | MACE at 1 month                                                                                                                                   |
| PIXEL                                     | SA, OL, MC          | Registry—PMA | 150              | 6 months | TVF at 30 days                                                                                                                                    |
| ASCENT                                    | R, DA, MC           | Pivotal—PMA  | 1,390            | 1 year   | TVF at 9 months                                                                                                                                   |
| ASCENT Registry                           | SA, OL, MC          | Postmarket   | 546              | 5 years  | TVF annual rate                                                                                                                                   |
| VISION                                    | SA, OL, MC          | Registry—PMA | 265              | 1 year   | TVF at 180 days                                                                                                                                   |
| <i>Selected Drug-Eluting Stent Trials</i> |                     |              |                  |          |                                                                                                                                                   |
| SPIRIT I (everolimus)                     | R, MC               | Feasibility  | 60               | 5 years  | In-stent late loss at 180 days                                                                                                                    |
| SPIRIT II (everolimus)                    | R, MC               | Preapproval  | 300              | 5 years  | In-stent late loss at 180 days                                                                                                                    |
| SPIRIT III (everolimus)                   | R, MC               | Pivotal      | 1,002            | 5 years  | In-stent late loss at 8 months                                                                                                                    |
| SPIRIT USA (everolimus) (pending)         | OL, SA, MC          | Postmarket   | 5,000            | 5 years  | Stent thrombosis rates annually through 5 years as defined (ARC) and composite rate of cardiac death and any myocardial infarction (MI) at 1 year |
| RAVEL (sirolimus)                         | R, MC               | Feasibility  | 238              | 1 year   | In-stent late loss at 6 months                                                                                                                    |
| SIRIUS (sirolimus)                        | R, MC               | Pivotal      | 1,058            | 5 years  | TVF at 9 months                                                                                                                                   |
| CYPHERTM (sirolimus)                      | OL, SA, MC          | Postmarket   | 2,070            | 1 year   | MACE at 30 days, 6 months, and 1 year                                                                                                             |
| e-SELECT Registry (sirolimus)             | OL, SA, MC          | Postmarket   | 15,000           | 3 years  | Acute, subacute and late stent thrombosis and MACE at 1, 6, 12, 24, and 36 months                                                                 |
| TAXUS IV (paclitaxel)                     | R, MC               | Pivotal      | 1,314            | 5 years  | TVR at 9 months                                                                                                                                   |
| TAXUS V (paclitaxel)                      | R, MC               | Pivotal      | 1,156            | 5 years  | TVR at 9 months                                                                                                                                   |
| TAXUS VI (paclitaxel)                     | R, MC               | Pivotal      | 446              | 5 years  | TVR at 9 months                                                                                                                                   |
| ARRIVE I (paclitaxel)                     | OL, SA, MC          | Postmarket   | 2,585            | 2 years  | TAXUS stent related cardiac events at 1 year                                                                                                      |
| ARRIVE II (paclitaxel)                    | OL, SA, MC          | Postmarket   | 5,016            | 2 years  | TAXUS stent related cardiac events at 1 year                                                                                                      |
| ENDEAVOR II (zotarolimus)                 | R, MC               | Pivotal      | 1,197            | 4 years  | TVF at 9 months                                                                                                                                   |
| ENDEAVOR IV (zotarolimus)                 | R, MC               | Pivotal      | 1,548            | 5 years  | TVF at 9 months                                                                                                                                   |
| PROTECT (zotarolimus)                     | R, MC               | Preapproval  | 8,800            | 5 years  | ARC definite or probable stent thrombosis at 3 years                                                                                              |
| ENDEAVOR (zotarolimus)                    | OL, SA, MC          | Postmarket   | 5,300            | 5 years  | ARC definite and probable stent thrombosis annually for 5 years and cardiac death/MI annually for 5 years                                         |

<sup>a</sup>Abbreviations: MC = multicenter; N/R = not reported; OL = open label; R = randomized; SA = single arm; SC = single center.  
<sup>b</sup>50 lead in subjects and 270 additional subjects enrolled.

rates and requirements for postprocedural intervention [13–16]. Vascular geometry influences distributions of wall shear stress (WSS), as evidenced by branching and curvature that produce regions of low WSS. These alterations adversely affect the preferential flow environment of intravascular cells and correlate with sites of neo-intimal hyperplasia [17–19]. Studies have shown that the use of a stent with thinner struts is also associated with a significant reduction of angiographic and clinical restenosis after coronary artery stenting [13, 20–22]. Although most currently approved stents are made of 316L stainless steel [23], other materials are also used, such as cobalt chromium in the Multi-Link VISION stent (ML VISION). ML VISION is more radiopaque and elicits a less vigorous host response because of thinner struts.

Drug-eluting stents are designed to deliver a drug locally that inhibits intimal thickening by interfering with pathways involved in inflammation, migration, proliferation, and/or secretion of the extracellular matrix. Both the drug and the delivery vehicle must fulfill pharmacological, pharmacokinetic, and mechanical requirements. Besides their biological effects, the drugs' chemical properties influence pharmacokinetic parameters and the possibilities for loading on a stent [24]. Tissue levels depend on lipophilic or lipophobic characteristics, molecular weight, and the degree of protein binding of the used drug. In the FDA-approved DES, drugs used in the DES are loaded into a polymer coating, which forms a reservoir for the drug. Drugs on DES inhibit smooth muscle cell proliferation by targeting cell cycle regulators and arresting the cycle. For example, everolimus, zotarolimus, and sirolimus (the “limus” drugs) differ from paclitaxel in their mechanism of action because of the point in the cell cycle at which the dividing cells are arrested. The limus drugs block an intracellular proliferation signal, which is triggered by IL-2 and other T-cell-specific growth factors. As a result, the cells are arrested in the G1 phase of the cell cycle. Paclitaxel, however, acts later in the cell cycle than the limus drugs: It stabilizes microtubules and has potent activity against proliferation, migration, and signal transduction [25]. Preclinical and clinical data have documented that low doses of everolimus, paclitaxel, sirolimus, or zotarolimus (used, respectively, on the XIENCE V, TAXUS, CYPHER, and ENDEAVOR stents) can achieve therapeutic arterial tissue concentration with insignificant concentrations in the systemic circulation, and that the drugs are safe. The ideal drug to prevent restenosis must have antiproliferative and antimigratory effects on smooth muscle cells, but on the other hand the stent design and the drug used must also minimize the reendothelialization duration [26]. As discussed in the section on next-generation DES, the greatest and quickest reendothelialization process was observed with XIENCE V when compared to TAXUS, CYPHER, and ENDEAVOR in preclinical models.

The drugs on the DES are delivered using a vehicle that must fulfill many requirements. Drug release must be controlled and predictable. The delivery vehicle (most often a polymeric carrier) must be biologically inert (e.g., nonthrombogenic, noninflammatory), sterilizable, must follow the changes in stent configuration during expansion and deployment and must be mechanically resistant to abrasion [27–29]. The development of a coating that meets all of these criteria has been challenging. Polymer coatings are needed as carriers for most drugs since these drugs do not adhere to the metallic stent surface. Numerous substances have been proposed as potential biomaterials for stent coatings [30]. Most of the polymers currently used in commercially available DES are proprietary. The safety and biocompatibility of

acrylic polymer and fluorinated copolymer coatings have been demonstrated based upon the XIENCE V everolimus-eluting coronary stent system (EECSS) biocompatibility testing per ISO 10993-1 and the long-term use of these polymers in medical implants.

Ideally, stent design should offer both effectiveness and safety measures and optimize flexibility, tractability, visibility, and biocompatibility [12]. While stent flexibility and tractability are highly dependent on mechanical stent design, radiological visibility and host biocompatibility rely on stent material.

### 10.1.2.1 Bare-Metal Stents

The concept of the stent grew directly out of interventional cardiologists' experience with angioplasty balloons in the first decade of use (1977–1987). The rates of restenosis with angioplasty balloons were of concern. Although the artery would be opened successfully using a balloon, in a small percentage of cases, the artery would collapse after the balloon was deflated. BMS stents, inserted after balloon angioplasty, were designed to address the issues met with balloon angioplasty. The stent itself was mounted on a balloon and could be opened once inside the coronary artery. In 1986, in Toulouse, France, Jacques Puel and Ulrich Sigwart inserted the first stent into a human coronary artery. In 1994 the first Palmaz–Schatz stent was approved for use in the United States. Over the next decade, several generations of BMS were developed, with each successive stent being more flexible and easier to deliver. Despite the fact that BMS stents virtually eliminated many of the complications of abrupt artery closure, restenosis typically after 6 months occurred in about 20% of cases, necessitating repeat procedures.

Current BMSs have emerged with advanced design characteristics imparting improved performance compared to prior generations. Stent characteristics of potential importance to procedural performance include metallic alloy content, stent crossing profile, strut thickness and configuration, flexibility, conformability, polish and texture, cell structure, radial support, scaffolding, side branch accessibility, fluoroscopic visibility, delivery balloon technology, and vessel surface coverage. Some of these features also aid in reducing long-term in-stent restenosis (ISR). The importance of strut thickness in reduction of ISR was studied prospectively in the ISAR-STEREO [31] and ISAR-STEREO-2 [21] trials, which compared angiographic and clinical restenosis rates between thick- and thin-strut stents, showed lower ISR rates by approximately 40% with thin-strut stents, with either similar or different strut designs. Similar findings were noted in a retrospective review by Briguori and co-workers [20]. Fluoroscopic visibility and metallic alloy composition are also related to strut thickness. Stainless steel stents become difficult to visualize as strut thickness approaches 0.05 mm. Cobalt alloy stents have superior visibility, as well as improved deliverability, radial strength, and flexibility. Several studies have evaluated BMS versus DES, but proving differences in clinical outcomes between DES and BMS has been problematic because of strut thickness. Speculation has been raised about the thick-strut bare-metal controls used in DES trials. For example, the BASKET trial enrolled 826 patients with 1281 treated lesions. In the BASKET trial patients received either CYPHER or TAXUS in the DES arm or VISION in BMS arm. The study found no difference at 18 months regarding cardiac death/MI between treatments. Interestingly, the BMS stent used in this trial

was the thin-strut VISION stent, which is also used in the everolimus drug-eluting stent, XIENCE V. As mentioned above, studies indicated that strut thickness is important in reducing the ISR [21, 31].

Several recent clinical studies [20–22] have been conducted comparing the performance of stents with varying strut thickness. These studies indicated that when two stents with different designs are compared, the stent with thinner struts elicits less angiographic and clinical restenosis than the thicker strut stent. The VISION stent system is a next-generation BMS that can achieve good acute performance (deliverability, visibility) with thinner strut technology. Because of excellent results obtained with the VISION stent, Abbott Vascular has used the VISION stent as the platform in the XIENCE V. The ML VISION family of stents is fabricated from a L605 cobalt chromium alloy that has a thinner strut stent (0.08 mm or 0.0032 inch) and has a crossing profile of 1.07 mm. The ML VISION stent has been on the market in the European Union since 2002 and the MINI VISION since 2003. The ML VISION and MINI VISION stents have also been on the market in the United States since 2004. Medtronic's Driver cobalt nickel stent, with a smooth, modular design, a strut thickness of 0.09 mm, and crossing profile of 1.09 mm is another future generation BMS. Registry data for the Driver [32] and VISION [33] stents in mostly type B1 and B2 lesions in 3.0–4.0 mm diameter (mean 3.04–3.07 mm) vessels document excellent (100%) device success rates, and remarkably low ischemia-driven target lesion revascularization (TLR) rates at 6 months (4.3% for the VISION and 3.4% for the Driver). Angiographic in-stent restenosis (ISR) rates were 15.7% for both stents at 6 months.

In summary, the latest available BMSs have advanced design characteristics related to strut thickness and cell configuration, metal alloy composition, flexibility, crossing profile, and resultant deliverability. The next generation of DES (e.g., XIENCE V) benefits from the use of these thinner strut BMS platforms. The future generation of BMS may incorporate biodegradable materials, such as magnesium alloys, which may reduce the TLR rates.

#### 10.1.2.2 Drug-Eluting Stents

The newest minimally invasive intervention alternative to open-heart bypass surgery is PCI with a DES. The first DES, a sirolimus-eluting stent, was approved by the FDA in 2003. The introduction of DES to interventional cardiology practice has resulted in significant improvement in the long-term efficacy of PCI. DES successfully combine mechanical benefits of BMS and stabilizing the lumen with direct delivery and the controlled elution of a drug to the injured vessel wall to reduce further the neointimal proliferation. The dramatic reduction in restenosis has resulted in the implantation of DES in the clinical practice and has rapidly expanded the spectrum of new DES generations [34].

**First-Generation DES** The first-generation DESs, which combine the mechanical scaffolding properties of metallic stents with the site-specific delivery of an antiproliferative agent, have proved to inhibit vascular responses to arterial injury and reduce restenosis. The polymer-regulated site-specific delivery of an antiproliferative agent has also been shown to inhibit tissue growth after coronary stent implantation and to improve long-term event-free survival compared with BMS [35, 36].



The first-generation DES, CYPHER, was evaluated in the early sirolimus trials RAVEL ( $n = 238$ ), E-SIRIUS ( $n = 352$ ), C-SIRIUS ( $n = 100$ ), and SIRIUS ( $n = 1058$ ) before FDA approval. These trials reported a lower incidence of restenosis in patients who had received sirolimus DES, up to 8 months after PCI [37–39]. The attenuation of neointimal hyperplasia in these trials correlated with fewer TLR, an important clinical marker of restenosis. Rates of death and MI were low and comparable between study arms. These observations were confirmed in the U.S. pivotal trial, SIRIUS. Rates of death and MI, at 9 months, were, respectively, 0.9 and 2.8% for the CYPHER arm compared to 0.6 and 3.2% in the TAXUS arm.

The first-generation DES, TAXUS, was evaluated in TAXUS I ( $n = 61$ ), TAXUS II ( $n = 536$ ), TAXUS IV ( $n = 1314$ ), TAXUS V ( $n = 1156$ ), and TAXUS VI ( $n = 446$ ) trials. These trials showed reductions for the incidence of restenosis and TLR in patients receiving a slow-release polymeric paclitaxel DES compared to BMS [40]. These trends were later confirmed for slow-release and moderate-release formulations by the TAXUS II trial [41]. The pivotal trial, TAXUS IV, showed that the slow-release polymeric paclitaxel DES was associated with significantly reduced rates of in-stent binary restenosis (5.5% compared to 24.4%) and TLR (3.8% compared to 14.6%) at 9 months [42]. Furthermore, no significant adverse events were experienced in DES-treated patients in these early trials. The objective of TAXUS V was to evaluate the SR polymeric paclitaxel DES in complex lesions [43]. Finally, TAXUS VI evaluated a moderate-release polymeric paclitaxel DES, also in complex lesions [44]. Consistent with earlier TAXUS trials, the polymeric paclitaxel DES demonstrated significant reductions in the incidence of restenosis and TLR (72 and 64%, respectively) compared to metallic stents.

The FDA approved a new DES, ENDEAVOR. The ENDEAVOR IV [45] trial compared the performance of ENDEAVOR to TAXUS. Results from the ENDEAVOR IV trial demonstrated that in-stent late-loss, in-segment late-loss, and in-segment angiographic binary restenosis (ABR), respectively, were 60, 57, and 47% higher in the ENDEAVOR arm compared to the TAXUS arm. TLR was 41% higher in the ENDEAVOR compared to TAXUS with comparable MACE and TVF rates between the two arms (Table 3).

Abbott Vascular evaluated first-generation DES in the ACTION, DELIVER, FUTURE, and ZoMaxx clinical trials prior to developing the next generation of DES, XIENCE V. The ACTION trial was a randomized, parallel, three-arm, single-blind, multicenter trial designed to evaluate the safety and performance of the MULTI-LINK TETRA-D, actinomycin-eluting coronary stent system at two different drug concentrations compared to a bare-metal control. Early results of the ACTION trial indicated higher target lesion revascularization rates than expected in DES arms, driven by restenotic events. The ACTION trial was terminated in March 2002.

The DELIVER trial, was a prospective, randomized, single-blinded, parallel-group, multicenter clinical investigation evaluating the RX ACHIEVE paclitaxel drug-coated coronary stent system (CCSS) compared to the bare-metal MULTI-LINK RX PENTA stent in the treatment of patients with de novo native coronary artery lesions. The primary endpoint was target vessel failure (TVF) at 270 days. The primary endpoint was not met as powered; superiority of the RX ACHIEVE paclitaxel-eluting CCSS was not met.

The ZoMaxx DES has a phosphorylcholine coating and releases the drug ABT-578 (zotarolimus), a sirolimus derivative. ZoMaxx I and ZoMaxx II assessed the safety and efficacy of the ZoMaxx stent system compared with the TAXUS Express [2] paclitaxel-eluting stent system (Boston Scientific, Natick, Massachusetts). Although these trials demonstrated that ZoMaxx is safe to be implanted, Abbott Vascular decided not to pursue further the development of the ZoMaxx stent since a next-generation DES by Abbott Vascular, XIENCE V, had demonstrated superiority in terms of late loss to both bare-metal stent and an FDA-approved DES, TAXUS at 8 months, in SPIRIT III (see next section).

In conclusion, clinical trials for first-generation DES have proved the effectiveness of the DES in reducing the rate of ISR by directly inhibiting the progression of the cell cycle and migration of vascular smooth muscle cells, preventing neointimal hyperplasia and injury-induced thickening of the arterial wall. These trials evaluating the first-generation DES were randomized trials with more complex design, and a larger sample size when compared to pivotal studies for approval of BMS by FDA. Surrogate primary endpoints such as angiographic late loss were permitted. However, the search for improved deliverability and greater efficacy has prompted development of the next generation of DES.

**Next-Generation DES** Next-generation DESs are being designed with the goal of enhanced safety and/or efficacy compared to first-generation devices. The Abbott Vascular DES XIENCE V EECS was approved in the U.S. on July 2, 2008. This is a next-generation DES, which has demonstrated superior clinical outcomes compared to TAXUS. XIENCE V, an everolimus-eluting stent, has been designed so that the drug is released from a thin ( $7.8\mu\text{m}$ ), nonadhesive, durable, biocompatible fluorinated copolymer coated onto a low profile, flexible cobalt chromium stent with a 0.0032-inch strut thickness. Several clinical studies have demonstrated that stents with thinner struts elicit less angiographic and clinical restenosis than the thicker strut stent [20–22]. The strut thickness of the ML VISION ( $81\mu\text{m}$ ) compares favorably to those of the metallic stent platforms used on the CYPHER ( $140\mu\text{m}$ ) and the TAXUS ( $132\mu\text{m}$ ) stents [22] (Fig. 1). The thin-strut design is one element of the XIENCE V EECS (based on the ML VISION stent design) that decreases clinical risk.

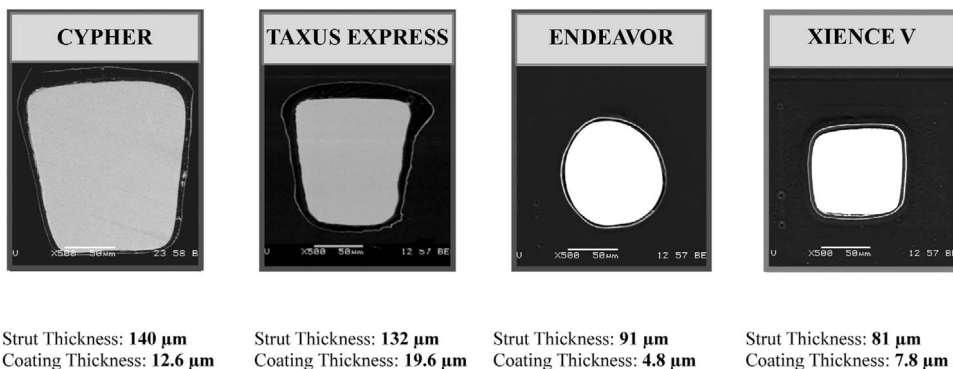


**FIGURE 1** Progression towards thinner struts in XIENCE. Albuluminal coating thickness is represented in all images in figure. Data on file at Abbott Vascular.

The XIENCE V stent uses the same stent design and L-605 alloy cobalt chromium material as the ML VISION and ML MINI VISION stents. Compared to the 316L stainless steel used in TAXUS and CYPHER [23], the cobalt chromium in the XIENCE V stent is more radiopaque, thinner, and elicits a less vigorous host response than stainless steel. L-605 cobalt chromium provides both excellent strength and fatigue resistance characteristics, which enables the XIENCE V stent to have significantly thinner struts than contemporary stainless steel stents while maintaining comparable radial strength. Thinner struts, cobalt chromium material, and the MULTI-LINK stent pattern were intended for increased flexibility and conformability, and therefore better deliverability.

More rapid endothelialization with XIENCE V in porcine model, which may be due to the thin struts, has also been reported in preclinical models (Fig. 2). Evaluations of scanning electron micrographs of the luminal surface of various DES demonstrated differences in endothelial cell coverage among DES. As shown in Figure 2, XIENCE V stents had the greatest coverage compared to CYPHER, TAXUS, and Endeavor. The faster reendothelialization associated with the XIENCE V is likely due, at least in part, to the thin stent struts of the XIENCE V. A thin stent strut is incorporated in neointima more rapidly and will require less neointima to completely cover the struts. Similarly, endothelial cells may migrate up and over to incorporate a thin strut more rapidly [46].

Following favorable results with XIENCE V in the SPIRIT FIRST ( $n = 60$ ) and SPIRIT II ( $n = 300$ ) clinical trials in Europe [47, 48], the large-scale SPIRIT III trial ( $n = 1002$ ) was performed to evaluate the XIENCE V in comparison to a widely used TAXUS in patients with CAD [49]. The XIENCE V clinical trials (SPIRIT FIRST, SPIRIT II, and SPIRIT III) were conducted in single- and dual-vessel disease and met their primary and major secondary endpoints with no adverse safety signals to date. The results of these clinical trials are provided in Table 2. As shown in Table 2, the SPIRIT family trials showed significant benefits to using XIENCE V over both commercially available BMS (VISION) and DES (TAXUS).

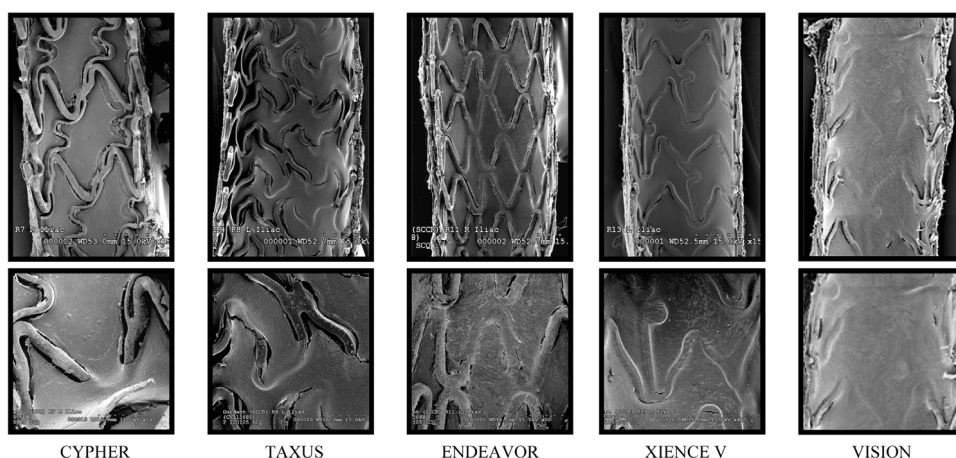


**FIGURE 2** Qualitative assessment of endothelial cell coverage (14-day rabbit iliac). Corresponding images of each stent in the top row are shown directly below in greater magnification of the above stent image.



**TABLE 3 Perspectives from the New DES Compared to TAXUS**

| Study                              | Stent                        | In-stent LL | In-seg. LL | In-seg.<br>ABR | TLR   | TVR   | MACE  | TVF   |
|------------------------------------|------------------------------|-------------|------------|----------------|-------|-------|-------|-------|
| <b>SPIRIT III<sup>a</sup></b>      | <b>XIENCE V<br/>vs.TAXUS</b> | ↓ 48%       | ↓ 50%      | ↓ 47%          | ↓ 46% | ↓ 29% | ↓ 43% | ↓ 22% |
| <b>ENDEAVOR<br/>IV<sup>b</sup></b> | <b>ENDEAVOR<br/>vs.TAXUS</b> | ↑ 60%       | ↑ 57%      | ↑ 47%          | ↑ 41% | ↑ 6%  | ↓ 2%  | ↓ 8%  |

<sup>a</sup>Clinical event rates for XIENCE V calculated from 284 day data.<sup>b</sup>ENDEAVOR IV results presented by M. Leon at TCT, 2007.

XIENCE V, as compared to TAXUS in the SPIRIT III trial, demonstrated (a) significant reductions in angiographic in-stent late loss ( $p = 0.006$ ), in-segment late loss ( $p = 0.004$ ), and in-segment %DS ( $p = 0.009$ ) at 8 months, (b) significant reduction in IVUS percent volume obstruction ( $p = 0.01$ ) without positive remodeling or late acquired incomplete apposition at 8 months, (c) nonsignificant trends toward less events for composite endpoints cardiac death and MI, and TVF at one year, (d) reductions in TLR and MACE rates at one year, and (e) comparable rates of thrombosis [49]. In the SPIRIT II and SPIRIT III RCT clinical studies, the XIENCE V has demonstrated not only noninferiority but superiority to TAXUS, in terms of the primary endpoints, in-stent and in-segment late loss, respectively. The SPIRIT IV clinical trial is a continued access study for continued access and was initiated in 2006 to further evaluate the safety and effectiveness of the XIENCE V EECSS and to enroll SPIRIT-III-like subjects in order to support the SPIRIT III major secondary endpoint (270-day TVF). The SPIRIT IV clinical trial is a single-blinded, multicenter clinical trial and will enroll approximately 3690 subjects at up to 70 sites in the United States. The primary endpoint of the trial is ischemia-driven MACE at 270 days. Patients will be followed out to 5 years. This study has enrolled approximately 3000 patients as of March 2008. SPIRIT IV will allow for analyzing subgroups of complex patients in a large randomized clinical trial. Furthermore, subjects enrolled in SPIRIT IV are SPIRIT-III-like subjects, so a true pooled analysis can be performed. In addition, postmarketing studies are currently being conducted internationally, which include SPIRIT V and SPIRIT WOMEN. A postapproval study, XIENCE V USA is currently also being planned in the United States. The results of a meta-analysis of SPIRIT II and SPIRIT III clinical studies obtained data at one year to support previous findings from SPIRIT II and III. The analysis indicated reductions in relative risk for both MACE and TVF, suggesting that XIENCE V has lower cardiac event rates in comparison with TAXUS, which was driven by lower TLR (XIENCE V FDA Panel, data are available at Abbott Vascular website; [www.abbottvascular.com/av\\_dotcom/url/home/en\\_us](http://www.abbottvascular.com/av_dotcom/url/home/en_us)). In the meta-analysis of the SPIRIT II and SPIRIT III RCT clinical studies, key clinical endpoints observed in XIENCE V arm were lower than those observed in the TAXUS arm. Results from several randomized clinical trials that compared commercially available DES and evaluated various clinical and angiographic endpoints have been published [45, 50–52]. The TAXi study ( $n = 202$ ) demonstrated safety of both TAXUS and CYPHER stents in various clinical and anatomical subsets [50]. Similarly, the REALITY trial confirmed the clinical safety of TAXUS in a larger subject population. In the REALITY trial ( $n = 1386$ ), the primary endpoint of in-lesion ABR rate at 8 months (primary endpoint) was comparable between TAXUS and CYPHER.



Additionally, there was no difference between the two DES in terms of MACE rates at one year [51]. In the SIRTAX trial, although sirolimus-eluting stents showed a 42.6% decrease in MACE at 9 months compared to paclitaxel-eluting stents, the rates of clinical and angiographic restenosis were low for both DES [52]. Although the first-generation DES used BMS as control stents, next-generation DES clinical trials were designed to evaluate (older) DES versus (newer) DES, which resulted in increased sample sizes and new statistical hypothesis testing. Surrogate angiographic endpoints were permitted as primary endpoints (but for longer time points) with the focus on rare safety events such as death, MI, and ST. Composite endpoints such as MACE and TVF were used in these clinical trials to evaluate both safety and effectiveness of next-generation DES.

**Future-Generation DES** Drug-eluting stents continue to evolve, with enhanced drugs, polymers, and delivery systems designed to improve outcomes for patients with CAD. Improved delivery systems allow access to lesions in vessels with excessive tortuosity. Abbott Vascular plans to evaluate a future everolimus-eluting coronary stent (EECS), which is a balloon-expandable stent fabricated from a single piece of medical-grade L-605 cobalt chromium alloy tubing, which is the same material used in the ML VISION and XIENCE V. This EECS will provide a more flexible, better delivery system profile while maintaining good radiopacity and strength. The sirolimus and paclitaxel stents, CYPHER SELECT Plus Stent (by Cordis), and TAXUS Element Stent (by Boston Scientific) are also designed with improved delivery system features and are currently on the market in the European Union (EU). In addition, clinical trials have recently begun for a sirolimus-eluting stent (by Conor and Cordis) using Conor's unique fully bioresorbable reservoir technology.

The bioabsorbable stent is a future-generation stent type. Bioabsorbable stents are an exciting class of future-generation stents because of their ability to be absorbed over time. Ideally, this allows the lifetime of the stent to be tailored to the clinical need of the disease or condition. In order to open the blocked vessel and offer mechanical support, the bioabsorbable polymeric stent must have high radial strength, low recoil, and a certain degree of flexibility while its degradation product must be nontoxic in body fluids. The bioabsorbable stent holds promise as a device system that will be able to perform its mechanical function (i.e., high radial strength and low recoil), facilitate stent placement, and deliver drugs for the prevention of restenosis. Furthermore, the stent should support the artery while the vessel heals and should gradually transfer the mechanical load to the tissue as the stent degrades over time. It is also worth noting that radiopacity is another consideration in stent design. The BVS (Bioabsorbable Vascular Solutions) stent was developed by Abbott Vascular. In March 2006, BVS started the world's first clinical trial of drug-eluting polymeric bioabsorbable stent called the ABSORB trial.

The BVS stent drug coating layer contains everolimus (antiproliferative agent) and poly-D, L-lactic acid (PDLLA) polymer at a 1:1 ratio for controlled drug release. Because of its unique stent pattern design, the BVS stent looks like a regular DES with great mechanical properties. The ABSORB trial enrolled 30 patients [53]. The procedure success rate was 100% and the device success rate was 93.5%. The safety of the device was confirmed in all 30 patients after 12 months follow-up. Only one non-Q-wave MI was reported that was related to the procedure of a nonischemic-

driven target lesion revascularization. There was no ischemia-driven target lesion revascularization, and no thrombosis occurred up to 12 months. Through 12 months, there were no protocol-defined or ARC-defined stent thrombosis [54]. Efficacy measurements of in-stent late loss was  $0.44 \pm 0.35$  mm with angiographic follow-up at 6 months and was mainly due to a mild reduction of the stent area (−11.8%) as measured by intravascular ultrasound (IVUS) [54]. At 24 months, an increase in lumen area on IVUS and optical coherence tomography (OCT), and restored vasomotion in response to vasoactive agents, has been demonstrated.

Apart from Abbott Vascular, Reva Medical Inc. (supported by Boston Scientific, Inc.) is developing its own bioabsorbable DES made from tyrosine-derived polycarbonate material and has started its clinical trial as of late 2007. Reva Medical uses a unique “slide-and-lock” geometry for its stent designs. When expanded, the stent elements slide from the compact state and lock into an expanded state, similar to safety lockouts on extension ladders. Thus the expansion is independent of material deformation and will provide steel-like scaffolding. The Reva stent is delivered by standard balloon deployment and is made radiopaque by a propriety method [55].

Although more long-term beneficial outcomes in controlled clinical studies have yet to be demonstrated, it can be anticipated that biodegradable stents offer the benefits of DES without leaving a metallic stent behind after the treatment of coronary vessel lesion.

As new approaches of stenting continue to evolve, clinical trial design requires new technology and perhaps novel endpoints to demonstrate the safety and effectiveness of these devices. For example, optical coherence tomography (OCT) and serial IVUS are used in the evaluation of biodegradable stents. More long-term safety data are required by the FDA for low-event rate endpoints resulting in larger sample size for these trials.

### 10.1.3 APPROPRIATE CONTROL GROUPS FOR PCI CLINICAL TRIALS

Coronary atherosclerotic disease has been dominated in the past by four main therapy options: CABG, optimal medical therapy (OMT), balloon angioplasty (POBA), and percutaneous coronary intervention (PCI) with BMS or DES. From 1960 to 1994, CABG surgical procedures dominated as the treatment for CAD. The CABG was eventually challenged with the introduction of angioplasty in 1977 [56]. The FDA approved the first BMS in 1996, which was followed by the first DES approval in 2003. According to the Centers for Disease Control and Prevention, the most common interventional cardiology procedure performed in 2007 was PCI with DES [57]. Several studies have compared the various treatment options of CAD, POBA, PCI, CABG, and optimal medical therapy. Despite the fact that comparative studies have been conducted, there is still the need for studies directly comparing PCI with DES to these treatments.

In the next paragraphs the key outcome of clinical trials comparing PCI to other CAD treatment options will be discussed.

#### 10.1.3.1 Coronary Artery Bypass Graft Surgery

Coronary artery bypass graft procedures are a common choice for treatment of advanced CAD. In patients with single-vessel disease though, PCI is the preferred



treatment because of the lower clinical risks associated with PCI [58]. Several studies compared PCI with BMS to CABG including the ERACI II, ARTS, SOS, and the AWESOME trials [59–63]. The results of these trials demonstrated that repeat revascularizations were higher in the PCI with BMS group compared to the CABG group in all trials. However, there was no significant difference in survival benefit between CABG and PCI with BMS. Overall results showed that PCI with BMS was a relatively safe alternative for patients who were not candidates for surgery.

When DESs were initially being introduced, primary endpoints included late lumen loss on angiographic follow-up, death, myocardial infarction, or reinterventions (RAVEL, SIRIUS). Results of these trials showed significant reduction in target lesion revascularization rates for DES. Since the first FDA approval of DES in 2003, numerous studies have shown the benefit of DES for reducing the rate of revascularization [35–41, 43, 44, 64]. As it is well established that DESs out-perform BMS at reducing revascularizations, studies comparing CABG to DES now can compare the rate of repeat revascularization as primary endpoints, such as in the SYNTAX trial.

Furthermore, trials currently underway comparing CABG to DES are designed to compare the rate of cerebrovascular accidents, as CABG have shown higher rates of CVA [64a]. The SYNTAX trial has been designed to compare PCI with DES (TAXUS) to CABG for the treatment of triple-vessel and left main disease [65]. The SYNTAX trial will compare the rate of major adverse cardiac and cerebral events (MACCE) between DES (TAXUS) and surgery with CABG. The study includes two nested registries for subjects ineligible for CABG who are treated with PCI and for subjects ineligible for PCI who are treated with CABG. Complete data from the primary objective of the SYNTAX trial is not available yet, but differences in CABG and PCI practice patterns have been observed between North America and Europe. Thus far, the rate of PCI intervention for triple-vessel disease in Europe is approximately double the rate from North America [65]. Initial reports from the SYNTAX trial in late 2008 showed similar composite rates for death, stroke, and MI in both arms, but reduced repeat revascularization rates in the CABG arm and reduced stroke rate in the PCI arm.

The FREEDOM trial is another current study to investigate CABG versus DES (CYPHER) in the treatment for subjects with CAD and diabetes mellitus [66]. The primary objective will be to assess the effectiveness of PCI compared to CABG (on or off pump) in the prevention of all-cause mortality, nonfatal MI, or stroke. Outcomes of mortality, MI, revascularization, health-related quality of life, and cost effectiveness will be addressed for a mean of 4 years.

The SYNTAX and FREEDOM trials will have completed follow-up in 2012 and 2011, respectively. The SYNTAX primary endpoint results are expected to be communicated in the fall of 2008. SYNTAX and FREEDOM will provide valuable information on CABG treatment versus DES treatment, but data for other type of DES would bring more information for choosing the best treatment options for CAD patients.

### 10.1.3.2 Drug Therapy

Pharmaceutical advances in lipid-lowering drugs, antidiabetic agents, antiplatelet therapy, and antihypertensive medications in the 1990s have provided patients and

physicians with drug therapy treatment options. The MASS II trial compared PCI, CABG, and OMT modalities in subjects with stable angina and multivessel CAD [67]. The primary endpoints were total mortality, Q-wave myocardial infarction (QWMI), or refractory angina requiring revascularization. In the MASS II trial, all three therapeutic strategies yielded comparable results with relatively low rates of mortality. The rates of additional revascularization and long-term events were found to be similar for PCI and OMT, but were the lowest for CABG. Devices used for catheter-based PCI treatments included stents, lasers, directional atherectomy, and balloon angioplasty. Since enrollment was completed before DESs were approved by the FDA, no DES was included in the PCI arm of the MASS II trial. Overall, BMS compared to DES represents up to a 60% higher rate of revascularization [68]. Despite this fact for patients with stable multivessel CAD, the MASS II trial showed that routine PCI is as good as CABG or OMT. The MASS II trial results suggest that patients with mild to moderate angina are good candidates for OMT, whereas a PCI or CABG would be appropriate for patients with symptoms that could not be controlled or were more severe.

The COURAGE trial compared outcomes of subjects initially treated by OMT with PCI versus OMT alone for significant CAD. The primary endpoint in the COURAGE trial was all-cause mortality and nonfatal MI, which may well have been a limiting factor. By selecting the primary endpoint of all-cause mortality, the rate of mortality would include the rate of death from any cause and not just a death due to cardiac events. By choosing this type of endpoint, the comparison of mortality rates would include additional noise and would not provide the most accurate comparison. Because the COURAGE trial was comparing cardiac interventions, a more appropriate endpoint might have been cardiac events. In fact, a more appropriate safety endpoint would have included all MI events rather than only the nonfatal MI. Predictably, by selecting this primary endpoint, there was no significant difference in the rates of death distinguishable between the OMT and PCI versus the OMT alone [69]. The trial enrolled  $\geq 97\%$  of PCI subjects treated with BMS compared to 3% PCI subjects treated with DES. The very low rate of DES subjects ( $< 3\%$ ) in the COURAGE trial does not allow any comparison of PCI with DES to OMT [56]. This observed bias is because the DES was approved by the FDA at a very late stage of patient enrollment in the COURAGE trial. An additional weakness of the COURAGE trial was that it was powered to show superiority of angioplasty versus OMT with a reduction in death and MI by 22% [56]; this has never been the clinical driver of stenting in stable patients, in whom the major benefit of interventional therapy is reduction of angina, and improved quality of life. Moreover, approximately 33.0% of the OMT patients crossed over to the PCI group at a mean time of 10.2 months for angioplasty, which suggests that OMT was not an adequate treatment. Another weakness of the COURAGE trial was that the investigators modified the definition of MI part way through the study because of low event rates [56]. Currently, no data are available from major trials for comparing PCI with DES to OMT. However, it was suggested that OMT may be preferred for those more stable subjects with less severe progression of CAD.

### 10.1.3.3 Plain Old Balloon Angioplasty

Earlier trials comparing PTCA to CABG used primary endpoints that focused on safety; the BARI trial, for example, examined all-cause mortality as a primary

endpoint [70]. The EAST trial also compared PTCA to CABG with a composite endpoint of death, QWMI, and evidence of ischemia at 3 years [71]. Overall findings from these trials did not differ significantly with primary endpoint results. However, revascularization rates were higher in PTCA subjects. Trials were then designed to compare POBA to medical therapy. RITA-II used a primary endpoint of death and definite MI rates at 5 years and showed comparable mortality rates. However, lower rates of restenosis were noted when angioplasty was used in conjunction with coronary stents, compared to using angioplasty alone [72]. The benefit of decreased restenosis led to less repeat procedures that reduced long-term cost of repeat revascularization [73]. As PCI with stenting was introduced, comparisons of POBA with PCI indicated that POBA was not as effective as PCI at reducing restenosis [5, 6]. Study designs then evolved to include composite endpoints including both survival and revascularization rates. Endpoints such as TVF and MACE were implemented. As stent technology expanded through the 1990s, its designs, materials, techniques, and deployment improved, such that stenting has dominated over balloon angioplasty over the last 15 years.

#### 10.1.3.4 Stenting

The FDA first approved PCI with BMS in 1996. It was later reported that 20% of subjects treated with BMS had experienced recurrent symptoms attributed to restenosis from intimal hyperplasia [74]. Despite several studies showing angiographic and clinical improvements with these catheter-based interventions, restenosis within the BMS stent continued to be a major problem that caused repeat revascularizations [5, 6, 59, 75]. Comparison of PCI with BMS to PCI with DES showed decreased rates of revascularization with DES. Currently, more than 15 analyses have been performed to evaluate the performance of DES versus BMS that have generally shown that DESs are safe and effective [76–91]. One of the largest meta-analyses comparing 19 different trials found that restenosis and MACE were greatly reduced with DES as compared with BMS [92]. A more recent study, published in 2008, showed no significant difference between DES and BMS groups in the rate of death or MI at 1-year follow-up [93].

However, the rate of revascularizations was significantly lower in the DES group, and current studies continue to support the benefit of DES compared to BMS. Trial designs comparing DES to BMS sufficiently demonstrated that safety was comparable between DES and BMS. In addition, by using composite endpoints including revascularization rates, DES showed to be a great improvement to BMS at reducing the need for repeat revascularizations. Studies continue to show comparable safety results and improved efficacy for PCI with DES compared to PCI with BMS. Future trials with next-generation DES versus BMS should continue to investigate the rates of ST and cardiac death, as trials have already shown revascularization rates were dramatically improved with use of the first-generation DES compared to BMS.

Many clinical trials have compared DES to each other. The comparison of two first-generation DES, CYPHER versus TAXUS in the REALITY [51], SIRTAX [52], and ISAR-DIABETES [79] clinical trials did not show superiority of TAXUS over CYPHER. There was no statistically significant difference in the rate of death, MI, and revascularization. ENDEAVOR IV compared the newly approved DES, ENDEAVOR, to TAXUS, and it was concluded that ENDEAVOR was noninferior to TAXUS for the primary endpoint of TVF. The SPIRIT III trial compared the next

generation of DES, XIENCE V, to TAXUS. The trial included the use of an FDA-suggested angiographic primary endpoint of in-segment late loss, in addition to a co-primary endpoint of TVF at 9 months. The follow-up results demonstrated superiority of XIENCE V to TAXUS for the primary endpoints, with lower in-segment LL. In addition, XIENCE V had fewer MIs and less repeat target lesion revascularization procedures than TAXUS, all of which suggest that XIENCE V is safe and effective for treating patients with CAD. Additional DES trials designed after the SPIRIT III trial will probably no longer use surrogate or angiographic endpoints, as the emphasis of the FDA has shifted to clinical and safety endpoints. The SPIRIT IV trial will include the rate of MACE at one year as the primary endpoint.

#### 10.1.4 PCI STUDY DESIGN ENDPOINTS

Clinical trials are designed to evaluate the safety and effectiveness of DES. Selected endpoints play pivotal roles in device approval and in their adoption for clinical use. These endpoints must serve several purposes. They must have both short- and long-term pathophysiological relevance to device performance, represent clinically meaningful events, be sufficiently defined, preferably through blinded processes, and be subjected to statistical analysis. Because of the intrinsic limitations in the ability to obtain histology, serial examinations, or other mechanistic details from human subjects, clinical endpoints for DES studies are bound to include certain arbitrary assumptions and will frequently vary across clinical trials as the result of different approaches to such assumptions [94].

##### 10.1.4.1 Clinical Endpoints

Endpoints for clinical trials evaluating devices should measure safety, effectiveness, and patient well-being.

Safety endpoints represent any adverse outcome during the course of the clinical evaluation and whether or not it is specifically related to the use of the device. Safety endpoints are measured by events such as death, MI, stent thrombosis, or composite endpoints such as MACE, TVR, death, and MI. Effectiveness endpoints refer specifically to maintenance of coronary artery luminal patency, which is assessed by measuring the relief of such flow-limiting obstructions, initially through structural mechanisms and later with preservation of the luminal dimension through inhibition of neointimal hyperplasia or restenosis. Effectiveness-driven clinical endpoints are designed to assess clinically significant restenosis, assessed objectively as a requirement for ischemia-driven repeat revascularization, either of the stented segment itself, TLR, or of the stented vessel or its side branches, TVR [94].

Effectiveness endpoints refer specifically to maintenance of coronary artery luminal patency. DESs are implanted for the treatment of obstructive CAD. Their effectiveness is assessed by measuring the relief of such flow-limiting obstructions, initially through structural mechanisms and later with preservation of the luminal dimension through inhibition of neointimal hyperplasia or restenosis. Effectiveness clinical endpoints are designed to assess clinically significant restenosis, assessed objectively as a requirement for ischemia-driven repeat revascularization, either of the stented segment itself, TLR, or of the stented vessel or its side branches, TVR

[95]. Target vessel failure, defined as any TVR, death, or MI attributed to the target vessel, is an even broader metric of failed effectiveness and adjusts for the potential bias introduced when patients who die or sustain MI before the end of the TLR endpoint time are considered to be free from TLR [94].

General considerations for DES evaluation should also include cardiovascular outcomes from the patient's perspective. The outcomes of endpoints should reflect the complex interplay between device performance, revascularization strategy, secondary prevention, and key patient descriptors [96]. Both the time course and the composite selected should characterize patient well-being related to the pathophysiology of the implanted DES device and its impact on underlying CAD outcome.

The most commonly measured clinical indicator of stent efficacy is TLR, which is defined as recurrent ischemia due to angiographic restenosis within the stent or its margins necessitating repeat revascularization with either PCI or coronary artery bypass graft surgery. TLR rates are typically 30–60% lower than the corresponding binary restenosis rates, suggesting discordance between the ischemic thresholds of angiographic and clinical measures of restenosis [95, 97]. Given the low rates of TLR after BMS implantation in the relatively noncomplex lesions typically studied in stent trials, more than 2000 enrolled patients are required in randomized trials to demonstrate a clinically relevant 30% reduction in TLR with DES. Moreover, given the low-frequency clinical event rates with DES, very large potential differences between stents (the “delta”) often are allowed in noninferiority trials to make comparative DES studies practical, degrading confidence that the clinical performances of two devices are indeed similar.

#### 10.1.4.2 Angiographic Endpoints

To reduce the sample size required in superiority and noninferiority trials, continuous angiographic indexes of long-term stent patency such as late LL and follow-up %DS have been proposed as surrogates of TLR for use in randomized studies [35, 96, 98]. However, previous DES studies have suggested that the distribution of individual LL measures is asymmetric with a rightward skew and that the relation between LL and TLR is nonlinear [36, 99]. Recently, Pocock et al. evaluated angiographic measures as valid surrogates for TLR and cutoff values that correspond with clinical efficacy [100]. Using data from 11 randomized trials, the investigators evaluated 4 angiographic measures used as surrogates for the clinical endpoint of TLR. They concluded that all four angiographic measures, instent and in-segment LL, as well as both assessments of %DS, provided a reliable estimation of TLR rates in all the trials and met the accepted criteria for surrogate endpoints. In the same report it was also concluded that %DS may have an advantage over LL as a surrogate endpoint. Therefore, Pocock et al. [100] suggested that angiographic surrogates can be used as the primary outcome measure in smaller patient populations compared to when TLR is the primary outcome, and larger trials should provide information on less frequent safety events.

#### 10.1.5 SAFETY

Clinically, safety after stenting can be measured by the rate of death, MI, or stent thrombosis or by the rate of composite endpoints such as TVF (composite of death,

MI, TVR) and MACE (composite of death, MI, TLR). Composite endpoints generated by the combination of individual endpoints provide additional statistical power to detect potentially meaningful differences between treatments. The individual components should each represent clinically meaningful events and should be linked by common elements of pathophysiology. The ARC consensus suggests two composite endpoints for DES trials, one that is device-oriented and one for overall patient-oriented clinical outcome. The device-oriented composite includes cardiac death, MI attributed to the target vessel, and TLR. The broader patient-oriented outcome composite includes all-cause mortality, any MI, and any revascularization (includes TLR), target vessel revascularization, or revascularization of nontarget vessels.

Many trials have evaluated the rates of ST after the PCI with DES versus BMS. However, there are some challenges to comparing ST rates between trials because of the differences in patient and lesion characteristics, the prevalence of diabetes mellitus (which ranged from 14 to 31%), the use of glycoprotein IIb/IIIa inhibitors (which ranged from 0 to 64%), and the use of various definitions for ST in trials performed up until 2006. Recently, the FDA, academia, and industry agreed to use ARC definitions in order to align and standardize the definitions used in clinical trials [94]. However, more than 15 studies evaluated the performance of DES versus BMS. There was no significant and consistent pattern of difference reported between the ST rates of DES and BMS. Meta-analyses of randomized trials performed with DES demonstrated that incidence of ST was not increased in patients receiving DES (0.58% for DES versus 0.54% for BMS) [86]. The overall rate of ST did not differ significantly between patients receiving sirolimus- or paclitaxel-eluting stents (0.57 versus 0.58%). In contrast, a significant relationship between the rate of ST and the stented length was found. In patients with DES, mean stented length was longer in those with ST. Per ARC definitions, the low incidence of ST was reported in the SPIRIT III trial and ST rates were low and comparable through one year for XIENCE V and TAXUS. Premature discontinuation of antiplatelet therapy has been identified as one of the predictors for ST [80]. Recent studies have reported a low incidence of DES thrombosis under prolonged therapy with aspirin plus thienopyridines, comparable to that of BMS, even in unstable clinical settings [101]. The most striking factor associated with DES thrombosis is the absence of treatment with ticlopidine/clopidogrel. In the ASPECT trial, the rate of DES thrombosis in patients receiving cilostazol instead of thienopyridines was 14.8% (4 of 27) [102]. In another study, 30% of patients discontinuing ticlopidine early after DES implantation suffered ST [103]. Furthermore, Park et al. [104] performed a trial in order to investigate the incidence, risk factors, and association of antiplatelet therapy interruption for the development of ST after DES implantation during long-term follow-up (median follow-up 19 months). In this study a total of 1911 consecutive patients with DES were enrolled. The results of this study demonstrated that the incidence of late ST was 0.6%, similar to that of metallic stents. The predictors of ST were premature antiplatelet therapy interruption, primary stenting in acute MI, and total stent length. For a rare event such as ST, using standardized definitions across postmarketing studies and registries may help to further identify predictor of ST. However, as the ST incidence is low, a large sample size is needed to accurately identify these predictors.



## 10.1.6 POSTMARKET REGISTRIES AND OFF-LABEL STENT USE

Drug-eluting stent commercialization and use is monitored by the Center for Device and Radiologic Health (CDRH), a branch of the FDA. Recently, the off-label market use of DES has been under scrutiny by the FDA. The issues surrounding the real-world use versus the approved FDA use will be discussed in the next two sections.

### 10.1.6.1 Real-World Issues

The FDA has mandated a postmarket safety surveillance plan that must include at least 2000 patients (21 CFR 820). The purpose of the surveillance is to assure that public safety is not compromised and that the product is being used in accordance with its approved indication. Postmarket registries also collect more data on low rate events, such as death, MI, and ST. More long-term follow-up on a larger subset of complex patients can also be accomplished in postmarket studies that could not be done in smaller randomized controlled trials. Complex patients with diabetes, renal dysfunction, acute MI, triple-vessel disease, bifurcations, left main coronary disease, heavily calcified lesions, and longer lesion segments are included in postmarket registries; these patients would otherwise be excluded from most preapproval randomized controlled trials [105]. Additionally, meta-analysis can be completed for indirect comparisons of treatments to determine study inconsistencies [106]. Moreover, the effect of antiplatelet therapy on ST and the compliance with dual antiplatelet therapy can be monitored in postmarket trials. In particular, the duration of clopidogrel use in combination with aspirin has not been established yet for DES. Three to six months of use was initially recommended for the first-generation DES. The American Heart Association/American College of Cardiology/Society for Cardiovascular Angiography and Interventions currently recommends up to 12 months of antiplatelet therapy with DES use for those patients not at high risk for bleeding [107]. Although the duration of clopidogrel appeared to be adequate for the selected patients in the original clinical trials, optimal duration of clopidogrel in more complex patients has not been defined. Many years of additional data are needed to characterize the effects of antiplatelet therapy on DES ST rates before conclusions can be drawn on the optimal antiplatelet therapy duration for DES patients.

Between 2002 and 2004, the CYPHER postmarketing surveillance program enrolled over 15,000 patients who received a CYPHER DES [108]. The results demonstrated safety with low rates of ST and TLR. ARRIVE-1 was a postmarket registry for TAXUS DES and enrolled over 2000 patients. The findings at 2 years showed that discontinuation of clopidogrel before 6 months was a significant predictor of ST at 1 and 2 years. In addition, rates of cardiac events at 6 months were similar for diabetic and nondiabetic patients [109]. The ARRIVE-2 registry was a postapproval study that enrolled over 4000 patients in the real-world setting. The ARRIVE-2 registry showed outcomes and usage patterns similar to those seen in ARRIVE-1 [105]. Boston Scientific pooled the data from these registries and was able to show that TAXUS DES was being implanted in more complex settings. These trials led to several ATLAS trials that aimed to expand the indications for use in these more complex procedures and patients [105]. Medtronic recently



received conditional approval for the ENDEAVOR DES and will initiate a post-approval study with a minimum of 5000 patients. In addition, Abbott Vascular is seeking approval for the XIENCE V DES in the United States and has a postmarket surveillance plan including real-world use trials in populations worldwide. The XIENCE V DES program will include up to 16,000 patients treated with XIENCE V in the United States, India, Japan, China, and Europe. The XIENCE V postmarket surveillance plan is designed to detect low rate events of cardiac death, MI, and ST. Furthermore, the XIENCE V program will monitor patients' long-term compliance with dual antiplatelet therapy.

Collectively, the XIENCE V, TAXUS, ENDEAVOR, and postmarket surveillance programs will provide a comprehensive data set for use in analyzing the low rate of events such as death, MI, and ST.

#### 10.1.6.2 Characterizing Off-Label Stent Use

Physicians may use a medical device off-label in good faith as the best available treatment option for the patient. The outcome of this off-label use can be a great benefit to the patient or may introduce considerable unnecessary risk or harm to the patient. Currently, the widespread off-label use of FDA-approved DES has prompted increased attention from regulatory agencies. Although the FDA does not regulate physician practice, it does establish guidelines for physicians to use products legally as indicated on the label. Monitoring safety- and efficacy-related information in these off-label use situations has become a considerable challenge for the manufacturer, physicians, and the FDA, making this dilemma a public health policy concern.

According to the FDA, it was estimated that 60% of DES use was off-label in 2006 [74]. Manufacturers and regulatory agencies can gather information regarding the breadth of off-label use through registries, postmarket surveillance/studies, device complaints, publications, payers, investigator-sponsored studies, professional societies, and scientific conferences. The manufacturer can gain considerable insight into the possibility of label expansion from this data because it may indicate that there are unmet needs in the market. In the case of DES, several off-label uses have prompted further studies to support approval of DES for more complex cases [105]. Some current off-label DES use involves bifurcation lesions, saphenous vein grafts, in-stent restenosis, total occlusions, acute MIs, and left main triple-vessel disease. Optimal medical therapy, BMS, and CABG therapies can also be compared against off-label DES use through explorative meta-analysis registries. A recent study compared BMS and DES off-label use versus BMS and DES on-label use. The results demonstrated that the risks of MI, death, or ST were lower for patients treated on-label versus off-label in both BMS and DES. There was also no difference observed between DES types [110]. In another recent trial comparing off-label use in DES to BMS, it was demonstrated that the use of DES was not associated with increased risk of death or MI at 1 year and that DES use continued to show lower revascularization rates in off-label use settings compared to BMS [93]. Without further data, the safety and efficacy of off-label use in all DES remains unknown.

The major disadvantage of off-label DES use is the potential for serious adverse outcomes to patients. The manufacturer may have initially sought FDA approval for an indication that was denied because of safety concerns. Physicians in everyday

practice may not be aware of the reasons why an indication was rejected or of this additional risk to their patients. Currently, there is a lack of process that can accurately analyze the outcomes of all off-label use. Patients not entered into a trial or registry may never receive long-term follow-up and their outcomes may remain unknown. The FDA has an obligation to obtain this information, yet at the same time it cannot condone the off-label practice that would generate the data needed for analysis. The manufacturer is restricted from promoting off-label use in practice. Physicians may use a DES for an unapproved indication instead of waiting for FDA approval, which can be a lengthy and complex process. The label expansion process can take as many or more resources as the initial device approval did [111]. Current FDA limitations for label expansion make this a costly and lengthy process, which ultimately may restrict the best treatment options for the patient. In the future, the industry and regulatory approving bodies must work toward a system that allows for rapid approval process while maintaining safety standards.

The practice of using DES for unapproved product indications is an increasing concern. Additional data are needed to assure the safety and efficacy of DES for the complex cases, lesions, and patients that are being treated off-label. Data should be gathered from randomized control trials, postapproval registries, or through national databases from physicians and manufacturers.

### 10.1.7 PERIPHERAL STENTS

Peripheral atherosclerotic occlusive disease or *peripheral vascular disease* (PVD) is exceedingly common in aging western societies; the prevalence of plaque in the femoral arteries of patients over the age of 70 may be as high as 74% [112]. Symptomatic PVD affects approximately 10% of the population over the age of 50, and as much as 20% of the population over the age of 70. Failure to maintain or restore adequate perfusion to the lower extremity results in nearly 250,000 lower extremity amputations annually.

Symptomatic PVD has traditionally been treated by surgical bypass grafting. First conceived in 1949 [113], and popularized in the 1980s through refinements in anesthesia and technique, bypass grafting was the accepted standard of care for nearly 50 years. Unfortunately, patient morbidity is considerable, as those requiring this invasive procedure are older and infirm, and the operation requires relatively long incisions and dissections through ischemic tissue. Hoping to circumvent the excessive morbidity of surgical bypass grafting, techniques of endovascular infrainguinal revascularization were developed in the 1980s. Percutaneous peripheral transluminal balloon angioplasty was first described in 1982, followed closely by reports of adjunctive stent implantation in 1985 [114, 115]. These techniques were adapted and enhanced over the ensuing two decades, including significant improvements in access, imaging, negotiation of chronic occlusions and, most notably, the wider application of self-expanding nitinol stent technology. Despite these refinements, maintenance of long-term patency following endovascular interventions is limited, as restenosis eventually complicates up to 50% of procedures during the first year (Fig. 3) [112, 116].

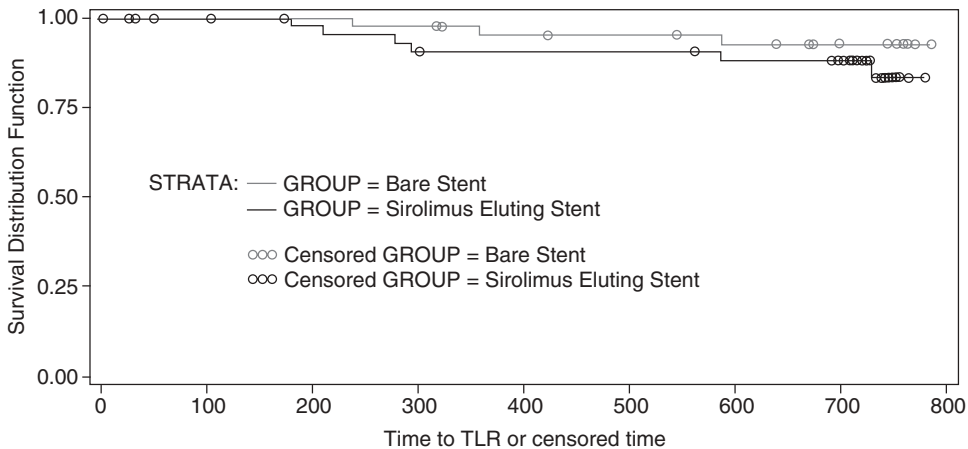
In the coronary circulation, the problem of restenosis was eventually circumvented through the development of coronary DESs [117]. Given their success, many



**FIGURE 3** Peripheral in-stent neointimal hyperplasia and restenosis. [From Scheinert, D., Scheinert, S., Sax, J., et al. (2005), *Prevalence and clinical impact of stent fractures after femoropopliteal stenting*, *J. Am. Coll. Cardiol.*, 45, 312–315.]

have theorized that this technology might also be useful in the peripheral arteries. The Sirocco stent (Cordis, a Johnson & Johnson Company, Miami Lakes, Florida) remains the first, and only, peripheral DES to be tested in clinical trials and reported in the scientific literature [118–120]. The Sirocco stent utilized the nitinol self-expanding S.M.A.R.T. stent as its platform, was loaded with  $90\mu\text{g}$  sirolimus/ $\text{cm}^2$  stent area using a 5- to  $10\text{-}\mu\text{m}$  copolymer matrix (total drug content  $\sim 1\text{ mg}$  per 80 mm stent), and delivered its drug load over a period of about 7 days [119]. A total of 93 patients were enrolled in the combined SIROCCO I and SIROCCO II clinical trials. Unfortunately, neither trial achieved its primary endpoint of a reduction in restenosis and, even after 4 years, there was no difference in any metric comparing patients treated with the bare S.M.A.R.T. stent versus the sirolimus-eluting Sirocco stent (Fig. 4) [120]. The development of the Sirocco stent was terminated; as of this writing, no peripheral DES is yet available for clinical use anywhere in the world.

Despite the disappointing results of the Sirocco studies, two other peripheral DESs are currently being tested in clinical trials. The Zilver PTX stent (Cook Inc.,



**Kaplan-Meier Analysis of Target Lesion Revascularization (n = 93 pts)**

| Interval ending day             | 0     | 30    | 90    | 150   | 210  | 270  | 330  | 360  | 450  | 540  | 600  | 630  | 660  | 720  | 750  |
|---------------------------------|-------|-------|-------|-------|------|------|------|------|------|------|------|------|------|------|------|
| <b>Bare stent (n = 46)</b>      |       |       |       |       |      |      |      |      |      |      |      |      |      |      |      |
| # Censored                      | 0     | 1     | 0     | 0     | 0    | 0    | 1    | 0    | 0    | 1    | 1    | 0    | 0    | 10   | 17   |
| # At risk                       | 46    | 45.5  | 44.5  | 44    | 42   | 42   | 41.5 | 39   | 39   | 38.5 | 37.5 | 36   | 36   | 31   | 17   |
| # Events                        | 0     | 0     | 0     | 0     | 2    | 0    | 2    | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 1    |
| Cumulative survival, %          | 100.0 | 100.0 | 100.0 | 100.0 | 95.5 | 95.5 | 90.9 | 90.9 | 90.9 | 90.9 | 88.5 | 88.5 | 88.5 | 88.5 | 83.8 |
| <b>Sirolimus stent (n = 47)</b> |       |       |       |       |      |      |      |      |      |      |      |      |      |      |      |
| # Censored                      | 0     | 1     | 0     | 0     | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 7    | 21   |
| # At risk                       | 47    | 46.5  | 45.5  | 44.5  | 44   | 44   | 42   | 41   | 39.5 | 39   | 38.5 | 37   | 36.5 | 32   | 17.5 |
| # Events                        | 0     | 0     | 0     | 0     | 0    | 1    | 0    | 1    | 0    | 0    | 1    | 0    | 0    | 0    | 0    |
| Cumulative survival, %          | 100.0 | 100.0 | 100.0 | 100.0 | 97.7 | 97.7 | 95.3 | 95.3 | 95.3 | 95.3 | 92.8 | 92.8 | 92.8 | 92.8 | 92.8 |

**FIGURE 4** Freedom from target lesion revascularization (TLR) in the SIROCCO II trial. Kaplan-Meier event analyses of target lesion revascularization: 92.8% for sirolimus versus 83.8% for control ( $p = 0.30$ ). Reprinted from Duda, S. H., Bosiers, M., Lammer, J., et al. (2006), Drug-eluting and bare nitinol stents for the treatment of atherosclerotic lesions in the superficial femoral artery: Long-term results from the SIROCCO trial, *J. Endovasc. Ther.*, 13, 701-710.

Bloomington, Indiana) contains approximately  $300\mu\text{g}$  paclitaxel (PTX)/ $\text{cm}^2$  stent area [121]. Similar to the ACHIEVE stent tested in the DELIVER clinical trial [122], paclitaxel is applied directly to the Zilver PTX stent; there is no polymer system serving to control its release. Interim clinical results from a small trial have recently been presented [123]. Sixty patients with symptomatic PVD were randomly assigned to undergo endovascular intervention with either the Zilver PTX stent ( $n = 28$ ) or percutaneous transluminal balloon angioplasty alone (PTA,  $n = 32$ ). Patients in whom PTA yielded suboptimal intraoperative results (16/32) were secondarily randomized to receive either a Zilver PTX stent or a bare nitinol Zilver stent. Although no angiographic results have been reported to date, the safety of this approach is suggested by the equivalent incidence of total events in the two groups after 6 months (incidence of death, revascularization and/or worsened Rutherford score in the PTA and Zilver PTX groups was 91 and 89%, respectively). Further testing of the safety and efficacy of the Zilver stent is ongoing in the 480-patient DESTINY trial using a similar design. Nearly half the patients have been recruited as of this writing.

Lastly, an everolimus-eluting nitinol stent has also recently entered the clinical phase of development. The Dynalink-E device (Abbott Vascular, Santa Clara, California) contains approximately 225 µg everolimus/cm<sup>2</sup> stent area, which is released slowly through incorporation in an ethylene vinyl alcohol (EVAL) copolymer. Enrollment was recently completed in the STRIDES First-in-Human clinical trial in Europe and initial results are expected in 2009.

In summary, endovascular intervention has become the treatment of choice for many patients with symptomatic PVD. Similar to percutaneous coronary intervention, it is hoped that the persistent and frequent problem of restenosis can be ameliorated through development of effective DES.

### 10.1.8 FUTURE TECHNOLOGY

#### 10.1.8.1 Future Stents

Although the development and testing of the first-generation DES focused to a considerable degree on efficacy parameters, including restenosis, recent concerns over late clinical events have prompted refinement of the design criteria for succeeding generations of coronary artery devices.

Most new investigational balloon-expandable DES systems have lowered crossing profiles by thinning stent struts with the use of a cobalt chromium alloy, while investigational self-expanding DESs often use nitinol as the platform material. To further lower crossing profile, guidewire-based stents are currently under development [124]. Fully biodegradable stents are also being developed, with deliverability and performance to be determined in future clinical trials. New bifurcation-dedicated stents will secure branch accessibility to offer better deliverability in complex lesion morphologies. As noted, experimentation in stent design is already realizing multiple-lesion stenting and the *in situ* customization of stent length [124].

First-generation DESs have significantly lowered the rate of TLR compared with BMS. Rather than simply targeting further reductions in restenosis rates, efforts to improve efficacy are shifting toward a lesion-specific approach, including the design of stents dedicated to bifurcation lesions. Another future direction is a disease-specific approach, using DES as local drug delivery devices. So conceived, DESs present the possibility of expanding options for the local pharmacologic modulation of lesions. Potential therapeutic uses still in the investigational stage include myocardial protection at the time of infarct and the stabilization of vulnerable plaque.

Identifying long-term safety issues with the first-generation DES has reignited clinical interest in the development of stents that are more biologically based, including fully biodegradable stents and stents using biomimetic and biodegradable polymers. Important performance criteria for future DES agents include more cell-type specificity, broader safety margins, and greater facility at promoting endothelialization and healing. In addition, future vascular biologic findings on the mechanisms of restenosis will shed light on the potential relationship between gene therapy and local agent delivery. Safety is the perpetual challenging issue in the field of medicine. For interventional cardiologists, the recent recognition of a possible increase in late

clinical events in patients with DES has forced a reexamination and redefinition of the requirements of the ideal DES. Although previous DES clinical trials placed great emphasis on efficacy, especially midterm angiographic and clinical results, these findings pale in comparison to ensuring long-term safety for our patients.

New therapeutic agents such as proteins, nucleic acids [small interfering ribonucleic acids (RNAs) and large deoxyribonucleic acid (DNA) plasmids], viral delivery vectors, and even engineered cell therapies require specific delivery designs distinct from traditional smaller-molecule approaches on DES. While small molecules are currently the clinical standard for coronary stenting, extension of the DES to other lesion types, peripheral vasculature, and nonvasculature therapies will seek to deliver an increasingly sophisticated armada of drug types. In parallel with PCI procedure evolution, new technologies such as OCT will be introduced for measuring safety and effectiveness of new devices. The application of these new technologies to PCI procedures may require validation of new safety and effectiveness measures.

#### 10.1.8.2 Cardiac Tissue

Another alternative method for increasing coronary vascularization is the transplantation of stem or progenitor cells [125, 126]. These cells not only produce a variety of growth factors and cytokines but participate structurally in the formation of new vascular tissue and myocytes.

Promising results from experimental studies promoted the initiation of clinical trials of stem cell technologies [127, 128]. Stem and progenitor cells are being tested in patients with both acute MI and chronic ischemic heart failure [125, 129–136]. Improved wall motion or increased perfusion were demonstrated in most study patients. BOOST is a randomized controlled clinical trial [135] in which 60 patients with ST-segment elevation were randomly assigned to either a control group that received optimum postinfarction medical treatment or a bone-marrow-cell group that received optimum medical treatment and intracoronary transfer of autologous bone marrow cells 4.8 days after percutaneous coronary intervention (PCI). After 6 months, mean global left ventricular ejection fraction (LVEF) had significantly increased in the bone-marrow-cell group, compared to the control group. However, most studies at present are limited by the small patient enrollment, and recently some papers reported that bone-marrow-derived hematopoietic stem cells do not transdifferentiate into cardiac myocytes in ischemic myocardium [137, 138].

#### 10.1.8.3 Genetic Advances

A number of vehicles are used to transfer DNA into heart tissue, including purified DNA, DNA/lipid complexes, adeno-associated viruses, or adenovirus [125, 139]. An advantage of this approach appears to be production of angiogenic factors that is localized and sustained but not indefinite [140]. Apparent disadvantages are the possibilities of vector toxicity or an immune response to the gene therapy vector.

The efficacy of gene transfer approaches to therapeutic angiogenesis is now being tested in clinical trials. Early uncontrolled open-label clinical trials have generally given positive results, although the possibility of a placebo effect has not been excluded [126, 139, 141–147]. Controlled phase II trials are providing positive but

not definitive results. This is promising since the patient population being studied has failed all other therapies and is likely to be refractory to intervention. However, most of the efficacy measures, studied to date, have been surrogate endpoints such as exercise tolerance time, angina, or perfusion. While these measures are useful in suggesting clinical efficacy, hard clinical endpoints such as mortality, MI, and the need for revascularization should be studied. Long-term follow-up data are also needed. An important observation is that the safety results of these trials indicate no major problems. Potential side effects such as worsening of atherosclerosis, retinopathy, or cancer have not been observed in clinical trials.

Two large phase III clinical trials (AGENT 3 and 4) were designed to evaluate further the safety and efficacy of Ad5FGF4. Both trials were designed as randomized, double-blind, and placebo-controlled trials [141]. In each trial, the recruitment goal was 450 patients, and these patients would be randomized to 3 groups (placebo group, adenovirus (Ad) human fibroblast growth factor (FGF) gene (Ad5FGF-4) at a dose of  $10^9$  vp group, and Ad5FGF-4 at a dose of  $10^{10}$  vp group). In January 2004, enrollment was stopped (416 patients in AGENT 3 and 115 patients in AGENT 4) because interim data analysis of AGENT 3 indicates that the studies will provide insufficient evidence of efficacy. However, enrolled patients' follow-up continues and final data will be presented in the near future.

### 10.1.9 CONCLUSIONS

The development of DES represents an exciting area of breakthrough technology, which has generated an enormous literature in parallel with widespread use in a short period of time. Interaction of innovative stent platforms, polymers, and molecular entities, as well as pharmaceutical adjuncts such as dual antiplatelet therapy, present a unique degree of complexity for systematic ongoing evaluation of these devices, their optimal use, and their real safety and performance results. Toward this end, clinical trials and DES industry programs have developed a broad variety of endpoint definitions, which differ across a heterogeneous array of cutoff values, timing of endpoint assessment, and outcome composites. The number of stents currently under investigation is substantial. They are all loaded with drugs that interfere with pathways in the process of inflammation and neointimal proliferation. However, the process of restenosis is a sequence of complex events that has been partly elucidated over the last 2 decades. Locally acting DESs provide the opportunity to interfere with the various mechanisms responsible for each step in the restenotic cascade [148], and a wide variety of different agents are currently available. Johnson and Johnson's sirolimus-eluting stents (SES), Boston Scientific's paclitaxel-eluting stents (PES), Medtronic's zotarolimus-eluting stents, and Abbott Vascular's XIENCE V everolimus-eluting stents have received FDA approval. Recently, in view of late adverse events after stent implantation, the FDA has called for long-term monitoring of safety outcomes in the postapproval context. It is therefore likely that postmarket programs for new DES will have larger sample size and longer follow-up in order to identify the true incidence of late stent thrombosis, a low frequency event. Such programs resemble traditional pharmacovigilance in their scope and size, which are likely to be global, and once implemented and successful, can result in restoration of confidence to the DES arena.



## REFERENCES

1. Gruentzig, A. R., King, S. B., 3rd, Schlumpf, M., and Siegenthaler, W. (1987), Long-term follow-up after percutaneous transluminal coronary angioplasty. The early Zurich experience, *N. Engl. J. Med.*, 316, 1127–1132.
2. Holmes, D. R. Jr., Vlietstra, R. E., Smith, H. C., et al. (1984), Restenosis after percutaneous transluminal coronary angioplasty (PTCA): A report from the PTCA Registry of the National Heart, Lung, and Blood Institute, *Am. J. Cardiol.*, 53, 77C–81C.
3. Baim, D. S. (2004), New devices for percutaneous coronary intervention are rapidly making bypass surgery obsolete, *Curr. Opin. Cardiol.*, 19, 593–597.
4. Baim, D. S., Kent, K. M., King, S. B., 3rd, et al. (1994), Evaluating new devices. Acute (in-hospital) results from the New Approaches to Coronary Intervention Registry, *Circulation*, 89, 471–481.
5. Fischman, D. L., Leon, M. B., Baim, D. S., et al. (1994), A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease. Stent Restenosis Study Investigators, *N. Engl. J. Med.*, 331, 496–501.
6. Serruys, P. W., de Jaegere, P., Kiemeneij, F., et al. (1994), A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. Benestent Study Group, *N. Engl. J. Med.*, 331, 489–495.
7. Fricke, F. U., Silber, S. (2005), Can PCI with drug-eluting stents replace coronary artery bypass surgery? A comparative economic analysis regarding both therapeutic options based on clinical 12-month data reflecting the German social health care insurance system, *Herz*, 30, 332–338.
8. Baim, D. S., Mehran, R., Kereiakes, D. J., et al. (2000), Postmarket surveillance for drug-eluting coronary stents: A comprehensive approach, *Circulation*, 113, 891–897.
9. Barragan, P., Rieu, R., Garitey, V., et al. (2000), Elastic recoil of coronary stents: A comparative analysis, *Catheter Cardiovasc. Interv.*, 50, 112–119.
10. Ormiston, J. A., Dixon, S. R., Webster, M. W., et al. (2000), Stent longitudinal flexibility: A comparison of 13 stent designs before and after balloon expansion, *Catheter Cardiovasc. Interv.*, 50, 120–124.
11. Kastrati, A., Dirschinger, J., Boekstegers, P., et al. (2000), Influence of stent design on 1-year outcome after coronary stent placement: A randomized comparison of five stent types in 1,147 unselected patients, *Catheter Cardiovasc. Interv.*, 50, 290–297.
12. Hoffmann, R., Jansen, C., Konig, A., et al. (2001), Stent design related neointimal tissue proliferation in human coronary arteries; an intravascular ultrasound study, *Eur. Heart J.*, 22, 2007–2014.
13. Kastrati, A., Mehilli, J., Dirschinger, J., et al. (2001), Restenosis after coronary placement of various stent types, *Am. J. Cardiol.*, 87, 34–39.
14. McLean, D. R., Eiger, N. L. (2002), Stent design: Implications for restenosis, *Rev. Cardiovasc. Med.*, 3(Suppl 5), S16–22.
15. Rogers, C., Edelman, E. R. (1995), Endovascular stent design dictates experimental restenosis and thrombosis, *Circulation*, 91, 2995–3001.
16. Yoshitomi, Y., Kojima, S., Yano, M., et al. (2001), Does stent design affect probability of restenosis? A randomized trial comparing Multilink stents with GFX stents, *Am. Heart J.*, 142, 445–451.
17. Ku, D. N., Giddens, D. P., Zarins, C. K., and Glagov, S. (1985), Pulsatile flow and atherosclerosis in the human carotid bifurcation. Positive correlation between plaque location and low oscillating shear stress, *Arteriosclerosis*, 5, 293–302.

18. Malek, A. M., Alper, S. L., Izumo, S. (1999), Hemodynamic shear stress and its role in atherosclerosis, *JAMA*, 282, 2035–2042.
19. Moore, J. E. Jr., Xu, C., Glagov, S., Zarins, C. K., and Ku, D. N. (1994), Fluid wall shear stress measurements in a model of the human abdominal aorta: oscillatory behavior and relationship to atherosclerosis, *Atherosclerosis*, 110, 225–240.
20. Briguori, C., Sarais, C., Pagnotta, P., et al. (2002), In-stent restenosis in small coronary arteries: Impact of strut thickness, *J. Am. Coll. Cardiol.*, 40, 403–409.
21. Pache, J., Kastrati, A., Mehilli, J., et al. (2003), Intracoronary stenting and angiographic results: strut thickness effect on restenosis outcome (ISAR-STEREO-2) trial, *J. Am. Coll. Cardiol.*, 41, 1283–1288.
22. Rittersma, S. Z., de Winter, R. J., Koch, K. T., et al. (2004), Impact of strut thickness on late luminal loss after coronary artery stent placement, *Am. J. Cardiol.*, 93, 477–480.
23. Serruys, P. W. (1998), *Handbook of Coronary Stents*, Martin Dunitz Ltd., London.
24. van der Hoeven, B. L., Pires, N. M., Warda, H. M., et al. (2005), Drug-eluting stents: Results, promises and problems, *Int. J. Cardiol.*, 99, 9–17.
25. Axel, D. I., Kunert, W., Goggelmann, C., et al. (1997), Paclitaxel inhibits arterial smooth muscle cell proliferation and migration in vitro and in vivo using local drug delivery, *Circulation*, 96, 636–645.
26. Wong, A., Chan, C. (2004), Drug-eluting stents: The end of restenosis? *Ann. Acad. Med. Singapore*, 33, 423–431.
27. Ong, A. T., Serruys, P. W. (2005), Technology Insight: an overview of research in drug-eluting stents, *Nat. Clin. Pract. Cardiovasc. Med.*, 2, 647–658.
28. Regar, E., Sianos, G., Serruys, P. W. (2001), Stent development and local drug delivery, *Br. Med. Bull.*, 59, 227–248.
29. Winslow, R. D., Sharma, S. K., Kim, M. C. (2005), Restenosis and drug-eluting stents, *Mt. Sinai. J. Med.*, 72, 81–89.
30. Tanabe, K., Regar, E., Lee, C. H., Hoye, A., van der Giessen, W. J., Serruys, P. W. (2004), Local drug delivery using coated stents: New developments and future perspectives, *Curr. Pharm. Des.*, 10, 357–367.
31. Kastrati, A., Mehilli, J., Dirschinger, J., et al. (2001), Intracoronary stenting and angiographic results: strut thickness effect on restenosis outcome (ISAR-STEREO) trial, *Circulation*, 103, 2816–2821.
32. Sketch, M. H. Jr., Ball, M., Rutherford, B., Popma, J. J., Russell, C., and Kereiakes, D. J. (2005), Evaluation of the Medtronic (Driver) cobalt-chromium alloy coronary stent system, *Am. J. Cardiol.*, 95, 8–12.
33. Kereiakes, D. J., Cox, D. A., Hermiller, J. B., et al. (2003), Usefulness of a cobalt chromium coronary stent alloy, *Am. J. Cardiol.*, 92, 463–466.
34. Kipshidze, N., Dangas, G., Tsapenko, M., et al. (2004), Role of the endothelium in modulating neointimal formation: Vasculoprotective approaches to attenuate restenosis after percutaneous coronary interventions, *J. Am. Coll. Cardiol.*, 44, 733–739.
35. Moses, J. W., Leon, M. B., Popma, J. J., et al. (2003), Sirolimus-eluting stents versus standard stents in patients with stenosis in a native coronary artery, *N. Engl. J. Med.*, 349, 1315–1323.
36. Stone, G. W., Ellis, S. G., Cox, D. A., et al. (2004), A polymer-based, paclitaxel-eluting stent in patients with coronary artery disease, *N. Engl. J. Med.*, 350, 221–231.
37. Morice, M. C., Serruys, P. W., Sousa, J. E., et al. (2002), A randomized comparison of a sirolimus-eluting stent with a standard stent for coronary revascularization, *N. Engl. J. Med.*, 346, 1773–1780.

38. Schampaert, E., Cohen, E. A., Schluter, M., et al. (2004), The Canadian study of the sirolimus-eluting stent in the treatment of patients with long de novo lesions in small native coronary arteries (C-SIRIUS), *J. Am. Coll. Cardiol.*, 43, 1110–1115.
39. Schofer, J., Schluter, M., Gershlick, A. H., et al. (2003), Sirolimus-eluting stents for treatment of patients with long atherosclerotic lesions in small coronary arteries: double-blind, randomised controlled trial (E-SIRIUS), *Lancet*, 362, 1093–1099.
40. Grube, E., Silber, S., Hauptmann, K. E., et al. (2003), TAXUS I: Six- and twelve-month results from a randomized, double-blind trial on a slow-release paclitaxel-eluting stent for de novo coronary lesions, *Circulation*, 107, 38–42.
41. Colombo, A., Drzewiecki, J., Banning, A., et al. (2003), Randomized study to assess the effectiveness of slow- and moderate-release polymer-based paclitaxel-eluting stents for coronary artery lesions, *Circulation*, 108, 788–794.
42. Ellis, S. G., Popma, J. J., Lasala, J. M., et al. (2005), Relationship between angiographic late loss and target lesion revascularization after coronary stent implantation: analysis from the TAXUS-IV trial, *J. Am. Coll. Cardiol.*, 45, 1193–1200.
43. Stone, G. W., Ellis, S. G., Cannon, L., et al. (2005), Comparison of a polymer-based paclitaxel-eluting stent with a bare metal stent in patients with complex coronary artery disease: A randomized controlled trial, *JAMA*, 294, 1215–1223.
44. Dawkins, K. D., Grube, E., Guagliumi, G., et al. (2005), Clinical efficacy of polymer-based paclitaxel-eluting stents in the treatment of complex, long coronary artery lesions from a multicenter, randomized trial: Support for the use of drug-eluting stents in contemporary clinical practice, *Circulation*, 112, 3306–3313.
45. Leon, M. (2007), The ENDEAVOR IV. TCT 2007, Washington DC.
46. Joner, M., Quee, S. C., Coleman, L., et al. (2006), Abstract 2460, Competitive comparison of reendothelialization in drug eluting stents, *Circulation*, 114.
47. Serruys, P., Ruygrok, P., Neuzner, J., et al. (2006), A randomized comparison of an everolimus-eluting coronary stent with a paclitaxel-eluting coronary stent: The SPIRIT II trial, *EuroIntervention*, 2, 286–294.
48. Serruys, P. W., Ong, A. T., Piek, J. J., et al. (2005), A randomized comparison of a durable polymer everolimus-eluting stent with a bare metal coronary stent: The SPIRIT first trial, *EuroIntervention*, 1, 58–65.
49. Stone, G. W. (2008), *SPIRIT III. EuroPCR*. Barcelona, Spain.
50. Goy, J. J., Stauffer, J. C., Siegenthaler, M., Benoit, A., Seydoux, C. (2005) A prospective randomized comparison between paclitaxel and sirolimus stents in the real world of interventional cardiology: The TAXi trial, *J. Am. Coll. Cardiol.*, 45, 308–311.
51. Morice, M. C., Colombo, A., Meier, B., et al. (2006), Sirolimus- vs paclitaxel-eluting stents in de novo coronary artery lesions: The REALITY trial: A randomized controlled trial, *JAMA*, 295, 895–904.
52. Windecker, S., Remondino, A., Eberli, F. R., et al. (2005), Sirolimus-eluting and paclitaxel-eluting stents for coronary revascularization, *N. Engl. J. Med.*, 353, 653–662.
53. Ormiston, J. A., Webster, M. W., Armstrong, G. (2007), First-in-human implantation of a fully bioabsorbable drug-eluting stent: the BVS poly-L-lactic acid everolimus-eluting coronary stent, *Catheter Cardiovasc. Interv.*, 69, 128–131.
54. Ormiston, J. A., Serruys, P. W., Regar, E., et al. (2008), A bioabsorbable everolimus-eluting coronary stent system for patients with single de-novo coronary artery lesions (ABSORB): A prospective open-label trial, *Lancet*, 371, 899–907.
55. [http://www.teamreva.com/heart\\_stents\\_tech.html](http://www.teamreva.com/heart_stents_tech.html) RMIW.
56. Kereiakes, D. J., Teirstein, P. S., Sarembock, I. J., et al. (2007), The truth and consequences of the COURAGE trial, *J. Am. Coll. Cardiol.*, 50, 1598–1603.

57. Anon. (2007), Prevalence of heart disease—United States, 2005, *MMWR Weekly*, 56, 113–118.
58. Bravata, D. M., Gienger, A. L., McDonald, K. M., et al. (2007), Systematic review: The comparative effectiveness of percutaneous coronary interventions and coronary artery bypass graft surgery, *Ann. Intern. Med.*, 147, 703–716.
59. Anon. (2002), Coronary artery bypass surgery versus percutaneous coronary intervention with stent implantation in patients with multivessel coronary artery disease (the Stent or Surgery trial): A randomised controlled trial, *Lancet*, 360, 965–970.
60. Morrison, D. A., Sethi, G., Sacks, J., et al. (2001), Percutaneous coronary intervention versus coronary artery bypass graft surgery for patients with medically refractory myocardial ischemia and risk factors for adverse outcomes with bypass: A multicenter, randomized trial. Investigators of the Department of Veterans Affairs Cooperative Study #385, the Angina with Extremely Serious Operative Mortality Evaluation (AWESOME), *J. Am. Coll. Cardiol.*, 38, 143–149.
61. Rodriguez, A., Bernardi, V., Navia, J., et al. (2001), Argentine Randomized Study: Coronary angioplasty with stenting versus coronary bypass surgery in patients with multiple-vessel disease (ERACI II): 30-day and one-year follow-up results. ERACI II Investigators, *J. Am. Coll. Cardiol.*, 37, 51–58.
62. Rodriguez, A. E., Baldi, J., Fernandez Pereira, C., et al. (2005), Five-year follow-up of the Argentine randomized trial of coronary angioplasty with stenting versus coronary bypass surgery in patients with multiple vessel disease (ERACI II), *J. Am. Coll. Cardiol.*, 46, 582–588.
63. Serruys, P. W., Ong, A. T., van Herwerden, L. A., et al. (2005), Five-year outcomes after coronary stenting versus bypass surgery for the treatment of multivessel disease: The final analysis of the Arterial Revascularization Therapies Study (ARTS) randomized trial, *J. Am. Coll. Cardiol.*, 46, 575–581.
64. Leon, M. B., Holmes, D. R., Jr Weisz, G., et al. (2006), Two-year outcomes after sirolimus-eluting stent implantation: Results from the Sirolimus-Eluting Stent in de Novo Native Coronary Lesions (SIRIUS) trial, *J. Am. Coll. Cardiol.*, 47, 1350–1355.
- 64a. Valgimigli, M. (2007), Impact of stable versus unstable coronary artery disease on 1-year outcome in elective patients undergoing multivessel revascularization with sirolimus-eluting stents, *JACC*, 49.
65. Kappetein, A. P., Dawkins, K. D., Mohr, F. W., et al. (2006), Current percutaneous coronary intervention and coronary artery bypass grafting practices for three-vessel and left main coronary artery disease. Insights from the SYNTAX run-in phase, *Eur. J. Cardiothorac. Surg.*, 29, 486–491.
66. Farkouh, M. E., Dangas, G., Leon, M. B., et al. (2008), Design of the Future Revascularization Evaluation in patients with Diabetes mellitus: Optimal management of Multivessel disease (FREEDOM) Trial, *Am. Heart J.*, 155, 215–223.
67. Hueb, W., Lopes, N. H., Gersh, B. J., et al. (2007), Five-year follow-up of the Medicine, Angioplasty, or Surgery Study (MASS II): A randomized controlled clinical trial of 3 therapeutic strategies for multivessel coronary artery disease, *Circulation*, 115, 1082–1089.
68. Pasceri, V., Patti, G., Speciale, G., Pristipino, C., Richichi, G., Di Sciascio, G. (2007), Meta-analysis of clinical trials on use of drug-eluting stents for treatment of acute myocardial infarction, *Am. Heart J.*, 153, 749–754.
69. Boden, W. E., O'Rourke, R. A., Teo, K. K., et al. (2007), Optimal medical therapy with or without PCI for stable coronary disease, *N. Engl. J. Med.*, 356, 1503–1516.

70. Anon. (1996), Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. The Bypass Angioplasty Revascularization Investigation (BARI) Investigators, *N. Engl. J. Med.*, 335, 217–225.
71. King, S. B., 3rd, Lembo, N. J., Weintraub, W. S., et al. (1994), A randomized trial comparing coronary angioplasty with coronary bypass surgery. Emory Angioplasty versus Surgery Trial (EAST), *N. Engl. J. Med.*, 331, 1044–1050.
72. Henderson, R. A., Pocock, S. J., Clayton, T. C., et al. (2003), Seven-year outcome in the RITA-2 trial: Coronary angioplasty versus medical therapy, *J. Am. Coll. Cardiol.*, 42, 1161–1170.
73. Ikeda, S., Bosch, J., Banz, K., Schneller, P. (2000), Economic outcomes analysis of stenting versus percutaneous transluminal coronary angioplasty for patients with coronary artery disease in Japan, *J. Invasive. Cardiol.*, 12, 194–199.
74. Uchida, T. (2006), Summary from the Circulatory Systems Devices Advisory Panel Meeting, Gaithersburg, Maryland, December 7–8.
75. Serruys, P. W., Unger, F., Sousa, J. E., et al. (2001), Comparison of coronary-artery bypass surgery and stenting for the treatment of multivessel disease, *N. Engl. J. Med.*, 344, 1117–1124.
76. Babapulle, M. N., Joseph, L., Belisle, P., Brophy, J. M., Eisenberg, M. J. (2004), A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents, *Lancet*, 364, 583–591.
77. Biondi-Zoccai, G. G., Agostoni, P., Abbate, A., et al. (2005), Adjusted indirect comparison of intracoronary drug-eluting stents: evidence from a metaanalysis of randomized bare-metal-stent-controlled trials, *Int. J. Cardiol.*, 100, 119–123.
78. Brunner-La Rocca, H. P., Kaiser, C., Pfisterer, M. (2007), Targeted stent use in clinical practice based on evidence from the Basel Stent Cost Effectiveness Trial (BASKET), *Eur. Heart J.*, 28, 719–725.
79. Dibra, A., Kastrati, A., Alfonso, F., et al. (2007), Effectiveness of drug-eluting stents in patients with bare-metal in-stent restenosis: meta-analysis of randomized trials, *J. Am. Coll. Cardiol.*, 49, 616–623.
80. Ellis, S. G. C. A., Grube, E., Popma, J., Koglin, J., Dawkins, K. D., and Stone, G. W. (2007), Incidence, timing, and correlates of stent thrombosis with the polymeric paclitaxel drug-eluting stent: A TAXUS II, IV, V, and VI meta-analysis of 3445 patients followed for up to 3 years, *J. Am. Coll. Cardiol.*, 49, 1043–1051.
81. Holmes, D. R. Jr., Moses, J. W., Schofer, J., Morice, M. C., Schampaeert, E., Leon, M. B. (2006), Cause of death with bare metal and sirolimus-eluting stents, *Eur. Heart J.*, 27, 2815–2822.
82. Kastrati, A., Dibra, A., Spaulding, C., et al. (2007), Meta-analysis of randomized trials on drug-eluting stents vs. bare-metal stents in patients with acute myocardial infarction, *Eur. Heart J.*, 28, 2706–2713.
83. Lagerqvist, B., James, S. K., Stenestrand, U., Lindback, J., Nilsson, T., Wallentin, L. (2007), Long-term outcomes with drug-eluting stents versus bare-metal stents in Sweden, *N. Engl. J. Med.*, 356, 1009–1019.
84. Mauri, L., Hsieh, W. H., Massaro, J. M., Ho, K. K., D'Agostino, R., Cutlip, D. E. (2007), Stent thrombosis in randomized clinical trials of drug-eluting stents, *N. Engl. J. Med.*, 356, 1020–1029.
85. Moreno, R., Fernandez, C., Calvo, L., et al. (2007), Meta-analysis comparing the effect of drug-eluting versus bare metal stents on risk of acute myocardial infarction during follow-up, *Am. J. Cardiol.*, 99, 621–625.

86. Moreno, R., Fernandez, C., Hernandez, R., et al. (2005), Drug-eluting stent thrombosis: results from a pooled analysis including 10 randomized studies, *J. Am. Coll. Cardiol.*, 45, 954–959.
87. Moses, J. W., Stone, G. W., Nikolsky, E., et al. (2006), Drug-eluting stents in the treatment of intermediate lesions: pooled analysis from four randomized trials, *J. Am. Coll. Cardiol.*, 47, 2164–2171.
88. Nordmann, A. J., Briel, M., Bucher, H. C. (2006), Mortality in randomized controlled trials comparing drug-eluting vs. bare metal stents in coronary artery disease: A meta-analysis, *Eur. Heart J.*, 27, 2784–2814.
89. Spaulding, C., Daemen, J., Boersma, E., Cutlip, D. E., and Serruys, P. W. (2007), A pooled analysis of data comparing sirolimus-eluting stents with bare-metal stents, *N. Engl. J. Med.*, 356, 989–997.
90. Stettler, C., Wandel, S., Allemann, S., et al. (2007) Outcomes associated with drug-eluting and bare-metal stents: A collaborative network meta-analysis, *Lancet*, 370, 937–948.
91. Stone, G. W., Moses, J. W., Ellis, S. G., et al. (2007), Safety and efficacy of sirolimus- and paclitaxel-eluting coronary stents, *N. Engl. J. Med.*, 356, 998–1008.
92. Roiron, C., Sanchez, P., Bouzamondo, A., Lechat, P., Montalescot, G. (2006), Drug eluting stents: An updated meta-analysis of randomised controlled trials, *Heart*, 92, 641–649.
93. Marroquin, O. C., Selzer, F., Mulukutla, S. R., et al. (2008), A comparison of bare-metal and drug-eluting stents for off-label indications, *N. Engl. J. Med.*, 358, 342–352.
94. Cutlip, D. E., Windecker, S., Mehran, R., et al. (2007), Clinical end points in coronary stent trials: A case for standardized definitions, *Circulation*, 115, 2344–2351.
95. Cutlip, D. E., Chauhan, M. S., Baim, D. S., et al. (2002), Clinical restenosis after coronary stenting: Perspectives from multicenter clinical trials, *J. Am. Coll. Cardiol.*, 40, 2082–2089.
96. Cutlip, D. E., Chhabra, A. G., Baim, D. S., et al. (2004), Beyond restenosis: Five-year clinical outcomes from second-generation coronary stent trials, *Circulation*, 110, 1226–1230.
97. Kereiakes, D. J., Kuntz, R. E., Mauri, L., Krucoff, M. W. (2005), Surrogates, substudies, and real clinical end points in trials of drug-eluting stents, *J. Am. Coll. Cardiol.*, 45, 1206–1212.
98. Spertus, J. A., Winder, J. A., Dewhurst, T. A., et al. (1995), Development and evaluation of the Seattle Angina Questionnaire: A new functional status measure for coronary artery disease, *J. Am. Coll. Cardiol.*, 25, 333–341.
99. Anon. (2000), Myocardial infarction redefined—a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction, *Eur. Heart J.*, 21, 1502–1513.
100. Pocock, S. J., Lansky, A. J., Mehran, R., et al. (2008), Angiographic surrogate end points in drug-eluting stent trials: A systematic evaluation based on individual patient data from 11 randomized, controlled trials, *J. Am. Coll. Cardiol.*, 51, 23–32.
101. Lemos, P. A., Lee, C. H., Degertekin, M., et al. (2003), Early outcome after sirolimus-eluting stent implantation in patients with acute coronary syndromes: Insights from the Rapamycin-Eluting Stent Evaluated at Rotterdam Cardiology Hospital (RESEARCH) registry, *J. Am. Coll. Cardiol.*, 41, 2093–2099.
102. Hong, M. K., Mintz, G. S., Lee, C. W., et al. (2003), Paclitaxel coating reduces in-stent intimal hyperplasia in human coronary arteries: A serial volumetric intravascular ultrasound analysis from the Asian Paclitaxel-Eluting Stent Clinical Trial (ASPECT), *Circulation*, 107, 517–520.



103. Pasceri, V., Granatelli, A., Pristipino, C., Pelliccia, F., Pironi, B., and Richichi, G. (2003), High-risk of thrombosis of Cypher stent in patients not taking ticlopidine or clopidogrel, *Am. J. Cardiol.*, 92.
104. Park, S. J., Shim, W. H., Ho, A. E., et al. (2003), A paclitaxel-eluting stent for the prevention of coronary restenosis, *N. Engl. J. Med.*, 348, 1537–1545.
105. Lasala, J. M., Stone, G. W., Dawkins, K. D., et al. (2006), An overview of the TAXUS Express, paclitaxel-eluting stent clinical trial program, *J. Interv. Cardiol.*, 19, 422–431.
106. Lumley, T. (2002), Network meta-analysis for indirect treatment comparisons, *Stat. Med.*, 21, 2313–2324.
107. King, S. B., 3rd, Aversano, T., Ballard, W. L., et al. (2007), ACCF/AHA/SCAI, 2007 update of the Clinical Competence Statement on Cardiac Interventional Procedures: A report of the American College of Cardiology Foundation/American Heart Association/American College of Physicians Task Force on Clinical Competence and Training (Writing Committee to Update the 1998 Clinical Competence Statement on Recommendations for the Assessment and Maintenance of Proficiency in Coronary Interventional Procedures), *Circulation*, 116, 98–124.
108. Urban, P. (2005), CYPHER. PCR, Paris.
109. Dobies, D. (2007), ARRIVE-1. PCR, Barrelogna.
110. Applegate, R. J., Sacrinty, M. T., Kutcher, M. A., et al. (2008), “Off-label” stent therapy, 2-year comparison of drug-eluting versus bare-metal stents, *J. Am. Coll. Cardiol.*, 51, 607–614.
111. Russell, M. E., Friedman, M. I., Mascioli, S. R., Stolz, L. E. (2006), Off-label use: An industry perspective on expanding use beyond approved indications, *J. Interv. Cardiol.*, 19, 432–438.
112. Hirsch, A. T., Haskal, Z. J., Hertzner, N. R., et al. (2006), ACC/AHA 2005 guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): executive summary. A collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (writing committee to develop guidelines for the management of patients with peripheral arterial disease), *J. Am. Coll. Cardiol.*, 1239–1312.
113. Kunlin, J. (1949), Le traitement de l’artère oblitérante par la greffe veineuse, *Arch. Mal. Coeur*, 42, 371–372.
114. Ring, E. J., McLean, G. K., Freiman, D. B. (1982), Selected techniques in percutaneous transluminal angioplasty, *Am. J. Roentgenol.*, 139, 767–773.
115. Palmaz, J. C., Sibbitt, R. R., Reuter, S. R., Tio, F. O., Rice, W. J. (1985), Expandable intraluminal graft: A preliminary study. Work in progress, *Radiology*, 156, 73–77.
116. Norgren, L., Hiatt, W. R., Dormandy, J. A., Nehler, M. R., Harris, K. A., Fowkes, F. G. R. (2007), Inter-society consensus for the management of peripheral arterial disease (TASC II), *J. Vasc. Surg.*, 45(S), S5A–S67A.
117. Serruys, P. W., Kutryk, M. J. B., Ong, A. T. L. (2006), Coronary artery stents, *N. Engl. J. Med.*, 354, 483–495.
118. Duda, S. H., Pusich, B., Richter, G., et al. (2002), Sirolimus-eluting stents for the treatment of obstructive superficial femoral artery disease: Six-months results, *Circulation*, 106, 1505–1509.
119. Duda, S. H., Bosiers, M., Lammer, J., et al. (2005), Sirolimus-eluting versus bare nitinol stent for obstructive superficial femoral artery disease: The SIROCCO II Trial, *J. Vasc. Interv. Radiol.*, 16, 331–338.



120. Duda, S. H., Bosiers, M., Lammer, J., et al. (2006), Drug-eluting and bare nitinol stents for the treatment of atherosclerotic lesions in the superficial femoral artery: Long-term results from the SIROCCO trial, *J. Endovasc. Ther.*, 13, 701–710.
121. Heldman, A. W., Ragheb, A. O. (2004), Paclitaxel persists in vessel wall after rapid delivery from nonpolymeric paclitaxel coating on self-expanding Nitinol stents (abstract), *Am. J. Cardiol.*, 94(suppl, 6A), 27E.
122. Lansky, A. J., Costa, R. A., Mintz, G. S., et al. (2004), Non-polymer-based paclitaxel-coated coronary stents for the treatment of patients with de novo coronary lesions: angiographic follow-up of the DELIVER clinical trial, *Circulation*, 109, 1948–1954.
123. Dake, M. D. (2007), *Interim Report on the Zilver PTX Clinical Trial*. International Society of Endovascular Therapy (ISET), Hollywood, FL.
124. Aoki, J., Rodriguez-Granillo, G. A., Serruys, P. W. (2005), Emergent strategies in interventional cardiology. *Rev. Esp. Cardiol.*, 58, 962–973.
125. Losordo, D. W., Dimmeler, S. (2004), Therapeutic angiogenesis and vasculogenesis for ischemic disease: part II: Cell-based therapies, *Circulation*, 109, 2692–2697.
126. Rafii, S., Lyden, D. (2003), Therapeutic stem and progenitor cell transplantation for organ vascularization and regeneration, *Nat. Med.*, 9, 702–712.
127. Kawamoto, A., Gwon, H. C., Iwaguro, H., et al. (2001), Therapeutic potential of ex vivo expanded endothelial progenitor cells for myocardial ischemia, *Circulation*, 103, 634–637.
128. Kawamoto, A., Tkebuchava, T., Yamaguchi, J., et al. (2003), Intramyocardial transplantation of autologous endothelial progenitor cells for therapeutic neovascularization of myocardial ischemia, *Circulation*, 107, 461–468.
129. Assmus, B., Schachinger, V., Teupe, C., et al. (2002), Transplantation of progenitor cells and regeneration enhancement in acute myocardial infarction (TOPCARE-AMI), *Circulation*, 106, 3009–3017.
130. Britten, M. B., Abolmaali, N. D., Assmus, B., et al. (2003), Infarct remodeling after intracoronary progenitor cell treatment in patients with acute myocardial infarction (TOPCARE-AMI): mechanistic insights from serial contrast-enhanced magnetic resonance imaging, *Circulation*, 108, 2212–2218.
131. Perin, E. C., Dohmann, H. F., Borojevic, R., et al. (2003), Transendocardial, autologous bone marrow cell transplantation for severe, chronic ischemic heart failure, *Circulation*, 107, 2294–2302.
132. Stamm, C., Westphal, B., Kleine, H. D., et al. (2003), Autologous bone-marrow stem-cell transplantation for myocardial regeneration, *Lancet*, 361, 45–46.
133. Strauer, B. E., Brehm, M., Zeus, T., et al. (2002), Repair of infarcted myocardium by autologous intracoronary mononuclear bone marrow cell transplantation in humans, *Circulation*, 106, 1913–1918.
134. Tse, H. F., Kwong, Y. L., Chan, J. K., Lo, G., Ho, C. L., and Lau, C. P. (2003), Angiogenesis in ischaemic myocardium by intramyocardial autologous bone marrow mononuclear cell implantation, *Lancet*, 361, 47–49.
135. Wollert, K. C., Meyer, G. P., Lotz, J., et al. (2004), Intracoronary autologous bone-marrow cell transfer after myocardial infarction: The BOOST randomised controlled clinical trial, *Lancet*, 364, 141–148.
136. Fuchs, S., Satler, L. F., Kornowski, R., et al. (2003), Catheter-based autologous bone marrow myocardial injection in no-option patients with advanced coronary artery disease: A feasibility study, *J. Am. Coll. Cardiol.*, 41, 1721–1724.
137. Balsam, L. B., Wagers, A. J., Christensen, J. L., Kofidis, T., Weissman, I. L., Robbins, R. C. (2004), Haematopoietic stem cells adopt mature haematopoietic fates in ischaemic myocardium, *Nature*, 428, 668–673.

138. Murry, C. E., Soonpaa, M. H., Reinecke, H., et al. (2004), Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts, *Nature*, 428, 664–668.
139. Yla-Herttuala, S., Alitalo, K. (2003), Gene transfer as a tool to induce therapeutic vascular growth, *Nat. Med.*, 9, 694–701.
140. Simons, M., Bonow, R. O., Chronos, N. A., et al. (2000), Clinical trials in coronary angiogenesis: issues, problems, consensus: An expert panel summary, *Circulation*, 102, E73–86.
141. Grines, C. L. (2004), The AGENT clinical trials programme, *Eur. Heart J. Suppl.*, 6, E18–E23.
142. Hedman, M., Hartikainen, J., Syvanne, M., et al. (2003), Safety and feasibility of catheter-based local intracoronary vascular endothelial growth factor gene transfer in the prevention of postangioplasty and in-stent restenosis and in the treatment of chronic myocardial ischemia: Phase II results of the Kuopio Angiogenesis Trial (KAT), *Circulation*, 107, 2677–2683.
143. Henry, T. D., Annex, B. H., McKendall, G. R., et al. (2003), The VIVA trial: Vascular endothelial growth factor in Ischemia for Vascular Angiogenesis, *Circulation*, 107, 1359–1365.
144. Kastrup, J., Jorgensen, E., and Ruck, A. (2003), Euroinject Trial. Late breaking clinical trials session. American College of Cardiology, 2003, Chicago, *J. Am. Coll. Cardiol.*, 41, 1603.
145. Seiler, C., Pohl, T., Wustmann, K., et al. (2001), Promotion of collateral growth by granulocyte-macrophage colony-stimulating factor in patients with coronary artery disease: A randomized, double-blind, placebo-controlled study, *Circulation*, 104, 2012–2017.
146. Simons, M., Annex, B. H., Laham, R. J., et al. (2002), Pharmacological treatment of coronary artery disease with recombinant fibroblast growth factor-2: double-blind, randomized, controlled clinical trial, *Circulation*, 105, 788–793.
147. Grines, C. L., Watkins, M. W., Helmer, G., et al. (2002), Angiogenic Gene Therapy (AGENT) trial in patients with stable angina pectoris, *Circulation*, 105, 1291–1297.
148. Virmani, R., Farb, A. (1999), Pathology of in-stent restenosis, *Curr. Opin. Lipidol.*, 10, 499–506.



# 10.2

---

## *Clinical Trials Involving Oral Diseases*

**Bruce L. Pihlstrom,<sup>1</sup> Bryan Michalowicz,<sup>1</sup> Jane Atkinson,<sup>2</sup> and Albert Kingman<sup>2</sup>**

*<sup>1</sup>School of Dentistry, University of Minnesota, Minneapolis, Minnesota*

*<sup>2</sup>National Institute of Dental and Craniofacial Research, National Institutes of Health,  
Bethesda, Maryland*

### **Contents**

|          |                                                                      |     |
|----------|----------------------------------------------------------------------|-----|
| 10.2.1   | Introduction                                                         | 436 |
| 10.2.2   | Prominent Design Issues in Oral Health Clinical Trials               | 436 |
| 10.2.2.1 | Split-Mouth and Cross-Over Designs                                   | 437 |
| 10.2.2.2 | Single-versus Multicenter Trials                                     | 437 |
| 10.2.2.3 | Use of Active versus Negative Controls                               | 438 |
| 10.2.2.4 | Identification of Clinically Meaningful Endpoints                    | 438 |
| 10.2.2.5 | Independent Sampling (Sites versus Subjects) and Multilevel Modeling | 439 |
| 10.2.2.6 | Randomization and Potential Bias Due to Confounding Factors          | 440 |
| 10.2.2.7 | Examiner Training and Calibration                                    | 440 |
| 10.2.2.8 | Trial Duration                                                       | 441 |
| 10.2.3   | Prominent Statistical Issues in Oral Health Clinical Trials          | 441 |
| 10.2.3.1 | Clinical Significance                                                | 442 |
| 10.2.3.2 | Confidence Intervals versus <i>p</i> Values                          | 442 |
| 10.2.4   | Regulatory Issues                                                    | 443 |
| 10.2.5   | Management of Oral Health Clinical Trials                            | 443 |
| 10.2.5.1 | Roles and Responsibilities of Trial Personnel                        | 443 |
| 10.2.5.2 | Data Coordination Center                                             | 444 |
| 10.2.5.3 | Data Collection                                                      | 445 |

|                                                             |     |
|-------------------------------------------------------------|-----|
| 10.2.5.4 Recruitment                                        | 445 |
| 10.2.5.5 Provision of Standard of Care                      | 447 |
| 10.2.5.6 Use of Nontreatment and Delayed Treatment Controls | 448 |
| 10.2.6 Caries Clinical Trials                               | 449 |
| 10.2.7 Periodontal Clinical Trials                          | 450 |
| 10.2.8 Dental Implant Clinical Trials                       | 451 |
| 10.2.9 Other Oral Conditions                                | 453 |
| 10.2.9.1 Xerostomia                                         | 453 |
| 10.2.9.2 Mucosal Candidiasis                                | 454 |
| 10.2.9.3 Mucosal Diseases                                   | 454 |
| References                                                  | 455 |

## 10.2.1 INTRODUCTION

Dentistry is rapidly entering an era of evidence-based practice, and individual patients and society are demanding that clinical practice and public health policy decisions be based on the best available science. Well-designed randomized oral health clinical trials (RCTs) are needed because they provide the highest level of scientific evidence [1]. It is very difficult for oral health practitioners, researchers, and sponsors who have not been directly involved in large oral health RCTs to truly understand their complexity, challenges, and expense. Moreover, few oral health investigators have appropriate training or clinical trial experience. These factors have contributed to the relative lack of large phase III multicenter oral health clinical trials.

Clinical trials in oral health have mainly focused on the diseases of dental caries and periodontal disease, but a limited number have involved other oral diseases and conditions. These include chronic facial pain, particularly associated with the temporomandibular joint, acute pain after extraction of third molars (wisdom teeth), xerostomia (dry mouth), and various oral mucosal diseases. With the advent of evidence-based dentistry, more clinical trials are needed to establish the efficacy *and* effectiveness of prevention and treatment for a wide variety of oral diseases and conditions.

Oral health trials share many things in common with clinical trials in general, but they have several unique challenges. This chapter focuses on phase III oral health clinical trials in terms of design and analyses, regulatory and management issues, and challenges of conducting trials involving specific oral diseases.

## 10.2.2 PROMINENT DESIGN ISSUES IN ORAL HEALTH CLINICAL TRIALS

Design issues of special interest in oral health clinical trials include: advantages and disadvantages of so-called split-mouth and cross-over designs, single versus multicenter trials, use of active versus negative controls, identification of clinically meaningful endpoints, independent sampling (sites versus subjects) and multilevel modeling, randomization and potential bias due to confounding factors, examiner training and calibration, and trial duration.

### 10.2.2.1 Split-Mouth and Cross-Over Designs

The parallel design in which subjects are randomly assigned to receive either an active or a control (placebo) intervention is the standard design used in clinical trials. However, oral health clinical trials often use a “split-mouth” design in which various segments of the dentition are randomly assigned to treatment and control arms. Furthermore, many toothpaste and oral rinse studies use a cross-over design that in its simplest form randomizes subjects to one of two arms of a trial, followed by a “washout” period, after which they are assigned to the other arm. An investigator testing the longevity or integrity of dental sealants, composite restorative materials, or dental implants may wonder which study design would be the most appropriate for the trial. One could either use a split-mouth, cross-over, or a parallel design.

The main weakness of the split-mouth design is that treatments may have effects on segments of the dentition other than those to which they were assigned; the outcome of a split-mouth design is the treatment effect plus the sum of these effects [2]. The split-mouth design removes the between-group variance in estimating treatment effect, but it may include effects from treatments administered to other parts of the dentition. As such, it has the potential of biasing the trial against finding differences between treatments. When considering use of this design, it is important to estimate the within-patient correlation of treatment effect; if it is small, little gain in efficiency is obtained by using the split-mouth design; if it is moderate to high, the gain in efficiency should be weighed against the potential disadvantages of the split-mouth design including potential bias, potential difficulty in subject recruitment (there must be similar disease in two quadrants of the mouth), and complexity of the statistical analysis [3]. A split-mouth design may be appropriate if the “carry-across” effect from one tooth site or treatment segment to another is minimal. Unfortunately, there is no easy method to test for a carry-across effect in a split-mouth study, and the investigator would have to be convinced that potential carry-cross effects in the trial are negligible. The parallel design avoids these potential sources of bias and facilitates straightforward statistical analysis. The main weakness of the parallel design compared to the split-mouth and cross-over is it will require many more subjects to achieve comparable statistical significance. In spite of this constraint, the parallel design is generally preferable for diseases such as periodontal disease and caries.

Cross-over designs could be appropriate for short-term trials investigating outcomes that are reversible, such as gingivitis, plaque, and even for some types of chronic pain. However, cross-over trials have limitations due to possible treatment carry-over effect from one period to the next period and the possibility of a patient learning effect that can bias the results. In an attempt to minimize these concerns, a washout period is introduced between active treatment periods. Overall, for plaque and gingivitis trials, the double-blind parallel randomized trial is preferred to the double-blind cross-over design, except possibly for a professionally applied treatment that does not leach into the saliva or affect the oral microflora [4].

### 10.2.2.2 Single versus Multicenter Trials

The need for generalizability of results and the number of patients required for a clinical trial are important considerations when designing clinical trials. Multicenter

trials are conducted to generate results that can be generalized to the population at large. For many dental diseases, community water fluoridation, cultural norms, tobacco use, diet, and many other factors may contribute to trial outcomes. Therefore, regardless of size, the outcome of a single-center trial may only be applicable to a particular sample of subjects. Another important reason for using multiple center sites is that there may be insufficient subjects at a single enrollment site that meet trial enrollment criteria. Adequate numbers of subjects are frequently not available for large dental implant and periodontal disease trials; multicenter studies are often needed to enroll enough subjects in these trials. Earlier caries clinical trials were conducted at several schools in a single community (single-site trial). This was possible because most of these trials focused on preventing, rather than treating dental caries. Moreover, caries was very common and it was cost effective to recruit sufficient numbers of qualifying subjects in schools. As caries rates have declined in developed countries, it is much more difficult to find subjects who are at risk for developing new caries. This has necessitated the use of much larger, multicenter caries prevention trials, often involving 800–1500 patients per group.

### 10.2.2.3 Use of Active versus Negative Controls

For ethical reasons the majority of recent caries clinical trials have used active rather than negative controls. This is because fluoride has proven to be effective in preventing dental caries and many feel that it is not ethical to include a nonfluoride (negative) control group. If the purpose of a trial is to demonstrate that a new agent or product is superior to the current standard (positive control) agent or product, the minimum difference ( $\Delta$ ) judged to be clinically meaningful between the new agent and the standard would be smaller than if a negative control group were used. Use of a positive control would also require increased sample size. Alternatively, because of improved taste, less cost, ease of application, or reduction in side effects, the purpose of a trial may be to demonstrate, for example, that a new fluoride agent is as good as the standard fluoride agent (positive control). In these trials, a noninferiority trial design generally aims to demonstrate that the new agent or product is not inferior to the standard. A value for clinical comparability or equivalence for the new agent must be defined ( $\Delta$ ) a priori in order to define how close the primary clinical outcome for the new agent ( $\mu_N$ ) must approximate the outcome using the standard agent ( $\mu_S$ ) for it to be considered at least as good as the standard ( $\Delta = \mu_N - \mu_S$ ). To establish efficacy of nonfluoride agents for caries prevention, use of negative controls may be possible. For these trials, the magnitude of the clinically important difference in the development of new caries over a specific time period must be defined. The rate of new caries over time is often called the caries increment.

### 10.2.2.4 Identification of Clinically Meaningful Endpoints

Outcomes in randomized clinical trials must be clearly specified, measurable, valid, and reproducible. It is also important that pivotal phase III clinical trials have clinically meaningful endpoints [5]. As defined by a 1999 joint National Institute of Health–Food and Drug Administration (NIH–FDA) workshop on biomarkers and clinical endpoints [6] a *clinical endpoint* is a characteristic or variable that reflects



how a patient feels, functions, or survives; a *surrogate endpoint* is a biomarker intended to substitute for a clinical endpoint and is expected to predict clinical benefit (or harm, or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence; and a *biomarker* is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or (pharmacologic) responses to a therapeutic intervention.

True validation of surrogate endpoints that are used in clinical trials in medicine is rare [5] and oral health clinical trials are no exception. Hujoel pointed out that there has been widespread use of periodontal variables such as probing depth, clinical attachment level, and bleeding on probing that have not been validated for use in clinical trials [7]. He suggested that true endpoints such as tooth loss, pain, bleeding after tooth brushing, and other subjective quality of life measurements should be used in pivotal trials of periodontal disease [7]. In a caries trial, reducing pain from caries would be more meaningful than reducing the level of *mutans streptococci*, and in a clinical trial of xerostomia change in subject-perceived oral dryness would be more meaningful than a statistically significant change in inflammatory salivary proteins. Surrogate endpoints can greatly increase efficiency and decrease costs of clinical trials, and more research is needed to validate their use in oral health clinical trials [8].

#### 10.2.2.5 Independent Sampling (Sites versus Subjects) and Multilevel Modeling

Because clinical measures of oral disease such as caries and periodontal disease can be obtained at many sites in the dentition, there has been confusion as to whether individual sites within subjects or subjects themselves should be the independent unit of analysis in clinical trials. A unique feature of many dental clinical trials is their complex data structure that frequently involves clustered observations of multiple measurements repeated over time within each subject. This introduces multiple levels of structural data dependency; it is clear that sites are correlated or clustered within subjects and that sites cannot be used as independent sampling units in statistical analysis [9].

Collection of a large number of data items from each subject is often necessary because dental disease is tooth- and even tooth-site-specific. That is, diseases such as caries and periodontal disease can progress at very few tooth sites in an otherwise healthy dentition. Unless all sites are evaluated, it is possible that localized progressive disease will remain undetected. The criteria for progressive disease and the protocol for delivering rescue or additional treatment should be specified beforehand in the trial's operations manual. With so many measures, it can be difficult to screen repeated measures to determine if disease is stable or progressing, and redundant checks are needed to compare measures collected over time. Web-based or electronic chair-side data capture can facilitate this process.

There are 32 teeth in the adult dentition. It is common to measure various periodontal parameters at 6 sites per tooth (mesiobuccal, midbuccal, distobuccal, distolingual, midlingual, and mesiolingual) and caries on 5 surfaces per tooth (occlusal, mesial, distal, buccal, and lingual). Therefore, it is possible to include measurements of up to 168 sites (excluding third molars = 28 teeth  $\times$  6 sites per tooth) in periodon-

tal trials and 128 sites in caries trials (excluding third molars there are 16 posterior teeth  $\times$  5 surfaces per tooth plus 12 anterior teeth  $\times$  4 surfaces per tooth). Since it is clear that sites cannot be treated as independent units of sampling, it is common to aggregate site-based data at the patient level. Depending on the specific trial, typical patient-level periodontal outcome measures have included (1) mean clinical attachment loss, (2) mean probing pocket depth, (3) number of sites with bleeding on probing, (4) extent of clinical attachment loss, and (5) extent of probing pocket depth (using a specific threshold, e.g., 3, 4, or 6 mm). Use of such aggregated data at the patient level may be appropriately used in conventional statistical analyses with the subject as the independent unit of observation. Although statistically correct, this may result in loss of detailed information at the site level; it is clear that the variance in periodontal parameters may be attributed to factors at both the site and the subject level [10]. In this regard, use of multilevel modeling (MLM) can be a valuable statistical method to assess disease progression over time at all levels of the natural hierarchy of tooth sites within subjects [11, 12]. Multilevel modeling of periodontal data has demonstrated that there is considerable variation at all levels of analysis and that MLM can be a more powerful research tool than single-level techniques for the analysis of hierarchical dental data [13].

#### 10.2.2.6 Randomization and Potential Bias Due to Confounding Factors

Common confounding factors that could lead to selection or confounding bias in oral health clinical trials include age, sex, socioeconomic status, educational level, smoking, diet, oral hygiene, exposure to fluoride and baseline caries level (for caries), use of medications that influence salivary flow, and alcohol use (for oral cancer). Smoking should be considered a strong confounder or effect modifier in periodontal disease, and in most cases it is helpful to separately compare treatment arms among smokers and nonsmokers [14].

Proper attention to randomization procedures assures comparability between treatment groups in clinical trials by minimizing selection and confounding bias [15]. However, small sample clinical trials are prone to imbalances in known and unknown factors that may affect treatment outcome [16]. For relatively small sample trials, use of stratified randomization to achieve balance between test and control arms for factors that may affect treatment outcome should be considered. As noted by Schulz and Grimes, proper randomization takes little time and effort but results in major rewards in terms of scientific accuracy and credibility [15].

#### 10.2.2.7 Examiner Training and Calibration

As stated by Fleiss in the very first sentence of his classical text *The Design and Analysis of Clinical Experiments* [17]: “The most elegant design of a clinical study will not overcome the damage caused by unreliable or imprecise measurement.” Because of the complexity of oral disease measurements, examiner training and calibration are essential components of any oral health clinical trial. For example, it is common for oral health clinical trial examiners to record measurements on 28 teeth in *each subject's mouth* for probing depth, clinical attachment level, gingivitis, dental plaque, and dental calculus ( $n = 168$  sites for each parameter) plus tooth mobility ( $n = 28$ ) for a total of 868 separate measurements per patient at each evalu-

ation visit. This is an enormous task and to minimize examiner fatigue and subject discomfort, only essential data that are needed to measure trial outcome should be collected.

Proper training, calibration, and certification of examiners to prespecified standards of discrimination and standardization are essential. The use of multiple examiners makes this task even more complex, time-consuming, and costly because inter- *and* intraexaminer variability must be minimized by training and repeated calibration. Examiners should be trained and calibrated using patients whose severity and extent of disease is similar to the trial population because measures of examiner reliability may be artificially inflated if patients used in calibration exercises are healthier than study participants. For periodontitis trials, calibration statistics such as percent agreement within a specified threshold will be artificially high if subjects used for calibration have only early or localized disease. Examiners should be recalibrated during the trial using a subsample of subjects (e.g., 5%) to ensure consistent evaluation of trial outcomes. Interim assessments are important because experienced examiners trained at the beginning of a trial may slowly revert to their original techniques or methods over time [18]. It is important that measurement error be expressed in terms that are meaningful to the clinician while retaining statistical validity [19]. There is within-subject clustering of agreement for various oral disease measures and failure to account for dependence among site-level measurements can result in a false sense of precision in examiner reliability estimates [20]. Because examiner calibration and reliability is such a major issue in oral health clinical trials, details of examiner training and reproducibility data should be reported in publications.

#### 10.2.2.8 Trial Duration

As with all clinical trials, the duration of oral health clinical trials depends on many factors including the specific disease being investigated, trial design, numbers and enrollment criteria of subjects to be enrolled, the intervention, and the primary outcome variable. It has been recommended that gingivitis trials be at least 6 months in duration [21] and that periodontitis trials be 9 months [19]. Cross-over designs would require a trial at least twice as long as a two-arm split-mouth or parallel design. Caries trials should be 3 years in duration and include at least one interim examination in addition to an assessment at baseline and at the end of the trial [22]. Ideally, clinical outcomes should be monitored for at least 3 years in trials of direct dental restorations (amalgams, resin-based composites) and for 5 years in trials of indirect restorations (crowns, inlays). In reality, outcomes are rarely assessed after 3 years, even though the materials are designed to last much longer in clinical practice.

### 10.2.3 PROMINENT STATISTICAL ISSUES IN ORAL HEALTH CLINICAL TRIALS

Obtaining multiple measurements for each subject in many dental clinical trials provides the opportunity to analyze data at the subject, tooth, and site level. Analysis will depend on the primary purpose of the trial and trial design. Most dental clinical

trials use a parallel design with randomization of treatment to subjects and a statistical analysis focused on a subject-level primary outcome measure. Outcomes could also be analyzed at the individual site or tooth level, but within-mouth correlations must be considered in these analyses. In this regard, generalized estimating equations (GEE) [23, 24] and other methods such as an incidence density method [23, 25], Poisson regression models [26], survival methods [24, 27, 28], and specific non-parametric procedures for ordinal-scaled outcome measures have been proposed [29, 30]. Using GEE is one example of a marginal approach, that is, a population average method that provides estimates at the population level, rather than for individual subjects. Various conditional approaches can be used if the focus is on individual subjects; in dental clinical trials, a conditional model is probably more appropriate. A mixed-model approach can be used to account for repeated correlated observations within subjects.

### 10.2.3.1 Clinical Significance

Phase III RCTs must have sufficient sample size to yield results that are statistically *and* clinically significant. Clinical significance is a matter of judgment and may be defined as the minimal magnitude of difference between test and control interventions that would cause clinicians to change their standard of clinical practice or that would inform public health policy. In other words, it answers the question “Is the difference between the groups large enough to be worthwhile?”

It is often difficult to reach consensus on what constitutes clinical significance. It becomes a trade-off between the size of the intervention effect that is considered to be clinically superior and the size that is considered to be equivalent. It also may involve a compromise between what is desirable and what is practical for testing in a clinical trial. In any event, to properly estimate sample size for a trial, clinical significance for the outcome variable must be defined a priori. For example, if a 3-year phase III RCT demonstrated that subjects using an inexpensive over-the-counter (OTC) caries prevention product had a clinically meaningful (i.e., 20% reduction) average reduction in 0.6 new carious surfaces per subject compared to appropriate control subjects, it would have major public health significance and would substantially impact the OTC market for caries prevention. Using the formula from Pocock [31], if we assume a standard deviation in 3-year caries incidence of  $\sigma = 3$  surfaces, a difference in caries increment ( $\Delta$ ) = 0.6 surfaces,  $\alpha = 0.05$ , and  $1 - \beta = 0.80$  (power), the sample size for each arm may be estimated at 395. That is,  $2 \times 395 = 790$ , or almost 800 subjects would be needed for the trial. The actual number of subjects required for screening would be much larger because of the need to identify subjects at high risk for developing dental caries.

### 10.2.3.2 Confidence Intervals versus *p* Values

Two basic statistical methods are used to assess the role of chance in estimation and hypothesis testing. The confidence interval (CI) is used in estimation and the *p* value is used in hypothesis testing. A 95% CI is the default value for most clinical trials. It is a frequency-based interval estimate of the unknown true difference between treatment groups based on what would happen under multiple replications of the same study. A desirable property of the 95% CI is that its range or width represents a measure of the precision of the estimate in addition to its significance. For example,

suppose a 3-year caries trial compared two fluoride toothpastes (500ppm vs. 1000ppm fluoride) and reported mean (SE) caries DMFS increments of 1.75 (0.1) and 1.55 (0.08), respectively, and that the difference of 0.20 (SE = 0.13;  $p = 0.12$ ) had an associated 95% CI of  $-0.06$  to  $+0.46$ . There was no statistically significant difference in the effectiveness of the two toothpastes because  $p = 0.12$ . It is important to note that the probability value ( $p = 0.12$ ) does not provide an estimate of precision. The clinical significance of the trial outcome may be better appreciated by an inspection of the 95% CI for the difference in DMFS increments. The statistical significance of the difference between treatment outcomes can be seen in the 95% CI because the value 0 (representing no difference between fluoride toothpastes) lies within the confidence interval estimate. Moreover, in this study, a fairly high level of precision of the estimated difference was achieved because the confidence interval was fairly narrow ( $-0.06$  to  $+0.46$ ). Sample size would need to be increased if a higher level of precision (narrower CI) is desired.

#### 10.2.4 REGULATORY ISSUES

Most materials used to restore tooth structure are considered medical devices by regulatory agencies such as the U.S. FDA. One specific center within the FDA, the Center for Devices and Radiological Health, regulates products such as dental restorative materials, electric toothbrushes, and dental implants [32]. Each device is grouped by medical specialty and assigned a class (class I, II, or III) based on the level of control needed to ensure safety and effectiveness of the device. Standards to guide the manufacture and performance of various dental materials have been developed by two large organizations, the dental committee of the International Organization for Standardization (ISO), and the American Dental Association, which develops performance standards in conjunction with the American National Standards Institute. Each publishes its own set of standards [33, 34], which are used by dental manufacturers worldwide.

Most individuals seeking FDA clearance for new dental materials must submit a 510(k) application. Many of these products will be classified as class I or class II devices. Class I devices may be exempt from premarket review provided that the device is appropriately registered with the agency. Most class II devices require demonstration of substantial equivalence to another legally U.S.-marketed device. Substantial equivalence means that the new device is at least as safe and effective as the predicate [35]. In some instances, clinical data is needed for FDA clearance of the new device, especially if the manufacturer asserts differences in claims or intended use. Because many new dental materials are similar to older materials, they are classified as class II devices. As such, phase II trials are sufficient and they are not tested in phase III clinical trials.

#### 10.2.5 MANAGEMENT OF ORAL HEALTH CLINICAL TRIALS

##### 10.2.5.1 Roles and Responsibilities of Trial Personnel

Each staff member involved in a trial must understand, accept, and be capable of performing his or her role. Specific roles and responsibilities should be assigned

before the trial begins and monitored regularly throughout the trial. While clinical data may be collected by qualified personnel (e.g., dental hygienists), licensure as a dentist is generally required to diagnosis oral disease. For example, in a clinical trial of periodontal disease, it is common for a dental hygienist to perform clinical assessments and collect dental plaque samples or oral biospecimens.

Typically, however, a licensed dentist must also examine each subject or review the clinical and radiographic findings to establish that the patient has the condition or disease specified by the protocol. For caries trials, a nondentist may determine the presence of cavitated lesions (areas of pitted or soft tooth) or photograph and assess fluorescent images of teeth, but a licensed dentist is usually required to establish a diagnosis and provide (or supervise) the trial intervention. Institutional review boards may look unfavorably on a study in which the patient is not examined, diagnosed, and monitored by a dentist.

Therapists must be competent and experienced in treating or managing the disease or condition under study. Patients enrolled in clinical trials should receive the same quality of care that is delivered in nonresearch clinical settings. Dental hygienists or general dentists typically deliver nonsurgical periodontal therapy and in caries trials, treatment is usually provided by licensed and experienced dentists. For severe periodontal disease, treatment involving surgery, or for other interventions outside of the expertise of general dental practitioners, treatment should be provided by specialists or dentists with advanced training or experience.

Study masking or blinding is an important feature of clinical trials that serves to minimize bias during data collection and assessment [16]. Masking can be defined in terms of the participant, the therapist, and the examiner. Treatment of oral disease tends to be procedurally oriented, and therapists in most oral health trials are not masked. Masking participants also can be difficult in some trials. Consider a trial that compares a nontraditional, minimally invasive treatment for dental caries to conventional restorative care (i.e., dental fillings). Because conventional care involves removal of tooth decay with a high speed drill and requires local anesthesia, subjects receiving the minimally invasive therapy would also need to receive local anesthesia (regardless of the need) to maintain masking. For periodontitis trials, it can be difficult to mask examiners because some treatments may produce more gingival recession (resulting in more exposed tooth root) than others. An experienced clinical examiner could unintentionally note such changes following some treatments. To minimize potential bias, patients should be instructed to refrain from discussing treatment or treatment-related signs or symptoms with trial examiners. Patient complaints, such as tooth sensitivity, can be discussed with the study coordinator or lead investigator, provided these individuals do not collect, record, or discuss data with examiners.

#### **10.2.5.2 Data Coordination Center**

Trial coordination and data management are essential to the success of any clinical trial. Large multicenter trials need a data coordinating center (DCC). The DCC helps design the trial, collects data from all centers, manages data, provides day-to-day guidance on protocol implementation, provides trial monitoring, oversees examiner training and calibration, ensures data quality, and analyzes data. Often, the trial statistician is part of the DCC. In most contemporary clinical trials, the



DCC facilitates data input at individual clinical centers by providing computer software and information technology (IT) support.

### 10.2.5.3 Data Collection

A primary focus of the investigative team should be to collect quality data. Rigorous monitoring is needed to ensure that data are collected according to trial protocol. Study personnel should carefully review study forms before the end of each subject visit so that errors or omissions can be identified and immediately rectified. Data collected on paper forms should be submitted to the DCC in a timely manner. The DCC should review data collection forms for completeness and send queries to enrollment sites when errors are detected or suspected. Missing data or data entry errors should be corrected as soon as possible because a long duration between error identification and correction leads to greater likelihood for error.

Oral health clinical trials frequently collect information about a participant's medical health. If medical information is collected as part of a trial, it should be collected by staff who are well versed in medical terminology and who are able to verify its accuracy and completeness. If data is abstracted from medical records, this task should be performed by someone who is familiar with the medical record format. In our experience, trained nurses working in the medical care facility are best suited to perform these tasks. Using "in-house" nurses also helps protect patient privacy because medical records do not leave the facility and because nurses are usually well trained in privacy issues.

### 10.2.5.4 Recruitment

Even the best designed and most innovative clinical trials will fail if recruitment efforts fail. Investigators should not underestimate the amount of effort needed for successful subject recruitment. Like all trials, oral health trials should have detailed recruitment plans in place before it begins. Investigators often overestimate the prevalence of a condition or disease in the population they plan to study. The "take all comers" environment of clinical practice often causes clinicians to overestimate the number of available patients for trials. Recruitment and participation in oral health clinical trials have been discussed elsewhere [36–41]. It must be emphasized that recruitment will be more successful when trials have relatively few exclusion criteria. Limiting exclusion criteria also increases the likelihood that trial results will be generalizable and embraced by community practitioners.

Institutional review board-approved fliers and advertisements can be posted in busy public places or played on radio or television. While this may yield large numbers of initial inquiries, enrollment success with these methods is often relatively low because people typically have only a cursory understanding of their oral health. For example, enrollment in periodontal disease trials is usually limited to people with disease severity that can only be determined by a periodontal examination. Mass media advertisements for periodontal trials often attract people with a wide variety of periodontal conditions such as gingival recession due to toothbrush abrasion, gingivitis, and even nonbacterial inflammatory conditions such as desquamative gingivitis. If mass media advertising is used, appropriate attention and



resources must be committed to screen potential subjects for trial participation. This can include initial telephone screening for inclusion and exclusion criteria, followed by a clinical examination. Recruitment from dental clinics is often productive and trials can be advertised by posting fliers in waiting areas and patient examination rooms. However, it is critical that investigators establish an ongoing relationship with clinic staff. We have found that simply posting fliers or placing pamphlets in dental clinics is unproductive. Successful recruitment requires that study personnel have regular and substantive contact with clinic personnel to remind them of the trial and its importance. Importantly, recruitment of patients from clinics or private offices should be a “seamless” process. Clinic staff should make patients aware of the study and of the potential benefits of participating. However, once a patient expresses interest in the trial, study personnel, not a clinical staff member, should guide the patient through the consent, screening, and enrollment processes. Because clinic staff are very busy providing patient care, they cannot be expected to perform burdensome tasks such as scheduling trial appointments, distributing enrollment forms, or collecting data. Milgrom et al. [40] reported that recruitment for oral health trials can be enhanced by using concurrent recruitment methods, accurately estimating the costs and time required for recruitment, and regular monitoring recruitment efforts. They also found that newspaper public service advertisements and campus posters were the most effective means for recruiting subjects with periodontitis in a university setting [41].

Active clinical research centers may regularly screen patients recruited from print and broadcast advertisements to identify subjects for possible participation in clinical research. It must be stressed that screening volunteers for research generally requires institutional review board approval. Those who are screened should be encouraged to seek needed dental care if relevant trials are not planned for the near future because some patients for financial or other reasons might defer treatment in the hope they will be enrolled in a trial.

Unfortunately, most dental diseases are quite prevalent in minority and underserved populations [42]. It is critical that research be conducted in partnership with these populations so that disparities in oral health can be addressed. Research investigators should partner with community leaders from underserved populations to make sure that people from underserved populations are made aware of trials that are available for their participation. Moreover, barriers for participation by underserved populations need to be addressed. For example, investigators should consider establishing satellite clinics in underserved communities or provide taxi vouchers if transportation to the main study site is an impediment to trial participation. Provision of on-sight child care and extended or nighttime clinic hours may also make participation in clinical trials more feasible.

While clinical trials must satisfy the Belmont principles of respect, beneficence, and justice, patients and communities may demand more from investigators for their participation. Educational programs on disease prevention may help investigators obtain community trust and reduce the burden of oral diseases. For trials that recruit in underserved communities, study personnel should provide information about community or low-cost clinics where participants can receive care for non-study-related oral diseases. Despite elaborate advertisement and recruitment efforts, enrollment in clinical trials will not be successful if people perceive they are not being treated with respect and compassion. Everyone involved is more likely to be

engaged and enthusiastic if they are treated with respect, appreciate the importance of the trial, and are not unduly burdened as a result of their participation.

### 10.2.5.5 Provision of Standard of Care

Participants randomized to the control group in RCTs often receive “standard” therapy; these are termed positive-controlled trials. Alternatively, control subjects may be monitored rather than treated. These are termed negative-controlled trials. Occasionally, trials will include positive and negative controls. In planning a positive-controlled trial, it is necessary for investigators to decide what constitutes the “standard of care.” In planning a positive controlled trial, it is necessary for investigators to decide what constitutes the “standard of care” [43].

For dental caries, common and scientifically proven methods to remove caries include lasers, high-pressure water jets, and traditional high-speed rotary drills [44]. The definition of standard of care can impact how a new treatment compares to the standard in terms of clinical outcomes, but also how patients perceive the new treatment. Consider a trial designed to explore patient preferences for a new restorative technique. The new treatment may compare differently if the standard involves the use of a noisy high-speed drill and placement of a dental amalgam filling versus a method in which tooth decay is removed with a laser and replaced with a tooth-colored filling. Because both methods are used in clinical practice, each could satisfy the legal definition of standard of care.

Professional societies typically do not attempt to define standard of care but instead may publish practice guidelines or parameters of care [45]. Nonetheless, the standard of care in dentistry includes an examination to identify disease so it may be appropriately treated. Patients with oral disease should be referred for treatment if it is not part of the trial protocol, and those with oral signs and symptoms of medical disorders should be referred to a health care provider for assessment and care. Examples would include patients with oral disease suggestive of uncontrolled diabetes, or those with mucosal lesions that are suggestive of conditions such as pemphigoid, pemphigus, lichen planus, or a variety of blood dyscrasias.

The American Academy of Periodontology’s (AAP) parameter of care for gingivitis [46] recommends that patients with gingivitis receive education, personalized oral hygiene instruction, and a tooth scaling. It is also recommended that the use of antimicrobial and antiplaque agents or devices be considered for patients who are unable to clean their teeth adequately and that overhanging tooth restorations, ill-fitting prostheses, or other conditions that retain dental plaque be corrected. In gingivitis trials, the test intervention should be compared to treatment that is consistent with these guidelines.

The AAP’s parameters of care for chronic periodontitis [47, 48] clearly outlines the importance of a personalized oral hygiene instruction and the provision of nonsurgical and maintenance care. In positive-controlled periodontitis trials, control subjects should receive treatment that is consistent with these guidelines. For trials involving periodontal surgery, patients should be managed according to these guidelines either before trial enrollment or before they are randomized to receive surgical treatment. This is important because periodontal surgery is ineffective in the presence of inadequate plaque control [49].

### 10.2.5.6 Use of Nontreatment and Delayed Treatment Controls

When considering the use of placebo or untreated controls in oral health clinical trials, the consequences of not providing care or delaying it must be carefully considered. The ethics of using untreated controls may vary according to the oral disease being studied. For example, the ethics of using untreated controls differ in trials of gingivitis versus trials of premalignant oral lesions. Gingivitis is an early form of periodontal disease that does not affect the supporting structures of the teeth; it is reversible with simple oral hygiene procedures such as tooth brushing and flossing and does not cause destruction of periodontal support. People with gingivitis could be included in a delayed treatment group, provided there is monitoring to ensure that there is no progression to periodontitis and loss of tooth support. Conversely, delaying treatment of premalignant oral lesions could put patients at risk for developing cancer and would raise serious ethical concerns.

Chronic periodontal disease is a slowly progressing disease that can lead to tooth loss. Tooth loss generally occurs over many years and the risk for progressive disease is relatively low, even in untreated patients over a 3-year period [50]. In a recent study of pregnant women with periodontitis, it was found that only 0.28% of sites in untreated control subjects had progressive disease over 20 weeks [51]. Despite the low risk for disease progression, it is important to monitor subjects in clinical trials and provide early rescue treatment as soon as evidence of progressive disease is found. For nonsurgical periodontitis trials, untreated control subjects with progressive disease should receive scaling, root planning, and oral hygiene instruction. Depending on the patient's periodontal status and oral hygiene, previously treated subjects may be re-treated with mechanical therapy alone, receive local or systemic antimicrobials, or be treated surgically.

Oral health trials of long duration sometimes use community-based controls. In these trials, initial treatment and follow-up care is administered to test subjects while control subjects are referred to community dentists for treatment and follow-up care. These trials are uncommon in oral diseases because the trials tend to be of rather short duration. However, oral disease prevention trials may require long-term follow-up and use of community-based controls should be considered. A concern with this approach is the lack of available dental care in some communities. If a trial recruits from low socioeconomic or underserved populations who have limited access to oral health care, requiring control subjects to seek care in community settings may not be ethical.

In all oral health trials, untreated or delayed treated control subjects should receive follow-up evaluations at the same time intervals as test subjects. Because patients with periodontal disease are typically seen for follow-up care every 3–4 months, participants in periodontitis trials also should be examined and receive follow-up care at this interval. Examinations should include full-mouth periodontal probing with radiographs as needed. For dental caries trials, participants should be examined at least semiannually; more frequently for high-risk subjects. Depending on the specific trial, these examinations may include visual assessment, clinical probing, tooth transillumination, and use of other imaging systems. Intraoral radiographs should be obtained as indicated by caries risk.

### 10.2.4 CARIES CLINICAL TRIALS

Dental caries is a chronic, slowly progressing infectious disease caused by acid produced from microorganisms and fermentable carbohydrates. Dental caries can affect enamel, dentin, and cementum; clinically, it occurs on a continuum from initial mineral loss to complete tooth destruction [52]. Caries can occur on the coronal or root surfaces of the teeth and on smooth surfaces or occlusal pits and fissures. Historically, the primary purpose of caries prevention and/or treatment has been to restore function and prevent tooth loss.

Traditional methods for assessing caries focused on cavitated lesions using the Radike or the World Health Organization (WHO) DMFS indices [53]. The Radike DMFS index diagnoses active caries (D) at the cavitated level, whereas the WHO index allows for four different levels of active caries, scored as D<sub>1</sub>, D<sub>2</sub>, D<sub>3</sub>, and D<sub>4</sub>. D<sub>1</sub> and D<sub>2</sub> scores indicate noncavitated lesions, whereas D<sub>3</sub> and D<sub>4</sub> are gradations of cavitated lesions. For comparability, the Radike DMFS index is labeled D<sub>3</sub>MFS to indicate assessment of cavitated lesions. Past evidence of caries is evaluated by counting restorations (F), and missing teeth (M). In caries clinical trials, incidence (new disease or caries increment) is calculated as the difference between the final and initial D<sub>3</sub>MFS scores. Historically, caries trials were conducted over 3 years and included a baseline, final, and at least one interim D<sub>3</sub>MFS scoring. The 3-year caries mean increments were compared for the test groups using linear models, after adjusting for important covariates. A common summary measure was the percentage reduction in caries incidence for a test agent compared with a reference (control) agent. D<sub>3</sub>MFS indices tend to overestimate caries experience because teeth, especially in adults, can be missing or filled for reasons other than caries. Despite its limitations, the D<sub>3</sub>MFS index has been widely used in caries trials. It is sufficiently sensitive to detect differences in relative efficacy of a wide variety of fluoride delivery systems, including water fluoridation, fluoride mouth rinses, dental office fluoride application, and fluoride dentifrices.

Dental caries has declined in many developed countries over the past few decades [54], and the rate of caries progression through dentin has slowed [55, 56]. As a result, caries activity has extended into and beyond late adolescence. While fluoride can enhance enamel remineralization, a caries-resistant outer shell of enamel can sometimes mask dentinal caries, making caries difficult to diagnose without using radiographs. As a result of changes in caries rates, prevention and treatment trials are more expensive, which may increase the possibility that newer, potentially superior anticaries agents may never be tested in phase III trials.

Since the 1980s, additional methods have been developed for diagnosing caries at the precavitated state. These include modifications of the visual-tactile method, such as the D<sub>1</sub>MFS [57], the International Caries Diagnostic Assessment System (ICDAS) [58] and sophisticated technologies such as fiber-optic transillumination (FOTI) and digital fiber-optic transillumination (DIFOTI) [59], quantitative light-induced fluorescence (QLF) [60], DIAGNOdent [61], and electrical conductivity measurement (ECM) [62]. Although these methods need to be more fully validated before they are accepted for use in pivotal clinical trials, they have the potential to improve diagnosis of precavitated carious lesions. Use of newer and more sensitive diagnostic methods also could change the focus of therapy from “drilling and filling”

decayed teeth to remineralizing or reversing early caries lesions. It could also reduce the length, size, and cost of caries trials.

A major issue that has plagued caries trials is high rates of subject dropout. Earlier caries trials commonly reported findings based on a per-protocol analysis rather than using some method to impute data for dropouts. Use of a per-protocol analysis that ignores subject dropouts risks introduction of unknown bias. For example, if a test mouth rinse, which has an unpleasant taste and stains the teeth, is compared to a pleasant tasting, nonstaining control rinse, test subjects who use the rinse regularly may be more likely to drop out of the study than irregular users because of the unpleasant side effects. This would introduce a bias against finding a favorable effect. By using data imputation methods, such as an intention to treat analysis, one may minimize the effects of unknown bias in a trial. There are many other methods available to impute missing data from dropouts that are covered elsewhere in this text.

Future caries trials, regardless of whether they score dental caries at the precavitated or cavitated level, will require refined analytical methods such as proposed at the Glasgow Conference on Dental Caries [22]. The new primary caries measures are ordinal scaled variables that are refinements of the traditional DMF scoring system. One new diagnostic system being investigated is the ICDAS method [58], which accounts for possible remineralization or “healing.” Previous analytical approaches did not consider that carious lesions could be reversed, and necessarily considered observed longitudinal reversals as examiner error. Today, investigators must distinguish “real” reversals from those due to examiner error. In the future, more emphasis on examiner training and calibration will be required to estimate the frequency of these reversals during caries trials.

Clinical performance of dental materials used to replace carious tooth structure is usually judged with U.S. Public Health Service (USPHS) Guidelines or Ryge criteria developed by Cvar and Ryge in 1971 [63]. These criteria have been modified and expanded over the last 40 years for evaluation of new dental materials. Criteria in the current system include color match, quality of margins, cavosurface marginal discoloration, anatomic form, recurrence of caries around the margin of the restoration, surface texture, postoperative sensitivity, proximal contact, and occlusal contacts [64]. Calibration training materials for investigators evaluating clinical outcomes of dental materials are available online [65].

## 10.2.7 PERIODONTAL CLINICAL TRIALS

Periodontal disease is an inflammatory disease that affects the surrounding and supporting tissues of the teeth (periodontium). It is usually divided into the two broad categories of gingivitis and periodontitis. These common infectious diseases are caused by pathogenic microflora in the dental biofilm (dental plaque) that forms daily on the teeth [66]. Gingivitis, the mildest form of periodontal disease affecting 50–90% of adults worldwide [67], is readily reversible by simple, effective oral hygiene. Estimates from a large survey based on partial mouth examinations indicated that about 22% of U.S. adults had mild periodontitis and 13% had moderate or severe disease [68].

A variety of interventions have been tested in clinical trials involving gingivitis and periodontitis ranging from nonsurgical mechanical therapy and attempts to improve oral hygiene to surgical and pharmacologic approaches. Outcomes generally include indices of dental plaque and gingival inflammation for gingivitis [21] and measures of periodontal pocket depth, clinical attachment level, and radiographic measures of alveolar bone support for periodontitis. The American Dental Association (ADA) published guidelines for clinical trials of products designed to control plaque and gingivitis [21]. For ADA acceptance, a product must demonstrate efficacy in two trials, each of which must include a placebo arm and follow subjects for at least 6 months. The guidelines further specify outcome measures for gingivitis and plaque as measures of safety including effects on the oral soft tissues, teeth, toxicology, and microbiology.

The ADA also has published guidelines for periodontitis trials [69] and others have published suggestions for the conduct of periodontitis trials [70]. These recommend that new treatment methods should be compared to basic periodontal therapy consisting of thorough scaling and root planing, oral hygiene instruction, and regular maintenance care. Moreover, it has been suggested that the primary response variable should be clinical attachment level, but that it is important to document changes in probing depth since this is a meaningful measure to many clinicians [19]. Gingival inflammation and bleeding should be used as secondary response variables. Radiographic measures of disease may be useful as primary response variables if safe, reproducible, and valid methods to assess change are utilized. Microbiological monitoring should be a secondary response variable because of numerous questions concerning sampling methodology, quantitative expression of data, and meaningful interpretation in terms of relevance to disease activity. It has also been recommended that the length of periodontitis trials be a minimum of 9 months if claims of superiority or equivalency are made compared to basic periodontal therapy [19]. As with most other diseases and conditions, there is a lack of validated surrogate outcomes for periodontal clinical trials and much more research is needed in this area [8].

### 10.2.8 DENTAL IMPLANT CLINICAL TRIALS

Dental implants have been used to replace missing teeth for over 20 years [71]. Usually, implants are made of commercially pure titanium or titanium alloys. Compared to smooth or machined surfaces, roughened surfaces increase surface area and promote earlier and more complete osseointegration [72]. Most implants are placed in the alveolar bone and are termed endosseous implants. Although dental implants can be used to retain maxillofacial prostheses and provide anchorage during orthodontic treatment, this discussion is limited to situations in which implants are used to replace natural teeth.

Implant failures are classified according to time after initial placement in bone. Early failures, occurring weeks or months after the implant is placed, result from excessive surgical trauma, postoperative infection, or early micromovement of the fixture. Late failures, occurring months or years after artificial tooth or teeth are attached to the implant(s), can be caused by peri-implantitis or occlusal trauma. Peri-implantitis is a destructive inflammatory disease caused by many of the same



microorganisms associated with periodontitis of natural teeth. Occlusal trauma occurs when the biting force on dental implants is excessive, causing loss of the intimate contact between the alveolar bone and the implant. This loss of contact may manifest as decreased radiographic evidence of bone support, as a peri-implant radiolucency, or clinical mobility of the implant.

Dental implant studies have several unique features. Much of the information regarding implant success comes from case series and retrospective studies rather than randomized controlled trials. Relatively few trials have been reported that compare two or more implant treatments or designs. The success of implants depends in part on the amount and density of the surrounding alveolar bone. The highest success rates tend to be in the anterior region of the mandible [73] where there is often abundant dense cortical bone for implant anchorage. Therefore, randomization in dental implant trials should take into account jaw location of implants and other possible confounding factors such as number of implants placed, occlusal force distribution, and type of prosthesis supported by the implant(s). This is especially true for trials that have relatively few subjects (even hundreds of subjects) because randomization without appropriate stratification may result in an unequal distribution of important selection and confounding factors among the treatment groups.

Implant trials generally follow participants for 1–5 years, which is longer than most oral health clinical trials. The longer duration is dictated in part by the U.S. FDA but also by the need to evaluate longer-term stability of the bone-to-implant interface. Implant metal or surface treatments that are biocompatible over the short term may have relatively high long-term failure rates if, for example, the oxide or metal surface corrodes or fractures over time. Longer follow-up is also needed to evaluate outcomes after the prosthesis (artificial teeth) are fixed to the implant(s). A variety of outcomes are used in dental implant trials to judge efficacy, including implant and prosthesis survival, annual rate of radiographically determined bone loss, and prosthetic complications (e.g., screw loosening, implant fracture, soft tissue inflammation). While most trials track implant or prosthesis survival as a primary outcome, use of patient-centered outcomes should be considered because the ultimate goal is to improve patient function, comfort, and esthetics. It may also be important to evaluate if specific implant designs require more frequent follow-up care related to, for example, loosening or breakage of screws or other devices that attach prosthesis to implants. Because implants can remain immobile and asymptomatic until bone loss approaches their apex, trials occasionally compare crestal bone loss rates rather than implant survival. The annual rate of bone loss around a “healthy” functioning implant is less than two-tenths of a millimeter a year [74] and “failing” implants typically exhibit a higher rate of loss. By longitudinally comparing sequential, standardized radiographs over time, it may be possible to compare bone loss rates between implant designs. However, at the present time, the rate of alveolar bone loss is an unproven surrogate marker for implant survival. Although less powerful than a randomized clinical trial in terms of design and interpretation, a nonrandomized observational prospective cohort study could be used to compare bone loss to historical “normal” rates of bone remodeling. Occasionally, investigators use composite outcomes to compare implant treatments, where “events” are defined as combinations of implant loss, excessive alveolar bone loss adjacent to the implant, and persistent inflammation or pain. As with periodontitis trials, investigators have studied associations between various biochemical or microbiological measures and



implant survival [75], but there is currently insufficient data to support use of these surrogates in clinical trials. Clinical trials can also be designed to compare various treatments for peri-implantitis. Because peri-implantitis affects the soft tissue cuff surrounding the metal fixture, outcomes for these trials are similar to those used in periodontitis trials, and include probing depth, gingivitis indices, and radiographic measurements of alveolar bone.

With the increasing awareness that dental implants are a predictable treatment option for replacing missing teeth, there is a growing interest in exploring the use of dental implants as an option for addressing extensive dental treatment needs. These trials would randomly assign patients with extensive dental treatment needs to have either: (1) conventional dental treatment involving periodontal, endodontic, and restorative care or (2) tooth extraction and replacement with dental implants. For these trials, it will be necessary to evaluate patient-centered outcomes and cost-benefit ratios of various treatments using methods such as those reported by Zitzmann et al. [76].

## 10.2.9 OTHER ORAL CONDITIONS

### 10.2.9.1 Xerostomia

Saliva is an important component of the system that maintains oral health. It modulates oral microbial ecosystems, plays a critical role in preventing dental caries, lubricates oral tissues, and facilitates swallowing. Patients with decreased saliva production have an increased risk for dental caries, tooth loss, mucosal candidiasis, and decreased quality of life. While many bodily diseases are associated with decreased salivary output (or salivary flow rate), the most pronounced decreases are found in patients with Sjögren's syndrome and individuals who have received radiation to the head and neck for treatment of cancer.

Sjögren's syndrome is an autoimmune exocrinopathy typified by an intense lymphoplasmocytic infiltration of the salivary and lacrimal glands. Affected individuals have partial to complete loss of their ability to make saliva and tears, but the precise processes responsible for this loss of function are yet to be defined. Medications for Sjögren's syndrome can be classified into three groups: (1) drugs that stimulate remaining functional salivary and lacrimal tissue (cholinergic agonists) [77, 78], (2) anti-inflammatory agents that attempt to decrease autoimmune activity, thereby restoring glandular function [79], and (3) lubricating agents that improve patient comfort [79, 80]. The two agonists approved for use in Sjögren's syndrome in the United States are pilocarpine HCl (Salagen) and cevimeline (Evxoxac). In most clinical trials, treatment is evaluated using subjective patient reports of changes in oral and/or ocular dryness and objective measures that quantify saliva and/or tear production. Occasionally, trials involving anti-inflammatory agents measure the amount of lymphoplasmocytic infiltration in the minor salivary glands using a gradation scale [81]. Other outcomes might include levels of serum autoantibodies and other markers of autoimmune disease that are commonly found in these patients.

Cholinergic agonists [82] and mucosal tissue lubricating agents are often used to manage patients with radiation-induced xerostomia. Only small studies have tested the efficacies of different lubricating agents for relief of oral dryness [83] and trials

to *prevent* radiation-induced xerostomia have involved amifostine [84] and other strategies to exclude the salivary glands from the radiotherapy field. These prevention trials need to evaluate patients at least one year postradiotherapy to determine long-term efficacy.

### 10.2.9.2 Mucosal Candidiasis

The oral mucosa may become infected with opportunistic fungal pathogens, including *Candida* species. These infections are most common in symptomatic HIV-infected patients, patients who have certain types of immunodeficiencies (primarily T cell and phagocyte deficiencies) and individuals undergoing cancer therapy. Both systemic and topical antifungal agents have been tested and approved as therapies for mucosal candidiasis and most phase III trials have been conducted in people with HIV infection [85]. However, candidiasis trials have not used consistent methods to diagnose infection and judge cure. Moreover, erythematous mucosal lesions caused by fungal infections may be ignored because of diagnostic difficulties associated with these infections. The most widely used diagnostic criteria in Phase III trials is a positive culture for *Candida* and presence of a white patch on the oral mucosa that rubs off, leaving bleeding or redness. A “clinical cure” may be defined as the absence of clinical lesions, with or without negative culture. If the trial employs cultures as an outcome measure, a negative culture for *Candida* is termed “mycologic cure.”

### 10.2.9.3 Mucosal Diseases

The two most common mucosal diseases treated by dentists are aphthous ulcerations and oral lichen planus. These diseases are believed to be mediated by infiltrating T lymphocytes; both can be idiopathic or develop after exposure to certain medications; and both may be chronic and have a major impact on quality of life. Clinical trials of treatment and prevention of these diseases usually enroll subjects with the most severe forms of disease. Since both diseases are mediated by lymphocytes, most drugs used to treat these diseases have been topical or systemic anti-inflammatory agents.

Clinical trials of oral lichen planus (OLP) usually require that patients have histopathologic evidence of OLP as an entry criterion. Various scales using objective methods to assess therapeutic response are available to semiquantify the clinical appearance and mucosal surface area affected by the lesions [86]. The primary outcome of many studies involves symptomatic scoring such as the amount of oral pain or burning sensation. The therapeutic portion of trials may be 4 weeks, with additional follow-up for one month [87, 88]. Unfortunately, most patients with symptomatic OLP of moderate severity have persistent disease; therefore, long-term outcomes (i.e., disease severity at 6 months) should be assessed.

Oral aphthae typically persist for 7–14 days after eruption. Patients with the most severe forms of the disease continually develop new lesions, which can cause significant oral pain and limit food intake. The most common outcome measures in trials for recurrent aphthous ulcerations include amount of oral pain, duration of an individual ulceration, and number of new ulcerations in a defined time period [88]. Therefore, trials should be of sufficient duration (4–6 months) to determine the impact of the therapy on development of new lesions.

## REFERENCES

1. ADA (2007), American Dental Association Policy Statement on Evidence-Based Dentistry; available February 28, 2008, <http://www.ada.org/prof/resources/positions/statements/evidencebased.asp>; accessed January 2009.
2. Hujoel, P. P., and DeRouen, T. A. (1992), Validity issues in split-mouth trials, *J. Clin. Periodontol.*, 19, 625–627.
3. Hujoel, P. P. (1998), Design and analysis issues in split mouth clinical trials, *Community Dent. Oral Epidemiol.*, 26, 85–86.
4. Chilton, N. W., and Fleiss, J. L. (1986), Design and analysis of plaque and gingivitis clinical trials, *J. Clin. Periodontol.*, 13, 400–410.
5. Fleming, T. R., and DeMets, D. L. (1996), Surrogate end points in clinical trials: Are we being misled? *Ann. Intern. Med.*, 125, 605–613.
6. Downing, D. G., Ed. (2000), *Biomarkers and Surrogate Endpoints: Clinical Research and Applications*, Proceedings of an NIH–FDA Conference, April 15–16, 1999, Bethesda, MD, Elsevier, New York.
7. Hujoel, P. P. (2004), Endpoints in periodontal trials: The need for an evidence-based research approach, *Periodontol. 2000*, 36, 196–204.
8. Barnett, M. L., and Pihlstrom, B. L. (2004), Methods for enhancing the efficiency of dental/oral health clinical trials: Current status, future possibilities, *J. Dent. Res.*, 83, 744–750.
9. Laster, L. L. (1985), The effect of subsampling sites within patients, *J. Periodontal. Res.*, 20, 91–96.
10. Goodson, J. M. (1986), Clinical measurements of periodontitis, *J. Clin. Periodontol.*, 13, 446–460.
11. Tu, Y. K., Gilthorpe, M. S., Griffiths, G. S., et al. (2004), The application of multilevel modeling in the analysis of longitudinal periodontal data—part I: Absolute levels of disease, *J. Periodontol.*, 75, 127–136.
12. Tu, Y. K., Gilthorpe, M. S., Griffiths, G. S., et al. (2004), The application of multilevel modeling in the analysis of longitudinal periodontal data—part II: Changes in disease levels over time, *J. Periodontol.*, 75, 137–145.
13. Gilthorpe, M. S., Griffiths, G. S., Maddick, I. H., and Zamzuri, A. T. (2000), The application of multilevel modelling to periodontal research data, *Community Dent. Health*, 17, 227–235.
14. Hyman, J. (2006), The importance of assessing confounding and effect modification in research involving periodontal disease and systemic diseases, *J. Clin. Periodontol.*, 33, 102–113.
15. Schulz, K. F., and Grimes, D. A. (2002), Generation of allocation sequences in randomised trials: Chance, not choice, *Lancet*, 359, 515–519.
16. Friedman, L. M., Furberg, C. D., and DeMets, D. L. (1996), *Fundamentals of Clinical Trials*, Mosby, St. Louis.
17. Fleiss, J. L. (1986), *The Design and Analysis of Clinical Experiments*, Wiley, New York.
18. Kingman, A. (1986), A procedure for evaluating the reliability of a gingivitis index, *J. Clin. Periodontol.*, 13, 385–391.
19. Pihlstrom, B. (1992), Issues in the evaluation of clinical trials of periodontitis: A clinical perspective, *J. Periodontal. Res.*, 27, 433–441.
20. Hill, E. G., Slate, E. H., Wiegand, R. E., Grossi, S. G., and Salinas, C. F. (2006), Study design for calibration of clinical examiners measuring periodontal parameters, *J. Periodontol.*, 77, 1129–1141.

21. ADA (2007), American Dental Association Council on Scientific Affairs. Acceptance Program Guidelines Chemotherapeutic Products for Control of Gingivitis, 1997: available from [http://www.ada.org/ada/seal/standards/guide\\_chemo\\_ging.pdf](http://www.ada.org/ada/seal/standards/guide_chemo_ging.pdf); accessed January 2009.
22. Pitts, N. B., and Stamm, J. W. (2004), International Consensus Workshop on Caries Clinical Trials (ICW-CCT)—Final consensus statements: Agreeing where the evidence leads, *J. Dent. Res.*, 83 (Spec No C), C125–128.
23. Beck, J. D., Lawrence, H. P., and Koch, G. G. (1997), Analytic approaches to longitudinal caries data in adults, *Community Dent. Oral Epidemiol.*, 25, 42–51.
24. Mancl, L. A., Hujoel, P. P., and DeRouen, T. A. (2004), Efficiency issues among statistical methods for demonstrating efficacy of caries prevention, *J. Dent. Res.*, 83 (Spec No C) C95–98.
25. Caplan, D. J., Slade, G. D., Biesbrock, A. R., Bartisek, R. D., McClanahan, S. F., and Beck, J. D. (1999), A comparison of increment and incidence density analyses in evaluating the anticaries effects of two dentifrices, *Caries. Res.*, 33, 16–22.
26. Hujoel, P. P., Isokangas, P. J., Tiekso, J., Davis, S., Lamont, R. J., DeRouen, T. A., and Makinen, K. K. (1994), A re-analysis of caries rates in a preventive trial using Poisson regression models, *J. Dent. Res.*, 73, 573–579.
27. Hannigan, A. (2004), Using survival methodologies in demonstrating caries efficacy, *J. Dent. Res.*, 83 (Spec No C), C99–102.
28. Hannigan, A., O'Mullane, D. M., Barry, D., Schafer, F., and Roberts, A. J. (2001), A re-analysis of a caries clinical trial by survival analysis, *J. Dent. Res.*, 80, 427–431.
29. Imrey, P. B., and Kingman, A. (2004), Analysis of clinical trials involving non-cavitated caries lesions, *J. Dent. Res.*, 83 (Spec No C), C103–108.
30. Katz, B. P., and Huntington, E. (2004), Statistical issues for combining multiple caries diagnostics for demonstrating caries efficacy, *J. Dent. Res.*, 83 (Spec No C), C109–112.
31. Pockock, S. J. (2007), *Clinical Trials: A Practical Approach*. Chapter 9. The Size of a Clinical Trial, Wiley, New York.
32. Title 21, Food and Drugs. Subchapter H, Medical Devices, Part 872 Dental Devices. <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=872>, accessed January 2009.
33. International Organization for Standardization. ISO Standards: TC 106 Dentistry; available at: [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_tc\\_browse.htm?commid=51218](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=51218); accessed January 2009.
34. American Dental Association Dental Products Specifications; available at: [http://www.ada.org/prof/resources/standards/products\\_specifications.asp](http://www.ada.org/prof/resources/standards/products_specifications.asp); accessed January 2009.
35. USFDA Center for Devices and Radiological Health Standards. Premarket Notification 510(K); available at: <http://www.fda.gov/cdrh/devadvice/314.html>; accessed January 2009.
36. Witham, M. D., and McMurdo, M. E. (2007), How to get older people included in clinical studies, *Drugs Aging*, 24, 187–196.
37. Walton, J. N., and MacEntee, M. I. (2008), Screening and enrolling subjects in a randomized clinical trial involving implant dentures, *Int. J. Prosthodont.*, 21:210–214.
38. Weinstein, P., Milgrom, P., and Sanghvi, H. (1995), Recruitment issues: Errors of omission in dental research, *J. Dent. Res.*, 74, 1028–1029.
39. Shaya, F. T., Gbarayor, C. M., Yang, H.-K., Agyeman–Duah, M., and Saunders, E. (2007), A perspective on African American participation in clinical trials, *Contemp. Clin. Trials.*, 28, 213–217.

40. Milgrom, P. M., Hujoel, P. P., Weinstein, P., and Holborow, D. W. (1997), Subject recruitment, retention, and compliance in clinical trials in periodontics, *Ann. Periodontol.*, 2, 64–74.
41. Weinstein, P., Milgrom, P., and Sanghvi, H. (1995), Recruitment issues: Errors of omission in dental research, *J. Dent. Res.*, 74, 1028–1029.
42. Edelstein, B. L. (2002), Disparities in oral health and access to care: Findings of national surveys, *Ambul. Pediatr.*, 2, 141–147.
43. Graskemper, J. P. (2004), The standard of care in dentistry: Where did it come from? How has it evolved? *J. Am. Dent. Assoc.*, 135, 1449–1455.
44. Yip, H. K., and Samaranayake, L. P. (1998), Caries removal techniques and instrumentation: A review, *Clin. Oral. Investig.*, 2, 148–154.
45. American Dental Association (2007), Dental Practice Parameters; available at: <http://www.ada.org/prof/prac/tools/parameters/index.asp>; accessed January 2009.
46. American Academy of Periodontology (2000), Parameter on plaque-induced gingivitis, *J. Periodontol.*, 71, 851–852.
47. American Academy of Periodontology (2000), Parameter on chronic periodontitis with advanced loss of periodontal support, *J. Periodontol.*, 71, 856–858.
48. American Academy of Periodontology (2000), Parameter on chronic periodontitis with slight to moderate loss of periodontal support, *J. Periodontol.*, 71, 853–855.
49. Nyman, S., Lindhe, J., and Rosling, B. (1977), Periodontal surgery in plaque-infected dentitions, *J. Clin. Periodontol.*, 4, 240–249.
50. Lindhe, J., Haffajee, A. D., and Socransky, S. S. (1983), Progression of periodontal disease in adult subjects in the absence of periodontal therapy, *J. Clin. Periodontol.*, 10, 433–442.
51. Michalowicz, B. S., Hodges, J. S., DiAngelis, A. J., Lupo, V. R., Novak, M. J., Fevguson, J. E., Buchanan, W., Bofill, J., Papapanou, P. N., Mitchell, D. A., Matseoane, S., and Tschida, P. A. (2006), Treatment of periodontal disease and the risk of preterm birth, *N. Engl. J. Med.*, 355, 1885–1894.
52. Thylstrup, A., and Fejerskov, O. (1994), Clinical and pathological features of dental caries, in Thylstrup, A., Fejerskov, O., Eds. *Textbook of Clinical Cariology*, Munksgaard, Copenhagen.
53. World Health Organization (1979), *A Guide to Oral Health Epidemiological Investigations*, WHO, Geneva.
54. Whelton, H. (2004), Overview of the impact of changing global patterns of dental caries experience on caries clinical trials, *J. Dent. Res.*, 83 (Spec No C), C29–34.
55. Wenzel, A., Pitts, N., Verdonchot, E. H., Kalsbeek, H., and Wenzel, A. (1993), Developments in radiographic caries diagnosis, *J. Dent.*, 21, 131–140.
56. Kidd, E. A., Ricketts, D. N., and Pitts, N. B. (1993), Occlusal caries diagnosis: A changing challenge for clinicians and epidemiologists, *J. Dent.*, 21, 323–331.
57. Kingman, A., and Selwitz, R. H. (1997), Proposed methods for improving the efficiency of the DMFS index in assessing initiation and progression of dental caries, *Community Dent. Oral Epidemiol.*, 25, 60–68.
58. Ismail, A. I., Banting, D., Eggertsson, H., Ekstrand, K. R., Ferreira-Zandona, A., Longbottom, C., Pitts, N. B., Reich, E., Ricketts, D., Selwitz, R., Sohn, W., Topping, G. V., and Zero, D. (2005), Rationale and evidence for the International Caries Detection and Assessment System (ICDAS II). In: G. K. Stookey, ed., *Clinical Models Workshop: Remin-Demin, Precavitation, Caries*. Indiana University School of Dentistry, Indianapolis, IN.

59. Yang, J., and Dutra, V. (2005), Utility of radiology, laser fluorescence, and transillumination, *Dent. Clin. North Am.*, 49, 739–752, vi.
60. Stookey, G. K. (2004), Optical methods—quantitative light fluorescence, *J. Dent. Res.*, 83 (Spec No C), C84–88.
61. Lussi, A., Hibst, R., and Paulus, R. (2004), DIAGNOdent: An optical method for caries detection, *J. Dent. Res.*, 83 (Spec No C), C80–83.
62. Longbottom, C., and Huysmans, M. C. (2004), Electrical measurements for use in caries clinical trials, *J. Dent. Res.*, 83 (Spec No C), C76–79.
63. Cvar, J. F., and Ryge, G. (2005), Reprint of criteria for the clinical evaluation of dental restorative materials. 1971, *Clin. Oral Investig.*, 9, 215–232.
64. Hickel, R., Roulet, J. F., Bayne, S., Heintze, S. D., Mjor, I. A., Peters, M., Rousson, V., Randall, R., Schmalz, G., Tyas, M., and Vanherle, G. (2007), Recommendations for conducting controlled clinical studies of dental restorative materials, *Int. Dent. J.*, 57, 300–302.
65. Clinical Evaluation of Dental Restorations (CER) for Teaching and Research, University of Michigan; available at: <http://www.dent.umich.edu/cer/>; accessed December 2007.
66. Pihlstrom, B. L., Michalowicz, B. S., and Johnson, N. W. (2005), Periodontal diseases, *Lancet*, 366, 1809–1820.
67. Albandar, J. M., and Rams, T. E. (2002), Global epidemiology of periodontal diseases: an overview, *Periodontol.* 2000, 29, 7–10.
68. Albandar, J. M., Brunelle, J. A., and Kingman, A. (1999), Destructive periodontal disease in adults 30 years of age and older in the United States, 1988–1994, *J. Periodontol.*, 70, 13–29.
69. American Dental Association Council on Scientific Affairs. Acceptance Program Guidelines: Determination Efficacy in Product Evaluation 1999; available at: [http://www.ada.org/ada/seal/standards/guide\\_efficacy.pdf](http://www.ada.org/ada/seal/standards/guide_efficacy.pdf); accessed December 2007.
70. Imrey, P. B., Chilton, N. W., Pihlstrom, B. L., Proskin, H. M., Kingman, A., Listgarten, M. A., Zimmerman, S. O., Ciancio, S. G., Cohen, M. E., and O’Agostino, R. (1994), Proposed guidelines for American Dental Association acceptance of products for professional, non-surgical treatment of adult periodontitis. Task Force on Design and Analysis in Dental and Oral Research, *J. Periodontal. Res.*, 29, 348–360.
71. Adell, R., Eriksson, B., Lekholm, U., Branemark, P. I., and Jemt, T. (1990), Long-term follow-up study of osseointegrated implants in the treatment of totally edentulous jaws, *Int. J. Oral Maxillofac. Implants*, 5, 347–359.
72. Puleo, D. A., and Thomas, M. V. (2006), Implant surfaces, *Dent. Clin. North Am.*, 50, 323–338, v.
73. Ferrigno, N., Laureti, M., Fanali, S., and Grippaudo, G. (2002), A long-term follow-up study of non-submerged ITI implants in the treatment of totally edentulous jaws. Part I: Ten-year life table analysis of a prospective multicenter study with 1286 implants, *Clin. Oral. Implants. Res.*, 13, 260–273.
74. Eliasson, A., Eriksson, T., Johansson, A., and Wennerberg, A. (2006), Fixed partial prostheses supported by 2 or 3 implants: a retrospective study up to 18 years, *Int. J. Oral Maxillofac. Implants*, 21, 567–574.
75. Paolantonio, M., Di Placido, G., Tumini, V., Di Stilio, M., Contento, A., and Spoto, G. (2000), Aspartate aminotransferase activity in crevicular fluid from dental implants, *J. Periodontol.*, 71, 1151–1157.
76. Zitzmann, N. U., Marinello, C. P., and Sendi, P. (2006), A cost-effectiveness analysis of implant overdentures, *J. Dent. Res.*, 85, 717–721.



77. Vivino, F. B., Al-Hashimi, I., Khan, Z., Leveque, F. G., Salisbury, P. L. III, Tran-Johnson, T. K., Muscoplat, C. C., Trivedi, M., Goldlust, B., and Gallagher, S. C. (1999), Pilocarpine tablets for the treatment of dry mouth and dry eye symptoms in patients with Sjogren syndrome: A randomized, placebo-controlled, fixed-dose, multicenter trial. P92-01 Study Group, *Arch. Intern. Med.*, 159, 174–181.
78. Petrone, D., Condemi, J. J., Fife, R., Gluck, O., Cohen, S., and Dalgin, P. (2002), A double-blind, randomized, placebo-controlled study of cevimeline in Sjogren's syndrome patients with xerostomia and keratoconjunctivitis sicca, *Arthritis. Rheum.*, 46, 748–754.
79. Brennan, M. T., Shariff, G., Lockhart, P. B., and Fox, P. C. (2002), Treatment of xerostomia: A systematic review of therapeutic trials, *Dent. Clin. North. Am.*, 46, 847–856.
80. von Bultzingslowen, I., Sollecito, T. P., Fox, P. C., Daniels, T., Jonsson, R., Lockhart, P. B., Wray, D., Brennan, M. T., Carrozzo, M., Gandera, B., Fujibayashi, T., Navazesh, M., Rhodus, N. L., and Schiodt, M. (2007), Salivary dysfunction associated with systemic diseases: Systematic review and clinical management recommendations, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S57.e1–15.
81. Greenspan, J. S., Daniels, T. E., Talal, N., and Sylvester, R. A. (1974), The histopathology of Sjogren's syndrome in labial salivary gland biopsies, *Oral Surg. Oral Med. Oral Pathol.*, 37, 217–229.
82. Johnson, J. T., Ferretti, G. A., Nethery, W. J., Valdez, I. H., Fox, P. C., Ng, D., Muscoplat, C. C., and Gallagher, S. C. (1993), Oral pilocarpine for post-irradiation xerostomia in patients with head and neck cancer, *N. Engl. J. Med.*, 329, 390–395.
83. Chambers, M. S., Posner, M., Jones, C. U., Biel, M. A., Hodge, K. M., Vitti, R., Armstrong, I., Yen, C., and Weber, R. S. (2007), Cevimeline for the treatment of postirradiation xerostomia in patients with head and neck cancer, *Int. J. Radiat. Oncol. Biol. Phys.*, 68, 1102–1109.
84. Shiboski, C. H., Hodgson, T. A., Ship, J. A., and Schiodt, M. (2007), Management of salivary hypofunction during and after radiotherapy, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S66.e1–19.
85. Baccaglini, L., Atkinson, J. C., Patton, L. L., Glick, M., Ficarra, G., and Peterson, D. E. (2007), Management of oral lesions in HIV-positive patients, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S50.e1–23.
86. Al-Hashimi, I., Schifter, M., Lockhart, P. B., Wray, D., Brennan, M., Migliorati, C. A., Axell, T., Bruce, A. J., Carpenter, W., Eisenberg, E., Epstein, J. B., Holmstrup, P., Jontell, M., Lozada-Nur, F., Nair, R., Silverman, B., Thongprasom, K., Thornhill, M., Warnakulasuriya, S., and van der Waal, I. (2007), Oral lichen planus and oral lichenoid lesions: Diagnostic and therapeutic considerations, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.*, 103 (Suppl), S25.e1–12.
87. Chainani-Wu, N., Silverman, S. Jr., Reingold, A., Bostrom, A., McCulloch, C., Lozada-Nur, F., and Weintraub, J. (2007), A randomized, placebo-controlled, double-blind clinical trial of curcuminoids in oral lichen planus, *Phytomedicine*, 14, 437–446.
88. Thornhill, M. H., Baccaglini, L., Theaker, E., and Pemberton, M. N. (2007), A randomized, double-blind, placebo-controlled trial of pentoxifylline for the treatment of recurrent aphthous stomatitis, *Arch Dermatol.*, 143, 463–470.





# 10.3

---

## *Dermatology Clinical Trials*

**Maryanne Kazanis, Alicia Van Cott, and Alexa Boer Kimball**

*Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts*

### **Contents**

|          |                                                     |     |
|----------|-----------------------------------------------------|-----|
| 10.3.1   | Introduction                                        | 461 |
| 10.3.2   | Overall Study Design Considerations in Dermatology  | 464 |
| 10.3.2.1 | Drug Therapy Trials                                 | 464 |
| 10.3.2.2 | Device Studies: Phototherapy and Laser Trials       | 465 |
| 10.3.2.3 | Determining Study Size                              | 467 |
| 10.3.2.4 | Patient Selection and Recruitment Methods           | 467 |
| 10.3.2.5 | Patient Compliance                                  | 469 |
| 10.3.2.6 | Disease and Quality of Life Measurement Instruments | 470 |
| 10.3.2.7 | Source Documentation in Dermatology                 | 472 |
| 10.3.2.8 | Clinical Trial Registration                         | 474 |
| 10.3.3   | Summary                                             | 474 |
|          | References                                          | 475 |

### **10.3.1 INTRODUCTION**

Dermatology is a relatively young medical specialty [1, 2]. For many years diseases of the skin were not considered processes involving an individual organ but rather the manifestation of greater alterations within the whole human body. As a result of this school of thought, skin conditions fell under the practice of general medicine until the early eighteenth and nineteenth centuries. It was then that European physicians became increasingly curious about skin diseases and began to consider the skin as an organ that could be independently studied, diagnosed, and treated.

During this time, an analytical approach to dermatology emerged, and physicians increasingly relied on clinical observations to uncover cause-and-effect relationships between various exposures and skin diseases.

In America, it was not until the late nineteenth century that dermatology was established as a clinical discipline [2]. It was then that American dermatologists began to study in Europe, the academic epicenter for the specialty at the time. These physicians would learn about progress made in the field and would model their American practices after those of their European counterparts. During this period, although progress was made in establishing dermatology as a specialty in America, there was little emphasis placed on dermatological research. The scarcity of research in the field prompted the formation in 1937 of two organizations: the American Academy of Dermatology and the Society for Investigative Dermatology [2–4]. One year later, the Society for Investigative Dermatology published its first *Journal of Investigative Dermatology*. Both the creation of these societies and the publication of this journal were the first of their kind, as no previous journals or organizations had focused primarily on dermatology research.

Today, there are over 200 dermatology specialty journals and the importance of clinical research in the field is unquestionable [5]. In the last 15 years, the approach of medical practice in general has become more evidence based, and there has been a corresponding shift in clinical dermatology [6]. The evidence garnered from various research trials has significantly shaped the way physicians care for their patients, and the results of clinical research studies often guide dermatologists in treatment decision making. In this setting, the accurate design and execution of dermatology clinical trials is paramount.

Given the current diversity within the specialty, there is a large range of dermatology topics under study (Table 1). Although the majority of trials are therapeutic in nature, there are also quality of life studies and research projects investigating the response of skin disease to environmental exposures and other stressors. The structure of therapeutic trials in dermatology varies based upon what disease and treatment method is being studied. Dermatological therapeutic treatments include topical, oral, and subcutaneous drugs, phototherapy, lasers, and other devices. Each of these therapies presents unique considerations when designing and implementing clinical trials, and later these will be discussed in further detail.

Dermatology is a wonderful field for conducting clinical trials, as there are many advantages to studying the skin. First, it is an accessible organ that can be easily

**TABLE 1** Examples of Common versus Rare Debilitating Diseases in Dermatology

| Most Common Diseases          | Rare Diseases                                 |
|-------------------------------|-----------------------------------------------|
| Acne vulgaris                 | Epidemolysis bullosa                          |
| Actinic keratosis             | Blistering diseases: Pemphigus and pemphigoid |
| Nonmelanoma skin cancer       | Toxic epidermal necrolysis                    |
| Contact and atopic dermatitis | Cutaneous T-cell lymphoma                     |
| Cutaneous fungal infections   | Ichthyosis                                    |
| Herpes simplex and zoster     | Scleroderma                                   |
| Human papillomavirus          | Erythrodermic psoriasis                       |
| Melasma                       | Lichen planus                                 |
| Psoriasis                     | Pityriasis rubra pilaris                      |
| Acne rosacea                  | Pyoderma gangrenosum                          |

seen, touched, measured, and sampled. This is a benefit not achievable with other organ systems and allows for unique treatment options, readily observed patient responses, and less invasive study procedures. Further, given the symmetry of dermatological lesions, subjects may serve as their own controls, which is advantageous since it requires the recruitment of fewer subjects to a given trial. Lastly, the duration of clinical trials is generally shorter than in other areas of medicine. Endpoints may be easily recognizable as early as 12 weeks, which makes the skin an ideal organ system to study, especially in pilot programs. Overall, these characteristics permit rapid trials with straightforward treatment schedules and endpoints. In addition, dermatology is one of the only areas of medicine in which study subjects may evaluate their specific responses to treatment, allowing investigators to collect data on patient perceptions of improvement. These data are important when considering quality of life issues and other psychosocial factors that are associated with disease states. When a clinical trial has successful findings, these unique attributes of studying the skin have the potential to not only influence overall patient care but also to greatly impact patients' self-esteem and sense of well-being [7, 8]. An additional benefit to conducting dermatology clinical trials is that diseases of the skin manifest in both adults and children, so there is a demographic of patients from which to draw a study subject population.

Of course, there are multiple challenges associated with conducting dermatology clinical trials. Unlike other medical specialties that use biological markers or other objective tests as indicators of health or disease, many dermatology measurements are inherently subjective. Two physicians may differ in their assessments of the same patient simply because of varying degrees of interpretation. This interobserver variability can potentially lead to different results reporting and reductions in standardization across study centers and clinical facilities.

A second issue involved in conducting dermatology clinical trials is the influence of market constraints on the selection of diseases being studied [6]. Given the high costs associated with conducting studies, the pharmaceutical industry often sponsors research. The focus of industry sponsors is of course affected by fiscal concerns, which sometimes means that rare, serious skin diseases are studied less frequently than more common and sometimes less debilitating disorders.

The progress made in dermatology because of clinical trials research is significant [9]. Acne vulgaris is just one example of a condition that has seen drastic treatment changes because of clinical research [10]. Over the past century, many conclusions regarding the causes and remedies for acne have been made. Previously, sulfur and mercurials were two of the topical treatments used for acne, but today new acne treatments have been derived based on the understanding of acne's pathogenesis. Systemic therapy has evolved to include the use of oral antibiotics, and a number of topical drugs including various antimicrobials and retinoids are now the mainstay of treatment.

Similar advances have been made in psoriasis treatment. The earliest therapy for this skin disease was coal tar. Although today this is rarely used as primary therapy, many preparations for psoriasis do contain coal tar. Clinical trials have provided advancements allowing topical corticosteroids and phototherapy to emerge as methods for improving psoriatic lesions. Current dermatology psoriasis therapies include highly successful intravenous and injectable biologics, many of which are antibodies.

Advances in acne and psoriasis are just two examples of the progress made in dermatology because of the results of clinical trials. The future advancement of dermatology depends greatly on the ability to demonstrate improvements in patient care through clinical research [6]. The remainder of this chapter will focus on the guidelines, considerations, and other factors involved in the effective design and operation of these studies.

### 10.3.2 OVERALL STUDY DESIGN CONSIDERATIONS IN DERMATOLOGY

The aim of this section is to give a broad overview of some of the distinctive aspects involved in designing and conducting dermatology clinical trials.

#### 10.3.2.1 Drug Therapy Trials

The study design used in a dermatology clinical trial largely depends on the particular skin condition under investigation. The gold standard of therapeutic dermatology trials for common skin conditions is the randomized controlled trial [6]. Some drug studies compare placebo to a new medication of interest in order to evaluate its efficacy and safety in treating a particular disease. In dermatology clinical trials, placebos may be in the form of a vehicle (i.e., cream, ointment, lotion, foam, etc.) or in the form of a pill, subcutaneous injection, or infusion. Placebo, however, is really a misnomer when designing a study of an active topical versus its vehicle control. The effect of moisturizing the skin with any topical has beneficial effects on many skin conditions ranging from wrinkles to eczema to psoriasis and can easily result in improvements in the 20–30% range on many scoring scales. Care must be taken when designing these studies to make sure that the sample size is large enough to compensate for this effect.

In a blinded study, it is imperative that the investigator remains unaware of the treatment assignment. In trials that use placebo, avoiding unblinding is not an issue since the vehicle or other placebo can generally be made to look exactly like the actual drug under study.

However, there are a great number of dermatology trials that compare a drug already established in treating a condition to another, newer drug. In these instances, when trials are double blinded or investigator blinded, the potential exists for the subject or investigator to easily distinguish between the two drugs under study, especially in the case of topical treatments. For example, in a study comparing an ointment to a lotion, the study subject is aware of the treatment assignment based on the type of topical she or he is given. As a result, additional considerations are required in the study design to protect against investigator unblinding. In these trials, the protocol should state that study coordinators or other study staff members are responsible for drug administration and collection, drug application instruction, and any other aspects of the trial that involve transactions with the medication(s) under study. Furthermore, protocols should include guidelines for study staff to educate subjects to not discuss with the investigator the type of study medication they are using. For example, enrolled subjects should use the term “study medication” or “study drug” when discussing their participation with the investigator, rather than descriptions such as ointment, lotion, gel, and the like. In these trials,

subject and study staff education is of utmost importance in order to avoid potential investigator unblinding and bias. Study protocols should clearly outline medication dispensing, application, drug collection, and education procedures in accordance with these considerations.

Dermatologists also conduct open-label clinical trials, where both the investigator and patient are aware of the treatment assignment. By and large, open-label trials in dermatology are used to study rare diseases, as the subjects enrolled to these trials in the dermatology setting often require immediate therapy, rendering placebo use unsuitable. The guidelines mentioned above to avoid investigator unblinding do not apply here, given the nature of the study design. However, it is of utmost importance that the investigator remain as impartial as possible when measuring outcomes in this patient population, so as to avoid biasing study results. Whenever possible, the same investigator should be involved in the evaluation of disease throughout an open-label trial in order to minimize additional differences resulting from interobserver variability. Scoring scales, photography use with subsequent evaluation by a blinded observer, and patient perception of improvement assessments may play particularly important roles in these studies. Another valuable element to ensure objectivity in open-label trials is to have multiple assessors conducting subject evaluations or to design a trial where the treating physician is different from the investigator appraising subjects' response to treatment. These features should be incorporated into the study design where applicable to maintain impartiality.

#### **10.3.2.2 Device Studies: Phototherapy and Laser Trials**

The earliest trials involving lasers and phototherapy compared actual therapy to simulated treatment for placebo comparison. However, today very few, if any, trials are done using sham light and laser sources since the efficacy of phototherapy and laser treatment to ameliorate certain disease symptoms has been well established. Most current studies done in these areas investigate whether a particular light source or laser type is superior to another for a given indication, whether a light source or laser is better than no treatment, or whether a light source or laser type works better with or without adjuvant drug therapy. The latter typically involve placebo use for comparison. The double-blind or investigator-blinded study design frequently involves randomizing subjects to treatment regimens and assessing efficacy and monitoring adverse effects of treatment. For example, one group of subjects might be randomly assigned to treatment with conventional ultraviolet B (UVB) phototherapy and compared to another group of subjects that undergoes narrow-band UVB phototherapy to assess which offers better disease response and patient safety. In phototherapy trials of this type, the risk of investigator unblinding is rarely an issue since technicians typically administer light treatment to subjects. This is useful since the investigator may remain unaware of the subjects' treatment assignment and conduct assessments in an unbiased manner. Laser clinical trials differ in this respect since, especially in new studies, the investigator is often involved in the treatment of subjects. Therefore, one way to avoid potential bias in the design of these trials is to state that the treating physician should not be the investigator who is conducting subject skin assessments. Additionally, photography may be incorporated into clinical trials of this type to provide objective evidence of skin modifications.

Some laser and phototherapy studies use symmetrical target lesions on the same subject for the purpose of left–right comparisons. Left–right comparisons involve treating two similar target areas on opposite sides of the body and comparing outcomes. For example, if an investigator is interested in seeing if a topical drug works better than placebo in combination with a light source, a double-blind study might be done where one side of the body is treated with a placebo vehicle, and the opposite side of the body is treated with the topical drug. Both areas would be exposed to the light source and outcomes would be measured to see if the drug did increase the efficacy of the phototherapy treatment. Although this method requires the recruitment of fewer individuals and is advantageous for gathering data on how two particular treatments affect the same subject, this approach does allow for greater chance of error. Subsequently, the study staff must take great care to meticulously record accurate drug/placebo application, outcome measurements, and drug safety assessment information. Protocols using this study design should contain guidelines to avoid error, and source documents should be thoughtfully designed and well labeled in order to capture correct information.

Additionally, there are open-label trials done with phototherapy and lasers in dermatology. For example, some of the light studies involving biologics are open label. Similar to therapeutic trials that use open-label design, these studies typically involve subjects with more advanced disease and require investigators to be as impartial as possible in their subject assessments. Photography and other more objective methods of recording outcomes may be useful in these trials to correlate with physician assessments, and these should be considered when designing such a study.

Finally, for all dermatology studies, regardless of therapy type, timelines for follow-up visits should be outlined in the study protocol and occur at the given time intervals or within the “visit window” identified in the study protocol. The visit window is the amount of time around a designated study visit that a subject has to complete the visit without it being considered a protocol deviation. For example, if the protocol designates week 2 as occurring on day 15 and there is a 3-day visit window, then the day 15 visit can technically occur anytime between days 14 and 16 and still be within study protocol. The same holds true for the laboratory assessments for a particular visit. If, for example, blood specimens were not able to be obtained at the week 2 visit, and it occurred on day 14, then the subject has until day 16 to return to the clinic for a blood draw before being considered a protocol deviation.

Although not obtaining all study data at a given visit is not advisable, on occasion it may be unavoidable. At follow-up visits, all medical exams, investigator assessments, subject questionnaires, laboratory collection, and redispensing of study drug should occur as outlined in the study protocol. If a study drug could potentially be associated with reproductive risks, the protocol should state that women of reproductive age have regular urine pregnancy tests done while on the study and that subjects of both sexes who are sexually active report their method of contraception. In addition, adverse events and concomitant medication taken should be reviewed with the study subject. A copy of the subject’s diary should be taken to include as a part of the source documentation if it is to be redispensed to the subject and proper study drug administration instructions should be reviewed with the subject. It is advisable to ask subjects how they take their study drug in order to eliminate



study staff coaxing and ascertain exactly how subjects are taking their study drug. It is also advisable to include a study checklist in the source documentation to verify that all study-related information is collected at follow-up visits.

### 10.3.2.3 Determining Study Size

One of the first steps in developing a clinical trial is deciding what number of subjects to enroll. Preliminary studies typically offer a good starting point in this process by providing information on the anticipated size needed to obtain a desired outcome [11, 12]. However, when deciding upon a final sample size, thought must be given to both statistical and clinical significance.

For clinical trials that address common skin conditions and diseases, patient recruitment is typically not a difficult issue, and obtaining large sample sizes for trials is uncomplicated. However, dermatologists conducting these trials should consider that too large a sample size may lead to a situation of a statistically significant finding that holds no real clinical relevance [11]. This is because as sample sizes grow exceedingly larger, even the smallest differences that are not necessarily useful to patient care may be detected. Further, in dermatology it is especially important to prospectively consider the hypothesized central outcome of a trial, given that patient perceptions of improvement play a significant role in evaluating clinical significance. For example, a study that has the statistical power to show that drug A for acne is slightly better than drug B will only really be clinically useful if the subjects on trial felt that their acne indeed improved on drug A [12, 13].

Conversely, for rare, more serious dermatological diseases, the opposite challenge exists. Recruiting adequate numbers of study participants within a reasonable time frame proves to be logistically and economically difficult [11]. Often, these smaller trials may fail to statistically conclude that a relationship exists, even if one does, simply because the study sample size is not large enough. As a result, it is often the most devastating diseases that are the hardest to research. Investigators conducting trials with these patient populations should be aware of the challenges and thoughtfully plan subject recruitment strategies to ensure adequate subject enrollment.

Similar problems exist for detecting differences in rare outcomes or uncommon side effects between groups. For example, even if a sample size is large enough to determine efficacy, it may still be too small to ascertain useful data on safety and dosing schedules [11]. Therefore, it is necessary in these instances to appropriately design a trial that will have a large enough number of subjects to ascertain reliable safety and outcome results.

### 10.3.2.4 Patient Selection and Recruitment Methods

At the outset of a clinical trial, it is important for the investigator to define a study's target population. In dermatology, a study population is defined in terms of any combination of the following characteristics: age, gender, race, complexion, skin phototype (type I pale white skin to type VI dark brown or black skin), previous photodamage, duration of disease, and severity of disease. The disease under study dictates which characteristics are of significance. Of note, several major dermato-

logic diseases occur predominantly or prominently in the pediatric population ranging from eczema to acne. Designing studies for this age group, of course, requires additional considerations, ranging from assenting patients, minimizing discomfort, and designing ethically appropriate studies for this age group.

All required baseline characteristics should be clearly outlined in the protocol. This is imperative in understanding which treatments work in a given patient population and is especially useful when translating study results to the patient care setting since a more standardized subject population yields a greater ability to detect true differences in study outcomes. Scoring instruments, which will be discussed in later sections, are used to grade subject disease severity and are helpful tools that may help determine which subjects should be enrolled to a particular trial. Protocols should state exactly which scoring scale or grading method should be used to assess disease at baseline and throughout the study.

Inclusion/exclusion subject criteria should contain directives regarding the use of other therapies to treat the skin disorder under study. Typically, studies require no use of such medications for a period of 14–30 days before the start of study drug. This time frame is often referred to as the “wash-out” period. The wash-out period ensures that all subjects enrolled to a trial have comparable baseline off-treatment characteristics. Enrollment criteria should also include information regarding past therapies used to treat the skin disorder under study. In some cases, it may be necessary to exclude a subject from a study if it is deemed that a prior therapy used could affect study participation and outcomes. Additionally, a protocol should specifically state if moisturizers, cleansers, or other skin products may be used by enrolled subjects. This is because various emollients and cleansers contain ingredients that may improve certain dermatological conditions and potentially affect clinical outcomes.

Dermatology is unique in that people who have skin conditions are able to identify their own diseases. As a result, subject recruitment outside of the clinic setting is a common and successful approach. Unlike other areas of medicine that recruit subjects from clinic patient populations, dermatology clinical trials are often advertised to the general public in the newspaper or in flyers. This method provides a great advantage to subject enrollment strategies since it reaches a much greater sample population.

Given that the sample population is from the general public, potential subjects who respond to trial advertisements must be screened thoroughly. Their medical histories and medications must be completely reviewed prior to enrollment to assure participant safety and subject eligibility. In addition, it is the responsibility of the investigator and study staff to assess subject compliance. For phototherapy trials, especially, where subjects are required to undergo treatments as often as three times per week, it is imperative that participants are willing to cooperate with study protocol procedures. Monetary or other incentives may be offered to increase patient recruitment and compliance. However, these incentives as well as any advertisements used in the study must be approved by the ethics committee of the participating institution to ensure that they coincide with federal regulations.

Regardless of how the subject is recruited, she or he must sign an informed consent form (ICF) before any study procedures are performed, as mandated by 21 Code of Federal Regulations, CFR 50.20. This may or may not occur at the screening visit. Subjects must be given ample time to consider if they want to participate in

the research study and should be allowed time to discuss the study with their physician and families. Furthermore, the ICF should clearly outline any potential risks associated with treatment, including reproductive risks for men and women.

#### 10.3.2.5 Patient Compliance

Monitoring patient compliance in clinical trials is essential to the acquisition of real outcome data and the conduction of a sound study [14–16]. It is imperative that subjects adhere to protocol visit schedules and that they comply with a stated treatment program. Prior to enrollment, subjects are asked to read the informed consent form, which describes what is entailed in study participation and gives information on potential benefits and risks. Subjects have the opportunity to ask questions and are only enrolled if they are amenable to study procedures. In most dermatology drug trials, studies are designed to be brief with few visits. This typically translates to increased patient compliance because it is less intrusive to the lives of subjects. However, for phototherapy trials, assessing and maintaining patient compliance becomes more of an issue since subject lives may be affected more by numerous, frequent study visits.

There are several ways to maintain an acceptable level of compliance among subjects. First, the simple act of calling patients to confirm appointments helps to remind them about their study visits. Second, short wait times and efficient execution of study visits help to make participation less of a chore for subjects. It is important to consider that subjects have issues that come up that may preclude them from maintaining a strict study schedule. Although these should be kept to a minimum, exceptions can be made for patient scheduling conflicts, and these should be reported to the investigational review board (IRB) of record and/or the study sponsor.

In addition to ensuring that subjects are seen at visits according to protocol, it is also important to make certain that they are properly using their study medication. This involves constant education of study subjects at visits. In dermatology clinical trials, for example, application techniques for topical drugs should be reviewed at each visit. Protocols must clearly and specifically outline how subjects should use medication and emphasize drug application reeducation by study staff. Whenever possible, especially in the case of topical and injected drugs, the first treatment should take place in the clinic so that proper dosing and drug administration practices may be established. In addition, for oral, subcutaneous, and topical treatment administrations that subjects perform at home, drug compliance logs should be completed by the subjects and collected at each visit in order to monitor patient adherence. Study protocols should specify that subjects return all medication at the end of study or at stated intervals throughout the trial to assess consumption. For oral and subcutaneous drugs, this is useful in ensuring that patients did indeed take the medication. An added benefit in collecting study medication for trials involving topical drugs is that bottles and containers may be weighed and the actual amount of drug used by each patient may be collected as data. Typically, drug containers are measured at dispensing, then again when returned by the subject at a time interval determined in the protocol. The difference in the two weights yields the amount used by the subject. When comparing patient outcomes, this information is valuable in evaluating subject responses in relation to the amount of drug applied to the skin.

Additionally, this measurement is helpful for assessing drug safety and efficacy since too much of a study drug could cause adverse reactions, while too little may be associated with inefficacy.

Subjects must also be instructed on the amount of time they have to take a dose of study drug, if one is missed. This is referred to as the “dosing window.” It is important that subjects are instructed to call the study staff if they do miss a dose of study drug and to record the missed dose in a study diary. In addition, the protocol should indicate that subjects should be instructed to record in their study diary any adverse events experienced and concomitant medications taken. The diary should be brought with each subject to all study visits. All of the above instructions should be reviewed at all study visits with the subject to make sure that the study protocol is being followed.

Common adverse events experienced in topical clinical trials include erythema, pruritus, irritation, and scale at application sites. These events are expected with most topicals in a small to moderate percentage of patients depending on the ingredients being studied. Some studies will therefore assess for these changes at each visit proactively and separately from the other adverse events in order to ensure uniformity of data collection.

It is important to note that study subjects may differ from “real” patients in that they are more motivated to adhere to treatment schedules and visits [6]. Monetary incentives combined with a general interest in seeking out investigational treatment make subjects in research studies a typically more compliant patient population. For dermatologists involved in clinical trials, it is useful to keep this in mind since a treatment or visit schedule that is inconvenient for study subjects has a high probability of translating into decreased acceptance by dermatology patient populations.

10.3.2.6 Disease and Quality of Life Measurement Instruments

Capturing appropriate data for dermatology clinical trials almost always involves the use of scoring instruments (Tables 2 and 3). These have been developed, tested, and established as reliable ways to obtain data points [17]. Scoring instruments serve

TABLE 2 Common Scoring Systems Used in Dermatology Clinical Trials

| Disease                                      | Scoring System                                        |
|----------------------------------------------|-------------------------------------------------------|
| Acne vulgaris                                | Acne lesion counts (papules, pustules, comedones)     |
|                                              | Leeds acne grading technique                          |
|                                              | Pillsbury scale                                       |
|                                              | Cook’s acne grading scale method                      |
| Atopic dermatitis                            | Scoring for atopic dermatitis (SCORAD)                |
|                                              | Eczema area scoring index (EASI)                      |
| Psoriasis                                    | Psoriasis area and severity index (PASI)              |
|                                              | Psoriasis severity index (PSI)                        |
|                                              | Lattice system—Physician’s global assessment (LS-PGA) |
|                                              | Overall lesion assessment (OLA)                       |
| Acne rosacea                                 | National Rosacea Society Rosacea Clinical Scorecard   |
| Nonspecific skin disease scoring instruments | Physician’s global assessment (PGA)                   |
|                                              | Dermatology index of disease severity (DIDS)          |

**TABLE 3 Acne Grading Systems**

| Name of Grading System                     | Scale                                                                                                                                                                               | Method                                                                                                                                                              |
|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Leeds acne grading technique               | Number of inflammatory and noninflammatory lesions present                                                                                                                          | Based on counting lesions and grouping them into noninflammatory and inflammatory categories                                                                        |
| Pillsbury scale                            | 1 (mildest) to 4 (severe)                                                                                                                                                           | Based on number of inflammatory and noninflammatory lesions                                                                                                         |
| Cook's acne grading scale method           | 0 (least severe) to 8 (most severe)                                                                                                                                                 | Uses photography to document acne severity                                                                                                                          |
| Global acne grading system                 | Grade: 0 (no lesions) to 4 ( $\geq 1$ nodule)<br>Global score: 0 (no acne) to $\geq 39$ (very severe acne)                                                                          | Takes into account location of lesions, and each location is given a factor rate. Factor rate $\times$ grade = local score. The Global score = sum of local scores. |
| Grading scale for overall disease severity | 0 (clear, no inflammatory lesions) to 6 (severe, numerous comedos, papules, and pustules, with larger inflamed lesions extending over much of the face, erythema may be pronounced) | Based on descriptive definitions of disease severity. Investigator selects the number that most closely identifies the extent of the subject's acne.                |

as standardized assessment methods to assure the reliability of clinical trials. The use of grading scales offers better consistency in reporting subject responses and allows for direct comparisons among clinical sites and studies. Global assessments, severity scores, and/or other more specific assessments are used in dermatology clinical trials. In general, global assessments are conducted by the study investigator at regular intervals to evaluate the response of total body disease. Severity scales are used to monitor changes in disease states such as pigmentation, pruritus, and induration that correlate with certain levels of disease involvement. Dermatology scales and scoring systems frequently incorporate body surface area (BSA) calculations in order to provide more accurate descriptions of the extent of disease.

The disease under study dictates the use of a particular scoring system. For example, there is the Psoriasis Area and Severity Index (PASI) for psoriasis and the Scoring for Atopic Dermatitis (SCORAD) for atopic dermatitis. These disease-specific scoring instruments will be considered in detail in later sections. Because many of the scoring systems are subjective, the most reliable way to maintain consistency in disease measurement is to adequately educate investigators and study staff in reporting guidelines and to ensure that within each research site, whenever possible, the same investigator conducts subject assessments throughout the trial. Keeping the number of evaluators to a minimum increases the integrity of reporting due to less interassessor variability.

In addition to disease-scoring systems, dermatology clinical trials often monitor patient perceptions of improvement, patient global disease assessments, and quality of life data [18]. These instruments should be administered at baseline and at the end of study or at another study time point. The administration schedule of these

instruments and the determination of which measurements are used should be clearly indicated in the protocol.

Instruments that assess subject perception of disease vary in the types of questions asked since the data collected depends upon which disease is under study. Generally, questions ask subjects to grade the severity of their disease and rate the level of disease-associated symptoms they have experienced. However, to measure how the life activities of subjects are affected by skin disease across disease states, the majority of therapeutic trials in dermatology use the Dermatology Life Quality Index (DLQI) [19]. The DLQI is not disease specific, and subjects answer by choosing “very much,” “a lot,” “a little,” “not at all,” or “not relevant” in response to 10 questions. Topics on the questionnaire range from how skin disease affects work and extracurricular activities to interpersonal relationships and sexual activity. For pediatric trials that enroll subjects between the ages of 5 and 16 years, there is the Children’s Dermatology Life Quality Index (CDLQI), which includes questions regarding how skin disease affects friendships, school work, sports, and other pediatric life issues.

Although not as widely used, the Skindex-29 is another measurement of skin disease’s impact on the quality of life. Skindex-29 is a shorter version of the original 61-question Skindex quality of life questionnaire. The newer version has been shown to be easier and faster to complete and a more effective measure of patients’ quality of life than its earlier form. It is composed of 30 questions that address three subscales: emotion, function, and symptoms. Patients answer questions on a scale of “0 = Never” to “4 = All the time,” and the higher the score, the greater impact the disease has on quality of life.

Typically, these quality of life measurements are administered at baseline and then again after an interval of treatment in order to assess whether a given therapy has improved subjects’ quality of life (Table 4). Given that dermatological diseases are visible, subjects often have stress related to how people respond to them and interact with them. This can have a major affect on how patients feels about themselves and their disease. Consequently, the quality of life measurements are extremely useful methods for predicting how successful a particular treatment is at ameliorating the psychosocial aspects of having skin disease [19].

### 10.3.2.7 Source Documentation in Dermatology

The importance of meticulous data collection in clinical trials cannot be over-emphasized. The design of source documents that capture all necessary information for a given trial is vital to the successful execution of a clinical study. This section explains several types of source documents that are useful in conducting dermatology clinical trials. Given the visual characteristic of the specialty and the limited availability of objective testing methods for monitoring disease, dermatology clinical trials typically rely on diagrams, photography, and the use of target area measurements in documenting subject response to treatment.

Target area measurements are typically used to monitor the response of one area of the skin. At baseline, the investigator determines a target area and measures its dimensions using a ruler. The dimensions of the target area are recorded at baseline and at subsequent study visits as a way of monitoring the skin’s response to therapy. It is important to designate whether the greatest dimensions should be measured

**TABLE 4 Quality of Life Metrics Used in Dermatology Clinical Trials**

|                                                                                  |                                                                                                                                                                      |
|----------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Facit fatigue scale                                                              | Measures how skin disease affects a subject's energy level and ability to complete daily activities.                                                                 |
| Beck depression inventory                                                        | Measures how skin disease affects a subject's appetite, appearance, sexual desire, health, and weight.                                                               |
| EuroQoL-5D (EQ-5D)                                                               | Measures how skin disease affects a subject's mobility, self-care, usual activities, level of pain/discomfort, and anxiety/depression.                               |
| EuroQoL-Visual analog scale                                                      | Feeling "thermometer" where patients rate subjective health status.                                                                                                  |
| Dermatology life quality index (DLQI)                                            | Measures how skin disease affects interpersonal relationships, sexual activity, work, and extracurricular activities.                                                |
| Children's dermatology life quality index (CDLQI)                                | Measures how skin disease affects friendships, school work, participation in sports, and other pediatric life issues.                                                |
| Psoriasis disability index                                                       | Measure how psoriasis affects the quality of life, level of discomfort/disability and associated stress of subjects with this skin condition.                        |
| Psoriasis index of quality of life (PSORIQoL)                                    |                                                                                                                                                                      |
| Psoriasis life stress inventory                                                  |                                                                                                                                                                      |
| Skindex-29                                                                       | Measures the effect of skin disease on subjects' emotions, symptoms, and day-to-day functioning.                                                                     |
| Quality of life index for atopic dermatitis (QoLIAD)                             | Measures the effect of atopic dermatitis on subjects' work and school activities, social functioning, and self-esteem.                                               |
| Melasma quality of life scale (MELASQOL)                                         | Measures the effect of melasma on subjects' daily functioning and level of disease-associated impairment.                                                            |
| Dermatology-specific quality of life instrument for contact dermatitis (DSDL-CD) | Measures how contact dermatitis affects subjects' level of discomfort, interpersonal functioning, self-care activities, work or school performance, and self-esteem. |
| Dermatology-specific quality of life instrument for acne DSQL-acne               | Measures how acne vulgaris affects subjects' level of discomfort, interpersonal functioning, self-care activities, work or school performance, and self-esteem.      |

as lesions may not be uniformly shaped. Monitoring target areas offers the added benefit of assessing other changes in the skin, namely adverse reactions to therapy. Any changes seen in the skin in the target area may be documented by the investigator in writing and followed over the course of a trial.

A major limitation of using target lesions when other areas of the body are also being treated is that they may not capture what is happening clinically elsewhere for a patient. Some skin diseases respond differentially to treatment based on body region. For example, psoriasis on the legs is often substantially harder to treat than other body regions and may lag in terms of improvement.

Body diagrams are a useful method for capturing the location of disease involvement. Investigators illustrate affected skin by using a writing instrument to shade in areas of disease involvement on a blank body diagram. By completing new front and back body diagrams at baseline and at regular intervals throughout a study, investigators may monitor a subject's overall response to disease and have a better clinical picture as to how different body areas respond to a particular treatment.

Photography plays a similar role in monitoring treatment response but warrants additional considerations regarding consistency and subject confidentiality. Study



protocols that incorporate photography should include specific guidelines explaining exactly how photographs should be taken for the trial. An outline of photography procedures should contain parameters to be used for all subjects at all study sites. For example, to maintain consistency the same backdrop should be used for all photos, a color chart and patient identification card should be in the photograph's field of view, the same film, camera, and camera focal length should be used for all photos, the same number of photographs should be taken at the same study intervals, and all subjects should be dressed and positioned for photographs similarly. In addition, appropriate measures must be taken to protect subject confidentiality. Photography of subjects' faces should be avoided unless this area is being monitored for study treatment response. Only designated study staff and the study sponsor (when applicable) should have access to the photographs. In cases where a subject's face is photographed, no subject photos should be used in publications prior to the removal of identifying characteristics, for example, the blacking out of a subject's eyes. Any use of photography in a clinical trial must be included in the study's informed consent form, as it is never appropriate to take photos of study subjects without their understanding and written permission.

### 10.3.2.8 Clinical Trial Registration

In January 2005, the International Committee of Medical Journal Editors (ICMJE) issued a statement requiring all phase II, III, and IV clinical trials to be registered in order to be considered for publication in any of their 11 member journals [20, 21]. This directly affects the dermatology clinical trials process since the major dermatology journals are associated with the ICMJE.

The ICMJE provided requirements for the registry database, including that it must offer free public access, must be supervised by a not-for-profit organization, have the ability to be searched electronically, and have a mechanism to confirm registration data legitimacy. The purpose for establishing such a registry is to prevent the discerning coverage of clinical trial results in medical journals [20]. In the past 50 years, nearly half of all clinical studies have not been published, leading to incomplete knowledge about therapeutic effectiveness [21]. Some of this stems from the preference of medical journals to publish positive results, which demonstrate the effectiveness of a new treatment, or those that show noninferiority, with two treatments showing equivalent efficacy. This bias consequently has led to the underpublication of negative or inconclusive findings and to the nondissemination of these study conclusions. The concern is that this underreporting leads to incomplete understanding of therapeutic efficacy, potential duplication of research efforts, and reduced integrity in study design and methodology. The latter issue is of considerable importance since it is paramount that investigators maintain the original study question and aims throughout a clinical trial. Registration would ensure increased veracity by precluding the potential for outcomes to be altered as data are interpreted.

### 10.3.3 SUMMARY

Collectively, the prevalence of skin problems surpasses other medical conditions such as hypertension, obesity, and cancer [22, 23] and has profound effects on health

and quality of life. What makes the skin such a unique and interesting organ to study in clinical trials is that the endpoints are often easily and quickly discernable and that the therapeutic modalities are so varied, ranging from topicals to immunotherapy to devices. The overall relative health of many skin disease patients also reduces the chances of confounding illnesses affecting the safety or interpretation of events in studies. In general, the field has also made tremendous progress in terms of methodology and scoring systems, and advances in imaging techniques will likely improve our ability to detect and measure change in the future. Lastly, while many studies, of course, do not succeed, we have seen tremendous successes in the past decade, with subjects experiencing dramatic improvements in both their health and quality of life, an experience that is greatly gratifying for investigators and subjects alike.

## REFERENCES

1. Stuttgen, G. (1996), Historical observations. *Dermatology, Clin. Dermatol.*, 14(2), 135–142.
2. Baer, R. L. (1994), Historical overview of the evolution of investigative dermatology (1775–1993), *J. Invest. Dermatol.*, 103(1), 3–6.
3. Holubar, K., and Wolff, K. (1989), The genesis of American investigative dermatology from its roots in Europe, *J. Invest. Dermatol.*, 92(4 Suppl), 14S–21S.
4. Potter, B. S. (2003), Bibliographic landmarks in the history of dermatology, *J. Am. Acad. Dermatol.*, 48(6), 919–932.
5. Delamere, F. M., and Williams, H. C. (2001), How can hand searching the dermatological literature benefit people with skin problems? *Arch. Dermatol.*, 137(3), 332–335.
6. Williams, H. (2001), Dowling oration 2001. Evidence-based dermatology—a bridge too far? *Clin. Exp. Dermatol.*, 26(8), 714–724.
7. Ben-Gashir, M. A., Seed, P. T., and Hay, R. J. (2004), Quality of life and disease severity are correlated in children with atopic dermatitis, *Br. J. Dermatol.*, 150(2), 284–290.
8. Weiss, J., Shavin, J., and Davis, M. W. (2003), Improving patient satisfaction and acne severity in patients with mild to moderate acne: The BEST study, *Cutis.*, 71(2 Suppl), 3–4.
9. Katz, S. I. (2001), Dermatological research in the 21st century: Our fantastic future, *J. Dermatol.*, 28(11), 599–601.
10. Witkowski, J. A., and Parish, L. C. (2001), Reflections on dermatology and projections for the 21st century, *Clin Dermatol.*, 19(1), 31–34.
11. Bhardwaj, S. S., Camacho, F., Derrow, A., Fleischer, A. B., Jr., and Feldman, S. R. (2004), Statistical significance and clinical relevance: The importance of power in clinical trials in dermatology, *Arch. Dermatol.*, 140(12), 1520–1523.
12. Lachin, J. M. (1981), Introduction to sample size determination and power analysis for clinical trials, *Control. Clin. Trials*, 2(2), 93–113.
13. Sheps, S. (1993), Sample size and power, *J. Invest. Surg.*, 6(6), 469–475.
14. Carroll, C. L., Feldman, S. R., Camacho, F. T., and Balkrishnan, R. (2004), Better medication adherence results in greater improvement in severity of psoriasis, *Br. J. Dermatol.*, 151(4), 895–897.
15. Zaghoul, S. S., Cunliffe, W. J., and Goodfield, M. J. (2005), Objective assessment of compliance with treatments in acne, *Br. J. Dermatol.*, 152(5), 1015–1021.

16. Zaghoul, S. S., and Goodfield, M. J. (2004), Objective assessment of compliance with psoriasis treatment, *Arch. Dermatol.*, 140(4), 408–414.
17. Gelmetti, C., and Colonna, C. (2004), The value of SCORAD and beyond. Towards a standardized evaluation of severity? *Allergy*, 59(Suppl 78), 61–65.
18. Feldman, S. R., Fleischer, A. B., Jr., Reboussin, D. M., Rapp, S. R., Exum, M. L., Clark, A. R., and Nurre, L. (1996), The self-administered psoriasis area and severity index is valid and reliable, *J. Invest. Dermatol.*, 106(1), 183–186.
19. Lewis, V., and Finlay, A. Y. (2004), 10 years experience of the dermatology life quality index (DLQI), *J. Investig. Dermatol. Symp. Proc.*, 9(2), 169–180.
20. Deangelis, C. D., Drazen, J. M., Frizelle, F. A., DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J., Schroeder, T. V., Sox, H. C., and Van Der Weyden, M. B. (2005), Clinical trial registration: A statement from the international committee of medical journal editors, *Arch. Dermatol.*, 141(1), 76–77; discussion 75.
21. Kimball, A. B., and Weinstock, M. A. (2005), Mandatory registration of clinical trials: A major step forward for evidence-based medicine, *J. Am. Acad. Dermatol.*, 52(5), 890–892.
22. Johnson, M. L. (2004), Defining the burden of skin disease in the united states—a historical perspective, *J. Investig. Dermatol. Symp. Proc.*, 9(2), 108–110.
23. The Lewin Group (2005), *The Burden of Skin Diseases*, Lewin Group, Falls Church, VA.

# 10.4

---

## *Emergency Clinical Trials*

**Joaquín Borrás-Blasco,<sup>1</sup> Andrés Navarro-Ruiz,<sup>2</sup> and Consuelo Borrás<sup>3</sup>**

<sup>1</sup>*Pharmacy Service, Hospital de Sagunto, Sagunto, Spain*

<sup>2</sup>*Pharmacy Service, Hospital General Universitario de Elche, Elche, Spain*

<sup>3</sup>*Department of Physiology, University of Valencia, Valencia, Spain*

### **Contents**

|          |                                                     |     |
|----------|-----------------------------------------------------|-----|
| 10.4.1   | General Information on Clinical Trials              | 477 |
| 10.4.2   | Clinical Trials Involving Emergency Patients        | 479 |
| 10.4.2.1 | Subject Recruitment                                 | 480 |
| 10.4.2.2 | Emergency Clinical Research in Pediatric Population | 481 |
| 10.4.3   | Ethical Issues in Emergency Clinical Trials         | 483 |
| 10.4.3.1 | Informed Consent                                    | 484 |
| 10.4.4   | Types of Clinical Trials in Emergency Departments   | 487 |
| 10.4.4.1 | Cardiovascular Alterations                          | 487 |
| 10.4.4.2 | Central Nervous System                              | 491 |
| 10.4.4.3 | Sepsis                                              | 492 |
| 10.4.5   | Conclusions                                         | 494 |
|          | References                                          | 494 |

### **10.4.1 GENERAL INFORMATION ON EMERGENCY CLINICAL TRIALS**

Emergency medicine is a global discipline that functions as a cornerstone for secondary disease prevention; therefore it is one of the many tools for implementing primary disease prevention programs [1].

In general, clinical trials involve the administration of the study drug to healthy human volunteers or to patients under the supervision of a qualified investigator, usually a physician, pursuant to a reviewed protocol. Human clinical trials are typically conducted under protocols that detail the objectives of the study, as well as

the parameters used to monitor the safety and efficacy criteria that are going to be evaluated. Moreover, it is very important to recognize that each clinical trial must be conducted under the auspices of an institutional review board (IRB) that considers, among other factors, ethical factors, the safety of human subjects, the possible liability of the institution, and the informed consent disclosure that must be obtained from participants of the clinical trial. Although the goal of this is to obtain safety and efficacy data, the overriding consideration in the studies is the safety of the study participants.

Research in the specialty of emergency medicine continues to blossom because this department is often considered as a desirable place to conduct clinical research due to the broad and undifferentiated spectrum of acute conditions encountered [2]. For this reason, there are some specific aspects when studying emergency medicine that have to be taken into account. One of them is the emphasis on ultra-rapid diagnosis and treatment to save the critically ill emergency patient [3]. Moreover, it is necessary to evaluate the use of a particular treatment; clinicians have to establish the parameters to be evaluated at the onset. These parameters include objective response rate, survival, disease-free survival, or duration of response. Toxicity is also commonly identified. The parameters of interest for a specific trial design are defined before the initiation of the study and analyzed at the completion.

Another characteristic of emergency clinical trials is that usually physicians working in the emergency department do not stay for a long period of time, and therefore it is especially critical for them to standardize the protocol to be used, as well as the ease of using it, for the physicians who are incorporated in the supervision of the trial [4].

Human clinical trials typically are conducted in three sequential phases, although sometimes they are overlapped. Phase I clinical trials represent the initial administration of the investigational drug to a small group of healthy human subjects or, more rarely, to a group of selected patients with the targeted disease or disorder [5]. Normally, this type of clinical trial is not developed in the emergency setting. Phase II clinical trials involve a small sample of the actual intended patient population and seek to assess the efficacy of the drug for specific targeted indications, to determine dose-response and the optimal dose range, and to gather additional information relating to safety and potential adverse effects. In the emergency room, these types of clinical trials involve mainly cardiovascular diseases such as acute ischemic stroke, acute stroke, cardiopulmonary resuscitation, acute myocardial infarction, and refractory angina patients with left ventricular dysfunction [6–9].

Once an investigational drug is found to have some efficacy and an acceptable safety profile in the targeted patient population, phase III clinical trials are initiated to establish further clinical safety and efficacy of the investigational drug in a broader sample of the general patient population. Phase III trials are randomized comparisons of two or more treatment options, often comparing a standard treatment to a new or more complex therapy. Emergency research phase III trials are related to several clinical conditions such as myocardial infarction, acute coronary syndromes, surgery, chronic obstructive pulmonary disease, hemorrhages, surgical pain, sepsis, and traumatic brain injuries [10, 11].

Phase IV trials are generally considered as “postregistration” trials, that is, trials of products that have already a marketing authorization.

It is also very important for the development of clinical research in emergency medicine to design and draw anticipated conclusions in order to note whether the trial is prospectively randomized with concurrent controls or is a clinical trial with historical controls. Proponents of randomized trials believe that one is more certain of equality between the two groups by a concurrent randomization process. This will reduce the bias of selecting control subjects from a historical pool and will also reduce the influence of improvements in management or different physicians who treat the patients in different periods of time. Furthermore, due to the specific characteristics of these clinical trials in emergency medicine, all of them require a special review and approval of trial proposals by an ethics committee.

In conclusion, there are some specific aspects that have to be taken into account when conducting clinical trials in emergency medicine due to the special conditions mentioned above.

#### 10.4.2 CLINICAL TRIALS INVOLVING EMERGENCY PATIENTS

Clinical trials involving emergency patients should be prospective trials, designed to evaluate emergency care treatments, answer specific questions, and more likely to answer the question correctly. In emergency medicine research it is mandatory that the prospective IRB reviews and approves the protocol designed for the study before it starts.

There are numerous potential barriers to conduct randomized controlled trials with emergency patients because of the specific characteristics mentioned below [12, 13]. Low incidence rates of emergency patient events require pooling of centers to conduct research, that is, in many occasions the trials have to be multicentric depending on the pathology investigated. A large number of emergency patients are required to attain diverse and representative study samples. In order for the precise development of a clinical trial, an infrastructure is needed to test the efficacy of treatments, as well as the transport and care that precede the arrival of emergency patients to the hospital emergency departments. A specific mechanism is needed to study the process of transferring research results to treatment settings.

A related challenge has been the difficulty for even large multicenter networks to gather sufficiently large sample sizes necessary to study treatments in the emergency medicine. Therefore, clinical trials in the emergency department need an infrastructure capable of overcoming inherent barriers to emergency medicine research. This infrastructure provides the capacity to conduct multicenter research studies to support research collaboration among medical investigators [14]. It should provide the test ability to collect, transfer, manage data from all sites, and determine availability, completeness, and agreement of core data from electronic and chart review. Despite these well-designed infrastructure multiinstitutional prospective trials, emergency medicine is complex due to:

- Training for investigators and staff is more difficult than those from single-center studies. This is especially critical for emergency departments where the physicians do not usually work for long periods.

- In the setting of emergency situations, randomization is not a simple process. Since researchers do not know when, and may not know where, the emergency is going to occur; it may be technically difficult to effect a randomization [15].
- Protocol development much more explicit because of the high rate of mobility of the physicians in this department.
- Quality control for maintaining high standards of quality and data management.
- Uniform standards for clinical research. Strong emphasis on training and standardization and include “good clinical practice” training.
- Site monitoring requirements of all centers regardless of whether observational or interventional study.
- Data transmission and security. It is necessary to involve epidemiologists and statisticians at the very beginning of the study. It is also essential to rely on independent data center/data manager. Paper data forms should be simple electronic, Web-based data entry of de-identified data, with logic and range checks. Original data forms/source documents at local site. Double data entry should be considered as a quality check—controversial queries for missing or inconsistent data.
- Multicenter studies are becoming more common, and the variability in local IRB assessments can be problematic if IRBs have different standards.
- Costs are high, and budgets tend to underfund training, site monitoring.
- Keep study simple, focused, and well budgeted.
- Accuracy measurement of predictor and outcome variable.

#### 10.4.2.1 Subject Recruitment

There is a paucity of clinical trials in emergency patients. This is partly due to difficulties in recruiting participants. Insufficient enrollment of patients is the most common reason for discontinuing emergency medicine studies. Failure to recruit members of specific populations to trials that require adequate representation reduces the external validity of the study results [16]. Frequently, the origin of these problems lies in the ethical issues surrounding clinical trials in this population [17]. In fact, recruitment is one of the most significant challenges to completing a research study on time and on budget. In most trials in emergency research, recruitment takes longer, is more expensive, and produces fewer subjects than planned. Subject recruitment can be difficult in the emergency department environment mainly because when patients present with acute symptoms, the emergency physicians are focused on life-saving interventions, not on study enrollment [18]. Though causes of recruitment difficulty have been documented in medical trials, little is known concerning recruitment in emergency medicine studies. Many reasons for this poor recruitment rate for emergency studies include:

- The subject recruitment should be made following established guidelines for protection of participant’s rights. The IRB now has an obligation to consider a strict inclusion and exclusion criteria document about subject recruitment.



- In life-threatening situations, relatives present many inconveniences for their relative to participate in a clinical study.
- Limited size of the emergency population, specially patients in critical situations.
- Plans for sampling and recruitment.
- The increased bureaucracy, reporting, and costs of clinical trials after the implementation of the European Clinical Trials Directive and U.S. Food Drug Administration (FDA) guidance for emergency research has led to a sharp decline in the initiation of clinical trials and in participant recruitment throughout Europe and North America [19].
- Regulations for clinical trials: Researchers have always been concerned about the effects that the regulations might have on clinical trials in emergency situations in patients with impaired consciousness. Examples include research on cardiac arrest, stroke, head trauma, spinal cord injury, gunshot wounds, major trauma, cardiac arrest, and poisoning. These regulations impose the need for prior consent from a personal or professional legal representative before a patient could be recruited into a clinical trial [20]. In most cases the regulations are only applicable to trials of drugs. Nonmedicinal trials and trials of clinical care continue to be governed by common law [21]. Federal regulations allow for waiver of consent in situations where there is more than minimal risk to participants, provided there is a prospect of direct benefit to participants and a number of other conditions are met. The conditions include the following:
  - The study could not be carried out without the waiver.
  - Consultation with community representatives occurs before the study starts.
  - Public disclosure is made before and after the study.
  - A therapeutic window is defined, and the researcher commits to try to locate a surrogate/legally authorized representative who can give consent within that window before proceeding to waive consent [22].

Recently, an amendment of the United Kingdom's Medicines for Human Use (Clinical Trials) Regulations allows unconscious patients in emergency situations to be enrolled in clinical trials without prior consent provided it has been approved by the appropriate ethics committee [23].

These difficulties could involve inadequate recruitment, which could reduce the ability to detect treatment differences.

In conclusion, strategies for the recruitment of the patients susceptible to be included in an emergency clinical trial are fundamental. Successful recruitment depends on multiple factors such as population, potential subjects, pathology, available alternatives, transportation availability, population demographics, and general or specific interest in medical research or science within the community.

#### **10.4.2.2 Emergency Clinical Research in Pediatric Population**

The well-known difficulties of conducting studies in young patients, along with the limited economic returns to pharmaceutical companies for pediatric drugs, have led

to a scarcity of pediatric studies and therefore a scarcity of knowledge about drug safety and efficacy in children [24]. The lack of scientifically evaluated medicines for children has been recognized as an area that requires correction [25]. Moreover, clinical trials in children have special ethical constraints [26].

Experimentation on children is considered by many to be unethical, resulting in difficulties to obtain critical safety data. Clinical trials are subject to detailed scrutiny by the various regulatory bodies. They have recently recognized the need for pharmaceutical companies to invest in pediatric medicines. The majority of marketed drugs are either not labeled, or inadequately labeled, for use in pediatric patients [27]. This is because, as we have stated before, there is a paucity of clinical trial works in children, which leads to the frequent use of off-label and unlicensed medications in this very vulnerable group [28].

The pediatric population is constituted by a wide range of individuals of substantially varied physical size, weight, and stage of physiological development. The differences between adults and children mandate strategies to improve this situation rather than continually relying on extrapolation from adult studies. In the protocols for clinical trials for pediatric emergency room it is mandatory that:

- Each trial should answer one primary study question.
- Sample size estimation based on error rates and clinically important difference must be calculated.
- Allocation concealment, randomization, and blinding are critical to prevent selection and ascertainment bias [29]. Clinical trial design must be practical, and a trial is more likely to succeed if a simple design is utilized, with minimal interference with school and work.
- Standardization of the intervention is important for the results validity, and study centers should be selected on the basis of the trial objective. The effects of dropouts, cross-overs, and missing data should be handled with an intention-to-treat analysis.
- Sample selection, informed consent, and valid outcome measures are specific challenges in the pediatric population [29].

Efforts should be made to increase the number of well-designed, randomized controlled trials in pediatric emergency patients. Therefore, new strategies have to be established between the pharmaceutical industry, the European Union (EU), and the U.S. regulatory bodies and pediatric centers to facilitate effective trials work. In this regard, The EU decided to support the development of a European register of clinical trials in children as part of the Fifth Framework Thematic Programme “Quality of Life” in 2002. DEC-NET (the European register of clinical trials on medicines for children—drug evaluation in children), supported by the EU, was created to provide the scientific community with a flexible tool for promoting communication and collaboration among researchers, disseminating clinical trial results, and facilitating patient access and recruitment to trials [30]. The project DEC-NET currently involves members of four countries: France, Italy, Spain, and the United Kingdom. DEC-NET fits International Committee of Medical Journal Editors criteria, is free of charge, and is designed to be used by the general public and health professionals [31]. It is unique in that it is the first population-oriented

clinical trial register. Such a register represents a useful source for planning new studies, promoting communications and collaborations between researchers, facilitating patient access and recruitment into trials, preventing trial duplication and inappropriate funding, and identifying the therapeutic needs of children that remain neglected. It will also allow active monitoring of new or evolved knowledge of drug therapies [32].

### 10.4.3 ETHICAL ISSUES IN EMERGENCY CLINICAL TRIALS

The ethical aspects of an emergency clinical trial cannot be separated from the scientific objectives. Segregation of ethical issues from the full range of study design components demonstrates a flaw in understanding the fundamental nature of research involving human subjects. Recognition that good science and good ethics are inextricably bound together in clinical research is today's reality. Thus, the ethical conduct of a clinical trial does not end with the formulation of study design or the obtainment of a signature on the informed consent form [33].

To improve the study design and conciliate science and ethics, some authors recommend that every protocol include a specific section in which the ethical considerations are exposed: the design and conduct of the study, selection and recruitment of participants, and the conduct of the consent process be presented, discussed, and justified, not in terms of ethics-based regulatory requirements but in true fulfillment of the investigator's responsibilities (as a scientist and moral duty) to ensure that the proposed study is sound, ethically and scientifically [34].

Patients attending an emergency department are entitled to assume that they will be treated in accordance with the best available evidence. The emergency physician has a responsibility to expand the fund of knowledge that underpins best evidence by means of clinical studies of new technologies and interventions. However, there are inherent difficulties in undertaking clinical research in an emergency department setting. Patients are usually frightened, often in pain, and sometimes have impaired cognition. Practitioners use case-based reasoning to apply bioethics to these clinical situations, usually giving most weight to patients' autonomy and values but also incorporating other relevant bioethical principles, including those encompassed in professional oaths and codes. Emergency clinicians must be able to recognize bioethical dilemmas, have action plans based on their readings and discussions, and have a method through which to apply ethical principles in clinical settings [35].

Other aspects that can help for the ethical consideration of the emergency clinical trials could be the randomization to treatment, which must occasionally be completed in the context of a clinical emergency. Furthermore, patients take on trust the fact that the clinician will act in a professional manner and not misuse the authority of his or her position. The researchers must be committed to the principle that self-promotion must not take precedence over the best interests of the patient.

The moral obligation of emergency researchers to do well, beneficence, must be tempered by nonmaleficence. That is, the expected bad effects of an intervention must be weighed against the anticipated benefits. The type, amount, and probability of harm should be effectively communicated to subjects, and the physician should

take account of the fact that what constitutes a harmful effect, or a beneficial effect, may be particular to the individual subject. Justice and equality of access are inalienable rights that must be respected in participants and nonparticipants alike [35]. The researcher must be a good communicator, demonstrating openness and effective listening. Benefits may flow from a research study in favor of the patient or to society. When discussing the benefits of a proposed study, one must distinguish between therapeutic and nontherapeutic research. Researchers are obliged to consider the potential benefit to the patient when balancing potential risks associated with any protocol. Researchers are entitled to expect that their work is regarded as honest until shown to be otherwise.

Although it is uncommon, misconduct occurs in most countries. Several causes of misconduct have been described, including pressure to publish, financial inducements, and simple vanity. Misconduct in researchers must not be tolerated and must be reported promptly. There should be a named person within the organization who is responsible for the receipt of complaints of the misconducted research [36].

In order to ensure the good practices for emergency clinical trials, an IRB was established. The IRB was created to maintain ethical standards of practice in research, to protect subjects of research from harm, to preserve the subjects' rights and to provide reassurance to the public that this is being done. It is not intended to impede good research but facilitate it, approving studies of good quality, not badly planned or poorly designed protocols. The IRB has numerous protection responsibilities that include initial and continuing review of the study protocol and related documents, review of the documentation of informed consent, and review of reports of unanticipated problems and of adverse events [37].

In emergency clinical research the IRB will make an assessment of the risks and benefits represented by the proposal, taking into account the likelihood and severity of any risk to subjects and patients and weighing this against the potential for benefit. IRBs should ensure that a monitoring plan exists for each individual study site and that there is a data-monitoring program for multisite studies. Second, the IRB should certify investigators' understanding of compliance with regulations governing the subjects' safety during the trial, including adverse reaction reporting and the collection of relevant study documents. Finally, the IRB should review the safety monitoring board reports and query investigators as needed to determine whether additional safeguards are necessary [38]. These recommendations seem reasonable because the structure and function of IRBs would enable them to perform these roles better than the other entities involved in the monitoring of clinical trials. In addition, the suggested roles would add an efficient framework to the role that IRBs can play in the safety monitoring of human subjects [39].

#### **10.4.3.1 Informed Consent**

"Informed consent" means the knowledgeable consent of an individual, or his or her legally authorized representative, who is able to exercise free power of choice without undue inducement or any form of force, fraud, deceit, duress, or other form of constraint or coercion. This document will be used with subjects or their legally

authorized representatives in feasible situations. The information in the consent document will be also used when providing an opportunity for a family member to object to a subject's participation in the clinical investigation.

One of the most important ethical dilemmas in emergency department research is the difficulty of obtaining genuine informed consent from research subjects [40]. Patients with life and death emergencies such as those with head injuries and impaired consciousness are unable to give informed consent to participate in clinical trials. Nevertheless, controlled clinical trials are essential in such situations to identify effective ways to prevent death and long-term disability.

On the other hand, emergency department research on new treatments for cardiopulmonary arrest, neurologic emergencies, and major trauma is challenging, controversial, and necessary [41]. The critical nature of the illness requires immediate treatment; emergency medicine research often takes place when patients are in distress or even unconscious, making a thoughtful discussion of risks and benefits impossible. This would particularly affect vulnerable people exposed to acute, life-threatening conditions where available treatments are unproved or unsatisfactory, and the collection of valid scientific evidence, which may include evidence obtained through randomized, placebo-controlled investigations, is necessary to determine the safety and efficacy of particular interventions [42].

Randomized, controlled trials should be designated as minimal risk, allowing IRBs to approve their conduct with a waiver of informed consent if obtaining it is not feasible. These clinical situations are: the patients will not be able to give informed consent as a result of their medical condition; the intervention involved in the research must be administered before consent from the subjects' legally authorized representatives is feasible; or there is no reasonable way to identify prospectively the individuals likely to become eligible for participation in the research [43].

In these clinical trials the potential sources of risk that must be considered are physical risk from study treatments, the loss of individualized care, risk from non-therapeutic components of the research protocol, and the psychological impact of participation. The risks of research participation should be considered in comparison with the risk of nonparticipation; for example, the risks specific to research participation should be considered separately from the risks inherent to the treatment of the potential research participant's underlying condition. Therefore the risk–potential benefit balance for every arm of the study is at least as favorable as the risk–potential benefit balance for any therapy available outside of the research protocol. Moreover, participation in the study does not present added medical risk above the risk inherent to the research participant's underlying condition. The incremental risk of such a clinical trial would not exceed the threshold of minimal risk [44].

Participation in an randomized, controlled trial carried out without the waiver from informed consent requirements may pose no more than minimal risk when:

- Genuine clinical equipoise exists [45].
- All of the treatment options included in the research study fall within the current standard of care.
- There is no currently available treatment with a more favorable risk–benefit profile than the treatments included in the study.

- The nontherapeutic components of the research are safely under the minimal risk threshold.
- Research protocol provides sufficient latitude for treating physicians to individualize care when appropriate.

Moreover, it is very important to consider the potential negative psychological impact for research participants or their families in risk assessment [46]. However, a patient involved in a research protocol that has been reviewed for scientific and ethical quality is more protected than the one who receives an unproved treatment at the discretion of a provider who believes that it may help. Abboud et al. [47] reported that some patients in the emergency department would prefer to receive an unproved, potentially helpful drug if they were in cardiac arrest, neurologic emergencies, or major trauma, but a much smaller proportion would participate in a randomized trial comparing a new drug with a placebo. Patients were also less willing to participate in studies that evaluated highly invasive interventions [47].

In 1996, the FDA regulations were revised to include the report 21 CFR 50.24, which allows for an exception from informed consent requirements for a very limited class of research in emergency settings [19].

In this regard, European Union 2001/20/EC directive is a cornerstone of a Europe-wide harmonization of the provisions governing clinical trials and can be expected to foster and facilitate multinational clinical research. Article 5 outlines the conditions for research in incapacitated patients unable to give informed consent. The article, however, is framed to address the needs of individuals who are incapacitated for long periods, many even permanently [48].

A clinical trial can only be done if informed consent of the legal representative has been obtained. The directive may not only affect unconscious people. Patients with acute myocardial infarction have been enrolled in clinical trials so far. Many of these have severe pain on admission and receive treatment with opiates: Can they give informed consent, particularly those with cardiogenic shock. Research in the acute care setting is already difficult and this directive will make it even more difficult [49].

Both, the U.S. federal regulations of 1996 [19] and the 2001 European Union directive combined have resulted in an extremely limited number of emergency research studies that have been completed or are ongoing under an exception from informed consent [50].

On the other hand, poor readability of informed consent forms has been a persistent problem in clinical research. The low educational attainments of many patients with mental illness might suggest a still greater problem in emergency room settings. This problem is even worse in emergency department setting due to the stress of the patient, the severity of illness, or even the conscious level of the patient. This fact not only affects the patient but also his or her family. In most cases, a considerable level of education, consciousness, and calmness is necessary to understand the informed consent documents used to enroll subjects in emergency medicine research. Mader and Playe [51] reported that the length and complexity of the consent forms increased as risk to the subject increased. In conclusion, informed consent documents used in emergency medicine research may be too complex for the average patient to understand [51].

Hence, it is desirable to design a consent form to meet all of the regulatory requirements while maintaining a level of reading comprehension suitable for the general population. This is a difficult task for investigators, methods of reducing the complexity of forms, as part of improving the overall consent process, are much needed. IRBs are charged with safeguarding potential research subjects with limited literacy but may have an inadvertent role in promulgating unreadable consent forms [52].

#### 10.4.4 TYPES OF CLINICAL TRIALS IN EMERGENCY DEPARTMENTS

In this section, we describe the most common randomized clinical trials carried out in emergency medicine. Diseases with an acute process and high morbidity and mortality are prime candidates to develop research in the emergency department. Observational studies provide useful but potentially biased information about the effect of interventions. Because of their ability to deal with bias, randomized clinical trials are the optimum vehicle to obtain a true estimate of the effect of interventions, including emergency research.

We have selected those randomized clinical trials that include patients suffering from cardiovascular alterations or central nervous system affections conducted in the emergency department.

##### 10.4.4.1 Cardiovascular Alterations

**Atrial Fibrillation** Atrial fibrillation is the most prevalent arrhythmia in hospital emergency departments and is a serious disease associated with a high mortality rate. However, the management of atrial fibrillation in this scenario is variable and frequently inadequate. Some randomized controlled trial involved patients diagnosed as having atrial fibrillation in the emergency department, and their main objective was to compare drugs strategies.

Dermican et al. [53] compared the effectiveness of intravenous diltiazem and metoprolol in the management of rapid ventricular rate in atrial fibrillation. Forty patients met the inclusion criteria. All the enrolled patients received information about the study and signed written informed consent. The authors concluded that diltiazem and metoprolol were safe and effective for the management of rapid ventricular rate in atrial fibrillation. However, they concluded that diltiazem allowed patients to recover the rate of control effect earlier and that the percentage of ventricular rate decrease was higher than with metoprolol [53].

Another example was published by Davey and Teubner [54] who conducted a prospective, randomized, double-blind, placebo-controlled trial in an adult emergency department, enrolling patients suffering from rapid atrial fibrillation, to examine the safety and efficacy of magnesium sulfate infusion, in addition to usual care, for acute rate reduction in patients with atrial fibrillation, and a rapid ventricular response rate. In all 199 patients were randomized, 102 received magnesium sulfate and 97 received placebo. The trial results showed that the addition of magnesium sulfate to standard rate reduction therapies enhances rate reduction and conversion to sinus rhythm in patients with rapid atrial fibrillation [54].



Thomas et al. [55] carried out a study to assess the efficacy and safety of rapid high-dose intravenous infusions of amiodarone and sotalol for heart rate control and rapid reversion to sinus rhythm in patients who came to the emergency department with recent-onset symptomatic atrial fibrillation. One hundred and forty patients gave informed consent and were randomized. They received 1.5 mg/kg of sotalol infused in 10 minutes, 10 mg/kg of amiodarone in 30 minutes, or 500 µg of digoxin in 20 minutes. Electrical cardioversion was attempted for patients who did not convert to sinus rhythm within 12 hours. This clinical trial provided evidence that the rapid infusion of sotalol or amiodarone in patients with symptomatic recent-onset atrial fibrillation results in rapid control of ventricular rate. Even with high-dose rapid infusions, all 3 agents are associated with a poor overall reversion rate within 12 hours. Almost all patients were returned to sinus rhythm with the combination of pharmacological therapy and electrical cardioversion [55].

Kim et al. [56] performed a prospective, randomized pilot study of a traditional approach of hospital admission versus an accelerated emergency-department-based strategy with low-molecular-weight heparin and early cardioversion to sinus rhythm, in a cohort of patients with uncomplicated atrial fibrillation of a single university hospital. The primary endpoints were length of stay and total actual direct costs. Eighteen patients were randomized over a 15-month period. The new strategy resulted in a shorter length of stay at potentially lower cost [56].

This kind of assays has to be conducted in a very short period of time, therefore, it is extremely important to design them very accurately and control very sharply their follow-up in order to satisfy the proposed objectives. Especially, the protocol assay and patient recruitment and informed consent have to be strictly defined before starting the clinical trial.

**Myocardial Infarction** Advances in the treatment of myocardial infarction such as coronary care units, patient education, improved emergency response systems, and reperfusion therapy has reduced mortality. Thus, trials using mortality alone as an endpoint may no longer be feasible even in survivors of a heart attack, unless a very high risk group is studied. This has led to increased use of combination endpoints, such as cardiovascular mortality plus nonfatal myocardial infarction. Development of randomized clinical trials for management of ST-elevation myocardial infarction developed by a community-based emergency medical service, emergency department, or cardiovascular service is recommended.

Moreover, due to the involvement of numerous medical services in clinical trials, there are many types of clinical trials concerning different therapeutic approaches for the treatment of acute myocardial infarction [57]. To identify the characteristics of these assays, we have selected a study carried out to explore the effect of treating with glycoprotein IIb/IIIa receptor antagonist patients suffering from thrombolysis ST-segment elevation myocardial infarction, as an example of acute myocardial infarction clinical trials in the emergency environment design.

**Thrombolysis ST-Segment Elevation Myocardial Infarction** The importance of the dissolution and prevention of thrombosis in treating patients with ST-segment elevation myocardial infarction has motivated the development of novel therapies targeting platelet aggregation and thrombus formation [58].

Several randomized clinical trials have been developed in the emergency setting using glycoprotein IIb/IIIa receptor antagonists.

Abciximab is an antibody fragment that dose-dependently inhibits platelet aggregation and leukocyte adhesion by binding to the glycoprotein (GP) IIb/IIIa, vitronectin, and Mac-1 receptors [59]. The beneficial effects of using abciximab with stent placement in patients with acute MI compared to stent alone were demonstrated in the double-blind ADMIRAL study [60]. Three hundred patients with acute myocardial infarction were enrolled in a double-blind fashion either to abciximab plus stenting (149 patients) or placebo plus stenting (151 patients) before they underwent coronary angiography. The ethics review board of the Pitié-Salpêtrière Hospital approved the protocol, and the study was conducted in accordance with the Declaration of Helsinki. Signed informed consent was obtained from the patients. Clinical outcomes were evaluated 30 days and 6 months after the procedure. The angiographic patency of the infarct-related vessel and the left ventricular ejection fraction were evaluated at 24 hours and 6 months. The results of the trial reported that compared with placebo, early administration of abciximab in patients with acute myocardial infarction improves coronary patency before stenting, as well as the success rate of the stenting procedure, the rate of coronary patency at 6 months, left ventricular function, and clinical outcomes [60].

The large CADILLAC study randomly assigned 2082 patients with acute myocardial infarction to undergo PTCA alone (518 patients), PTCA plus abciximab therapy (528), stenting alone with the MultiLink stent (512), or stenting plus abciximab therapy (524), using a 2-by-2 factorial design [61]. The study was approved by the institutional review board or ethics committee at each participating center, and consecutive, eligible patients provided signed informed consent. In this study, both primary objectives were met; stenting alone was superior to percutaneous transluminal coronary angioplasty, and stenting alone was not inferior to percutaneous transluminal coronary angioplasty plus abciximab [61].

In this regard, a recent meta-analysis of eight randomized trials conducted on abciximab as adjunctive therapy to mechanical revascularization for ST-elevation myocardial infarction showed that abciximab, as adjunctive therapy to both primary angioplasty and fibrinolytic therapy for STEMI, is associated with a significant reduction in short-term reinfarction rate, whereas the benefits in reducing mortality are observed only in association with primary angioplasty [62]. Abciximab is not associated with an increased risk of intracranial hemorrhage (except in combination with fibrinolysis in elderly patients), whereas a higher risk of major bleeding complications is observed only in association with fibrinolysis [62].

Eptifibatide is an intravenously administered glycoprotein IIb/IIIa receptor antagonist that acts at the final step of the platelet aggregation pathway. Two clinical trials showed that concomitant administration of eptifibatide in patients undergoing elective PCI reduced thrombotic complications [63]. One of them was the IMPACT-II (Integrilin to Minimize Platelet Aggregation and Prevent Coronary Thrombosis II), a multicenter, double-blind, placebo-controlled trial, enrolling 4010 patients undergoing elective, urgent, or emergency coronary intervention. Patients were assigned one of three treatments: placebo ( $n = 1328$ ), a bolus of 135  $\mu\text{g/kg}$  eptifibatide followed by an infusion of 0.5  $\mu\text{g/kg/min}$  for 20–24 hours ( $n = 1349$ ), or 135  $\mu\text{g/kg}$  eptifibatide bolus with a 0.75- $\mu\text{g/kg/min}$  infusion ( $n = 1333$ ). All patients provided informed consent. Their results showed that the treatment strategy with eptifibatide

135/0.5 reduced rates of early abrupt closure and ischemic events at 30 days [63]. The other study was the ESPRIT (Enhanced Suppression of the Platelet IIb/IIIa Receptor with Integrilin Therapy), which is a randomized, placebo-controlled trial to assess whether a novel, double-bolus dose of eptifibatide could improve outcomes of patients undergoing coronary stenting. A total of 2064 patients undergoing stent implantation in a native coronary artery were enrolled. Immediately before percutaneous coronary intervention, patients were randomly allocated to receive eptifibatide, given as two 180 µg/kg boluses 10 minutes apart and a continuous infusion of 2.0 µg/kg/min for 18–24 hours, or placebo, in addition to aspirin, heparin, and a thienopyridine. The protocol was approved by each of the respective institutional review boards, and informed consent was obtained from all patients. The authors concluded that routine glycoprotein IIb/IIIa inhibitor pretreatment with eptifibatide substantially reduces ischemic complications in coronary stent intervention and is better than a strategy of reserving treatment to the bailout situation [64].

Moreover, the PURSUIT (Platelet Glycoprotein IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy) trial, which included 10,948 patients with non-ST-elevation acute coronary syndromes, tested the hypothesis that inhibition of platelet aggregation with eptifibatide would have an incremental benefit, beyond that of heparin and aspirin, in reducing the frequency of adverse outcomes. Patients with ischemic chest pain within the previous 24 hours and who had either electrocardiographic changes indicative of ischemia (but not persistent ST-segment elevation) or high serum concentrations of creatine kinase MB isoenzymes were enrolled in the study. They were randomly assigned, in a double-blind manner, to receive a bolus and infusion of either eptifibatide or placebo, in addition to standard therapy, for up to 72 hours (or up to 96 hours if coronary intervention was performed near the end of the 72-hour period). Because this was the first large-scale study using higher doses of eptifibatide than those previously used, it was specified in the protocol that the study would be stopped for the lower-dose group after the independent data safety and monitoring committee had conducted an interim review of safety data, provided the higher dose had an acceptable safety profile. After 3218 patients had been randomly assigned to treatment groups, the committee recommended dropping the lower dose. Eptifibatide significantly reduced the primary endpoint of death and nonfatal myocardial infarction at 30 days compared with placebo [65]. Studies are now evaluating eptifibatide in high-risk patients with non-ST-elevation acute coronary syndromes (NSTEMI-ACS) and a planned early invasive strategy in the EARLY-ACS (Eptifibatide Administration prior to Diagnostic Catheterization and Revascularization to Limit Myocardial Necrosis in Acute Coronary Syndrome) trial and in patients with primary PCI for STEMI in comparison to abciximab in the Eptifibatide versus Abciximab in Primary PCI for Acute Myocardial Infarction trial. After the completion of these trials, the value of eptifibatide in patients undergoing PCI in different indications could be determined based on the evidence thanks to the trials carried out [66].

Tirofiban is a small, synthetic nonpeptide, competitive GP IIb/IIIa antagonist with high specificity and high affinity for the GP IIb/IIIa receptor.

Multicenter phase III open-label, multinational investigator-driven clinical trial of single high-bolus dose tirofiban versus abciximab and sirolimus-eluting versus bare metal stent in acute myocardial infarction was performed. The objective of the study

was to evaluate, with a 2-by-2 factorial design, the safety/efficacy profile of 4 interventional strategies of reperfusion: tirofiban given at high-bolus dose (bolus of 25 µg/kg over 3 minutes), followed by an infusion of 0.15 µg/kg/min for 18–24 hours versus abciximab and sirolimus-eluting stent, as compared to bare-metal stent implantation in primary percutaneous coronary intervention. In a head-to-head comparison, tirofiban 10-µg/kg bolus followed by a 0.15-µg/kg/min infusion was found to be inferior to the standard dose of abciximab in patients undergoing percutaneous coronary intervention. Insufficient platelet inhibition with low-dose tirofiban may likely explain these results. Subsequently, a high-bolus dose of tirofiban followed by standard infusion was tested and evidence suggests that in this dosing tirofiban may be as effective as abciximab and have a comparable safety profile [67].

In conclusion, glycoprotein IIb/IIIa receptor antagonists inhibit the binding of ligands to activated platelet GP IIb/IIIa receptors and, therefore, prevent the formation of platelet thrombi. Glycoprotein IIb/IIIa inhibition to PCI for treatment of STEMI has substantially lowered the incidence of recurrent ischemic events and has improved early survival.

#### 10.4.4.2 Central Nervous System

Clinical research involving patients with neurologic emergencies has increased in the emergency department. We have highlighted randomized clinical trials studying therapeutic treatments for stroke.

**Stroke** The National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study was a randomized, double-blind trial of intravenous recombinant tissue plasminogen activator (t-PA) for ischemic stroke, when treatment was begun within 3 hours of the onset of stroke. The trial was divided in two parts: Part 1 (in which 291 patients were enrolled) tested whether t-PA had clinical activity, as indicated by an improvement of 4 points over baseline values in the score of the National Institutes of Health stroke scale (NIHSS) or the resolution of the neurological deficit within 24 hours of the onset of stroke. Part 2 (in which 333 patients were enrolled) used a global statistic test to assess clinical outcome at 3 months, according to scores on the Barthel index, modified Rankin scale, Glasgow outcome scale, and NIHSS. Informed consent was obtained for all patients. This study showed that treatment with intravenous t-PA within 3 hours of the onset of ischemic stroke improved clinical outcome at 3 months, despite an increased incidence of symptomatic intracerebral hemorrhage [68].

Randomized double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischemic stroke (ECASS II) has assessed the safety and efficacy of intravenous thrombolysis with alteplase (0.9 mg/kg body weight) within 6 hours of stroke onset. The trial protocol was reviewed and approved by local independent ethics committees or institutional review boards according to the regulatory requirements of the participating country, and carried out in accordance with the ethical principles of the Declaration of Helsinki. Informed consent was obtained from each patient (or from his or her legally authorized representative) before enrollment in the study. The results do not confirmed a statistical benefit for alteplase. The trend toward efficacy should be interpreted in the light of evidence from previous trials. Despite the increased risk of intracranial hemorrhage,

thrombolysis with alteplase at a dose of 0.9 mg/kg in selected patients may lead to a clinically relevant improvement in outcome [69].

The ATLANTIS study was a randomized controlled trial whose objective was to test the efficacy and safety of rt-PA in patients with acute ischemic stroke when administered between 3 and 5 hours after symptom onset. Its current approved use is limited to within 3 hours of symptom onset, which restricts a lot the number of patients who can be treated, since most stroke patients present more than 3 hours after symptom onset. This study found no significant rt-PA benefit on the 90-day efficacy endpoints in patients treated between 3 and 5 hours. All patients or their legal representatives signed an informed consent, which was previously approved by the correspondent institutional review board [70]. The risk of symptomatic ICH increased with rt-PA treatment. These results do not support the use of intravenous rt-PA for stroke treatment beyond 3 hours [70].

Hacke et al. [71] performed a meta-analysis of the recent large stroke trials, including NINDS [68], ECASS [69], and ATLANTIS [70] in order to draw conclusions about the best use of tissue plasminogen activator (t-PA) for ischemic stroke [71]. This study demonstrates that onset to treatment time is critically associated with improved functional outcome [71].

As opposed to the post hoc analysis of the NINDS trial, where an effect of time was not found in the prespecified analysis of the 90- and 180-minute groups, this larger meta-analysis sample nicely demonstrates a greater likelihood of good outcome, OR 2.8, for treatment in the 0- to 90-minute time frame versus OR of 1.6 and 1.4 for the 90- to 180- and 180- to 270-minute time frame. Interestingly, time from symptom onset did not influence the likelihood for hemorrhage conversion. The meta-analysis solidifies the rationale for the investment of significant public health effort in the development of emergency medicine service and hospital-based teams to treat stroke rapidly [72].

A study to investigate the feasibility and safety of a combined intravenous and intra-arterial approach to recanalization for ischemic stroke has been recently performed: 81 subjects with a baseline NIHSS  $\geq 10$  were enrolled. The median time to initiation of intravenous rt-PA was 142 minutes as compared with 108 minutes for placebo and 90 minutes for rt-PA-treated subjects in the NINDS rt-PA Stroke Trial ( $p < 0.0001$ ) [73]. The 3-month mortality in IMS II subjects was 16% as compared with the mortality of placebo (24%) and rt-PA-treated subjects (21%) in the NINDS rt-PA stroke trial. The rate of symptomatic intracerebral hemorrhage in IMS II subjects (9.9%) was not significantly different than that for rt-PA-treated subjects in the NINDS t-PA stroke trial (6.6%). IMS II subjects had significantly better outcomes at 3 months than NINDS placebo-treated subjects for all endpoints (OR  $\geq 2.7$ ) and better outcomes than NINDS rt-PA-treated subjects as measured by the Barthel index and global test statistic [73].

#### 10.4.4.3 Sepsis

Several studies have demonstrated the importance of early and adequate antimicrobial therapy in reducing the mortality and morbidity of patients with severe sepsis. About 6 to approximately 17% of empirical antibiotic selection were judged to be inappropriate according to subsequent microbiology and higher antimicrobial susceptibility as a result. It reflects the diversity in the presentations of infectious

diseases and the limited microbiological reports available from the first-line emergency physicians. Timely diagnosis and selection of appropriate antibiotics/treatment in the treatment of those patients is a challenge for an emergency physician more than ever before.

Rivers et al. [74] performed a prospective, randomized study to evaluate if the efficiency of early goal-directed therapy before admission to the intensive care unit effectively reduces the incidence of multiorgan dysfunction, mortality, and the use of health care resources among patients with severe sepsis or septic shock. This study was approved by the institutional review board for human research and was conducted under the auspices of an independent safety, efficacy, and data-monitoring committee: 263 patients were enrolled, 130 were randomly assigned to early goal-directed therapy and 133 to standard therapy. The result of the trial showed that hospital mortality was 30.5% in the group assigned to early goal-directed therapy, as compared with 46.5% mortality in the group assigned to standard therapy. Furthermore, mean APACHE II scores were significantly lower, indicating less severe organ dysfunction, in the patients assigned to early goal-directed therapy than in those assigned to standard therapy. The authors concluded that early goal-directed therapy provides significant benefits with respect to outcome in patients with severe sepsis and septic shock [74].

Another study tested the hypothesis that in the setting of undifferentiated symptomatic hypotension, the presence of hyperdynamic left ventricular function (LVF) on focused ED echocardiography would be a specific finding for sepsis as the etiology of shock [75]. This clinical trial was preplanned as a secondary analysis of 184 patients enrolled in a randomized clinical trial to investigate the role of an ultrasound protocol in evaluating the etiology of undifferentiated hypotension in the emergency department. Written informed consent was obtained from all patients. A final diagnosis of septic shock was made in 38% (39/103) of patients. Of 103 patients 17 had hyperdynamic LVF with an interobserver agreement of  $\kappa = 0.8$ . Hyperdynamic LVF had a positive likelihood ratio of 5.3 for the diagnosis of sepsis and was a strong independent predictor of sepsis as the final diagnosis with an odds ratio of 5.5 [95% confidence index (CI) 1.1–45] [75]. They conclude that among emergency department patients with nontraumatic undifferentiated symptomatic hypotension, the presence of hyperdynamic LVF on focused echo is highly specific for sepsis as the etiology of shock [75].

Drotrecogin alfa (activated), or recombinant human activated protein C, produced dose-dependent reductions in the levels of markers of coagulation and inflammation in patients with severe sepsis [76]. PROWESS, a randomized, double-blind, placebo-controlled, multicenter trial, was conducted to evaluate whether treatment with drotrecogin alfa activated reduced the rate of death from any cause among patients with severe sepsis. The institutional review board at each center approved the protocol, and written informed consent was obtained from all participants or their authorized representatives. A total of 1690 randomized patients were treated (840 in the placebo group and 850 in the drotrecogin alfa activated group). The mortality rate was 30.8% in the placebo group and 24.7% in the drotrecogin alfa activated group. However, the incidence of serious bleeding was higher in the drotrecogin alfa activated group than in the placebo group [77]. More studies are required to evaluate the possible beneficial effect of this drug. In this regard, Vincent et al. [78] performed the ENHANCE trial, a multiple-country, single-arm, open-label, trial in



order to provide further evidence for the efficacy and safety of drotrecogin alfa (activated) treatment in severe sepsis. Patients with known or suspected infection, three or four systemic inflammatory response syndrome criteria, and one or more sepsis-induced organ dysfunctions were recruited: 2434 adults entered, 2378 received drotrecogin alfa (activated), and of these, 2375 completed the protocol. Appropriate informed consent was obtained from all patients or their legal representative. The 28-day all-cause mortality was approximately the same as in the PROWESS trial (25.3% vs. 24.7%). However, patients in ENHANCE had increased serious bleeding rates compared with patients in the drotrecogin alfa (activated) arm of PROWESS. Increased postinfusion bleeding suggested a higher background bleeding rate. Intracranial hemorrhage was more common in ENHANCE than PROWESS. The authors concluded that ENHANCE provides supportive evidence for the favorable benefit–risk ratio observed in PROWESS and suggests that more effective use of drotrecogin alfa (activated) might be obtained by initiating therapy earlier [78].

### 10.4.5 CONCLUSIONS

Lately some clinical trials have been conducted in emergency departments, thus showing that the emergency-department-associated difficulties can be overcome. In spite of this there is a very short time frame between the admission of the patient and his or her inclusion in the clinical trial. With a well-designed protocol and a strict follow-up of it, as well as with very delimited functions of each physician, it is feasible to conduct a clinical trial in the emergency department, of course, without forgetting to obtain previously the signed informed consent.

To sum up, few clinical trials have been developed up to now in the emergency departments due to the associated difficulties. However, when possible, they have been demonstrated to be very useful for the physicians in order to know which is the best strategy when a patient arrives at the emergency department, especially in diseases related to the cardiovascular or central nervous system. For clinical trials in the emergency department to be feasible, it is crucial to have previously the designed protocol (which has to be very precise and, of course, approved by the corresponding committees), having the informed consent (clear and strictly redacted), the physicians must be trained for the protocol to be conducted, and the infrastructures have to be adequate for conducting the specific protocol.

### REFERENCES

1. Anderson, P., Petrino, R., Halpern, P., et al. (2006), The globalization of emergency medicine and its importance for public health, *Bull. World Health Organ*, 84, 835–839.
2. Walker, D. M., Tolentino, V. R., and Teach, S. J. (2007), Trends and challenges in international pediatric emergency medicine, *Curr. Opin. Pediatr.*, 19, 247–252.
3. Arnold, J. L., and Corte, D. F. (2003), International emergency medicine: Recent trends and future challenges, *Eur. J. Emerg. Med.*, 10, 180–188.
4. Vaslef, S. N., Cairns, C. B., and Falletta, J. M. (2006), Ethical and regulatory challenges associated with the exception from informed consent requirements for emergency



- research: From experimental design to institutional review board approval, *Arch. Surg.*, 141, 1019–1023.
5. Patterson, S. D., and Jones, B. (2007), A brief review of Phase 1 and clinical pharmacology statistics in clinical drug development, *Pharm. Stat.*, 6, 79–87.
  6. Boijsolie, C. R., Sharkey, S. W., Cannon, C. P., et al. (1995), Impact of a thrombolysis research trial on time to treatment for acute myocardial infarction in the emergency department, *Am. J. Cardiol.*, 76, 396–398.
  7. Soran, O., Kennard, E. D., Bart, B. A., et al.; IEPR investigators (2007), Impact of external counterpulsation treatment on emergency department visits and hospitalizations in refractory angina patients with left ventricular dysfunction, *Congest. Heart Fail.*, 13, 36–40.
  8. Mehta, S. R., Steg, P. G., Granger, C. B., et al.; ASPIRE Investigators (2005), Randomized, blinded trial comparing fondaparinux with unfractionated heparin in patients undergoing contemporary percutaneous coronary intervention: Arixtra Study in Percutaneous Coronary Intervention: a Randomized Evaluation (ASPIRE) Pilot Trial, *Circulation*, 111, 1390–1397.
  9. Saver, J. L., Kidwell, C., Eckstein, M., et al.; FAST-MAG Pilot Trial Investigators (2004), Prehospital neuroprotective therapy for acute stroke: Results of the Field Administration of Stroke Therapy-Magnesium (FAST-MAG) pilot trial, *Stroke*, 35, e106–108.
  10. Corneli, H. M., Zorc, J. J., Majahan, P., et al.; Bronchiolitis Study Group of the Pediatric Emergency Care Applied Research Network (PECARN) (2007), A multicenter, randomized, controlled trial of dexamethasone for bronchiolitis, *N. Engl. J. Med.*, 357, 331–339.
  11. Gibson, C. M., Kirtane, A. J., Murphy, S. A., et al.; TIMI Study Group (2006), Early initiation of eptifibatide in the emergency department before primary percutaneous coronary intervention for ST-segment elevation myocardia infarction: Results of the Time to Integrilin Therapy in Acute Myocardial Infarction (TITAN)-TIMI 34 trial, *Am. Heart J.*, 152, 668–675.
  12. Doney, M. K., and Macias, D. J. (2005), Regional highlights in global emergency medicine development. *Emerg. Med. Clin. North Am.*, 23(1), 31–44.
  13. Alagappan, K., Schafermeyer, R., Holliman, C. J., et al. (2007), International emergency medicine and the role for academic emergency medicine, *Ann. Emerg. Med.*, 14, 451–456.
  14. Barsan, W. G., Pancioli, A. M., and Conwit, R. A. (2004), Executive summary of the National Institute of Neurological Disorders and Stroke Conference on Emergency Neurologic Clinical Trials Network, *Ann. Emerg. Med.*, 44, 407–412.
  15. Hallstrom, A. P., and Paradis, N. A. (2005), Pre-randomization and de-randomization in emergency medical research: New names and rigorous criteria for old methods, *Resuscitation*, 65, 65–69.
  16. Roberts, I., Shakur, H., Edwards, P., et al. (2005), Trauma care research and the war on uncertainty, *BMJ*, 331, 1094–1096.
  17. Lemaire, F. (2006), The inability to consent in critical care research: Emergency or impairment of cognitive function, *Intensive Care Med.*, 32, 1930–1932.
  18. Anon. (2006), ER: The gateway for neurologic emergency trials? *Ann. Neurol.*, 60, A12–14.
  19. Food and Drug Administration (1996), Protection of human subjects; informed consent; final rule, *Fed. Reg.*, 61, 51498–51531.
  20. Shakur, H., Roberts, I., Barnetson, L., et al. (2007), Clinical trials in emergency situations, *BMJ*, 334, 165–166.

21. Department of Health (2001), *Reference Guide to Consent for Examination or Treatment*, DOH, London.
22. U.S. DHHS (2006), Guidance for Institutional Review Boards, Clinical Investigators and Sponsors Exception from Informed Consent Requirements for Emergency Research Draft Guidance, U.S. Department of Health and Human Services, Food and Drug Administration, Good Clinical Practice Program, Center for Biologics Evaluation and Research, Center for Drug Evaluation and Research Center for Devices and Radiological Health, Office of Regulatory Affairs, July.
23. Anon. (2006), Medicines for Human Use (Clinical Trials) Amendment (No. 2) Regulations 2006, Statutory Instrument 2006 No 2984.
24. Choonara, I. (2000), Clinical trials of medicines in children [editorial], *BMJ*, 321, 1093–1094.
25. Conroy, S., Choonara, I., Impicciatore, P., et al. (2000), Survey of unlicensed and off label drug use in paediatric wards in European countries. European Network for Drug Investigation in Children, *BMJ*, 320, 79–82.
26. Bush, A. (2006), Clinical trials research in pediatrics: Strategies for effective collaboration between investigator sites and the pharmaceutical industry, *Paediatr. Drugs*, 8, 271–277.
27. Ernest, T. B., Elder, D. P., Martini, L. G., et al. (2007), Developing paediatric medicines: Identifying the needs and recognizing the challenges, *J. Pharm. Pharmacol.*, 59, 1043–1055.
28. Cuzzolin, L., Atzei, A., and Fanos, V. (2006), Off-label and unlicensed prescribing for newborns and children in different settings: A review of the literature and a consideration about drug safety, *Expert Opin. Drug Saf.*, 5, 703–718.
29. Kan, P., and Kestle, J. R. (2007), Designing randomized clinical trials in pediatric neurosurgery, *Childs Nerv. Syst.*, 23, 385–390.
30. Bonati, M., and Pandolfini, C. (2005), DEC-net Collaborative Group. Pediatric clinical trials registry, *CMAJ*, 172, 1159–1160.
31. Bonati, M., and Pandolfini, C. (2005), DEC-NET Collaborative Group. More on compulsory registration of clinical trials: Complete clinical trial register is already reality for paediatrics, *BMJ*, 330, 480.
32. Jacqz-Aigrain, E., Zarrabian, S., Pandolfini, C., et al. (2006), A complete clinical trial register is already a reality in the paediatric field, *Therapie*, 61, 121–124.
33. Steinbrook, R. (2002), Improving protection for research subjects. *N. Engl. J. Med.*, 346, 1425–1430.
34. Quinn, S. C. (2004), Ethics in public health research: Protecting human subjects: The role of community advisory boards, *Am. J. Public Health*, 94, 918–922.
35. Nee, P. A., and Griffiths, R. D. (2002), Ethical considerations in accident and emergency research, *Emerg. Med. J.*, 19, 423–427.
36. Royal College of Physicians (1991), Fraud and misconduct in medical research: Causes, Investigations and Prevention. Report of a Working Party, Royal College of Physicians, London.
37. Silverman, H. (2007), Ethical issues during the conduct of clinical trials, *Proc. Am. Thorac. Soc.*, 4, 180–184.
38. Morse, M. A., Califf, R. M., and Sugarman, J. (2001), Monitoring and ensuring safety during clinical research, *JAMA*, 285, 1201–1205.
39. Silverman, H. (2007), Ethical issues during the conduct of clinical trials, *Proc. Am. Thorac. Soc.*, 4, 180–184.
40. Salzman, J. G., Frascione, R. J., Godding, B. K., et al. (2007), Implementing emergency research requiring exception from informed consent, community consultation, and public disclosure, *Ann. Emerg. Med.*, 50, 448–455.

41. Watters, D., Sayre, M. R., and Silbergleit, R. (2005), Research conditions that qualify for emergency exception from informed consent, *Acad. Emerg. Med.*, 12, 1040–1044.
42. Morris, M. C., Nadkarni, V. M., Ward, F. R., et al. (2004), Exception from informed consent for pediatric resuscitation research: Community consultation for a trial of brain cooling after in-hospital cardiac arrest, *Pediatrics*, 114, 776–781.
43. Richardson, L. D. (2005), The ethics of research without consent in emergency situations, *Mt. Sinai J. Med.*, 72, 242–249.
44. Morris, M. C., and Nelson, R. M. (2007), Randomized, controlled trials as minimal risk: An ethical analysis, *Crit. Care Med.*, 35, 940–944.
45. Freedman, B. (1987), Equipoise and the ethics of clinical research, *N. Engl. J. Med.*, 317, 141–145.
46. Sugarman, J., Kass, N. E., Goodman, S. N., et al. (1998), What patients say about medical research, *IRB*, 20, 1–7.
47. Abboud, P. A., Heard, K., Al-Marshad, A. A., et al. (2006), What determines whether patients are willing to participate in resuscitation studies requiring exception from informed consent? *J. Med. Ethics*, 32, 468–472.
48. EU (2001), Directive 2001/20/EC of the European parliament and of the council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the member states relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use, *Official J. Eur. Comm.*, L121, 34–44; <http://www.eortc.be/Services/Doc/clinical-EU-directive-04-April-01.pdf>.
49. Singer, E. A., and Mullner, M. (2002), Implications of the EU directive on clinical trials for emergency medicine, *BMJ*, 324, 1169–1170.
50. Cone, D. C., and O'Connor, R. E. (2005), Are US informed consent requirements driving resuscitation research overseas? *Resuscitation*, 66, 141–148.
51. Mader, T. J., and Playe, S. J. (1997), Emergency medicine research consent form readability assessment, *Ann. Emerg. Med.*, 29, 534–539.
52. Paasche-Orlow, M. K., Taylor, H. A., and Brancati, F. L. (2003), Readability standards for informed-consent forms as compared with actual readability, *N. Engl. J. Med.*, 348, 721–726.
53. Demircan, C., Cikrikler, H. I., Engindeniz, Z., et al. (2005), Comparison of the effectiveness of intravenous diltiazem and metoprolol in the management of rapid ventricular rate in atrial fibrillation, *Emerg. Med. J.*, 22, 411–414.
54. Davey, M. J., and Teubner, D. (2005), A randomized controlled trial of magnesium sulfate, in addition to usual care, for rate control in atrial fibrillation, *Ann. Emerg. Med.*, 45, 347–353.
55. Thomas, S. P., Guy, D., Wallace, E., et al. (2004), Rapid loading of sotalol or amiodarone for management of recent onset symptomatic atrial fibrillation: A randomized, digoxin-controlled trial, *Am. Heart J.*, 147, E3.
56. Kim, M. H., Morady, F., Conlon, B., et al. (2002), A prospective, randomized, controlled trial of an emergency department-based atrial fibrillation treatment strategy with low-molecular-weight heparin, *Ann. Emerg. Med.*, 40, 187–192.
57. Topol, E. J. (2003), Current status and future prospects for acute myocardial infarction therapy, *Circulation*, 108 (16 Suppl 1), III6–13.
58. Kandzari, D. E. (2006), Evolving antithrombotic treatment strategies for acute ST-elevation myocardial infarction, *Rev. Cardiovasc. Med.*, 7 (Suppl 4), S29–37.
59. Ibbotson, T., McGavin, J. K., and Goa, K. L. (2003), Abciximab: An updated review of its therapeutic use in patients with ischemic heart disease undergoing percutaneous coronary revascularisation, *Drugs*, 63, 1121–1163.

60. Montalescot, G., Barragan, P., Wittenberg, O., et al.; ADMIRAL Investigators (2001), Abciximab before direct angioplasty and stenting in myocardial infarction regarding acute and long-term follow-up. Platelet glycoprotein IIb/IIIa inhibition with coronary stenting for acute myocardial infarction, *N. Engl. J. Med.*, 344, 1895–1903.
61. Stone, G. W., Grines, C. L., Cox, D. A., et al.; Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications (CADILLAC) Investigators (2002), Comparison of angioplasty with stenting, with or without abciximab, in acute myocardial infarction, *N. Engl. J. Med.*, 346, 957–966.
62. De Luca, G., Suryapranata, H., Stone, G. W., et al. (2005), Abciximab as adjunctive therapy to reperfusion in acute ST-segment elevation myocardial infarction. Meta-analysis of randomized trials, *JAMA*, 293, 1759–1765.
63. Anon. (1997), Randomised placebo-controlled trial of effect of eptifibatide on complications of percutaneous coronary intervention: IMPACT-II. Integrilin to Minimise Platelet Aggregation and Coronary Thrombosis-II, *Lancet*, 349, 1422–1428.
64. ESPRIT Investigators (2000), Enhanced suppression of the platelet IIb/IIIa receptor with integrilin therapy. Novel dosing regimen of eptifibatide in planned coronary stent implantation (ESPRIT): A randomised, placebo-controlled trial, *Lancet*, 356, 2037–2044.
65. PURSUIT Investigators (1998), Inhibition of platelet glycoprotein IIb/IIIa with eptifibatide in patients with acute coronary syndromes. Platelet glycoprotein IIb/IIIa in unstable angina: Receptor suppression using integrilin therapy, *N. Engl. J. Med.*, 339, 436–443.
66. Zeymer, U. (2007), The role of eptifibatide in patients undergoing percutaneous coronary intervention, *Expert Opin. Pharmacother.*, 8, 1147–1154.
67. Valgimigli, M., Bolognese, L., Anselmi, M., et al. (2007), Two-by-two factorial comparison of high-bolus-dose tirofiban followed by standard infusion versus abciximab and sirolimus-eluting versus bare-metal stent implantation in patients with acute myocardial infarction: Design and rationale for the MULTI-STRATEGY trial, *Am. Heart J.*, 154, 39–45.
68. Anon. (1995), Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, *N. Engl. J. Med.*, 333, 1581–1587.
69. Hacke, W., Kaste, M., Fieschi, C., et al. (1998), Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischemic stroke (ECASS II). Second European–Australasian Acute Stroke Study Investigators, *Lancet*, 352, 1245–1251.
70. Clark, W. M., Wissman, S., Albers, G. W., et al. (1999), Recombinant tissue-type plasminogen activator (Alteplase) for ischemic stroke 3 to 5 hours after symptom onset. The ATLANTIS Study: A randomized controlled trial. Alteplase Thrombolysis for Acute Noninterventional Therapy in Ischemic Stroke, *JAMA*, 282, 2019–2026.
71. Hacke, W., Donnan, G., Fieschi, C., et al.; ATLANTIS Trials Investigators, ECASS Trials Investigators, NINDS rt-PA Study Group Investigators (2004), Association of outcome with early stroke treatment: Pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials, *Lancet*, 363, 768–774.
72. Hanley, D., and Hacke, W. (2005), Critical care and emergency medicine neurology in stroke, *Stroke*, 36, 205–207.
73. IMS II Trial Investigators (2007), The Interventional Management of Stroke (IMS) II Study, *Stroke*, 38, 2127–2135.
74. Rivers, E., Nguyen, B., Havstad, S., et al. (2001), Early goal-directed therapy in the treatment of severe sepsis and septic shock, *N. Engl. J. Med.*, 345, 1368–1377.

75. Jones, A. E., Craddock, P. A., Tayal, V. S., et al. (2005), Diagnostic accuracy of left ventricular function for identifying sepsis among emergency department patients with non-traumatic symptomatic undifferentiated hypotension, *Shock*, 24, 513–517.
76. Hartman, D. L., Bernard, G. R., Helterbrand, J. D., et al. (1998), Recombinant human activated protein C (rhAPC) improves coagulation abnormalities associated with severe sepsis, *Intensive Care Med.*, 24 (Suppl 1), S77–S77.
77. Bernard, G. R., Vincent, J. L., Laterre, P. F., et al.; Recombinant human protein C Worldwide Evaluation in Severe Sepsis (PROWESS) study group (2001), Efficacy and safety of recombinant human activated protein C for severe sepsis, *N. Engl. J. Med.*, 344, 699–709.
78. Vincent, J. L., Bernard, G. R., Beale, R., et al. (2005), Drotrecogin alfa (activated) treatment in severe sepsis from the global open-label trial ENHANCE: Further evidence for survival and safety and implications for early treatment, *Crit. Care Med.*, 33, 2266–2277.



# 10.5

---

## *Gastroenterology*

**Lise Lotte Gluud<sup>1</sup> and Jørgen Rask-Madsen<sup>2</sup>**

<sup>1</sup>*Copenhagen Trial Unit, Cochrane Hepato-Biliary Group, Copenhagen, Denmark*

<sup>2</sup>*Department of Medical Gastroenterology, Herlev Hospital, University of Copenhagen, Herlev, Denmark*

### **Contents**

|         |                                                             |     |
|---------|-------------------------------------------------------------|-----|
| 10.5.1  | Preface and Introduction to Evidence-Based Gastroenterology | 501 |
| 10.5.2  | Definitions and Classification of Clinical Trials           | 502 |
| 10.5.3  | Observational Studies                                       | 502 |
| 10.5.4  | Randomized Controlled Trials                                | 505 |
| 10.5.5  | Blinding in Randomized Controlled Trials                    | 508 |
| 10.5.6  | Sample Size Calculations and Statistical Power              | 509 |
| 10.5.7  | Follow-up and Attrition Bias                                | 511 |
| 10.5.8  | Systematic Reviews                                          | 512 |
| 10.5.9  | Publication Bias and Related Biases                         | 513 |
| 10.5.10 | Concluding Remarks                                          | 513 |
|         | References                                                  | 514 |

### **10.5.1 PREFACE AND INTRODUCTION TO EVIDENCE-BASED GASTROENTEROLOGY**

Traditionally, clinical decisions were based on experience, but sometimes experience may be misleading. We are often prone to remember the best and the worst cases of a disease. The natural course of diseases is fluctuating and humans have the capacity to recover spontaneously. The unsystematic collection of data and limitations in human processing limits the credibility of recommendations based only on experience. Accordingly, evidence-based medicine combining clinical experience with research evidence is gradually replaced by traditional experience-based



practice. Evidence-based gastroenterology involves literature searches and quality assessments to identify the most valid research. The internal validity of a trial refers to the credibility of the results and depending on its quality and risk of bias. Low-quality trials have a considerable risk of bias and low internal validity. High-quality trials with adequate bias control have a high internal validity. With a low internal validity of a trial, the assessment of its external validity becomes irrelevant because the results lack credibility. If the internal validity is adequate, assessment of the external validity is still necessary before the trial results are used in clinical practice. The external validity of clinical trials depends on the extent to which the results may be extrapolated. The external validity depends on the characteristics of patients included, treatment regimens, and the trial setting. In clinical practice, treatments are often used for larger patient groups with less strict selection criteria. The results of a trial with highly selective inclusion criteria performed at a specialized unit may, therefore, be difficult to extrapolate and use in general clinical practice.

### 10.5.2 DEFINITIONS AND CLASSIFICATION OF CLINICAL TRIALS

Research in gastroenterology deals with diseases of the digestive system, including the esophagus, stomach, intestine, liver, gallbladder, and pancreas. In clinical gastroenterology most trials assess the effects of various drugs, but interventional procedures may also be subject to clinical trials, such as efficacy and safety of endoscopic therapy and primary as well as secondary prevention. Large, high-quality randomized clinical trials and systematic reviews of several randomized clinical trials are considered the gold standard. However, the clinician often has to judge a number of clinical trials with inconsistent results and numerous methodological deficiencies. If no randomized clinical trials are available, observational studies may be considered the best available source of evidence.

### 10.5.3 OBSERVATIONAL STUDIES

The classical observational study designs are case-control studies, cohort studies, case series, and case reports. Cohort studies follow a group of patients (cohort) pro- or retrospectively. Case-control studies are based on patients with certain diseases or genetic characteristics (cases) and controls without the specific property but otherwise similar prognostic profile. Cohort studies start out with the identification of the intervention, for example, a group of patients undergoing stricturoplasty for Crohn's disease [1], whereas case-control studies initially identify a group of patients with a certain disease, for example, cases with colorectal cancer and matched controls without the disease, before attempts are made to detect the influence of intervention, for example, use of antidepressants [2]. Case series describe the outcome of a certain group of patients, for example, patients who were treated for presumed exacerbation of Crohn's disease but were subsequently found to have underlying small-bowel carcinoma [3]. Case reports describe unusual cases of individual patients with rare diseases, unusual adverse events, or treatment effects, for example, development of subfulminant hepatitis B after treatment with infliximab for Crohn's disease [4].

Prospective cohort studies are generally considered the most valid observational design. If a disease is rare, or if the course of a disease is protracted, performing a prospective cohort study may, however, not be feasible. For example, the development of cirrhosis and liver failure after infection with hepatitis B or C takes several years [5]. In that case, retrospective cohort studies or case-control studies may be considered.

Although randomized controlled trials are the gold standard for evaluation of intervention effects, this does not mean that observational studies have little or no value. However, if interventions have dramatic effects in line with penicillin for pneumonia as compared with no treatment, randomized controlled trials are not required. Observational studies may also provide important information when we are unable to perform randomized controlled trials, for example, if assessing behavior or if requiring extremely large samples of patients for the assessment of rare adverse events. In other cases, randomized trials may be unethical, for example, if the causative agent is considered potentially harmful. Furthermore, the evidence generated in observational studies may also be useful if complementary or the basis of randomized controlled trials. Although important information may sometimes be contained in observational studies, the information should only be used with due care considering the risk of biases.

One of the most important roles of observational studies lies in the postmarketing surveillance of adverse events. Rare adverse events may be difficult to detect in randomized controlled trials because very large samples are necessary. One example is terlipressin, which has been assessed for treatment of hepatorenal syndrome [6]. Today, only few randomized controlled trials have addressed this question. None of the trials detected serious adverse events. However, one case report suggested that terlipressin may be associated with worsening of cerebral hyperemia [7]. Another case report suggested that terlipressin may be associated with acute ST-segment elevation myocardial infarction [8]. Therefore, performing observational studies that are sufficiently large to detect the risk of serious adverse events seems highly important. One example is the introduction and widespread use of laparoscopic cholecystectomy in the 1990s, which was associated with a dramatic increase in the incidence of bile duct injuries [9]. Several studies have reviewed the outcomes of surgical management of bile duct injuries, but data on the current incidence of injuries are scarce. The completion of a large prospective cohort on the outcomes of laparoscopic cholecystectomy today may provide important information to patients as well as health care workers. The information may also be used in the planning of future randomized controlled trials that are considered necessary to determine whether patients fare better with or without the intervention.

Another important role for observational studies lies in the evaluation of intervention effects on long-term outcomes. The practical difficulties in maintaining long-term prospective randomized controlled trials can make the task impossible. One example of such a situation is the treatment of chronic hepatitis C with antiviral substances, for example, interferon and ribavirin [10]. Chronic hepatitis C has a protracted course. To determine whether antiviral therapy affects morbidity or mortality, patients need to be followed for decades. This may logistically be nearly impossible. Furthermore, the risk of attrition bias is considerable and increases with the duration of follow-up. Finally, the efficacy of today's recommended antiviral treatment for chronic hepatitis C (interferon or peg interferon combined with riba-

virin) may for ethical reasons be prohibitive for preserving a nontreatment control group. Several randomized controlled trials have been performed, but all have focused on biological response, that is, clearance of the hepatitis C virus RNA (ribonucleic acid) from the blood. None of these trials have established whether the demonstrated efficacy of treatment translates into a meaningful effect on clinical outcomes [10]. Therefore, we have to rely on observational studies to determine the outcome of treatment.

As previously described, a retrospective study design, which is used in some observational studies, increases the risk of bias due to factors that change with time, recall bias, and differential measurement errors [11, 12]. However, one of the most important limitations in prospective as well as retrospective observational studies lies in the risk of selection bias. In observational studies, prognostic factors determine whether patients are allocated to intervention or control groups. This is known as confounding by indication, which may lead to systematic differences between comparison groups [13]. When systematic differences exist, estimates of intervention benefits may be incorrect. It is impossible to determine the size of the effect of the differences on the estimated intervention effects, why separate evaluations of individual trials or studies are necessary. In some cases, differences may be derived from analyses comparing groups of several observational studies and randomized controlled trials. One example may be found in a systematic review on artificial and nonartificial support systems for acute and acute-on-chronic liver failure [14]. The review included observational studies as well as randomized controlled trials. Formally, the inclusion criteria used in the different studies and trials were comparable. All patients who were allocated to the control groups received comparable standard medical regimens. When the results of the individual randomized controlled trials were combined in a meta-analysis, the support systems did not appear to reduce mortality significantly compared with standard medical therapy. Although inclusion criteria and the characteristics of included patients at baseline were similar, the control group mortality rates were significantly higher in the observational studies than in the randomized controlled trials. No significant difference was found when comparing the mortality rates of patients allocated to the intervention groups. Accordingly, the observational studies found a statistically significant positive effect of the intervention as compared with the randomized controlled trials. This difference suggests that the observational studies had a skewed allocation of patients with the worst prognosis to control groups. Thus, the observational studies may have overestimated the intervention benefit. The strength of the association warrants separate evaluations in different situations, although observational studies are generally more susceptible to bias than randomized controlled trials. In a similar review, considerable differences were found between estimated intervention effects in comparisons of observational studies and randomized controlled trials [13].

Analyses of all comparisons showed that the effect estimates in the observational studies ranged from an underestimation of effect by 76% to an overestimation of effect by 160% [13]. In another methodological study, odds ratios generated by 168 observational studies and 240 randomized controlled trials within 45 different topics were compared [15]. All trials and studies were included in a meta-analysis with binary outcomes. Overall, the observational studies tended to generate larger summary odds ratios suggesting a more beneficial effect of the intervention. Bias associated with nonrandom allocation has also been analyzed in a review with

studies that compared randomized controlled trials with observational studies [16]. The review reported that nonrandom allocation was related to overestimation as well as underestimation of treatment effects. The variation in the results of observational studies was increased due to haphazard differences in case mix between groups.

Four strategies for case-mix adjustment were subsequently evaluated by generating nonrandomized studies from two large randomized controlled trials [16]. Participants were resampled according to allocated treatment, center, and period. None of the strategies adjusted adequately for bias in historical or concurrent controlled studies. Logistic regression was found to increase bias due to misclassifications and measurement errors in confounding variables as well as differences between conditional and unconditional odds ratio estimates of treatment effects.

#### 10.5.4 RANDOMIZED CONTROLLED TRIALS

Random allocation means that all patients have a known chance of being allocated to one of the intervention groups and that the allocation of the next patient is unpredictable [17]. To keep the allocation of patients unpredictable, both adequate generation of an allocation sequence and adequate concealment of allocation are required. The allocation sequence may consist of random numbers generated by computers or tables. Allocation concealment may consist of randomization through independent centers or serially numbered identical sealed packages. If the next assignment is known, enrollment of certain patients may be prevented or delayed to ensure that they receive the treatment that is believed to be superior. Theoretically, serially numbered sealed envelopes may provide adequate allocation concealment, although there is some evidence suggesting that this may not be true. In some cases, envelopes were opened before or after patients were excluded [18]. In other cases envelopes have been transilluminated [19]. The adequacy of using serially numbered sealed envelopes therefore seems debatable.

In theory, adequate randomization is necessary to obtain adequate control of bias. On the other hand, empirical evidence is necessary to determine whether the effect of randomization is only hypothetical or whether randomization affects the results and subsequently the conclusions drawn from randomized controlled trials. To address this question, two methodological studies of cohorts of clinical trials estimated the association between the methods of randomization and the effects of intervention [20, 21]. All trials were included in meta-analyses. One study included trials from the field of obstetrics and gynecology [20] and one included trials from a variety of disease areas [21]. Neither of the studies revealed a significant association between allocation sequence generation and intervention effects, but both showed that allocation concealment methods were significantly associated with estimated intervention effects. The analyses made in both methodological studies illustrated that inadequate allocation concealment was associated with significantly more positive estimates of intervention effects. These results suggest that inadequate allocation concealment may lead to exaggerated intervention effects. On the other hand, the results may also show bias due to selective publication of small or low-quality trials with positive results [22, 23]. Because there is no defined gold standard, it may be difficult to determine whether the trials with or the trials without

adequate allocation concealment were more correct. To address these concerns, a subsequent study used very large randomized controlled trials as a reference group [24]. The reference group included randomized controlled trials with more than 1000 participants. Each of the included trials, that is, the trials in the reference group and the smaller trials, were included in the meta-analyses. Each meta-analysis contained at least one large trial. Subsequently, analyses were performed to compare the results of trials in the reference groups with the results of the smaller trials. The analyses showed that, on average, the small trials without adequate allocation sequence generation or allocation concealment overestimated intervention benefits. The results of the small trials with adequate generation of the allocation sequence or adequate allocation concealment were not significantly different from the results of the trials in the reference group.

These results support the importance of adequate randomization to bias control in randomized controlled trials. Similar subsequent methodological studies of randomized controlled trials have also examined the association between randomization and intervention effects [25]. The evidence generated in these studies was combined in a random effects meta-analysis including the summary results from the individual studies. The results suggested that odds ratios were about 12% more positive in trials without adequate allocation sequence generation as compared to odds ratios in trials with adequate allocation sequence generation. Therefore, trials reporting inadequate methods may overestimate intervention benefits due to inadequate bias control. Similarly, odds ratios generated in trials without adequate allocation concealment were about 21% more positive than odds ratios generated in trials with adequate allocation concealment. Both components of randomization process therefore seem to be important. However, the meta-analyses on the effect of allocation sequence generation and allocation concealment also found that there was a considerable heterogeneity between studies. The heterogeneity may be related to the disease area, the type of intervention, trial inclusion criteria, and the classification of adequate randomization in the individual methodological studies. The variation suggests that caution is required when making inferences and recommendations for assessment of bias control. Using the described components as exclusion criteria, for example, disregarding all trials without adequate allocation concealment is not justified. Simply reducing the estimated intervention effect in randomized controlled trials that do not describe adequate allocation concealment seems also problematic. Thus, the quality of individual trials and meta-analyses has to be evaluated separately.

Many trials are described as randomized without reporting randomization methods. A number of cohort studies of randomized controlled trials suggest that the proportion of trials with adequate allocation sequence generation ranges from 1 to 52% (median 37%) [25]. The proportion with adequate allocation concealment ranges from 2 to 39% (median 25%). Some of the variation may depend on the disease areas evaluated or different classifications of randomization methods in the cohort studies, or other factors [26, 27].

Whether the lack of reported randomization methods reflect the actual conduct of a published randomized controlled trial is difficult to establish. In theory, inadequate reporting may hide important flaws in the design of a trial. On the other hand, a high-quality randomized controlled trial may be overlooked because of inaccurate reporting. In a methodological study, the reported methods used for allocation

concealment were extracted from the published reports of 105 randomized controlled trials [28]. The reported allocation concealment methods in full-text publications were subsequently compared with the information on allocation concealment obtained through direct communication with the author(s). The results of the study showed that several trials had adequate allocation concealment methods that were not described in the published report [29]. Another methodological study reached a different conclusion [30]. This study compared the reported descriptions of the methods used for allocation concealment in publications and protocols of 102 randomized controlled trials. The analyses showed that most of the trials with unclear allocation concealment in the published trial reports also had unclear allocation concealment in the protocol. The evidence concerning the discrepancies between the conduct and report of randomized controlled trials is equivocal. Additional evidence is needed to clarify this question.

In 1999, an observational study including all 235 randomized controlled trials published in the journal *Hepatology* from the initiation in 1981 through August 1998 was published [31]. Only 52% of the included trials reported adequate generation of the allocation sequence. The proportion of trials reporting adequate allocation concealment was 34%. In a similar observational study of all 383 randomized clinical trials published in the journal *Gastroenterology* as original articles from 1964 to 2000 were reviewed [26]. The individual authors in all of the included publications had described the trials as being randomized. However, only 42% of the trials reported adequate generation of the allocation sequence. The proportion of trials reporting adequate allocation concealment was 39%. Unlike the study on randomized controlled trials published in *Hepatology*, the reporting quality improved significantly in the mid-1990s.

Nevertheless, both studies found that there was still room for improvement. Whether these findings were specific to these two journals was evaluated in a subsequent study including 616 hepato-biliary randomized controlled trials published from 1985 to 1996 in 12 different MEDLINE-indexed journals [27]. All trials were described as randomized by the individual authors. However, the reported generation of the allocation sequence was only described in 48% of the included trials. Fifty-two percent of the trials did not include a description of the methods used for generation of the allocation sequence. In 38% of the included trials, the allocation concealment was described adequately. In the remaining 62% of the trials, the allocation concealment was not described. A number of analyses were performed to evaluate potential predictors for adequate reporting of the randomization methods, that is, allocation sequence generation and allocation concealment. These analyses focused on the importance of funding and disease area. Based on the published reports, 47% of the trials received no external funding, 30% were funded by non-profit organizations, and 23% were funded by for-profit organizations. The proportion of trials with adequate allocation sequence generation and allocation concealment was not significantly different among trials funded by profit or nonprofit organizations. When these trials were combined and compared with the trials not reporting external funding, the analyses showed that trials with external funding were significantly more likely to report adequate generation of the allocation sequence. The proportion of trials with adequate allocation concealment was insignificantly different in the two groups. Further analyses revealed that also the proportion of trials with funding was insignificantly different within different disease areas. However,



the proportions of trials with adequate generation of the allocation sequence or adequate allocation concealment were significantly associated with the disease area. Trials dealing with interventions for some disease areas reported adequate randomization methods significantly more often than trials in other areas. Several other aspects, including the sample size and number of clinical sites, have been suggested as potential predictors of the quality of the reported randomization. However, additional studies in this area are still warranted to establish the different patterns.

### 10.5.5 BLINDING IN RANDOMIZED CONTROLLED TRIALS

In randomized controlled trials, the term *blinding* refers to keeping participants, health care providers, data collectors, outcome assessors, or data analysts unaware of the assigned intervention. Double blinding may refer to blinding of participants and health care providers, investigators, data collectors, judicial assessors, or data analysts. The specific methods used to maintain blinding in trial reports are often missing. Furthermore, many researchers disagree on the correct definition of double blinding. It is therefore recommended that, in individual trials, the term double blinding is provided with clear information about who were blinded and how the blinding was performed. Sometimes the nature of the intervention precludes double blinding, but blinded outcome assessment and data analyses are usually possible. To ensure adequate double blinding, the interventions compared must be similar. If an intervention is compared to no intervention, an identical placebo must be used. Any difference in taste, smell, or appearance may destroy blinding. One example may be found in a randomized controlled trial on the effect of nicotine gum on smoking cessation [32]. The trial compared the effect of nicotine gum versus a placebo gum. The authors tried to disguise the placebo gum by preparing wrappings, which suggested that the contents included nicotine. However, Wrigley's chewing gum was used as placebo. It seems likely, therefore, that participants correctly guessed whether they belonged to the intervention group or the control group.

Another example of a break in blinding is found in a randomized controlled trial on ascorbic acid for the common cold [33]. The trial was described as using a double-blind randomized design. Participants were employees of the National Institutes of Health. As no established effective treatment was known, a placebo containing lactulose was used to maintain blinding. However, the results of the trial turned out to be questionable because many participants tasted their capsules and guessed in which group they were.

Although there is a clear theoretical association between blinding and the control of bias, empirical evidence is required to determine the actual size and direction of the association. Six methodological studies of randomized controlled trials have analyzed the association between double blinding and intervention effects [25]. Each of the randomized controlled trials was included in meta-analyses assessing binary outcomes. Analyses were subsequently performed to compare odds ratios in randomized controlled trials with or without blinding. Two of these studies revealed that randomized controlled trials without double blinding overestimated intervention effects compared to randomized double-blind trials. The remaining four methodological studies found no significant association between blinding and estimates



of intervention benefits. To combine the empirical evidence generated in the methodological studies, a random effects meta-analysis was performed. This meta-analysis showed no significant differences between odds ratios of intervention effects in groups of double-blind trials compared to trials without double blinding. The meta-analysis did, however, reveal a considerable difference between study variation that may be related to the nature of the disease or the intervention. The meta-analyses included trials from various disease areas including cardiology, gynecology, obstetrics, psychiatry, and smoking cessation. Accordingly, the interventions assessed included diagnostic measures and drugs as well as surgical procedures. The variation may reflect that some interventions are difficult to blind. If we perform double-blind trials on, for example, drugs associated with adverse events blinding may be ineffective. The type of outcome may be equally important. Hard outcomes may be less prone to assessment bias than subjective outcomes. Therefore, trials evaluating the effect of drugs on, for example, mortality may be less susceptible to bias than trials evaluating the effect of drugs on pain. The effect of blinding is highly unpredictable, and separate analyses of the effect of blinding in individual trials and meta-analyses are warranted.

In a cohort study including 616 hepato-biliary randomized controlled trials published during 1985–1996, only 34% were double blind [27]. The proportion of double-blind trials varied significantly in different disease areas. Trials on interventions for gallstones were significantly less often double blind as compared to, for example, trials on portal hypertension. To some extent, the variation reflects that some interventions are more difficult to blind than others. One example is endoscopic procedures. Some trials have attempted to perform “sham” endoscopy, but maintaining the blinding in such cases is obviously difficult. Blinding the effect of drugs may also be difficult if there are specific characteristic effects (e.g., lowering of blood pressure and heart frequency when using  $\beta$  blockers for prevention of bleeding esophageal varices). Characteristic adverse effects associated with treatment may also cause a break in the blinding. If the maintenance of adequate blinding is questionable, it may be relevant to test the possibility of a break in blinding before performing the trial [34]. The results of such pretrial assessments may be used in the development of the final trial protocol. In some cases, we do not need empirical evidence to establish that blinding is impossible for patients or investigators, for example, when assessing endoscopic procedures or interventions such as liver support systems [14]. Trying to maintain blinding may only result in making the trial more complicated. However, it is always possible to maintain some form of blinded outcome assessment or blinded data analyses. Unfortunately, very few studies on randomized controlled trials in gastroenterology report on these aspects.

### 10.5.6 SAMPLE SIZE CALCULATIONS AND STATISTICAL POWER

Random error may occur in any direction and subsequently lead to false-positive (type I error) or false-negative results (type II error). In randomized controlled trials, the risk of random error depends on the sample size and the size of the intervention effect. The larger the sample size and the intervention effect, the smaller the risk of random error. Accordingly, small trials on interventions with moderate

effects have a substantial risk of being subject to random error and consequently to produce false-positive or false-negative conclusions. Large randomized controlled trials on the same interventions with moderate effects have a lower risk of random error than the small trials. Small trials on interventions with substantial effects have a smaller risk of generating results that lead to false-negative or false-positive conclusions than small trials on interventions with moderate intervention effects. One way to determine the risk of random error is to calculate confidence intervals or levels of statistical significance. Confidence intervals and the level of statistical significance provide an estimate of the error that may occur and thus reflect the precision of the statistical estimates. However, these estimates do not tell whether they are clinically significant or meaningful.

Confidence intervals are related to the concept of the statistical power. The larger the confidence interval, the less power a study has to detect differences between outcomes in groups of patients allocated to the intervention or the control group. Sample size calculations are required in randomized trials because inadequate statistical power can lead to false-negative results. The calculations should account for the minimum relevant treatment effect, acceptable probabilities of type I and II errors and losses to follow-up [35]. The first parameter, that is, the minimum relevant treatment effect, is adjustable and sensitive. If you reduce the relevant difference by half, four times as many patients are needed. The risk of a type I error ( $\alpha$ ) is usually set to 5% and the risk of a type II error ( $\beta$ ) is usually set to 10 or 20%. Subsequently, the statistical power ( $1 - \beta$ ) is 90 or 80%. The power of a trial reflects the risk of overlooking intervention effects. Suppose you want to perform a trial on a specific treatment, for example, a drug that reduces mortality from 40 to 20% and you set the risk of a type I error ( $\alpha$ ) to 5% and include 90 patients both in the treatment arm and the control arm, your trial will have 80% power to detect the true treatment effect. If you repeat the trial a 100 times, 20 of the trials will overlook the true treatment effect. If you evaluate the same drug, but include 45 patients in each treatment arm instead, the sample size corresponds to a power of 55%. If you repeat the trial a 100 times, 45 of your trials will overlook the true treatment effect. If you search for evidence and identify 100 randomized controlled trials, the entire sample of trials must be evaluated. Including a subgroup of trials, for example, the trials overlooking the intervention effect, means that your conclusions become incorrect. In practice, looking at larger samples of randomized controlled trials may be done through a systematic review, making a meta-analysis of the trials.

In a cohort study including all 235 randomized controlled trials published in the journal *Hepatology*, from the initiation in 1981 through August 1998, were assessed [31]. All trials dealt with hepato-biliary diseases. Only 26% of the trials reported sample size calculations. In similar cohort studies of randomized controlled trials from various disease areas, sample size calculations were only reported in 8–38% of the trials included [25]. This is unfortunate because the sample size calculations provide crucial information about the reliability of the results. If the preset sample size remains unreported, it becomes difficult to evaluate whether the planned sample size was reached or whether the trial was extended beyond the planned size or was terminated at an arbitrary time point.

Two studies evaluated statistical power in trials with statistically insignificant outcomes [36, 37]. Both studies found that most trials had insufficient power to detect clinically relevant treatment effects. The relatively small sample size of randomized

trials suggests that only a few have the recommended statistical power [25]. For example, one observational study included 383 randomized controlled trials published in the journal *Gastroenterology* [26]. On average, the randomized controlled trials included 43 patients per intervention arm (standard error of the mean was 4 patients). If you perform a trial with 45 patients per intervention arm, the trial will have a 90% power to detect a reduction in rates of mortality from 60 to 25% (if the risk of a type I error is set to 0.05). However, very few interventions have such dramatic effects. If we repeat the trial, evaluating a drug that reduces mortality rates from 60 to 40%, the statistical power of the trial will decrease to 39%. Therefore, it seems likely that a number of trials overlooked clinically important significant intervention effects. In a randomized controlled trial including patients admitted with bleeding esophageal varices, the effect of emergency sclerotherapy with sodium tetradecyl sulfate was compared with octreotide infusion [38]. The trial included only 100 patients, although the sample size calculations reported by the authors showed that a sample size of 1800 patients was required. This means that the trial had a 5% chance of demonstrating a statistically significant result, if the treatment effect estimated—also by the authors—was correct. The authors concluded that emergency sclerotherapy and octreotide infusions were equally efficacious in controlling variceal hemorrhage. The conclusion is debatable considering the low statistical power of the trial. Here it is again important to remember that absence of evidence is no evidence of absence. The fact that no significant intervention effect was identified does not mean that it does not exist. It is possible that we simply used the wrong method to look for such an effect. A number of studies have evaluated the sample size of published randomized controlled trials on hepato-biliary and other gastroenterological diseases. Similarly, the average sample size of randomized trials in clinical gastroenterology suggests that effective interventions may have been disregarded based on insufficient grounds. Much larger trials are necessary to conclude that an intervention is ineffective or that two interventions are equally effective.

Sample size calculations are required before performing randomized controlled trials because inadequate statistical power may lead to false-negative results [39]. The calculations should account for the minimum relevant treatment difference, acceptable probabilities of type I and II errors, and losses to follow-up [35, 40]. The first parameter is adjustable and sensitive. If you reduce the relevant difference by half, four times as many patients are needed. The risks of a type I error ( $\alpha$ ) is usually set to 5%. The risk of a type II error ( $\beta$ ) is usually set to 10 or 20%. The corresponding power ( $1 - \beta$ ), which indicates the risk of overlooking an effect of the intervention, is usually set to 90 or 80%. In cohort studies of randomized controlled trials, sample size calculations were reported only in 8–38% of the included trials [25]. Without having preset the sample size beforehand, the reader is unable to assess whether the planned sample size was actually reached or whether the trial was extended beyond the planned size or perhaps terminated at an arbitrary time point.

### 10.5.7 FOLLOW-UP AND ATTRITION BIAS

Adequate follow-up is essential to avoid attrition bias. Nearly all clinical trials have some missing data due to losses to follow-up, which may affect the results, because they may be related to prognostic factors [41]. Development of methods to obtain

data on patients who are lost to follow-up and to account for losses to follow-up is important. Also, to achieve an adequate or fair interpretation of the results obtained, a clear description of follow-up is essential [42]. In a study on 235 hepato-biliary randomized controlled trials, the numbers or reasons for dropouts or withdrawals were not reported [31].

Several analytical strategies for dealing with missing data have been proposed [43]. One of the most popular methods is to perform the analysis by using the intention-to-treat principle, including all originally randomized patients. The alternative strategy, which is to perform a per-protocol analysis excluding data from patients who were lost to follow-up or had other protocol deviations [43, 44]. However, there are several problems associated with the design of per-protocol analysis. If, for example, an intervention has adverse effects causing dropouts lost to follow-up, the per-protocol analysis will overestimate the benefit of intervention. In a systematic review about randomized controlled trials on interferon and ribavirin for chronic hepatitis C, several patients were lost to follow-up in the individual trials [10]. Many of the protocol deviations and losses to follow-up reflect the occurrence of adverse events related to the use of interferon and ribavirin. An analysis, accounting for all patients randomized, was therefore required to obtain a valid result. If per-protocol analyses were performed, the intervention benefit may have been overestimated, which is why intention-to-treat analysis is generally the most reliable strategy for performing the analyses in systematic reviews and randomized controlled trials.

### 10.5.8 SYSTEMATIC REVIEWS

It is debatable whether large randomized controlled trials or systematic reviews provide the best evidence when comparing different interventions. Large randomized controlled trials are often considered the most reliable sources of evidence for assessment of intervention effects. However, a number of cohort studies of hepato-biliary and gastroenterological trials suggest that many trials are too small and have inadequate bias control [26, 31, 45]. The size of the individual trials suggests that several have inadequate statistical power and are too small for identification of significant intervention benefits. One way to overcome these problems is to perform a systematic review using meta-analysis of the identified randomized controlled trials. In meta-analyses, the results of individual trials are combined in a common statistical analysis to increase the statistical power of inferences. Traditional reviews often count the number of supportive trials and choose the view receiving most votes. This may lead to false-negative conclusions if trials are underpowered. In a systematic review, a meta-analysis of 12 randomized controlled trials was performed to compare the effect of interferon with ribavirin versus interferon alone for patients with chronic active hepatitis C, who had not previously responded to antiviral therapy [10]. The primary outcome was sustained clearance of the hepatitis C virus. The sample sizes of the included trials suggested that no single trial had sufficient statistical power to detect clinically relevant differences in treatment effects, while three trials in the systematic review showed a statistically significant difference, suggesting that interferon with ribavirin was the most effective treatment for viral clearance. A narrative review, counting the number of positive and negative trials, may, therefore, reach the conclusion that no significant differences in effect can be

identified. However, when the results of the individual trials were combined in a meta-analysis, the results clearly showed that adding ribavirin to interferon significantly increased the chance of achieving a virological response.

Performing a meta-analysis has other potential advantages than increasing statistical power. The combination of several trials in a meta-analysis increases the extent to which results can be generalized. Furthermore, systematic reviews and meta-analyses make it possible to identify publication bias, or other biases, and to evaluate the risk of overestimated intervention effects due to inadequate bias control. The main disadvantage of systematic reviews is related to their observational design. Subgroup analyses in systematic reviews generally require prospective evaluation [14]. Methods for identification and selection of trials are necessary to avoid bias. If bias in the individual randomized controlled trials remains undetected [46–48], the results of the meta-analysis may be false positive. Therefore, some systematic reviews may remain inconclusive (in spite of statistically significant results) if trials with inadequate bias control are the only available. In a systematic review of randomized controlled trials, the effect of antibiotics and nonabsorbable disaccharides on the development of hepatic encephalopathy was assessed [49]. The results of the meta-analyses of trials identified and included showed that antibiotics had a significantly more positive effect on hepatic encephalopathy as compared to nonabsorbable disaccharides. However, in the included trials the quality of bias control was inadequate and it was impossible to make recommendations useful for clinical practice, although the findings were both clinically and statistically significant.

### 10.5.9 PUBLICATION BIAS AND RELATED BIASES

Observational studies have demonstrated that clinical trials with a positive outcome (i.e., demonstrating a statistically significant superiority of the intervention) are significantly more likely to be published than negative trials showing no statistically significant differences between the experimental intervention versus the placebo or other comparative intervention [22]. Such selective publication of trials is known as publication bias. The risk of publication bias is related to the sample size. Small trials with negative results tend to remain unpublished, while, for example, hepatobiliary randomized controlled trials are cited more often if the reported results show a statistically significant difference. Publication bias and related biases affect the possibility of identifying the trial. Thus, trials with positive results are more likely to be identified and included in, for example, systematic reviews and meta-analyses. Such biases may influence our inferences about intervention effects so that the benefits are overestimated. Different proposals have been made to allow identification of publication bias in meta-analyses so that the results and conclusions may be adjusted for the risk of bias.

### 10.5.10 CONCLUDING REMARKS

Evidence-based medicine—although extremely time consuming—is gradually replacing the traditional experience-based clinical practice. The number of randomized controlled trials and observational studies performed within the fields of

gastroenterology and hepatology is steadily increasing. The first step for the evidence-based practitioner is to identify the relevant clinical trial. The second step is to assess the internal validity of the research identified, and the final step is to evaluate its external validity so that the evidence may be used in clinical decision making. A number of methodological studies including clinical trials from several disease areas have been made to develop general guidelines for the validity assessment. These guidelines also apply to clinical trials in gastroenterology. Sample size calculations, randomization methods, blinding, and follow-up are important components in the maintenance of adequate bias control, which is important both in the planning stage and during subsequent interpretation of the results. The general methodology is therefore relevant, not only for researchers but also for clinicians using evidence-based medicine in daily practice. Nevertheless, a number of surveys suggest that there may be considerable gaps between evidence and practice. One such survey included specialists in gastroenterology, hepatology, and internal medicine [50]. The survey showed that several treatments and diagnostic procedures were still used, although there was no significant evidence to support their use. On the other hand, some of the treatments based on statistically and clinically relevant beneficial effects in systematic reviews of randomized controlled trials were not used. Additional measures aiming at bridging the gap between evidence and clinical practice are therefore warranted. Also a number of initiatives have been performed to improve the quality of clinical trials, for example, the Baveno conferences on trials in portal hypertension [51]. Hopefully, similar initiatives will be made in other areas of gastroenterology and hepatology to improve the quality of health care.

## REFERENCES

1. Fearnhead, N. S., Chowdhury, R., Box, B., et al. (2006), Long-term follow-up of stricture-plasty for Crohn's disease, *Br. J. Surg.*, 93, 475–482.
2. Xu, W., Tamim, H., Shapiro, S., et al. (2006), Use of antidepressants and risk of colorectal cancer: A nested case-control study, *Lancet Oncol.*, 7, 301–308.
3. Shehendere, R. L., Thompson, N., Mansfield, J. C., et al. (2005), Adenocarcinoma as a complication of small bowel Crohn's disease, *Eur. J. Gastroenterol. Hepatol.*, 17, 1255–1257.
4. Millonig, G., Kern, M., Ludwiczek, O., et al. (2006), Subfulminant hepatitis B after infliximab in Crohn's disease: Need for HBV-screening? *World J. Gastroenterol.*, 12, 974–976.
5. Grønbaek, K., Krarup, H. B., Møller, H., et al. (1999), Natural history and etiology of liver disease in patients with previous community-acquired acute non-A, non-B hepatitis. A follow-up study of 178 Danish patients consecutively enrolled in The Copenhagen Hepatitis Acuta Programme in the period 1969–1987, *J. Hepatol.*, 31, 800–807.
6. Uriz, J., Gines, P., Ortega, R., et al. (2000), Terlipressin plus albumin infusion: An effective and safe therapy of hepatorenal syndrome. *J. Hepatol.*, 33, 43–48.
7. Shawcross, D. L., Davies, N. A., Mookerjee, R. P., et al. (2004), Worsening of cerebral hyperemia by the administration of terlipressin in acute liver failure with severe encephalopathy, *Hepatology*, 39, 471–475.
8. Lee, M. Y., Chu, C. S., Lee, K. T., et al. (2006), Terlipressin-related acute myocardial infarction: A case report and literature review, *Kaohsiung J. Med. Sci.*, 20, 604–608.



9. Lillemoe, K. D., Melton, G. B., Cameron, J. L., et al. (2000), Postoperative bile duct strictures: Management and outcome in the 1990s, *Ann. Surg.*, 232, 430–441.
10. Kjaergard, L. L., Krogsgaard, K., and Gluud, C. (2001), Interferon alfa with or without ribavirin for chronic hepatitis C: Systematic review of randomized trials, *BMJ*, 323, 1151–1155.
11. Sacks, H., Chalmers, T. C., and Smith, H., Jr. (1982), Randomized versus historical controls for clinical trials, *Am. J. Med.*, 72, 233–240.
12. White, E., Hunt, J. R., and Casso, O. (1998), Exposure measurement in cohort studies: The challenges of prospective data collection, *Epidemiol. Rev.*, 20, 43–56.
13. Randomization to protect against selection bias in healthcare trials (Cochrane Methodology Review). (2002), *In: The Cochrane Library*, Wiley, Chichester, UK.
14. Kjaergard, L. L., Liu, J. P., Als-Nielsen, B., et al. (2003), Artificial and bioartificial support systems for acute and acute-on-chronic liver failure: A systematic review, *JAMA*, 289, 217–222.
15. Ioannidis, J. P., Haidich, A. B., Pappa, M., et al. (2001), Comparison of evidence of treatment effects in randomized and nonrandomized studies, *JAMA*, 286, 821–830.
16. Deeks, J. J., Dinnes, J., D'Amico, R., et al. (2003), Evaluating non-randomized intervention studies, *Health Technol. Assess.*, 7, 1–173.
17. Altman, D. G. (1991), Randomization, *BMJ*, 302, 1481–1482.
18. Swinger, G. H., and Zwarenstein, M. (2000), An effectiveness trial of a diagnostic test in a busy outpatients department in a developing country: Issues around allocation concealment and envelope randomization, *J. Clin. Epidemiol.*, 53, 702–706.
19. Schulz, K. F. (1995), Subverting randomization in controlled trials, *JAMA*, 274, 1456–1458.
20. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273, 408–412.
21. Moher, D., Pham, B., Jones, A., et al. (1998), Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, 352, 609–613.
22. Easterbrook, P. J., Berlin, J. A., Gopalan, R., et al. (1991), Publication bias in clinical research, *Lancet*, 337, 867–872.
23. Cochrane Reviewers' Handbook 4.2.1 (updated December 2003) (2004), *In The Cochrane Library*, Wiley, Chichester, UK.
24. Kjaergard, L. L., Villumsen, J., and Gluud, C. (2001), Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses, *Ann. Intern. Med.*, 135, 982–989.
25. Gluud, L. L. (2006), Bias in clinical intervention research, *Am. J. Epidemiol.*, 163, 493–501.
26. Kjaergard, L. L., Frederiksen, S. L., and Gluud, C. (2002), Validity of randomized clinical trials in gastroenterology from 1964 to 2000, *Gastroenterology*, 122, 1157–1160.
27. Kjaergard, L. L., and Gluud, C. (2002), Funding, disease area, and internal validity of hepatobiliary randomized clinical trials, *Am. J. Gastroenterol.*, 97, 2708–2713.
28. Devereaux, P. J., Choi, P. T., El-Dika, S., et al. (2005), An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods, *J. Clin. Epidemiol.*, 57, 1232–1236.
29. Schulz, K. F., and Grimes, D. A. (2002), Blinding in randomized trials: Hiding who got what, *Lancet*, 359, 696–700.



30. Pildal, J., Chan, A. W., Hróbjartsson, A., et al. (2005), Comparison of descriptions of allocation concealment in trial protocols and the published reports: Cohort study, *BMJ*, 330, 1049–1052.
31. Kjaergard, L. L., Nikolova, D., and Gluud, C. (1999), Randomized clinical trials in hepatology: Predictors of quality, *Hepatology*, 30, 1134–1138.
32. Campbell, I. A., Lyons, E., and Prescott, R. J. (1987), Stopping smoking. Do nicotine chewing-gum and postal encouragement add to doctors' advice, *Practitioner*, 231, 114–117.
33. Karlowski, T. R., Chalmers, T. C., Frenkel, L. D., et al. (1975), Ascorbic acid for the common cold. A prophylactic and therapeutic trial, *JAMA*, 231, 1038–1042.
34. Walter, S. D., Awasthi, S., and Jeyseelan, L. (2005), Pre-trial evaluation of the potential for unblinding in drug trials: A prototype example, *Contemp. Clin. Trials*, 26, 459–468.
35. Pocock, S. J. (1996), *Clinical Trials—A Practical Approach*, Wiley, Chichester, UK.
36. Freiman, J. A., Chalmers, T. C., Smith, H., et al. (1978), The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials, *N. Engl. J. Med.*, 299, 690–694.
37. Moher, D., Dulberg, C. S., and Wells, G. A. (1994), Statistical power, sample size, and their reporting in randomized controlled trials, *JAMA*, 272, 122–124.
38. Sung, J. J., Chung, S. C., Lai, C. W., et al. (1993), Octreotide infusion or emergency sclerotherapy for variceal haemorrhage, *Lancet*, 342, 637–641.
39. Altman, D. G., and Bland, J. M. (1995), Absence of evidence is not evidence of absence, *BMJ*, 311, 485.
40. International Conference on Harmonisation Expert Working Group (1997), International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH harmonised tripartite guideline. Guideline for good clinical practice. 1997 CFR & ICH Guidelines, Barnett International/PAREXEL, Philadelphia.
41. Corrigan, J. D., Harrison-Felix, C., Bogner, J., et al. (2003), Systematic bias in traumatic brain injury outcome studies because of loss to follow-up, *Arch. Phys. Med. Rehabil*, 84, 153–160.
42. Egger, M., Jüni, P., Bartlett, C., et al. (2001), Value of flow diagrams in reports of randomized controlled trials, *JAMA*, 285, 1996–1999.
43. Montori, V. M., and Guyatt, G. H. (2001), Intention-to-treat principle, *CMAJ*, 165, 1339–1341.
44. Millis, S. R. (2003), Emerging standards in statistical practice: Implications for clinical trials in rehabilitation medicine, *Am. J. Phys. Med. Rehabil.*, 82, S32–S37.
45. Gluud, C., and Kjaergard, L. L. (2001), Quality of randomized clinical trials in portal hypertension and other fields of hepatology, in Franchis, R., Ed., *Portal Hypertension III. Proceedings of the Third Baveno International Consensus Workshop on Definitions, Methodology, and Therapeutic Strategies*, Blackwell Science, Oxford.
46. Egger, M., Smith, G. D., and Phillips, A. N. (1997), Meta-analysis: Principles, and procedures, *BMJ*, 315, 1533–1537.
47. Smith, D., and Egger, M. (2000), Meta-analysis. Unresolved issues and future developments, *BMJ*, 316, 221–225.
48. Egger, M., and Smith, G. D. (1997), Meta-analysis. Potentials and promise, *BMJ*, 315, 1371–1374.
49. Als-Nielsen, B., Gluud, L. L., and Gluud, C. (2004), Non-absorbable disaccharides for hepatic encephalopathy: Systematic review of randomized trials, *BMJ*, 328, 1046–1050.

50. Kurstein, P., Gluud, L. L., Willemann, M., et al. (2006), Agreement between reported use of interventions for liver diseases and research evidence in Cochrane systematic reviews, *J. Hepatol.*, 43, 984–989.
51. De Franchis, R. (1996), *Portal Hypertension II Proceesings of the Second Baveno International Consensus Workshop on Definitions, Methodology and Therapeutic Strategies*, Blackwell Science, Oxford.



# 10.6

---

## *Gynecology Randomized Control Trials*

**Khalid S. Khan,<sup>1</sup> Tara Selman,<sup>1</sup> and Jane Daniels<sup>2</sup>**

<sup>1</sup>*Birmingham Women's Hospital, Birmingham, United Kingdom*

<sup>2</sup>*Clinical Trials Unit and Academic Department of Obstetrics and Gynaecology,  
University of Birmingham, Birmingham, United Kingdom*

### **Contents**

|          |                                                                    |     |
|----------|--------------------------------------------------------------------|-----|
| 10.6.1   | Introduction                                                       | 519 |
| 10.6.2   | Drug Development Process in Gynecology                             | 520 |
| 10.6.2.1 | Phase I: Safety and Pharmacokinetics in Healthy Population         | 520 |
| 10.6.2.2 | Phases II and III: Efficacy and Effectiveness in Target Population | 520 |
| 10.6.2.3 | Phase IV: Long-Term Safety and Effectiveness                       | 520 |
| 10.6.3   | Expected Effect Size and Sample Sizes of Clinical Trials           | 521 |
| 10.6.4   | Avoidance of Systematic Biases                                     | 523 |
| 10.6.5   | Choice of Appropriate Outcome Measures                             | 525 |
| 10.6.6   | Choice of Appropriate Analysis                                     | 527 |
| 10.6.7   | Multicentered Trials                                               | 527 |
| 10.6.8   | Meta-analysis                                                      | 527 |
| 10.6.9   | Conclusion                                                         | 527 |
|          | References                                                         | 528 |

### **10.6.1 INTRODUCTION**

Developments in reproductive health are often low profile, but the sheer number of women with gynecological problems means that, if effective interventions exist, a massive overall health and financial benefit can be expected in the population. This chapter focuses on benign gynecology rather than gynecological oncology. Benign gynecological conditions such as chronic pelvic pain, heavy

menstrual bleeding, and subfertility may be treated with suboptimal therapies if there is not a commitment to the development of new drugs and a thorough evaluation of the relative merits of existing ones. There remains a burden of disease in women's health that could be alleviated if clinical care was built around evidence-based interventions.

Randomized control trials (RCTs) are widely accepted as a gold standard for scientific evaluation of all treatments. This is as true of gynecology as other specialties, yet there are much fewer trials in this speciality by comparison [1]. In addition to political and financial barriers to drug evaluation, there are methodological problems to be overcome. This chapter will highlight the issues regarding the design and analysis of RCTs that are pertinent to benign gynecology.

## **10.6.2 DRUG DEVELOPMENT PROCESS IN GYNECOLOGY**

The process of evaluation of new medicinal products will follow the same stepwise escalation of evaluative techniques as within other areas, but there are specific issues to consider at each phase within gynecology.

### **10.6.2.1 Phase I: Safety and Pharmacokinetics in Healthy Population**

The main difficulty for these trials is to find a "normal" population in gynecology. A large number of premenopausal women are either taking some form of hormonal contraception, so not undergoing normal menstruation, or actively trying to become pregnant, and therefore they would not want to expose themselves or their fetus to the risk of an intervention of unproven safety. Trials limited to postmenopausal women can only comment on the safety and bioavailability of that drug in that limited population.

### **10.6.2.2 Phases II and III: Efficacy and Effectiveness in Target Population**

Generally, phase II trials recruit a narrow population with the condition of interest, with no comorbidities, and compare the new intervention against placebo. There are ethical considerations regarding placebo where active, albeit not very effective, treatments exist. The acceptability of the route of administration may be of equal importance to its effect, especially if it affects compliance with a drug regime. So whereas oral contraceptives may be as efficacious as injectable alternatives, women may consider the latter preferable and more reliable.

### **10.6.2.3 Phase IV: Long-Term Safety and Effectiveness**

Interventions for some gynecological conditions, such as endometriosis, may require long-term use without a reduction in effectiveness or tolerability. Long-term follow-up of all trial patients is required to assess overall effectiveness without bias from participant withdrawal. Assessment of safety and teratogenicity may require use of systematic reviews to collate low-frequency events.

### 10.6.3 EXPECTED EFFECT SIZE AND SAMPLE SIZES OF CLINICAL TRIALS

In gynecology, it is realistic to expect small to moderate effects, even when compared to a placebo, to be clinically worthwhile. However, because the conditions are often extremely common, for example, menorrhagia or chronic pelvic pain, when aggregated over the population, even a small effect can result in a huge impact on women's quality of life and productivity. Hence trials have to be large to be able to show small differences in effect with sufficient power.

If the process of randomization is strictly performed, the two groups of an RCT should be equivalent, and then it follows that any difference in outcome should be due to chance or genuine treatment effect. This still leaves the difficulty of distinguishing what is due to chance. It is usually very easy to spot a highly successful drug in a common condition with serious outcomes such as myocardial infarction. However, this is not generally the case in gynecology, where the effect size is often moderate or small, especially when looking at pain outcomes and when the condition is chronic or self-limiting. In these cases trials need to be large. Small trials will sometime give nonsignificant results. This may be taken to mean that the treatment does not work, when in fact there were too few participants in the trial to demonstrate a small effect [2]. This is why estimates of a treatment's effectiveness are preferable to simply quoting  $p$  values, to highlight the degree of uncertainty around any estimate of effect.

For a given postulated treatment effect, the power of a trial is the probability that a significant result will be obtained in the trial, if the treatment effect is as predicted. Given the expense and immense effort, it becomes unacceptable in gynecology to run a trial that has only a 50, 60, or even 70% chance of spotting a genuine treatment effect. A "power" of 80% is, by convention, the minimum required to be confident of avoiding a falsely negative result and 90% power is preferred. Increasing the power of the trial to detect a difference means increasing the sample size. For example, if the number of women in a particular group with dysmenorrhea who eventually opt for a hysterectomy is 50%, in order to have a 90% chance of detecting whether a new treatment reduces this proportion by half, to 25%, a trial would need to recruit 150 women, whereas at 80% power, it would require 117 women. Unfortunately, intervention in gynecology is seldom that effective.

When anticipating the treatment effect, and hence estimating the sample size, one should consider what the minimum clinically important difference between the two treatments might be, or in other words, the degree of improvement that would lead to a change in the clinician behavior or acceptance by women. For example, if a 25% proportional increase in fertilization rate is considered beneficial and there is a 70% failure rate (or 30% fertilization rate), then to demonstrate this effect with a new therapy, such as low-dose aspirin, would require 256 women to achieve a 53% failure rate, as shown in Table 1. Note that this is not the same as reducing the failure rate by 25 percentage points from 70 to 45%, a proportionally much bigger drop requiring about 132 women. In a high prevalence condition such as dysmenorrhea [3], a reduction from 50% opting for hysterectomy to 45% would still be worthwhile, but would require over 4000 women and hence suddenly a very large trial, requiring an immense effort to complete.

**TABLE 1 Sample Sizes for a Range of Differences between Control and Experimental Arm Rates**

| Control Arm Rate | Proportional Reduction of Rate <sup>a</sup> |             |                       |             |                       |             |
|------------------|---------------------------------------------|-------------|-----------------------|-------------|-----------------------|-------------|
|                  | 50%                                         |             | 33%                   |             | 25%                   |             |
|                  | Experimental Arm Rate                       | Sample Size | Experimental Arm Rate | Sample Size | Experimental Arm Rate | Sample Size |
| 0.80             | 0.40                                        | 46          | 0.53                  | 94          | 0.60                  | 164         |
| 0.70             | 0.35                                        | 62          | 0.46                  | 132         | 0.53                  | 256         |
| 0.60             | 0.30                                        | 84          | 0.40                  | 194         | 0.45                  | 346         |
| 0.50             | 0.25                                        | 116         | 0.33                  | 262         | 0.375                 | 494         |
| 0.40             | 0.20                                        | 164         | 0.26                  | 390         | 0.30                  | 712         |
| 0.30             | 0.15                                        | 242         | 0.20                  | 588         | 0.23                  | 1246        |

<sup>a</sup>Assuming 80% power.**TABLE 2 Sample Sizes for Different Standardized Effect Sizes, Assuming Two Equal Groups**

| Power | Standardized Effect Size |      |      |
|-------|--------------------------|------|------|
|       | 0.5                      | 0.33 | 0.20 |
| 80%   | 128                      | 292  | 788  |
| 90%   | 172                      | 388  | 1054 |

When considering continuous outcome measures, the sample size calculation is not based on proportions but on anticipated differences between the mean and the outcome for each group, so a slightly different approach is taken. It is generally accepted that a standardized difference of 0.2 standard deviations can be considered to be a small effect size, 0.5 a medium effect and 0.8 a large effect. As differences between treatments are unlikely to be large, and small to moderate effects can be clinically significant, it is possible to calculate sample sizes for clinical trials based on standardized effect sizes. Table 2 shows the sample sizes for various effect sizes at 80 and 90% power.

If the standard deviation of the outcome measure is known, the difference in means can be predicted too. For example, if the mean length of stay in a hospital following hysterectomy is 5.3 days and standard deviation is 1.3, and the sample size was sufficient to detect a small to moderate effect size of 0.33 standard deviations, a trial of 292 women comparing hysterectomy with uterine artery embolization would be powered to detect increase or reduction in length of stay of at least  $1.3 \times 0.33 = 0.43$  days.

One of the alternatives used to avoid very large sample sizes, especially when comparing a new with a standard intervention, is to conduct a “noninferiority” trial. Here the objective is not to demonstrate superiority, which can require large numbers, but to establish that the effect of the experimental intervention in comparison to the standard is not more than some pre-stated small difference, called a noninferiority margin. Hence here clinicians determine the amount of noninferiority they are willing to accept. With the above dysmenorrhea example, if clinicians



were prepared to accept the new intervention does not increase the rate of hysterectomy by more than 5%, then the sample size would be 2460. As the direction of effect being assessed is one sided (the experimental intervention is not inferior to the control intervention), a one-sided hypothesis test is performed. There are limitations with this method: If the experimental group performs better than the control, it cannot be concluded to be superior as the trial was not designed to test this hypothesis and sample sizes are very dependent on the margin at which one accepts noninferiority. So it is important that the noninferiority margin selected is small enough to not exceed that which has clinical relevance, and that the standard intervention is already established as superior to placebo [4].

#### 10.6.4 AVOIDANCE OF SYSTEMATIC BIASES

In order to detect moderate, but meaningful, differences between treatments, it is important that causes of bias are minimized as much as possible. Biases almost universally tend to exaggerate the true effects of a treatment [5–7]. Figure 1 demonstrates the key biases possible in a badly designed trial. The importance of a sound randomization process to avoid selection bias cannot be overemphasised. This will ensure the comparability of the treatment groups at the start of the trial. An independent telephone service or coded drug containers are appropriate randomization

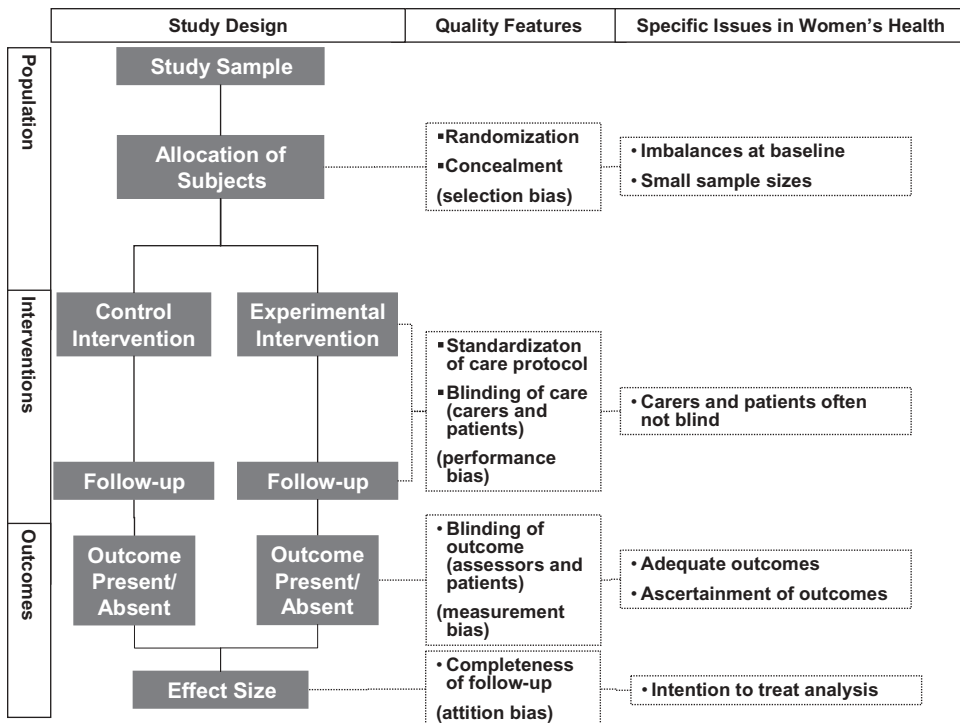


**FIGURE 1** Outline of trial, its quality features that minimize the risk of bias and issues specific to trials in benign gynecology.

methods; tossing a coin, odd and even hospital identification numbers, or allocating participants in order of arrival would not be. The point is not the degree of “randomness” but the extent to which the next allocation cannot be predicted, or in other words, the ability to conceal the allocation from the clinician recruiting the patient. For concealment to be maintained, the randomization process must be indefatigable. For example, if envelopes containing the randomized allocation are used, it is possible for the researchers in this case to manipulate the order of opening of the envelopes, or even resealing opened envelopes that do not contain the preferred allocation, thereby introducing selection bias. For the trial to be concealed, the randomized allocation should be provided by a third party once all eligibility criteria have been confirmed and the participant committed to the trial.

Even the strictest randomization process can be undermined if patients do not receive the treatment to which they are allocated. Postrandomization withdrawals can be minimized by careful screening of potential patients against the eligibility criteria. However, compliance with allocations may diminish if treatment is not instigated soon after randomization. This can be a particular problem in fertility studies, where it is not unusual to randomize women at the start of a cycle, but due to many reasons the embryo transfer procedure may not be carried out in that cycle, and alternative methods are chosen for subsequent cycles [4].

Blinding should not be confused with concealment. Blinding is a method in which trials can attempt to eliminate both performance and measurement biases, by keeping participants and clinicians unaware of the treatment allocation after randomization. Performance bias can occur if there are differences in groups related to co-interventions or supportive care, making it harder to disaggregate the effect of each intervention. Should this potentially be a problem, a standardized care protocol should be agreed upon for all patients and, of course, details of other interventions recorded to detect significant deviations between the groups. Detection or measurement bias can arise if there are differences in the way outcomes are measured or interpreted. Objective measurements are less prone to measurement bias than subjective measurements but may not capture the outcome of interest.

There are different levels of blinding:

- Single blind—Usually when the patient does not know to which treatment arm she has been allocated. Gynecological surgical procedures can be blinded from the woman by use of drapes to obscure her view or sham incisions.
- Double blind—When both clinician and patient do not know which treatment is being given, as in the typical placebo-controlled trial.
- Triple blind—Neither the clinician, patient, or the person performing the outcome assessment are aware of the treatment. This is the most difficult to achieve, but potentially some gynecological assessments, for example, urodynamics, could be blinded to allocation.

Patients should always be analyzed in the group to which they were randomized in an “intention to treat” analysis, as this helps retain the benefits achieved by randomization. If, for any reason, participants move from one arm of the trial into the other, all participants must be analyzed in the group to which they were originally allocated and not in their new group. While intention to treat analysis may underestimate the true treatment effect if it is “diluted” by cross-overs, it will more closely

reflect the clinical reality outside of the trial and provide an anticipated effect of pursuing a particular treatment policy.

Attrition bias can arise if there are systematic differences in the degree of follow-up of trial participants. This can arise if a more intensive follow-up for one treatment group is built into the protocol or where there are different rates of participant dropout between the groups. This can be a particular problem in some gynecological conditions such as chronic pelvic pain, where women may decide that the treatment is ineffective and opt out of the trial to seek alternative therapies. Trials where a drug has side effects as troubling as the primary complaint will experience a high degree of noncompliance. This has been observed in a trial of the levonogestrel-releasing intrauterine system when used for heavy menstrual bleeding. The device often has an unpredictable effect on the menstrual cycle in the first 6 months of use that women decide it is preferable to have it removed than wait to see if it has a beneficial effect on their periods.

In these circumstances, every effort should be made to continue to collect follow-up data on all participants, regardless of compliance, again so that an intention to treat analysis can be performed. However, it is inevitable that some women will withdraw their consent to provide data to the trial. One solution is to perform a per-protocol analyses excluding patients with missing data from the analyses. Alternatively, imputation of missing data may be considered. Possible imputation strategies include carrying forward the last observation to the missing time point or estimating the most likely outcome, based on the outcome of other participants in the trial. If patients with missing data are mainly outliers, the precision of the effect size by per-protocol analyses may be increased. But if losses to follow-up are related to prognostic factors, side effects, or lack of response to treatment, per-protocol analyses may overestimate the treatment effects [8]. Reasons for loss to follow-up should be recorded, if possible, to determine the degree of differential loss to follow-up and to establish whether such losses are random or will bias the results.

Unfortunately, following patients through to the prespecified end of the trial is a particular problem with trials of benign gynecology, as participants tend to be otherwise healthy, not have a life-threatening condition and are relatively young and mobile. To avoid, or at least make every attempt to reduce, the number of patients lost to follow-up, multiple identifiers and contact details for participants should be taken at the start of the trial. Follow-up by postal questionnaires direct to the participants is notoriously difficult to sustain and yet is the most appropriate method of collecting quality of life information. A meta-analysis of methods of improving the response rate to postal questionnaires identified a number of useful strategies, such as contact before the questionnaire is sent, provision of a prepaid envelope, and using a short, interesting questionnaire, in addition to obvious monetary incentives—all of these should be considered at the outset of the trial [9].

### 10.6.5 CHOICE OF APPROPRIATE OUTCOME MEASURES

Unlike cancer trials, where we are interested primarily in mortality, or obstetric trials where perinatal mortality is a suitable outcome measure, in gynecology the effect of a drug is often on disease-specific symptoms. One has to decide on the most

important outcome measure to use in determining the effectiveness of a drug, that is, whether the treatment is likely to have an effect on that outcome and how the outcome is to be measured.

Ideally, the primary outcome measure should be as important to the women as to the clinician or policy maker. Examples include measuring pain on a visual analog scale for dysmenorrhea [10], assessing sexual function by means of a specific questionnaire [11] in vulvodynia, or counting days of absence from work or usual activities for premenstrual dysphoria. Use of a generic quality of life questionnaire will allow the impact of a treatment to be directly compared to other conditions but tends to focus on physical dimensions of quality of life and may not be sufficiently sensitive to respond to the changes in different aspects of life quality experienced by treating benign gynecological conditions [12]. Disease or symptom-specific quality of life questionnaires capture the more subtle features of a benign condition, particularly those that are important to the patient. The number of available instruments is increasing, but not all exhibit sound psychometric and measurement characteristics; so they should be reviewed for quality before use [13].

Although the main aim of any trial is to determine the effect of a treatment on the primary outcome measure, there are usually other pertinent criteria that are also of interest, such as sexual function or health service resource usage. While the trial sample size is calculated in order to detect a clinically meaningful difference in the primary outcome, the trial may not have sufficient power to detect significant differences in secondary measures. If one looks at too many outcomes simultaneously, it is likely that one will turn out to be statistically significant, due to the play of chance, even in a trial of an ineffective intervention. Thus primary and secondary outcomes should be defined in advance and not chosen once the analysis has been done, on the basis of statistical significance.

Another issue in gynecological research is that the outcome of interest can occur a long time in the future, for example, the ultimate need for hysterectomy. It is often impractical to wait many years to answer such questions, so an alternative is to use a surrogate outcome that predicts the future outcome of interest, for example, using a laboratory marker [14]. Intermediate outcomes are also sometimes proposed as surrogates are for low-frequency events that would otherwise require prohibitively large trials to demonstrate an effect. However, this is not a wholly satisfactory method. If surrogate outcomes are used, it is necessary to ensure that they correlate with clinically relevant measures and they capture the whole of the clinically relevant effect. Hence, ideally, any surrogate should be previously validated. Polycystic ovary syndrome (PCOS) is a common cause of infertility; so the aim of any intervention would be to increase the chance of delivering a healthy baby. Intermediate outcomes are pregnancy and ovulation rates, and there are a host of biochemical and biometric surrogates. Yet the best correlation between surrogate outcomes and the desired clinical endpoint is the correlation between ovulation and pregnancy in women with PCOS taking clomiphene citrate [15] or metformin [16]. However, factors unrelated to PCOS, including male sperm quality and maternal age, will impact on live birth rate, as demonstrated by the meta-analysis of metformin in which there was no evidence of increased clinical pregnancy rate [16].

### 10.6.6 CHOICE OF APPROPRIATE ANALYSIS

Many of the common benign gynecological conditions are chronic and lack a definitive “endpoint.” Long-term observations of continuous outcomes such as severity of pain or menstrual blood loss are required. Choosing a particular time point for analysis of outcomes may be fairly arbitrary in the course of the condition. Measures that are collected at multiple time points from each trial participant are likely to be more closely related to each other than the variation between different participants. Multilevel modeling takes into account this hierarchy of data and gives the advantage of being able to estimate overall effect over time, utilizing all available information. An overall treatment effect, with confidence intervals can be estimated from the model.

### 10.6.7 MULTICENTERED TRIALS

The design, execution, and analysis of a trial might have done all that is possible to eliminate imprecision and biases. The results in this case will be reliable and “true” for the participants studied. But will such results also be generalizable to the wider population? One way to improve generalizability, or external validity, is to recruit a wide and heterogeneous population with the condition of interest, with few exclusion criteria. Another is to recruit from many clinical centers. This approach limits the effect of the peculiarities of single-center studies that make it difficult to replicate results in other settings. Increasing the number of recruitment centers also improves the ability to accrue large numbers rapidly within trials.

### 10.6.8 META-ANALYSIS

With the multitude of problems that can arise in gynecological trials, it is imperative that new data should be considered in relation to previous trials and the impact of the results discussed [17]. The best way to achieve this is through the use of meta-analysis to add the results of the trial to those already available. The larger the amount of evidence, the less likely there is to be overemphasis on any particular trial, which can give misleading findings. Investigation of subgroup effects is often not possible from published data meta-analysis due the manner in which studies are reported. Collection of individual patient data from the primary studies to perform a meta-analysis is the most reliable method of assessing the totality of the evidence overall and within subgroups, but this requires considerable effort to organize and goodwill on behalf of the original trial authors [18].

### 10.6.9 CONCLUSION

Trials in benign gynecology typically deal with chronic conditions where there is no definite endpoint but an expectation of gradual change on a continuous outcome measure. One exception is infertility studies where the outcome is a successful

pregnancy, but even here intermediate outcomes are often chosen. Most trials require the recruitment of large numbers to evaluate small to moderate effect sizes reliably. Outcome measures need to employ clinically important measures that impact life facets important to patients. In this regard, disease-specific quality of life tools require validation. Data-analytic approaches need to employ intention to treat principles using survival analyses when outcome is time dependent (infertility) or multilevel modeling of repeated measures in chronic conditions such as chronic pelvic pain. In summary, large, multicenter trials with simple entry criteria, robust randomization, consistent execution, and appropriate analysis are required.

## REFERENCES

1. Edwards, A., and Lilford, R. J. (2005), *National Clinical Trials Capacity Review*, NCCRC, UK.
2. Altman, D. G., and Bland, J. M. (1995), Statistics notes: Absence of evidence is not evidence of absence, *BMJ*, 311(7003), 485.
3. Latthe, P., Latthe, M., Say, L., et al. (2006), WHO systematic review of prevalence of chronic pelvic pain: A neglected reproductive health morbidity, *BMC Public Health*, 6(1), 177.
4. Daya, S. (2006), Methodological issues in infertility research, *Best Practice Res. Clin. Obstetrics Gynecol.*, 20(6), 779–797.
5. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273(5), 408–412.
6. Schulz, K. F., and Grimes, D. A. (2002), Allocation concealment in randomized trials: Defending against deciphering, *Lancet*, 359(9306), 614–618.
7. Schulz, K. F. (1995), Subverting randomisation in controlled trials, *JAMA*, 274(18), 1456–1458.
8. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273(5), 408–412.
9. Edwards, P., Roberts, I., Clarke, M., et al. (2002), Increasing response rates to postal questionnaires: Systematic review, *BMJ*, 324(7347), 1183.
10. Carlsson, A. M. (1983), Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale, *Pain*, 16(1), 87–101.
11. Thirlaway, K., Fallowfield, L., and Cuzick, J. (1996), The Sexual Activity Questionnaire: a measure of women's sexual functioning, *Qual. Life Res.*, 5(1), 81–90.
12. Lamping, D. L., Rowe, P., Clarke, A., et al. (1998), Development and validation of the Menorrhagia Outcomes Questionnaire, *Br. J. Obstet. Gynaecol.*, 105(7), 766–779.
13. Clark, T. J., Khan, K. S., Foon, R., et al. (2002), Quality of life instruments in studies of menorrhagia: A systematic review, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 104, 96–104.
14. Prentice, R. L. (1989), Surrogate endpoints in clinical trials: Definition and operational criteria, *Stat. Med.*, 8(4), 431–440.
15. Imani, B., Eijkemans, M. J., te Velde, E. R., et al. (2002), A nomogram to predict the probability of live birth after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility, *Fertil. Steril.*, 77(1), 91–97.

16. Lord, J. M., Flight, I. H. K., and Norman, R. J. (2003), Insulin-sensitising drugs (metformin, troglitazone, rosiglitazone, pioglitazone, D-chiro-inositol) for polycystic ovary syndrome, *Cochrane Database Syst. Rev.*, 2.
17. Young, C., and Horton, R. (2005), Putting clinical trials in context, *Lancet*, 366(9480) 107–108.
18. Clarke, M. J., and Stewart, L. A. (1994), Systematic reviews: Obtaining data from randomised controlled trials: How much do we need for reliable and informative meta-analyses? *BMJ*, 309(6960), 1007–1010.





# 10.7

---

## *Special Population Studies (Healthy Patient Studies)*

**Doris K. Weilert**

*Clinical Pharmacology, Quintiles, Inc., Kansas City, Missouri*

### **Contents**

|          |                                                           |     |
|----------|-----------------------------------------------------------|-----|
| 10.7.1   | Introductory Remarks                                      | 531 |
| 10.7.2   | General Considerations for All Special Population Studies | 532 |
| 10.7.3   | Geriatric Population                                      | 537 |
| 10.7.4   | Renal Impairment                                          | 539 |
| 10.7.4.1 | Design Considerations in RI                               | 541 |
| 10.7.4.2 | Special Considerations for Dialysis Patients              | 543 |
| 10.7.5   | Hepatic Impairment                                        | 546 |
| 10.7.6   | Women in Clinical Trials                                  | 550 |
| 10.7.7   | Ethnic Considerations                                     | 551 |
| 10.7.8   | Obesity                                                   | 555 |
| 10.7.9   | Conclusions                                               | 557 |
|          | References                                                | 557 |

### **10.7.1 INTRODUCTORY REMARKS**

After a new chemical entity (NCE) has successfully passed the hurdle of the first-in-man study(ies), the pharmacokinetic (PK) and pharmacodynamic (PD) characteristics of the NCE must be fully elucidated to support drug efficacy and drug safety claims for a regulatory submission. Clinical pharmacology/biopharmaceutics components include an understanding of the drug's (1) absorption, distribution, metabolism, and excretion (ADME) profile, (2) behavior at different dosage regimens

and/or dosage forms, (3) behavior under diverse or impaired physiological conditions (special populations), (4) potential interactions with coadministered medications or herbal supplements (DDI), and (5) specific PD characteristics that may affect the efficacy and/or safety of the drug. Clinical pharmacology/biopharmaceutics studies are generally conducted in healthy subjects unless the drug is a biological response modifier and cannot be studied in healthy subjects. Exceptions are special population studies which include subjects with physiological characteristics different from the standard healthy volunteer population or typical target patient population. PD studies may be conducted in either healthy subjects or a patient population depending on the PD endpoint of interest.

This chapter focuses on the diversity of physiological effects on NCE PK in people who distinguish themselves from the average adult patient population for which the NCE is being developed. The first section of this chapter provides a definition of criteria which define a special population and general study design considerations common to all special population studies. The subsequent sections address specific objectives, design considerations, statistical analysis, data interpretation, and regulatory requirements for studies conducted in geriatrics, renal impairment (RI), hepatic impairment (HI), women, ethnic, and obese populations.

### **10.7.2 GENERAL CONSIDERATIONS FOR ALL SPECIAL POPULATION STUDIES**

Special populations encompass a wide variety of people who are physiologically different from the average patient population for which the NCE is being developed and who are frequently not enrolled in sufficient numbers in routine phases I–III clinical trials with a PK sample collection component to allow appropriate characterization of NCE PK and potential risk factors. NCE exposure information in special populations (HI, RI, geriatrics, women, and race) is required as part of the drug registration package. The PK and/or PD data will be used to justify and negotiate the dosage regimen for populations at risk and are a critical component of the drug label. Depending on the PK/PD characteristics of the NCE, dose adjustments may be required in some populations while the use of the NCE may be restricted or contraindicated in others. Potential NCE PK/PD differences have been observed as a result of:

- Physiological progression in life (e.g., pediatric, adolescent, adult, or geriatric populations)
- Lifestyle (e.g., obesity, dietary habits)
- Disease state (e.g., renal or hepatic impairment)
- Gender (males versus females)
- Race (e.g., Japanese populations)
- Genetics/genomics (e.g., poor/extensive metabolizers, differences in receptor composition/affinity resulting in differences in efficacy or adverse-event profiles)

Most formal special population studies are conducted during the later part of phase II development or in parallel with phase III when the therapeutic dose range is

known and the likelihood is high that the NCE will be submitted for registration. Trials such as RI and HI studies are conducted in a multicenter setting, have slow recruitment rates (especially in the severely impaired populations), take a long time to complete, and tend to be expensive. For this reason, they are only conducted to complete the submission package, thus satisfying labeling requirements. Some studies are conducted early(ier) in the development program to either provide data for the subsequent phase II/III program or NCE development in another regions (e.g., Japan).

Dedicated/formal PK studies are routinely conducted in geriatrics, HI, or RI, while data on other special populations may come from integrated population pharmacokinetic (PopPK) analyses of clinical data. For example, formal gender comparison studies have become less common since the U.S. Food and Drug Administration (FDA) 1993 guideline to include women as part of the phase I clinical development and the advent of PopPK approaches [1]. When the nature of the NCE limits drug administration to the intended target population, the regulatory agencies recommend to include subpopulations in the integrated phase II/III development program or to conduct a smaller dedicated “phase I-like” trial in the patient subpopulation.

Study design considerations have many similarities across the various special populations (Table 1). Drug disposition in the special population (sometimes divided further into subpopulations) is compared to that of a control population which has undergone similar study procedures. Standard inclusion and exclusion criteria apply to all special population studies with some modification for elderly, renal, and hepatic impairment populations discussed below. Safety assessments are the same as in other later stage clinical PK study and consist of standard assessments such as physical examination, clinical laboratories (hematology, coagulation, serum chemistry, and urinalysis), vital signs, 12-lead electrocardiograms (ECGs), and adverse events (AEs) plus any safety parameters specific to the NCE. The frequency of safety assessments during the study conduct tends to be held to a minimum unless significantly higher exposure is expected or the drug is known to affect vital signs or cardiac conduction. In special cases, NCE-specific PD markers [e.g., central nervous system (CNS) tests, specific laboratories such as glucose or coagulation assessments] may be evaluated more extensively in order to determine whether PD differences are observed in the population of interest and whether the PD changes are a result of changes in drug disposition.

Most formal special population studies include a control group as one arm of the study design. While special population results can be compared to historical NCE PK data obtained in separate study(ies) of similar design, there is an inherent risk that the regulatory authorities do not agree with the selection of the control study and request other (less favorable) comparisons which then become part of the drug label. The regulatory authorities request control groups to be closely matched to either the special population or the intended patient population. Environmental or behavioral aspects such as diet and smoking habits may also be considered in the selection of the control population since these factors are known to affect NCE PK. It would be difficult to design one “fit-all” control study for the various special population comparisons. Based on the author’s experience, inclusion of a control group in each special population study is the preferred approach.

Special population studies are most commonly conducted as single-dose studies assuming the NCE PK is predictive of steady-state PK. Linear PK should be extended

**TABLE 1 Typical Design Features for Special Population Studies**

| Consideration For            | Geriatrics<br>>65 years                                                       | RI                  | HI                  | Gender           | Ethnic                           |
|------------------------------|-------------------------------------------------------------------------------|---------------------|---------------------|------------------|----------------------------------|
| Typical timing of study      | Phase II/III                                                                  | Phase III           | Phase III           | Phase II/III     | As needed in development program |
| General health status        | Healthy for age group                                                         | Underlying diseases | Underlying diseases | Healthy          | Healthy                          |
| Controls                     | <65 years                                                                     | Matched             | Matched             | Males            | Matched                          |
| Inclusion/exclusion criteria | Reasonably tight                                                              | Wide                | Wide                | Tight            | Tight                            |
| Comedications                | Limited                                                                       | Extensive           | Extensive           | Only "C/H"       | Only "C/H"                       |
| Confinement                  | Yes for SD                                                                    | Yes for SD          | Yes for SD          | Yes for SD       | Yes for SD                       |
| Meals                        | Fasted                                                                        | Fasted (+)          | Fasted (+)          | Fasted           | Fasted/other                     |
| PK blood sampling            | Extended                                                                      | Extended            | Extended            | Standard         | Standard (+)                     |
| PK urine sampling            | Sometimes                                                                     | Yes                 | Sometimes           | If key to NCE PK | If key to NCE PK                 |
| Safety assessments           | Minimal (+)                                                                   | Minimal (+)         | Minimal (+)         | Minimal          | Minimal (+)                      |
| Dose                         | Maximum therapeutic dose unless otherwise indicated                           |                     |                     |                  |                                  |
| Dosing frequency             | Single dose (SD) unless NCE or active metabolites have nonlinear PK           |                     |                     |                  |                                  |
| PD assessment                | If relevant to support NCE safety or to assess population-specific PD changes |                     |                     |                  |                                  |

**Note:**

*Matched:* Demographically matched to special population or indicated patient population.

*C/H:* Oral or depot contraceptives or hormonal replacement therapy; sometimes vitamins, herbal, and limited over counter medications permitted.

*Fasted:* For –10 hours to 4–5 hours postdose; (+) small snack may be required in more severely ill 1–2 hours prior to dose administration; (other) may receive other than standard meals.

*Extended:* For at least 24–48 hours beyond standard sampling times; (+) may need extension.

*Minimal:* (+) Two to three additional postdose vitals/ECG or more extensive if significant PK differences are expected.

to all analytes measured (parent drug and active/toxic metabolites). The single dose chosen can be the anticipated upper end of the therapeutic dose range, since age or disease state generally affects not  $C_{\max}$  but instead overall exposure [area under the curve (AUC)]. However, if the NCE has a narrow therapeutic window and/or a significant difference in exposure is expected, a dose in the lower to midrange of the therapeutic window may be more appropriate. For example, due to the lower body weight (WT) in the Japanese population, NCE  $C_{\max}$  may be different in addition to AUC, which may potentially require a lower dose in this population comparison.

A multiple-dose study may be required if the NCE (and/or metabolite) has either dose or time-dependent PK. Dose selection and dosing frequency should be carefully selected to avoid accumulation of the NCE and/or metabolites to unsafe levels. A dose escalation scheme may be employed to titrate to safe drug levels, or NCE PK parameters may be predicted by modeling the expected effect which the special population is thought to have on drug disposition.

Most of the special population studies, especially those following a single dose of NCE, are conducted under confinement. Subjects are screened and then return to the clinic either the day prior (preferred) or on the morning of the dose administration. Subjects are generally housed during the majority of PK sample

collections and may not be released until the study is completed (last PK sample collected). Healthy (male) volunteers are generally housed in a dormlike setting, while women and elderly tend to have a semiprivate setting. In these healthy populations it is acceptable that the study is conducted under restricted conditions and it is general practice to lock rest rooms (in the case of urine collection), limit access to water fountains, and have controlled access within the facility (e.g., to entertainment rooms). In RI and HI, subjects are managed in a hospital-like setting instead. Considering that many of the subjects are rather ill, every attempt is made to provide as much privacy and freedom (within study restrictions) to keep them comfortable.

General dietary considerations and restrictions are similar to those in any typical phase I study and should reflect the NCE requirements. In most cases, NCE PK is assessed under controlled fasted conditions (e.g., fasted for 10 hours prior to 4–5 hours following dose administration) to avoid introduction of additional variables into the study design. The RI and HI patients may require modifications to the composition of the meal (e.g., salt content) and/or permitted snacks. In a study comparing race (e.g., Japanese versus U.S. population), the content of meals may be purposefully different for the ethnic group of interest to maintain typical dietary habits as part of the factors influencing NCE PK.

The length and frequency of blood (and urine) collections depend on the expected outcome of the study. In geriatric, RI, and HI studies it is worthwhile to extend blood sampling for at least 24–48 hours beyond the sampling scheme employed in routine healthy volunteer studies, even if a minimal effect NCE PK is expected. If the special population is conducted as a multiple-dose study, the PK should be assessed both after single-dose administration and at steady state.

Special population studies are targeted to address potential “extreme” exposure conditions for the NCE which may alter levels of active/toxic metabolites relative to parent concentrations. Metabolites known to be involved in the efficacy and/or safety profile and presenting >10% of circulating parent levels should be considered for analysis. A 10% cutoff is a general guide only, since the decision to evaluate metabolites also depends on relative potency and/or the extent of protein binding of the metabolite compared to parent drug. To reduce costs associated with serial metabolite sampling, collections can be limited to three to four samples at times when the highest metabolite concentrations are expected.

Altered free drug concentrations can change NCE distribution and/or elimination, which may affect safety or efficacy. Blood samples should also be collected to assess protein binding if the NCE (and/or metabolites) are highly bound, the NCE is known to exhibit nonlinear binding, or its binding might be affected by metabolites. In most situations, protein binding is assessed in select samples encompassing the entire concentration range (e.g., around the time of maximum concentration followed by two to three more samples at representative times across the sampling period).

Pooled urine collection is of relevance when significant quantities of parent drug or metabolites are excreted into urine. If the renal elimination of NCE and metabolites is negligible, urine collections are not necessary with the exception of the RI study. Pooled urine intervals are generally collected for 0–12 and 12–24 hours postdose and in subsequent 24-hour intervals over the entire study period. More

frequent sampling during the initial elimination phase is only recommended for NCEs which are likely not quantifiable over a 12-hour collection interval.

Data are analyzed to describe the PK of the NCE and/or metabolites using either noncompartmental or compartmental methods [2]. The noncompartmental approach is simpler and potentially less subjective. Compartmental modeling approaches are useful if further simulations of the NCE PK are required. Typical PK parameters are time ( $t_{\max}$ ) to maximum concentration ( $C_{\max}$ ), area under the concentration-versus-time profile (AUC) collected over appropriate intervals, minimum concentration ( $C_{\min}$ ) over the dosing interval for multiple dose studies, NCE clearance (CL), volume parameters ( $V_d$ ), apparent terminal elimination rate constant ( $k_e$ ), and half-life ( $t_{1/2}$ ). Following a single dose, the estimation of AUC from time zero to infinity [AUC(0– $\infty$ )] should not exceed more than 20% for NCE (and 30% for metabolites) and samples should be collected sufficiently long to adequately characterize  $t_{1/2}$ , and  $k_e$  in the population of interest as well as the control group. For drugs that are highly bound, PK parameters are preferably expressed in terms of unbound concentrations. If urine data are collected in the study, the amount ( $A_e$ ) or fraction of dose ( $f_e$ ) excreted and renal clearance (CL<sub>R</sub>) are also assessed. Changes in CL<sub>R</sub> can be evaluated with regard to its relative impact on total clearance for the population of interest.

The PK parameters are presented descriptively for each treatment. For most studies measures of exposure (AUC and  $C_{\max}$ ) are compared as primary PK parameters across categorical population groups using an analysis of variance (ANOVA). The statistical analysis is performed on the log-transformed PK parameters and the model may include other demographic variables if applicable. The 90% confidence interval for the ratio of the least-squares (LS) treatment means is calculated for the various populations. Lack of population difference is established when the calculated confidence interval for the ratio of the LS means falls within the predefined confidence interval limit of the “no-effect” boundary for the study. Statistical methods may also be employed to test for significant differences in other relevant parameters ( $t_{\max}$ ,  $t_{1/2}$ ); however, these parameters are generally only summarized descriptively. Where differences in PK parameters can be described as a function of a continuous variable such as creatinine clearance (CL<sub>CR</sub>) in stage II RI studies, a statistical regression model may be more appropriate for data analysis.

Irrespective of the statistical approach, a no-effect boundary should be defined prior to study conduct which is used to assess the significance and clinical implications of any identified PK differences. The criteria should be based on clinical relevant changes in drug PK rather than the narrow statistical window of the 80–125% confidence interval traditionally applied in bioavailability/bioequivalence studies [3, 4]. A wider boundary is acceptable as long as the boundary can be justified with available clinical data. No-effect boundaries can also be defined for the slope of a regression analysis, for a population analysis, or for PK–PD modeling. If “clinically relevant” boundaries are not known prior to study start, the analysis plan should clearly differentiate between predefined narrow bioequivalence boundaries and clinically relevant changes in drug PK. If the 90% confidence interval for the LS mean ratio of the PK measurement falls within the predefined boundary, any changes in drug PK for the special population group can be considered not clinically relevant and do not require any dose adjustments. In the case of RI and HI studies, the small clinical sample size and/or high intersubject variability may preclude meeting tight



statistical criteria and result in statistical interpretation of confidence intervals which are inconclusive.

The data obtained in the special population studies must be evaluated in the context of all available PK, safety, and efficacy information to optimize drug therapy. PopPK analyses assess the contribution of factors to the intersubject variability in modeled NCE volume of distribution and clearance. These factors include demographic covariates [such as age, gender, race, body mass index (BMI) and/or WT, and height], physiological covariates (creatinine clearance, concurrent diseases, etc.), and lifestyle covariates (such as concomitant medications, smoking/drinking habits). The analysis model minimizes PK parameter variability in order of importance (i.e., statistical significance) of the various covariates. Results of PopPK analyses may further support the findings in special population studies, provide mechanistic explanations for apparent NCE PK differences, or alternatively show that statistically significant differences between populations may not have clinical relevance within the overall variability in the population estimates for the PK parameter.

Regulatory expectations have been formulated in various U.S., European, and International Conference on Harmonisation (ICH) guidance documents (see below) and provide detailed information on how the findings of the special population studies should be summarized in the drug label. The special population subsection within the clinical pharmacology section should briefly summarize any findings and describe any dosing adjustments or precautions required in the special population. If clinically relevant changes in drug disposition were observed, a statement should be included in the precautions/warnings section with further reference to the dosage and administration section. The latter should contain detailed instructions for the physician to adjust the NCE dose in the special population or to indicate that the NCE should not be used in the affected population. The reader is referred to the various guidance documents references in the sections below.

### 10.7.3 GERIATRIC POPULATION

The geriatric population has been arbitrarily defined as subjects/patients who are 65 years or older. While use of medications generally increases with age, many clinical trials have an upper age cutoff to exclude elderly subjects due to the concern that these subjects (a) are more frail and at potentially higher risk to have treatment-related AEs, (b) have underlying diseases that might affect the NCE PK and thus study objectives, (c) are receiving one or more comedications which are part of the studies' exclusion criteria, (d) have veins that are more difficult to access for blood collection, (e) require different housing conditions than young volunteers, and (f) may not be able to appropriately reason and express a choice when signing an informed consent [5].

As people age, the body undergoes physiological and composition changes [6]. The elderly tend to have lower lean body mass, lower total body water, and higher total body fat, which may affect the volume of distribution and half-life of highly lipophilic drugs. Gastrointestinal function changes with age, gastric pH increases, and adsorptive surface and motility decreases. Unless the NCE has a high first-pass extraction ratio, the absorption of drugs is unlikely to exhibit age-related changes in absorption. Glomerular filtration rate (GFR) and tubular secretion decrease with

age, which may increase half-life for drugs which are predominantly renally excreted [7]. Changes in hepatic metabolism are more complex. Liver mass as well as hepatic blood flow tend to decrease with age, which may affect drugs with low intrinsic clearance as well as those with high extraction ratio; however, there are conflicting reports in the literature whether an actual change in the intrinsic ability of the liver to clear drugs (true intrinsic clearance) exists in the geriatric population or whether the observed changes in metabolism are secondary to an array of related physiological changes [8].

Data in the geriatric population are a regulatory requirement unless the NCE is not indicated in this population. The FDA and ICH guidance documents outline the expectations of the regulatory community [9, 10]. The FDA has further issued a guidance document describing the content and format for geriatric labeling in U.S. submissions [11]. While the FDA focuses on the changes of age-associated conditions with regard to drug disposition characteristics (and that of active/toxic metabolites), the ICH guidances suggest that the elderly should be studied not only with a new NCE but also during the development of new formulations and new combinations of marketed drugs or for new indications which include geriatric patients in the indication. A separate study in the elderly may not be required for drugs with low systemic availability (e.g., some topical drugs) where age differences in PK are unlikely to be of significance.

In a formal elderly PK study, an appropriate number of male and female subjects 65 years of age and older is enrolled along with an equal number of young controls. Generally the number of subjects for each group does not exceed 20; however, the total number of subjects should be large enough to allow statistical comparisons. Since geriatric subjects are generally easy to enroll, a sample size which results in 80% power to detect a clinically significant difference between groups within the specified no-effect boundary is desired. The geriatric population may be further stratified into two to three age subgroups with an equal number enrolled in each age range. Some study designs are also stratified by gender with an equal number of young and elderly male/female subjects enrolled in each age group.

Subjects should be generally healthy within the criteria expected of a geriatric population. Wider inclusion criteria will be required on preadmission body mass index (BMI) values, vital signs, ECGs, and clinical laboratory evaluations to accommodate age-related physiological changes. Elderly are frequently taking multiple medications for various conditions (including hormone replacements, vitamins, and herbal supplements) and the inclusion/exclusion criteria must be specific enough to allow inclusion of a representative geriatric population while ensuring that the PK objectives of the study are not compromised. Geriatric PK studies are generally conducted under standard fasted conditions and meals are withheld for approximately 14 hours. The author has observed that the long fast can lead to a higher incidence of AEs (e.g., dizziness or nausea) in the elderly, who are less tolerant of extended fasting than healthy young subjects.

Pharmacokinetic sampling, safety collections, and data analyses follow the criteria outlined in the previous section. The AUC and  $C_{\max}$  are compared as primary PK parameters across categorical age groups using an ANOVA. Since age is a continuous variable, a statistical regression model may be employed if a graded age difference in NCE PK is apparent. If the NCE is excreted to a significant extent, it is advisable to collect urine data ( $A_e$  or  $f_e$  and  $CL_R$ ) and determine  $CL_{CR}$  in this

population. In the case of differences in drug disposition, the renal excretion data from the geriatric population can be correlated with the results obtained in the RI population for mechanistic interpretation of the results.

The elderly are more sensitive to CNS agents and may respond differently to cardiovascular agents than younger adults [12]. Phase I studies for CNS or cardiovascular drugs usually include a PD component. PD results can be evaluated using similar statistical methods as for categorical comparison of PK parameters. The objective for the PD component is the determination of whether PD changes occur and, if the case, whether the PD changes are a result of changes in drug disposition or whether the PD changes also occur in the absence of PK changes.

The PK results of definitive geriatric PK studies are frequently supplemented with the modeling results of phase II/III studies. These analyses may substantiate the findings of the definitive elderly PK study or alternatively may identify other factors that may have contributed to the apparent age-related changes in PK (e.g.,  $CL_{CR}$ ). At the current time, PopPK results are only considered supplementary in the registration package and have not replaced the formal PK study in the geriatric population.

#### 10.7.4 RENAL IMPAIRMENT

There are various kinds of kidney diseases. Chronic renal failure, in which the kidney can no longer cope with the load of endogenous and exogenous substances that must be excreted, is the most relevant with regard to affecting the PK of a NCE and/or its circulating metabolites [13]. Patients with RI can be categorized into two distinct subpopulations: those with various degrees of diminished renal function and those who require artificial means to remove the buildup of electrolytes, body waste, and drug product by dialysis methods such as hemodialysis, hemofiltration, hemoperfusion, and peritoneal dialysis. Patients with various degrees of diminished renal function are further categorized into distinct groups based on the severity of their disease. The FDA and European Medicines Agency (EMA) recommend the categorization of the RI population into five groups, as outlined in Table 2 [14, 15]. The use of categorical renal function groups provides a means of balancing patient enrollment across the RI spectrum and establishes uniformity in data presentation across various drug submissions.

PK data in the RI population are a regulatory requirement unless the NCE is not indicated in this population or a strong case can be made based on disposition

**TABLE 2 Categorization of RI Patients**

| Group | Description               | Estimated Creatinine Clearance (mL/min) <sup>a</sup> |
|-------|---------------------------|------------------------------------------------------|
| 1     | Normal renal function     | >80 mL/min                                           |
| 2     | Mild renal impairment     | 50–80 mL/min                                         |
| 3     | Moderate renal impairment | 30–50 mL/min                                         |
| 4     | Severe renal impairment   | <30 mL/min                                           |
| 5     | ESRD                      | Requiring dialysis                                   |

<sup>a</sup>EMA (2004) further normalizes the renal function groups to body surface area with creatinine clearance expressed in mL/min/1.73 m<sup>2</sup> [15].

Source: From [14].

and physiochemical properties of the drug that renal disease will not alter drug PK. FDA and EMEA guidelines specify when RI studies are indicated and how the RI data should be utilized in the registration package [14, 15]. If the target patient population is likely to include RI patients and the PK (or PD) of the NCE and/or its active/toxic metabolite is likely to be affected by changes in renal function, studies should be conducted to assess whether dose adjustments are required. RI studies are recommended for NCEs with (a) a narrow therapeutic index and (b) primary elimination of the NCE and/or active/toxic metabolites via the renal route (excretion and metabolism) or (c) a combination of high hepatic clearance and significant protein binding. An RI study is considered unnecessary if an NCE has a wide therapeutic window and (a) is primarily metabolized by the liver, excreted into the bile, or eliminated via the lungs, (b) is a topical drug that is not absorbed, or (c) is intended only for single-dose administration. Confirmation by the agency that a study in RI is not required for drug approval is advised to avoid a potential hold or rejection of a submission until appropriate renal impairment data are available.

The FDA and EMEA distinguish between a traditional approach assessing the NCE PK in patients with varying degrees of renal function (full-study design) and an adaptive two-stage approach (reduced/stage design) [14, 15]. The latter is recommended if the NCE and/or metabolite PK are not likely to require dose adjustments in RI. Stage I consists of a study comparing only the extremes, that is, healthy controls versus patients with severe RI (also see Table 2). If the PK difference between the two extremes does not warrant a dose adjustment, no further studies are needed. Alternatively, if PK in severe RI patients is altered to the extent requiring dose adjustments, patients with a lesser degree of RI should also be studied in a follow-up trial (stage II).

The regulatory agencies recommend RI studies in NCEs with narrow therapeutic margins even if an impact of RI on the NCE PK is unlikely, since RI may affect the disposition of drugs that are exclusively eliminated by nonrenal routes (highly metabolized or excreted into bile). Changes in drug elimination due to changes in protein binding are the most obvious causes of altered PK [16, 17]; however, diseases secondary to RI may also contribute to PK changes. The combination of physiological changes in RI can increase or decrease absorption by altering GI motility, transport mechanisms, and metabolism, blood flow to the kidney and other organs, and hepatic metabolism. Physiological explanations for these changes have been summarized by various authors [18–21]. Other underlying physiological changes secondary to RI include various forms of cardiovascular disorders such as hypertension, coronary artery disease, congestive heart failure, arrhythmias, lipid abnormalities, diabetes mellitus, peripheral neuropathy, and diabetic retinopathy [22–24]. Patients with severe RI (GFR < 30 mL/min) are those most affected by these physiological changes. Low urine production results in a higher accumulation of toxins in the circulation, which puts an additional strain on alternative routes of NCE elimination. Contrary to common belief, patients with end-stage renal disease (ESRD) who complete hemodialysis immediately prior to drug administration can have lower drug exposure for non-renally eliminated drugs compared to the other renal function groups [25]. While the reason for this finding is not well understood, one possible explanation is that the process of dialysis removes a large amount of endogenous material which otherwise would “clog up” the NCE’s elimination routes.

Studies of ESRD patients are conducted to determine whether dialysis may result in lower drug concentrations (i.e., patients might require replacement doses) and/or whether dialysis can be used in the treatment of overdose situations. A significant portion of the NCE is likely to be removed during dialysis if an NCE and/or active metabolite is water soluble, has a low nonrenal clearance, a low molecular weight (<500 Da), and low protein binding, and its volume of distribution is small (<1 L/kg) [20, 26]. On the other hand, if the NCE has a large unbound volume of distribution or high nonrenal clearance, dialysis is unlikely to contribute to removal of the NCE to any great extent.

#### 10.7.4.1 Design Considerations in RI

In the formal renal PK study, an appropriate number of male and female subjects with RI are included in the study along with an equal number of non-RI subjects who match the demographics of either the RI patients or the target population. Ideally, subjects should be known to the investigator long enough to establish the presence of reasonably stable renal function. For example, serum creatinine concentrations should not vary by more than approximately 0.4 mg/dL over a 3-month period prior to the initial screening. Other means of documenting stable serum creatinine concentration or stable creatinine clearance up to 6 months prior to the initial screening can be used as long as they are well specified. Since RI patients are harder to recruit due to restrictions in inclusion/exclusion criteria, the author recommends widening the age range to at least 70 years of age rather than the typical 65-year cutoff. The number of RI patients should be sufficiently large to allow appropriate statistical analyses, but the time it will take to enroll patients should also be considered.

In stage I, an equal number of patients with severe RI are enrolled along with an equal number of healthy controls. In the extended design (stage II), the different RI function groups generally contain a minimum of 8–10 subjects per group and one control group of similar size and demographic characteristics which brackets all RI groups. Patient enrollment should be closely monitored so that patients within each renal function group are evenly distributed (e.g., not all patients with mild renal impairment have a creatinine clearance  $CL_{CR}$  of >65 mL/min). This becomes important at the time of data analysis if PK parameters are regressed on  $CL_{CR}$ .

Creatinine clearance can be directly measured using serum and urine creatinine concentration or calculated by one of several formulas using serum creatinine concentrations only. If  $CL_{CR}$  is directly measured using both serum creatinine and urine creatinine, it is advised to collect urine for 24 hours and to draw the serum creatinine sample at the midpoint of the urine collection interval. In ESRD patients who produce urine, serum creatinine concentrations should be obtained at the beginning and end of the urine collection. The average of the two serum creatinine values provides a more accurate estimate since serum creatinine is likely to increase over time between dialyses. If  $CL_{CR}$  is measured using serum creatinine only, the FDA recommends using the Cockcroft–Gault formula in adult subjects and the formulas by Schwartz in children and infants from birth to one year of age (see Table 3) [27–29].

Inclusion and exclusion criteria should clearly define which underlying diseases are acceptable. Patients with any major illness (other than renal disease

**TABLE 3 Formulas to Calculate Creatinine Clearance from Serum Creatinine Concentrations**

| Group                            | Formula                                                                                                    | Reference |
|----------------------------------|------------------------------------------------------------------------------------------------------------|-----------|
| Adults (>12 years)               | $CL_{CR} = \frac{(140 - \text{age}) \times \text{weight}}{72 \times \text{Serum}_{CR}}$ (0.85 for females) | 27        |
| Infants less than 1 year         | $CL_{CR} = \frac{0.45 \times \text{length}}{\text{Serum}_{CR}}$                                            | 29        |
| Children 1–12 years <sup>a</sup> | $CL_{CR} = \frac{0.55 \times \text{length}}{\text{Serum}_{CR}}$                                            | 28        |

Note: Serum<sub>CR</sub>: serum creatinine in mg/dL; age in years; weight in kg; length in cm; CL<sub>CR</sub> in mL/min.

<sup>a</sup>Schwartz initially presented this equation with plasma creatinine for children up to 20 years of age [28].

and associated stable illnesses such as hypertension and diabetes mellitus) such as hepatic, GI, cardiovascular, adrenal, hematological, or obstructive airway disease should be excluded. Patients with mild hepatic and GI diseases should be evaluated carefully to determine if these might affect the pharmacokinetics of the NCE. All underlying diseases should not interfere with the study conduct or pose an unjustifiable risk to the study participant. Wider inclusion and exclusion criteria are required on preadmission vital signs, ECGs, and clinical laboratory evaluations to reflect the nature of the patients to be enrolled. Restrictive inclusion/exclusion criteria will affect the enrollment rate.

Subjects with RI will require concomitant medications. Patients whose medication is likely to interfere with the NCE PK should be excluded. Subjects receiving medications that are unlikely to interfere with the study medication (or vice versa) should receive their regular scheduled medication at a time different from the scheduled NCE dose whenever possible (e.g., withhold all nonessential medications  $\geq 12$  hours prior to and  $\geq 4$  hours following NCE administration). Patients receiving ranitidine and other H<sub>2</sub>-receptor antagonists may be asked to withhold their medications for  $\geq 36$  hours prior to dosing, if medically acceptable, and to use alternate measures (e.g., antacids) to control their symptoms. Antacids should be held for at least 1 hour prior to dosing and 2 hours postdose. For medication that must be administered near the time of the NCE administration, dosing should be scheduled either 1–2 hours prior to or 1–2 hours after administration of the study medication whenever possible. Some over-the-counter herbal supplements are associated with drug interactions; others may affect physiological parameters such as blood pressure, blood glucose, kidney function, and electrolyte balance in renal disease patients [30]. Patients requiring herbal supplements for medicinal purposes should be carefully evaluated prior to study enrollment.

The timing of meals in renal studies is similar to those in any typical phase I study and should reflect the NCE requirements. With progression of the disease, renal patients may require modifications to the composition of the meal (e.g., salt content) and/or permitted snacks. Patients with severe renal disease and with ESRD may have restrictions with regard to the volume of fluid intake, which needs to be taken into account in the design of the study.

Pharmacokinetic sample and safety collections follow the criteria outlined in Section 10.7.2 with the exception that urine samples are almost always collected in



RI studies. The FDA guidance recommends that unbound drug levels are determined in each sample unless the drug exhibits relatively low extent of binding (defined by the FDA as the extent of binding being less than 80%) and that PK parameters are calculated based on unbound concentrations [14]. In the stage I design, the primary exposure PK parameters ( $C_{\max}$  and AUC) are compared between RI patients and the control group using an ANOVA model with the appropriate confidence interval boundaries. Secondary parameters, including protein binding results, are summarized descriptively. In the stage II design, the primary PK parameters are most commonly regressed against the patients'  $CL_{CR}$ .

Since screening  $CL_{CR}$  values may have been obtained three to four weeks prior to dose administration, it is advised that  $CL_{CR}$  is obtained again on the day of pharmacokinetic sampling to avoid potential misclassification of the patient. In RI studies that the author has conducted, the “during study conduct”  $CL_{CR}$  estimates generally correlate well with those obtained during screening, but there may be changes in some patients in either direction of renal function. In some cases, patients “categorized” into a renal function group during screening fell into the adjacent renal function group during study conduct (e.g., a patient with  $CL_{CR}$  of 27 mL/min at screening had  $CL_{CR}$  of 33 mL/min during the actual study). If regression analysis is used to model the PK parameters to renal function, these small shifts do not affect the interpretation of the data, while categorical ANOVA analysis may be affected by shifting one subject from one category to another. The other PK parameters are generally presented descriptively.

Serum creatinine values are generally collected as part of the clinical laboratory package in phase II/III. The  $CL_{CR}$  can be calculated for the majority of patients and is available as a potential covariate for PopPK analysis. However, most phase II/III studies do not include a large enough population with decreased RI function and the  $CL_{CR}$  range tends to be too narrow to allow determination of meaningful differences. For this reason, the data generated in the definitive phase I study are not always supplemented by PopPK analyses. One of the exceptions would be the situation where the NCE cannot be studied outside the therapeutic target population. The regulatory agencies have requested that an attempt is made to include subpopulations in the integrated phase II/III development program and to utilize PopPK analysis approaches to elucidate the relationship between relevant PK parameters and renal function. However, the number of patients must be large (and diverse) enough to allow detection of PK changes requiring dose adjustments.

#### 10.7.4.2 Special Considerations for Dialysis Patients

The most common forms of renal replacement therapy in patients with ESRD are hemodialysis and continuous ambulatory peritoneal dialysis (CAPD). Both processes involve the principle of diffusion; however, hemodialysis is considered the most efficient artificial means of toxin and drug removal. During hemodialysis, blood is dialyzed against a dialysate medium across a permeable membrane. The extent of toxin and drug removal during hemodialysis is governed by the rate of blood flow through the dialyzer, the rate of dialysis flow, the surface area and type of dialysis membrane, and the dialysis medium [13]. In CAPD, the peritoneal membrane serves as the dialysis membrane to the dialysate in the peritoneal cavity. Toxin removal in



CAPD occurs through osmosis and diffusion from the small blood vessels in the walls of the abdominal cavity, the viscera, and the interstitium through the peritoneal membrane. For more information on CAPD, the reader is referred to reviews of advantages and disadvantages of CAPD and its influence on drug pharmacokinetics [31, 32]. Studies in CAPD patients are only recommended for the NCEs likely to be administered in this patient population.

Renal dialysis studies are conducted both “between” and “during” the scheduled dialyses regimen. The decision of what time in the dialysis cycle the PK information is collected depends on the objectives of the study. Determination of potential accumulation of the NCE and/or metabolites in the ESRD population will include patients between their scheduled dialysis while the use of hemodialysis as a means of drug removal would require drug administration during the patients’ hemodialysis regimen. The length of PK sampling is restricted between dialysis since ESRD patients undergo hemodialysis every 48–72 hours. PK sampling between dialysis regimens is rarely longer than 48 hours, which may introduce error into results for half-life and AUC assessments. Potential shorter PK sample collection should be taken into account when NCE PK data are compared across studies. It is recommended that drug administration be scheduled at least 1 hour after completion of the dialysis before dosing to allow for balance of the patient’s fluid level. PK sample collections, data analysis, and statistical comparisons for between-hemodialysis PK studies are performed as previously described and include categorical comparison of the primary exposure PK parameters to a control.

When assessing the NCE PK during hemodialysis, the timing of dose administration and potential fluid loss during analysis should be considered. Hemodialysis sessions generally last around 3–4 hours. NCE administration most often occurs prior to the start of the dialysis. Optimal results, especially if the dialysate is measured, will be obtained if the dialysis is started around the time of maximum exposure, since potential drug removal is proportionate to blood concentrations. Hemodialysis will remove accumulated water along with waste material. Determination of the difference in patient weight immediately before and following hemodialysis accounts for this fluid loss/weight loss and can be used as covariate or correction factor during data analysis and interpretation.

Collection of data regarding relevant equipment and conditions during hemodialysis is important since the extraction efficiency of the hemodialysis is affected by the type of dialyzer, dialysis membrane, and membrane surface area against which the blood is dialyzed. The ionic strength and type of the dialysate (dialysate potassium, dialysate calcium) and blood and dialysate flow rate will also affect the extraction ratio. Dialysis conditions are optimized to a given ESRD patient and are likely different between study subjects and study sites. The subject’s case report form should allow for documentation of all pertinent dialysis components.

The PK sample collections, data analysis, and statistical comparisons for during-hemodialysis PK studies include categorical comparison of the primary exposure PK parameters to a control using an ANOVA model with the appropriate confidence interval boundaries. However, AUC comparisons frequently focus on the partial AUC interval obtained during the hemodialysis since the dialysis window is so narrow. In addition, these studies typically include the estimate of dialysis clearance. Lee and Marbury have described a variety of methods to determine dialysis clearance [33]. Calculations may be based solely on blood/plasma concentrations or

can include dialysate concentration. Blood samples are collected during three stages of the study: (a) prior to start of dialysis, (b) during dialysis, and (c) during postdialysis treatment. During dialysis, blood samples can be removed directly from the lines entering (prefilter or arterial sample) and leaving (postfilter or venous sample) the dialysis chamber, while serial venous blood samples are collected prior to and following dialysis. The following information should be recorded in the subject's case record form to allow determination of dialysis clearance: (a) blood flow into the dialyzer, (b) hematocrit in the pre- and postfilter blood samples during dialysis to allow for conversion of blood dialysis clearance to plasma dialysis clearance, and (c) optional endogenous molecules such as blood urea nitrogen (BUN) and/or creatinine to assess the extraction efficiency of the hemodialyser (based on standard markers) and to permit comparison to the NCE and/or metabolites.

If the objective of the study is to determine percentage of dose removed during dialysis, collection of dialysate is required. Collection of dialysate may allow for a more accurate estimate of dialysis clearance than can be obtained from blood/plasma data [33]. It should be noted that dialysate can be a difficult matrix to assay for drug concentrations. Analytical method validation can be difficult to develop due to the high salt content of the dialysis medium. Dialysate flow rate should be collected as part of the study information. Although dialysis flow rate is specified as part of the dialysis procedures, measurement of the total dialysate volume is also recommended since it provides confirmation of the reported dialysis flow setting. Incorrectly recorded flow rates or variation in outputs of the dialyser can be easily detected if the preset flow did not deliver the total intended volume and PK calculations may be performed with an "adjusted" or "corrected" flow rate based on measured dialysate volume.

For the detailed derivation of dialysis clearance the reader is referred to the article by Lee and Marbury [33]. The authors discuss different direct methods such as arteriovenous (AV) difference, recovery rate, or cumulative dialysate methods as well as those derived from plasma pharmacokinetic parameters obtained between and during dialysis. If both plasma and dialysate data are available, data calculated by more than one approach can be compared.

In simplified form, the AV difference approach is based on a flow rate multiplied by the extraction efficiency of the hemodialyser (E). Blood dialysis clearance ( $CL_{HD-b}$ ) is thus calculated as

$$CL_{HD-b} = \frac{Q_{bi} [C_{bi} - C_{bo}]}{C_{bi}} \quad (1)$$

$$CL_{HD-b} = \frac{Q_{do} C_{do}}{C_{bi}} \quad (2)$$

where  $Q_{bi}$  is the blood flow entering and  $C_{bi}$  and  $C_{bo}$  the blood concentrations entering (i) and leaving (o) the dialysis chamber. The quotient term  $(C_{bi} - C_{bo})/C_{bi}$  is the extraction ratio based on blood/plasma concentration alone. Here  $Q_{do}$  is the dialysis flow leaving and  $C_{do}$  is the dialysate concentration leaving (o) the dialysis chamber. The assumption is made that ultrafiltration differential and backdiffusion are negligible compared to the blood and dialysis flows.

Since blood flow rather than plasma flow is measured as part of the study and the drug concentrations are derived from plasma samples,  $CL_{HD-b}$  needs to be converted to plasma dialysis clearance ( $CL_{HD-p}$ ) using

$$CL_{HD-p} = Q_b \{ [K_p * Hct + (1 - Hct)] \} E \quad (3)$$

where  $E = (C_{pi} - C_{po})/C_{pi}$  and  $C_{pi}$  and  $C_{po}$  are the plasma concentrations of the blood entering (i) and leaving (o) the dialysis chamber;  $K_p$  is the partition coefficient between RBCs and plasma and Hct is the hematocrit. If both the hematocrit and the partition coefficient of the NCE between RBCs and plasma are known, the assessment of  $CL_{HD-p}$  using the AV model is most accurate. If the  $K_p$  approaches zero (i.e., the analyte does not distribute into RBCs), the equation is reduced to reflect the physiological plasma flow:

$$CL_{HD-p} = Q_b (1 - Hct) E \quad (4)$$

Alternatively, assuming that  $K_p = 1$ , Equation (4) reduces to

$$CL_{HD-p} = Q_b E \quad (5)$$

Equation (5) is the most common equation used to assess  $CL_{HD-p}$  from plasma data in the absence of any other information. As an approximation,  $Q_b$  in Equation (5) can be corrected by the blood-to-plasma NCE concentration ratio ( $C_b/C_p$ ) to estimate plasma blood flow and  $CL_{HD-p}$ .

If dialysate data are available,  $CL_{HD-p}$  can be calculated by two basic methods: the recovery rate method and the cumulative dialysate method. In both methods, the assumption is made that the dialysis flow stays constant over the interval of observation. The percent recovery is calculated as the fraction of drug recovered in the dialysate relative to the administered dose. These methods are identical to those employed for the calculations of renal clearance and percent of dose excreted into urine. For this reason, the equations are not presented in this section. Since ESRD patients undergo hemodialysis only every two to three days and the NCE is likely administered more frequently, it might be of interest to assess the percent dialyzed value both for the “actual dose” studied during hemodialysis and for the likely “total dose” (sum of all the doses a patient would have received during one hemodialysis cycle) to determine if dose adjustment in ESRD patients is needed.

### 10.7.5 HEPATIC IMPAIRMENT

The liver is the main metabolizing organ in the human body. Hepatic function exhibits a slow decline with increasing age (decrease in liver weight and hepatic blood flow) [8]; however, these decreases are generally not considered of clinical relevance with respect to drug disposition. On the other hand, a wide etiology of pathophysiological mechanisms is known to cause a decrease in liver function and may alter the PK characteristics of drugs. HI tends to be a heterogeneous disorder due to both structural and functional abnormalities. In the western world including the United States, the most common cause of HI is chronic alcohol abuse leading

to liver cirrhosis. On a worldwide basis, viral infections (hepatitis B and C) are the main contributors to HI. Liver disease may also develop secondary to other diseases such as congestive heart failure or cancer and may not always be immediately recognizable.

In contrast to RI, there are no good endogenous markers which allow characterization of hepatic function (such as bilirubin, albumin, or prothrombin time), nor do model drugs which are exclusively cleared from the liver lend themselves as exogenous markers [34, 35]. While various clinical variables such as ascites, encephalopathy, nutritional status, peripheral edema, and histological liver assessments have been studied, none of the various clinical approaches has shown similar predictability of liver function as the use for serum creatinine (or  $CL_{CR}$ ) for RI. For pragmatic reasons, the Child–Pugh classification for alcoholic cirrhosis and portal hypertension has been adapted by both U.S. and European regulatory agencies as a means to allocate HI patients to hepatic function groups [36–37]. The Child–Pugh scale is based on the encephalopathy grade, degree of ascites, serum bilirubin and albumin concentrations, and prolongation in prothrombin time, as illustrated in Table 4. Based on the total number of points for each of the grading criteria, subjects are classified into mild (Child–Pugh grade A: 5–6 points), moderate (Child–Pugh grade B: 7–9 points), or severe (Child–Pugh grade B: 10–15 points) HI groups. A normal control group would have a grading of less than 5.

It should be understood that the Child–Pugh scale was not originally developed to categorize HI patients for the purpose of predicting drug PK. The FDA guidance provides alternative classification scales such as the Maddrey–Carithers discriminant function and Mayo risk scores for primary biliary cirrhosis as alternative clinical measures of liver function, but with the implementation of the FDA and EMEA guidelines, the Child–Pugh scale has become the classification system that is most commonly used [38, 39].

The PK data in the hepatic impairment population are a regulatory requirement unless the NCE is not indicated in this population or a strong case can be made based on disposition of the NCE that HI will not alter drug PK. The FDA and EMEA guidelines specify when HI studies are indicated and how the HI data should be utilized in the registration package [38, 39]. HI studies are indicated for almost all drugs likely to be used in subjects with HI. Drugs with >20% of hepatic

**TABLE 4** Categorization of HI Patients

| Points                                | 1        | 2       | 3        |
|---------------------------------------|----------|---------|----------|
| Encephalopathy grade                  | None (0) | 1–2     | 3–4      |
| Ascites                               | Absent   | Slight  | Moderate |
| Bilirubin, mg/dL                      | <2       | 2–3     | >3       |
| Albumin, g/dL                         | >3.5     | 2.8–3.5 | <2.8     |
| Prolongation in prothrombin time, sec | <4       | 4–6     | >6       |

*Note:* Encephalopathy grade:

- Grade 0: normal consciousness, personality, neurological examination, and electroencephalogram (EEG)
- Grade 1: restless, disturbed sleep, irritable or agitated, tremors, impaired handwriting, 5cps waves on EEG
- Grade 2: lethargic, time disoriented, inappropriate, asterixis, ataxia, slow triphasic waves on EEG
- Grade 3: somnolent, stuporous, place disoriented, hyperactive reflexes, rigidity, slower waves on EEG
- Grade 4: unrousable coma, no personality/behavior, decerebrate, slow 2–3cps delta waves on EEG

*Source:* From [38, 39].

metabolism or elimination (of parent drug and active/toxic metabolites) are considered to be extensively metabolized and a change in exposure may be clinically significant and thus require dose adjustment or contraindication in this population. If the extent of metabolism or biliary excretion is unknown, the agency assumes the drug to be highly metabolized, and hence a HI study is required. If the drug has a narrow therapeutic window, the HI study is recommended even if <20% of the drug and active/toxic metabolites are eliminated by the liver.

The guidelines consider only a few situations where a HI study may not be required (unless clinical concerns suggest otherwise): the NCE (a) is not to be used in patients with HI or (b) is intended for single-dose use only or (c) has a wide therapeutic window and <20% of parent and/or active/toxic metabolites are eliminated by the liver or (d) is gaseous or volatile and the drug and its active/toxic metabolites are primarily eliminated via the lung.

As in the case of RI, HI is associated with secondary physiological changes or disease states that may affect the NCE PK in this population. For this reason, it is prudent to obtain confirmation by the agency that a study in hepatic impairment is not required for drug approval to avoid a hold or rejection of a submission until appropriate hepatic impairment data are available.

The FDA and EMEA distinguish between two approaches in the design of HI studies [38, 39]. The extended approach assesses the NCE PK in patients with varying degrees of hepatic function compared to a control group. The reduced design only includes subjects with a Child–Pugh category of moderate which are compared to a matched control group. The latter is recommended if the NCE and/or metabolite PK are not likely to require dose adjustments in HI and subjects with a Child–Pugh category of mild are unlikely to provide a signal (e.g., clinically significant change in exposure). Under the reduced scenario, the findings in the moderate Child–Pugh category would be applied to those of the mild category, and dosing in the severe category would likely be contraindicated in the label. This approach has the disadvantage that significant findings in the moderate HI population would be extrapolated to the mild category and could result in contraindication of the drug in any subjects with HI. Under those circumstances, a second study in patients with mild HI may be required if this population is part of the target patient population, which could potentially delay the registration package.

The alternative is the full approach described by the agencies in which at least six subjects with HI are enrolled into each of the three Child–Pugh categories outlined in Table 4 along with a control population of similar or greater size. Six evaluable subjects per treatment arm appear to be a low number; however, subjects with HI are notoriously hard to enroll, especially if they fall into the moderate and severe HI categories. The number of subjects is therefore a compromise between what is feasible and what is minimally needed to obtain data likely to produce meaningful results. Since these studies are not intended to prove statistical equivalence between HI and healthy controls, the sample size in these trials are large enough to provide a signal for a clinically relevant change in NCE exposure. Control subjects should demographically match those in the HI population. In the reduced design, control subjects can be matched one to one, while in the full design their age and gender are matched across the demographic range of the three HI groups. For details with regard to the selection of HI subjects, inclusion/exclusion criteria, and study design the reader is referred to the review by Bonate and Russell [40]. Subjects should be

known to the investigator long enough to establish the presence of reasonably stable disease. They should not have clinically significant changes in their disease status over at least a 3-month period prior to the initial screening. Examples of clinical changes would be worsening clinical signs of hepatic impairment or worsening of total bilirubin or prothrombin time by more than 50%. If a previous diagnosis classified the hepatic dysfunction a result of cirrhosis, copies of documentation of any clinical information used to make the previous diagnosis should be available. Documentation may include clinical evidence of portal hypertension, a positive liver biopsy, hepatic ultrasound, computed tomography scan, and/or magnetic resonance imaging.

Inclusion and exclusion criteria should clearly define which underlying diseases are acceptable. Patients with any major illness (other than liver disease and associated stable illnesses) such as renal (e.g.,  $CL_{CR} < 50 \text{ mL/min}$ ), GI, cardiovascular, adrenal, hematological, or obstructive airway disease should be excluded. Patients with mild renal and GI diseases should be evaluated carefully to determine if these might affect the pharmacokinetics of the NCE. All underlying diseases should not interfere with the study conduct or pose an unjustifiable risk to the study participant. As in the case of RI, subjects with HI tend to require concomitant medications and similar study considerations apply.

The HI study is generally conducted as a single-dose study and the highest proposed therapeutic dose is selected unless safety concerns warrant a lower dose. In the full-study design, treatment administration can be structured to study the NCE PK and tolerability sequentially. Subjects with mild HI are dosed first. Once tolerability has been established, dose administration proceeds to the moderate HI group and then to the severe group. The study could be stopped after completion of any HI category. This design, while safer, will increase the duration of the study since moderate and severe HI populations are increasingly harder to enroll and enrollment of these populations cannot start until the previous category has completed the study.

The PK sample and safety collections as well as data analysis follow the criteria outlined in Section 10.7.2. The FDA guidance recommends that protein binding is performed for drugs that are highly extracted by the liver (extraction ratio  $> 0.7$ ) and are extensively bound to plasma proteins (fraction unbound  $< 10\%$ ). Unbound fraction should be determined at least at trough and maximum plasma concentration and clearance and volume parameters are appropriately expressed in terms of both unbound and total concentrations [38]. In both the extended and reduced study, the  $C_{max}$  and AUC are compared across categorical population groups using an ANOVA with the appropriate confidence interval boundaries. Secondary parameters including protein binding results are summarized descriptively.

The results of HI impairment studies are frequently difficult to interpret. Due to the small sample size, the study has low power resulting in wide 90% confidence intervals. As a result, the 90% confidence interval for the LS mean ratio of the PK measurement rarely falls within the no-effect boundary or the even more restrictive 80–125% if a no-effect boundary is not available. If the criteria are not met, it has to be concluded that HI has an effect on NCE PK. It is not always clear if the effect is clinically relevant. The FDA guidance states that dose adjustments should be reflected in the label in the case of an obvious effect, that is, in the case that the clearance of the NCE is significantly impaired. The definitions of obvious and



clinical significance need to be put in context with PK and safety data. Exposure–response relationship analysis with emphasis on the upper safety limits of exposure obtained from both phase II/III trials (representative of typical patient exposure) and phase I trials (high-exposure comparison) may be explored to defend the case for the labeling recommendations made to the regulatory agencies. NCI PK metabolism (enzymatic pathways, percent of drug eliminated by the liver, and disposition of metabolites by the liver) and the overall sensitivity of drug disposition to other physiological and external changes will also play a role in the decision if and how dose adjustments in the HI population are proposed. In spite of the difficulties in conducting and interpreting data of HI studies, these investigations are critical to the registration package to support dosing decisions in this population.

### 10.7.6 WOMEN IN CLINICAL TRIALS

Inclusion of women in clinical trials has drastically changed over the last 15 years. Up to the early 1990s, women were routinely excluded from phase I clinical trials using the argument that the benefit–risk ratio was too high to allow inclusion of women into all types of phase I trials, especially prior to completion of reproductive toxicity studies which, in the United States, are not required until women of childbearing potential are enrolled into phase III studies [41]. The definitive gender comparison was generally conducted in the later stages of drug development, unless the NCE characteristics required the data to support phase II or III clinical trials.

The FDA Guideline for the Study and Evaluation of Gender Differences in Clinical Evaluation of Drugs reversed the previous FDA policy which excluded women with childbearing potential from early drug studies unless it targets a serious or life-threatening disease [1]. The 1993 guideline requires women to be included in all phases of the clinical drug development program for agents that are indicated for women. The guideline states that the FDA can place an investigational new drug (IND) or specific studies on hold if a sponsor proposes to exclude women or men with reproductive potential because of a risk, or potential risk, or reproductive toxicity from the use of an investigational drug product. In 1998, the FDAMA Women and Minority Working Group Report was issued which reviewed the 1993 guideline and recommended not to generate further gender-related guidelines but to institute a permanent tracking system to gather, search, and evaluate demographic data and potential deficiencies with regard to the inclusion of women in clinical trials [42]. The ICH has a website dedicated to ICH and women which shares the results of a recent review of the existing ICH guidelines and regulatory experiences in the three ICH regions: the United States, European Union (EU), and Japan [43]. In the posting, the ICH concluded that the current ICH and regional guidelines should be consulted for guidance on demographic considerations and that there was no need for a separate ICH guideline on women as a special population in clinical trials. The U.S., EU, and Japanese surveys showed that, in general, women were adequately presented in pivotal trial populations and that some form of gender effect was generally conducted and expected, be it a subpopulation analysis and/or formal PK/PD studies. The EMEA published the ICH findings as part of the ICH Guidelines Step 5 [44].

In today's environment, women (both those of childbearing and non-childbearing potential) are routinely included in phase I clinical trials unless specific NCE



characteristics preclude women from participating. Inclusion/exclusion criteria are worded such that women must adhere to stringent criteria in order to avoid pregnancy over the duration of the study. Women who are pregnant or lactating are always excluded unless the study is specifically conducted in one of these populations. A negative (serum or urine) pregnancy test result prior to enrollment and on check-in prior to drug administration is required for women of childbearing potential or who are perimenopausal. For women of childbearing potential (including perimenopausal women who have had a menstrual period within one year) appropriate birth control [defined as a method which results in a low failure rate, i.e., less than 1% per year, when used consistently and correctly, such as implants, injectables, some intrauterine contraceptive devices (IUDs)], sexual abstinence, or a vasectomized partner during the entire duration of the study are stipulated. It should be noted that many clinical protocols no longer restrict birth control methods to females and also require a barrier contraceptive method for males.

The majority of phase I trials include women who are taking oral contraceptive medications provided that they do not change their regimens throughout the course of the study. Similarly, postmenopausal women are permitted to continue their hormone replacement therapy throughout the study duration. The important point is that the medication is not changed during the study conduct to avoid potential changes in NCE PK across treatments. Exceptions to permitting systemic birth control medication or hormone replacement therapy are the cases where (a) the NCE is known to affect efficacy/safety of the medication (e.g., NCE induces liver enzymes), (b) the NCE is evaluated in a drug–drug interaction study in which the other administered drugs may affect efficacy/safety of the medication, or (c) the contraception/hormonal replacement medication may interfere with the objectives of the study.

Since PK data in women are generated throughout the entire drug development program, PopPK analyses of phases I–III PK data have become accepted not only in the United States but also the EU and Japan and have essentially replaced formal PK studies as the venue to assess potential effects of gender on drug disposition. If an NCE requires a formal PK study in women, the number of male and female subjects enrolled is based on the intrasubject variability of the NCE and is sufficiently large to perform the statistical comparison of the primary PK endpoints. Inclusion/exclusion criteria are described above. PK sampling, analysis, and statistical approaches reflect those described in Section 10.7.2.

### 10.7.7 ETHNIC CONSIDERATIONS

When an NCE is introduced into a new region, there is always the concern that ethnic differences may affect the drug's efficacy or safety and that dosage adjustment may be required. Ethnic differences in drug disposition may derive from three different sources: differences in (a) subject size and weight, (b) metabolism and transport due to the diversity of the cytochrome P-450 (CYP) family and transport proteins, and (c) food consumption (dietary) habits. With regard to physiological differences, the Asian population tends to have a smaller body size and lower WT and BMI than a typical U.S. or European population, which can result in higher exposure ( $C_{\max}$  and AUC) in the Asian population. A recent example is the 25.6% higher AUC(inf)

of fosfluconazole in Japanese subjects compared to Caucasians, which could be explained by the difference in BW between the two populations [45].

The diversity of the human CYP family has been well established. At least 57 human CYP isoforms have been identified to date; however, only 10 of these isozymes contribute to drug metabolism. The isoforms CYP3A4, CYP2D6, and CYP2C9 are the main contributors. Most of the CYP isozymes exhibit some form of genetic polymorphism and have structurally altered alleles with reduced or deficient catalytic activity. Ethnic differences in the make-up and polymorphism of CYP isozymes have been extensively reviewed [46–49]. The frequency of the deficient allele expression varies across ethnic groups, and the differences are thought to be likely a result of regional evolutionary diversification and environmental factors. There are also instances of variant alleles that may only be expressed in some ethnic groups but not in others [50]. Genetic polymorphism is not restricted to drug-metabolizing enzymes but has also been reported for drug transporters with affinity to chemotherapeutic agents such the multidrug-resistant *p*-glycoprotein (MDR1), multidrug resistance protein 2 (MRP2), and a breast cancer resistance protein [51, 52]. Genetic differences in metabolism and transport can be considered intrinsic ethnic variations independent of any environmental factors.

Food consumption habits are considered external factors and vary across ethnic groups within and across regions. Dietary constituents have been reported to alter the activity of various isozymes of the CYP family, phase II enzymes, and transporter systems and thus contribute to both inter- and intrasubject variability in drug PK. Harris et al. provide an in-depth review of various food constituents and their effect on metabolism enzymes and transport systems [53]. Dietary effects have been described for CYP3A4 (e.g., grapefruit juice, Seville orange juice, red wine, St. John's wort, garlic), CYP12 (e.g., caffeine, grapefruit juice, jufeng grape juice, cruciferous vegetables, charbroiled meat), CYP2E1 (ethanol, watercress), and phase II enzymes (vegetables) and *p*-glycoprotein (fruit juices, St. John's wort). For this reason, foods that contain complex mixtures of phytochemicals such as herbs, spices, teas, fruits, and vegetables may affect the activity of enzymes or active transport mechanisms and contribute to ethnic variations in drug disposition.

In 1998, the ICH issued the Guideline E5: Ethnic Factors in the Acceptability of Foreign Clinical Data to provide direction to the pharmaceutical industry when drug submissions are considered in more than one region with likely ethnic diversity [54]. The ICH and FDA subsequently issued a follow-up questions-and-answer document to formally respond to common questions regarding the type and extent of bridging studies required to support submissions in new regions [55, 56]. While this guidance is directed towards regulatory strategies for a bridging data package including PD, efficacy, and/or safety bridging studies, it recognizes the importance of appropriate characterization of pharmacokinetics as part of the data package. Pharmacokinetic processes which are biologically or biochemically mediated have the potential to exhibit differences between ethnic groups [57]. If the NCE characteristics are known, it should be possible to identify drugs that are likely to exhibit race differences in PK. Drugs which have a wide therapeutic range, flat concentration-versus-effect profile, linear PK, minimal metabolism, high/nonvariable bioavailability, low protein binding, and low potential for drug–drug interactions fall into the classification of less likely to be sensitive to ethnic factors. Those with

opposite characteristics, metabolized by a single pathway or polymorphic enzyme, and likely coadministered with multiple medications are more likely to be sensitive to ethnic factors. The ICH considerations for NCE classification as either less likely or more likely to be sensitive to ethnic factors based on their PK characteristics are provided in Table 5.

With the emerging globalization of drug development, more drugs from the Western market (United States/Europe) enter the Asian market (such as Japan). For this reason, the comparison of these two populations has become of particular interest. Many of the newer drug applications are employing PopPK approaches to assess ethnic differences in drug disposition. Data may be compiled across one or more phase II/III studies or may also include appropriate PK data from phase I control populations. The data may come from clinical trials conducted either in the population relevant to the new region or in the new region. As long as the PopPK model is robust enough and a sufficiently large number of subjects of the desired ethnic group is available for the analysis, the analysis results should be able to discern race-related differences in PK. Recent examples of PopPK analysis including Japanese subjects are the modeling of PK parameters for (a) olmesartan, which has shown the absence of ethnic differences in drug clearance between Japanese and Westerners, (b) telmisartan, which contributed differences in food intake conditions across populations to the differences in telmisartan PK, and (c) darifenacin, which exhibited a lower bioavailability in Japanese males [58–60]. The above three examples illustrate the diversity of results across different NCEs.

Since formal studies in a different ethnic population are often part of the drug's transition into a new region, the design of the trial is often adapted to a specific

**TABLE 5 NCE Sensitivity to Ethnic Factors**

| Less Likely to be Sensitive to Ethnic Factors                                                                        | More Likely to be Sensitive to Ethnic Factors                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| Linear PK                                                                                                            | Nonlinear PK                                                                                                                                         |
| Flat PD (effect v. concentration) curve for both efficacy and safety in range of recommended dosage and dose regimen | Steep PD curve for both efficacy and safety (small change in dose results in large change in effect) in range of recommended dosage and dose regimen |
| Wide therapeutic dose range                                                                                          | Narrow therapeutic dose range                                                                                                                        |
| Minimal metabolism or metabolism distributed among multiple pathways                                                 | Highly metabolized, especially through single pathway, thereby increasing potential for drug–drug interaction                                        |
| High bioavailability, thus less susceptibility to dietary absorption effects                                         | Metabolism by enzymes known to show genetic polymorphism                                                                                             |
| Low potential for protein binding                                                                                    | Administration of prodrug, with potential for ethnically variable enzymatic conversion                                                               |
| Little potential for drug–drug, drug–diet, and drug–disease interaction                                              | High intersubject variation in bioavailability                                                                                                       |
| Nonsystemic mode of action                                                                                           | Low bioavailability, thus more susceptibility to dietary absorption effects                                                                          |
| Little potential for inappropriate use                                                                               | High likelihood of use in setting of multiple comedications                                                                                          |
|                                                                                                                      | High likelihood for inappropriate use, e.g., analgesics and tranquilizers                                                                            |

From [54].

need in the NCE's development program and may take on many forms. Regulatory guidelines do not pose any limits on the study design of formal PK comparisons across ethnic groups but rather focus on the type of data that are required for ethnic comparisons. Traditional ethnic PK studies may follow the design outlined in Section 10.7.2 in which a single high-end therapeutic dose of the NCE is administered in an equal number of the healthy ethnic group (e.g., Japanese) and healthy Western or Caucasian controls in a parallel-group design. Alternative approaches may include the administration of more than one dose strength to each subject in either a randomized or ascending-dose cross-over design [61–63]. Administration of multiple dose levels across the anticipated therapeutic dose range allows determination of dose proportionality of the NCE as well as comparative PK. The decision of whether to randomize or escalate doses may depend on the likelihood of ethnic differences in the NCE and the timing of the study within the overall development program. From a statistical perspective, a randomized cross-over design is preferred; however, safety considerations may outweigh the need for a balanced design.

Pharmacodynamic markers such as blood glucose levels to determine the relative changes (RC) in glycemia values may be employed to relate potential differences in NCE exposure to a clinically relevant pharmacological response. Thomsen et al. was able to demonstrate that the higher exposure to repaglinide, an insulin secretagogue, in Japanese subjects was associated with a greater decrease in the RC values in this population [61]. If the drug cannot be safely administered in healthy subjects, a formal PK study may be conducted in the target population. A recent report summarizes the results for cefacitabine, which was administered twice daily for 14 days to an approximately equal number of Caucasians and Japanese subjects with breast cancer [64]. Blood and urine data were collected on the first and last day of drug administration and data analysis followed standard approaches [ANOVA with 90% confidence interval (CI) for the ratio of the LS means]. When drug development is scheduled to occur simultaneously in different regions (e.g., United States and Japan), one consideration might be to include both a Western and Japanese population in a “first-in-man” dose tolerance study conducted in one of the two regions. PK, PD, safety, and tolerance data allow early determination of potential ethnic differences in NCE behavior and appropriate adjustment in a global development program.

The same inclusion/exclusion considerations apply for ethnic comparisons as for other special population comparisons. Where possible, subjects should be healthy and free of concomitant medications, and the study should be powered with a sufficient number of subjects to allow statistical comparisons. Control groups should be age and gender matched whenever possible. The standard age range may require minor adjustment if regional requirements stipulate a higher minimum age for a typical phase I population. The population should be representative of the new ethnic region. If the formal PK study is not conducted in the new ethnic region (e.g., Japanese subjects are studied in the United States), inclusion criteria should clearly define the appropriate population. The study may be restricted to only include first-generation (Japanese) subjects or may be extended to second (or later) generation(s) of the ethnic population if no generation differences are expected. Since ethnic differences in NCE PK are partially attributed to differences in diet and life-style, a decision needs to be made whether the two populations are to receive the same meals during their study conduct period or whether the subjects

should receive meals that are consistent with their typical diet. In the case of a comparison to Japanese subjects, inclusion of smokers may be required since a higher number of Japanese smoke and exclusion of smokers might limit enrollment or is not representative of a typical population in the new ethnic region.

In summary, PK data relevant to all ethnic populations targeted for treatment need to be part of a submission package. Both formal PK studies or PopPK analysis of clinical trials can support the registration package. How the data are generated will depend on the NCE characteristics and the NCE development program.

### 10.7.8 OBESITY

Obesity can be considered a nutritional disorder and is defined as a state of excess body fat and body weight in proportion to an individual's height. The weight-to-height ratio, BMI, is used to determine whether an individual is obese. An adult with a BMI in the range of 20–24.9 is considered normal, an adult with a BMI of 25–29.9 is considered overweight, and individuals with a BMI of 30 or greater are considered obese [65, 66]. The number of obese adults and children has tremendously increased over the last 10–20 years, especially in the Western nations. Obesity is now recognized as a global epidemic and a significant health risk. While no linear relationship exists, obesity is associated with hypertension and more severe cardiovascular diseases, dyslipidemia (e.g., high total cholesterol or triglyceride levels), type 2 diabetes, gallbladder disease, osteoarthritis, some cancers (endometrial, breast, and colon), pulmonary disorders, and sleep apnea/sleep-disordered breathing [66–68]. Obesity in children and adolescents between 6 and 19 years of age is increasing more rapidly than obesity in adults. This population is especially at risk for many comorbid conditions such as hypotension and diabetes at an early age, which will affect their long-term health and life expectancy [69].

Obese subjects are becoming a major target population for drugs in the future. Use of medications in this population is twofold: (1) drug products are marketed and in development to treat obesity and (2) this population requires medications to treat the secondary illnesses described above. In the first scenario, the medications are specifically developed for the obese population. Regulatory expectations for the clinical evaluation of obesity drugs are outlined in both U.S. and European guidance documents [70, 71]. In contrast to drug development in most other therapeutic indications, first-in-man and other early clinical pharmacology studies are usually performed in obese subjects who are otherwise healthy. Furthermore, the majority of subsequent clinical pharmacology and biopharmaceutic studies with the exception of specialized phase I studies such as RI and HI studies and formal QTc studies will be conducted in the obese population. For obesity drugs, this population is considered the target population and does not represent a special population.

For the second scenario, obese subjects can be considered a special population rather than a target patient population. Effect of obesity on drug disposition has been studied as early as the late 1980s when obesity was first recognized as an emerging nutritional disorder, and it was found that modifications in body weight can affect drug distribution into tissues [72, 73]. While slightly liposoluble molecules are less likely affected by increases in body fat, lipophilic drugs may have a higher volume of distribution in the obese population and therefore a longer half-life. For

this reason, dose adjustments may be different in obese subjects based on the physicochemical characteristic of the drug.

The effect of obesity on the PK in various drug classes, including anti-infectives, anticancer, CNS, obesity, anesthetics, and beta blockers, has been reviewed [74]. Drug disposition in the obese can be altered not only due an increase and change in body mass (lean WT plus fat WT) but also due to intrinsic physiological changes and concurrent disease. While drug absorption is unlikely to be affected by obesity, both drug distribution and elimination may be altered. Blood flow into fatty tissue is poor, hemodynamic conditions in obesity may be altered, and livers in obese subjects may suffer from fatty infiltration which may lead to HI. Obesity may be associated with RI, and exposure of drugs which are primarily eliminated by the kidney may increase. However, obese subjects have larger kidneys than nonobese subjects [75], and the clearance of some renally eliminated drugs is not substantially affected [74, 76]. The effect of obesity on CYP was reviewed by Kotlyar and Carson, who concluded that changes in CYP activity differed across isozymes [77]. CYP3A4 activity was reported to decrease while that of CYP2E1 appeared to increase. The authors considered the findings regarding other CYP isozymes (1A2, 2C9, 2C19, and 2D6) inconclusive.

Since the potential effect of underlying diseases and intrinsic changes other than body mass are difficult to assess, all dose recommendations to date have been related to weight (i.e., milligram-per-kilogram recommendations). Different weight algorithms have been employed. The most common approaches are total, lean, or ideal body weight [73]. Recently use of predicted normal body weight (PNWT) has been proposed [78]. The PNWT corrects for excess fat weight (in relationship to lean body weight) that is above the weight of fat normally expected and may therefore be a better standard weight descriptor to describe body size for dose adjustments in obese patients. However, it remains difficult to predict the impact of obesity on PK, and each drug may behave differently. For drugs with a narrow therapeutic index, a study in obese subjects may be indicated. Anticancer agents are one category of drugs where the understanding of the influence of obesity is patchy and the outcome of treatment in this population may be detrimental due to the narrow therapeutic window of these agents. Toxicity for obese patients is impacted by appropriate dose adjustments and outcome data have suggested that obese patients may have a lesser disease-free survival and overall survival [79]. Based on his review, Navarro [79] recommends prospective studies to address optimal chemotherapy dosing in the obese population.

While formal obesity PK studies are currently not a regulatory requirement, it might be advisable to obtain PK information in the obese population if (a) the drug is likely prescribed in this population, (b) the physiochemical properties of the NCE suggest potential differences in drug disposition, and (c) the drug has a narrow therapeutic window. The data can be derived from either PopPK approaches or a formal study. The design of a formal PK study follows that outlined for HI with categorical comparison of the obese population to matched controls. Two subpopulations have been defined for obesity: moderately obese subjects with BMI values in the range of 30–39 kg/m<sup>2</sup> and morbidly obese subjects with BMI values equal to or greater than 40 mg/m<sup>2</sup>. PK data are compared in equal numbers of a group of obese subjects and demographically matched nonobese subjects. Designs may include only moderately obese subjects or may be further stratified between moderately and morbidly obese subjects [76, 80]. The decision to include morbidly obese subjects



depends on the likelihood that these subjects are a substantial part of the target patient population. Inclusion/exclusion criteria for secondary diseases and comedications follow the same considerations described for RI and HI studies and will vary between the obesity categories. Since the onset of secondary illness is not linearly related to obesity, there is a sufficiently large pool of reasonably healthy subjects with moderate obesity; however, morbidly obese subjects will likely require lenient inclusion/exclusion criteria. The number of subjects enrolled in each of the treatment arms will depend on statistical and logistical (in the case of morbidly obese) considerations, while PK sampling and data analysis strategies are the same as described for the other special populations.

In the current environment, formal studies in obese subjects are rarely performed for registration purposes and the majority of data derive from PopPK analyses. PopPK approaches are appropriate if the database contains a sufficiently large number of obese subjects with a wide enough BMI range to allow comparison to nonobese subjects. Since morbidly obese subjects are unlikely to be included in standard phase II/III safety and efficacy studies, PopPK analysis will rarely provide PK information for these subjects and a formal PK study may be needed if data in this population are required for a particular registration.

### 10.7.9 CONCLUSIONS

Pharmacokinetic data in special populations are registration requirements if the target patient population is likely to include the special population(s). U.S. and European regulatory guidelines require formal PK studies in geriatric, RI, and HI impairment subjects while PK data in other special populations such as women, ethnic groups, and obese subjects may be obtained by using population analysis approaches. Exceptions to these rules can be found in the specific regulatory guidelines. Study designs for formal PK studies carry common features across all special populations with special considerations to age- or disease-related inclusion/exclusion criteria and comedication requirements that are less restrictive than those in healthy normal subjects.

Over the last 5–10 years, population analysis approaches have obtained greater acceptance in the international community and PopPK results have either supplemented PK data of formal studies or replaced those of formal studies. Population analysis approaches are powerful tools to discern intersubject variability in PK parameters and are only limited by the inclusion/exclusion criteria of clinical studies which may preclude enrollment of (sufficient numbers) of a special population of interest. While the use of population analysis approaches to understand the effect of demographic, physiological, genetic, or life-style covariates on drug disposition, efficacy, or safety response to drugs is likely to increase over the next years, formal PK studies for some populations such as RI and HI will continue to be required to the limited enrollment of these subjects in the phase II and III program.

## REFERENCES

1. Food and Drug Administration (1993), Guidance for Industry: Guidelines for the study and evaluation of gender differences in clinical evaluation of drugs. U.S. Department of



- Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, July.
2. Gibaldi, M., and Perrier, D. (1982), *Pharmacokinetics*, 2nd ed., Marcel Dekker, New York.
  3. Food and Drug Administration (2001), Guidance for Industry: Statistical approaches to establishing bioequivalence. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research; available at: <http://www.fda.gov/cder/guidance/3616fnl.pdf>.
  4. Food and Drug Administration (2003), Guidance for Industry: Bioavailability and bioequivalence studies for orally administered drug products—General considerations. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: <http://www.fda.gov/cder/guidance/5356fnl.pdf>.
  5. Abernethy, D. R., and Azarnoff, D. L. (1990), Pharmacokinetic investigations in elderly patients. Clinical and ethical considerations, *Clin. Pharmacokinet.*, 19(5), 89–93.
  6. Vestal, R. E., and Gurwitz, J. H. (2000), Geriatric pharmacology, in Curruthers, S. G., Hoffman, B. B., Melmon, K. L., and Nierenberg, D. W., Eds., *Melmon and Morelli's Clinical Pharmacology*, 4th eds, McGraw-Hill, New York, pp. 1151–1177.
  7. Rowe, J. W., Andres, R., Tobin, J. D., et al. (1976), The effect of age on creatinine clearance in men: A cross-sectional and longitudinal study, *J. Gerontol.*, 31, 155–163.
  8. Woodhouse, K. W., and Wynne, H. A. (1988), Age-related changes in liver size and hepatic blood flow. The influence on drug metabolism in the elderly, *Clin. Pharmacokinet.*, 15, 287–294.
  9. Food and Drug Administration (1989), Guidance for Industry: Guidance for the study of drugs likely to be used in the elderly. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research.
  10. International Conference on Harmonisation (ICH) (1993), Studies in support of special populations: Geriatrics (E7); available at: <http://www.ich.org/LOB/media/MEDIA483.pdf>.
  11. Food and Drug Administration (2001), Guidance for Industry: Content and format for geriatric labeling. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: <http://www.fda.gov/cber/gdlns/gerlab.pdf>.
  12. Abernethy, D. R. (2001), Drug therapy in the elderly, in Atkinson, A. J., Daniels, C. E., Dedrick, R. L., Grudzinskas, C. V., and Markey, S. P., Eds., *Principles of Clinical Pharmacology*, Academic, New York, pp. 307–317.
  13. Guyton, A. C. (1987), The body fluids and the kidneys—renal disease, in Guyton, A. C. Ed., *Human Physiology and Mechanisms of Disease*, 4th ed., W.B. Saunders, Philadelphia, pp. 286–290.
  14. Food and Drug Administration (1998), Guidance for Industry: Pharmacokinetics in patients with impaired renal function—study design, data analysis, and impact on dosing and labeling. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: <http://www.fda.gov/cder/guidance/1449fnl.pdf>.
  15. European Medicines Agency (2004), Note for Guidance on the evaluation of the pharmacokinetics of medicinal products in patients with impaired renal function. CPMP/EWP/225/02; available at: <http://www.emea.eu.int/pdfs/human/ewp/022502en.pdf>.

16. Zini, R., Riant, P., Barre, J., et al. (1990), Disease-induced variations in plasma protein levels: Implications for drug dosage regimens (Part I), *Clin. Pharmacokinet.*, 19, 147–159.
17. Zini, R., Riant, P., Barre, J., et al. (1990), Disease-induced variations in plasma protein levels: Implications for drug dosage regimens (Part II), *Clin. Pharmacokinet.*, 19, 218–229.
18. Touchette, M. A., and Slaughter, R. L. (1991), The effect of renal failure on hepatic drug clearance. *DICP: Ann. Pharmacother.*, 25, 1214–1224.
19. Turnheim, K. (1991), Pitfalls of pharmacokinetic dosage guidelines in renal insufficiency, *Eur. J. Clin. Pharmacol.*, 40, 87–93.
20. Lam, Y. W. F., Banerji, S., Hatfield, C., et al. (1997), Principles of drug administration in renal insufficiency. *Clin. Pharmacokinet.*, 32(1), 30–57.
21. Launay-Vacher, V., Storme, T., Izzedine, H., et al. (2001), Pharmacokinetic changes in renal failure, *Presse Med.*, 30, 597–604.
22. Landray, M. J., Thambyrajah, J., McGlynn, F. J., et al. (2001), Epidemiological evaluation of known and suspected cardiovascular risk factors in chronic renal impairment, *Am. J. Kidney Dis.*, 38(3), 537–546.
23. Salem, M. M. (2002), Pathophysiology of hypertension in renal failure, *Semin. Nephrol.*, 22, 17–26.
24. Wheeler, D. C. (2001), Lipid abnormalities in the nephrotic syndrome: The therapeutic role of statins, *J. Nephrol.*, 14(Suppl), S70–75.
25. Robbins-Weilert, D. (2004), Design, conduct, and analysis of studies in patients with renal impairment, in Bonate, P. L., and Howard, D. R., Eds., *Pharmacokinetics in Drug Development: Clinical Study Design and Analysis, Vol. 1*, American Association of Pharmaceutical Sciences, pp. 177–208.
26. Bonate, P. L., Reith, K., and Weir, S. (1998), Drug interactions at the renal level. Implications for drug development, *Clin. Pharmacokinet.*, 34(5), 375–404.
27. Cockcroft, D. W., and Gault, M. H. (1976), Prediction of creatinine clearance from serum creatinine, *Nephron*, 16, 31–41.
28. Schwartz, G. J. (1976), A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine, *Pediatrics*, 58, 259–264.
29. Schwartz, G. J. (1984), A simple estimate of glomerular filtration rate in full term infants during the first year of life, *J. Pediatr.*, 104, 849–854.
30. Dahl, N. V. (2001), Herbs and supplements in dialysis patients: Panacea or poison, *Semin. Dialysis*, 14(3), 186–192.
31. Bailie, G. R., and Eisele, G. (1992), Continuous ambulatory peritoneal dialysis: A review of its mechanism, advantages, complications, and areas of controversy. *Ann. Pharmacother.*, 26, 1409–1420.
32. Taylor, C. A., 3rd, Abdel-Rahman, E., Zimmerman, S. W., et al. (1996), Clinical pharmacokinetics during continuous ambulatory peritoneal dialysis, *Clin. Pharmacokinet.*, 31(4), 293–308.
33. Lee, C. C., and Marbury, T. C. (1984), Drug therapy in patients undergoing hemodialysis, *Clin. Pharmacokinet.*, 9, 42–66.
34. Oellerich, M., Burdelski, M., Lautz, H.-U., et al. (1990), Lidocaine metabolite formation as a measure of liver function in patients with cirrhosis, *Ther. Drug Monit.*, 12:219–226.
35. Figg, W. D., Dukes, G. E., Lesesne, H. R., et al. (1995), Comparison of quantitative methods to assess hepatic function: Pugh's clarification, indocyanine green, antipyrine, and dextromethorphan, *Pharmacotherapy*, 15, 693–700.

36. Pugh, R. N. H., Murray-Lyon, I. M., Dawson, J. L., et al. (1973), Transection of the oesophagus for bleeding oesophageal varices, *Br. J. Surg.*, 60, 646–649.
37. Zakim, D., and Boyer, T. D. (1996), Hepatology, in *A Textbook of Liver Disease*, W.B. Saunders Company, Philadelphia.
38. Food and Drug Administration (2003), Guidance for Industry: Pharmacokinetics in patients with impaired hepatic function—study design, data analysis, and impact on dosing and labeling. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research; available at: <http://www.fda.gov/cder/guidance/3625fnl.pdf>.
39. European Medicines Agency (2005), Guideline on the evaluation of the pharmacokinetics of medicinal products in patients with impaired hepatic function. CPMP/EWP/2339/02; available at: <http://www.emea.eu.int/pdfs/human/ewp/233902en.pdf>.
40. Bonate, P. L., and Russell, T. (2004), Design, conduct, and analysis of studies in patients with hepatic impairment, in Bonate, P. L., and Howard, D. R., Eds., *Pharmacokinetics in Drug Development: Clinical Study Design and Analysis, Vol. 1*, American Association of Pharmaceutical Sciences, pp. 149–175.
41. Food and Drug Administration (1997), Guidance for Industry: M3 Nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, and ICH; available at: <http://www.fda.gov/cder/guidance/1855fnl.pdf>.
42. FDAMA Women and Minorities Working Group Report (1998), posted by Woodcock, J. at: <http://www.fda.gov/cder/guidance/women.pdf>.
43. ICH and Women: Gender consideration in the conduct of clinical trials; available at: <http://www.ich.org/cache/compo/276-254-1.html>.
44. European Medicines Agency (2005), ICH Step 5: Gender consideration in the conduct of clinical trials. EMEA/CHMP/3916/2005; available at: <http://www.emea.eu.int/pdfs/human/ich/391605en.pdf>.
45. Sobue, S., Tan, K., Shaw, L., et al. (2004), Comparison of the pharmacokinetics of fosfluconazole and fluconazole after single intravenous administration of fosfluconazole in healthy Japanese and Caucasian volunteers, *Eur. J. Clin. Pharmacol.*, 60(4), 247–253.
46. Cascorbi, I. (2003), Pharmacogenetics of cytochrome P4502D6: Genetic background and clinical implications, *Eur. J. Clin. Investigation*, 33, 17–22.
47. Schwarz, U. I. (2003), Clinical relevance of genetic polymorphisms in the human CYP2C9 gene, *Eur. J. Clin. Investigation*, 33, 23–30.
48. Daly, A. K. (2004), Pharmacogenetics of the cytochromes P450, *Curr. Top. Med. Chem.*, 4(16), 1733–1744.
49. Ozawa, S., Soyama, A., Saeki, M., et al. (2004), Ethnic differences in genetic polymorphisms of CYP2D6, CYP3A5, CYP3A4, and MDR1/ABCB1, *Drug Metabol. Pharmacokinet.*, 19(2), 83–95.
50. Roy, J. N., Lajoie, J., Zijenah, L. S., et al. (2005), CYP3A5 genetic polymorphisms in different ethnic populations, *Drug Metabol. Disp.*, 33(7), 884–887.
51. Sakaeda, T., Nakamura, T., and Okumura, K. (2004), Pharmacogenetics of drug transporters and its impact on the pharmacotherapy, *Curr. Top. Med. Chem.*, 4(13), 1385–1398.
52. Bosch, T. M., Meijerman, I., Beijnen, J. H., et al. (2006), Genetic polymorphisms of drug-metabolizing enzymes and drug transporters in the chemotherapeutic treatment of cancer, *Clin. Pharmacokinet.*, 45(3), 253–285.
53. Harris, R. Z., Jang, G. R., and Tsunoda, S. (2003), Dietary effects on drug metabolism and transport, *Clin. Pharmacokinet.*, 42(13), 1071–1088.

54. International Conference on Harmonisation (ICH) (1998), Ethnic factors in the acceptability of foreign clinical data E5(R1); available at: <http://www.ich.org/LOB/media/MEDIA481.pdf>.
55. Food and Drug Administration (2004), Guidance for Industry: Ethnic factors in the acceptability of foreign clinical data. Questions and answers. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, ICH; available at: <http://www.fda.gov/cder/guidance/6200fnl.pdf>.
56. International Conference on Harmonisation (ICH) (2003), E5 Ethnic factors: Questions and answers; available at: <http://www.ich.org/LOB/media/MEDIA1194.pdf>.
57. Johnson, J. A. (1997), Influence of race or ethnicity on pharmacokinetics of drugs, *J. Pharm. Sci.*, 86(12), 1238–1233.
58. Yoshihara, K., Gao, Y., Shiga, H., et al. (2005), Population pharmacokinetics of olmesartan following oral administration of its prodrug, olmesartan medoxomil: in healthy volunteers and hypertensive patients, *Clin. Pharmacokinet.*, 44(12), 1329–1342.
59. Tatami, S., Yamamura, N., Sarashina, A., et al. (2004), Pharmacokinetic comparison of an angiotensin II receptor antagonist, telmisartan, in Japanese and western hypertensive patients using population pharmacokinetic method, *Drug Metab. Pharmacokinet.*, 19(1), 15–23.
60. Kerbusch, T., Wahlby, U., Milligan, P. A., et al. (2003), Population pharmacokinetic modeling of darifenacin and its hydroxylated metabolite using pooled data, incorporating saturable first-pass metabolism, CYP2D6 genotype and formulation-dependent bioavailability, *Br. J. Clin. Pharmacol.*, 56(6), 639–653.
61. Thomsen, M. S., Chassard, D., Evane, E., et al. (2003), Pharmacokinetics of repaglinide in healthy Caucasian and Japanese subjects, *J. Clin. Pharmacol.*, 43(1), 23–28.
62. Jhee, S. S., Lyness, W. H., Rojas, P. B., et al. (2004), Similarity of insulin detemir pharmacokinetics, safety, and tolerability profiles in healthy Caucasian and Japanese American subjects, *J. Clin. Pharmacol.*, 44(3), 258–264.
63. Van Giersbergen, P. L., and Dingemans, J. (2005), Comparative investigation of the pharmacokinetics of bosentan in Caucasian and Japanese healthy subjects, *J. Clin. Pharmacol.*, 45(1), 42–47.
64. Reigner, B., Watanabe, T., Schuller, J., et al. (2003), Pharmacokinetics of capecitabine (Xeloda) in Japanese and Caucasian patients with breast cancer, *Cancer Chemother. Pharmacol.*, 52(3), 193–201.
65. World Health Organization (1998), Report of a WHO consultation on obesity: Preventing and managing the global epidemic, WHO, Geneva, June 3–5.
66. Center for Disease Control and Prevention (CDC). Overweight and obesity; available at: <http://www.cdc.gov/nccdphp/dnpa/obesity/>.
67. Li, Z., Bowerman, S., and Heber, D. (2005), Health ramifications of the obesity epidemic, *Surg. Clin. N. Am.*, 85(4), 681–701.
68. Poirier, P., Giles, T. D., Bray, G. A., et al. (2006), Obesity and cardiovascular disease: Pathophysiology, evaluation, and effect of weight loss: An update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism, *Circulation*, 113(6), 898–918.
69. Center for Disease Control and Prevention (CDC) (2005), Children and teens told by doctors that they were overweight—United States, 1999–2002, *MMWR, Rep.*, 54(34), 848–849.

70. Food and Drug Administration (1996), Guidance for the clinical evaluation of weight-control drugs. Division of Metabolic and Endocrine Drug Products; available at: <http://www.fda.gov/cder/guidance/obesity.pdf>.
71. European Medicinal Agency (1997), Note for guidance on clinical investigation of drugs used in weight control. CPMP/EWP/281/96. December 1997; available at: <http://www.emea.eu.int/pdfs/human/ewp/028196en.pdf>.
72. Cheymol, G. (1988), Drug pharmacokinetics in the obese, *Fundament. Clin. Pharmacol.*, 2(3), 239–256.
73. Cheymol, G. (1993), Clinical pharmacokinetics of drugs in obesity: An Update, *Clin. Pharmacokinet.*, 25(2), 103–114.
74. Cheymol, G. (2000), Effects of obesity on pharmacokinetics. Implications for drug therapy, *Clin. Pharmacokinet.*, 39(3), 215–231.
75. Naeye, K. L., and Rowe, P. (1970), The size and number of cells in several visceral organs in human obesity, *Am. J. Clin. Path.*, 54, 251–253.
76. Dvorchik, B. H., and Damphousse, D. (2005), The pharmacokinetics of daptomycin in moderately obese, morbidly obese, and matched nonobese subjects, *J. Clin. Pharmacol.*, 45, 48–56.
77. Kotlyar, M., and Carson, S. W. (1999), Effects of obesity on the cytochrome P450 enzyme system, *Int. J. Clin. Pharmacol. Therap.*, 37(1), 8–19.
78. Duffull, S. B., Dooley, M. J., Green, B., et al. (2004), A standard weight descriptor for dose adjustment in the obese patient, *Clin. Pharmacokinet.*, 43(15), 1167–1178.
79. Navarro, W. H. (2003), Impact of obesity in the setting of high-dose chemotherapy, *Bone Marrow Transplant.*, 31(11), 961–966.
80. Sarich, T. C., Teng, R., Peters, G. R., et al. (2003), No influence of obesity on pharmacokinetics and pharmacodynamics of melagatran, the active form of the oral direct thrombin inhibitor ximelagatran, *Clin. Pharmacokinet.*, 42(5), 485–492.

# 10.8

---

## *Musculoskeletal Disorders*

**Masami Akai**

*Director, Rehabilitation Hospital,  
National Rehabilitation Center Japan, Saitama, Japan*

### **Contents**

|          |                                                  |     |
|----------|--------------------------------------------------|-----|
| 10.8.1   | Introduction                                     | 564 |
| 10.8.2   | Definition of Musculoskeletal Disorders          | 565 |
| 10.8.2.1 | Musculoskeletal Disorders                        | 565 |
| 10.8.2.2 | Aims of Treatment                                | 565 |
| 10.8.2.3 | Concept of “Construct”                           | 566 |
| 10.8.2.4 | Assessment of Function and Quality of Life (QOL) | 566 |
| 10.8.3   | Background of Clinical Trials                    | 567 |
| 10.8.3.1 | Meaning of Clinical Trial                        | 567 |
| 10.8.3.2 | Estimate and Testing                             | 567 |
| 10.8.3.3 | Various Biases                                   | 568 |
| 10.8.3.4 | Methods of Clinical Trials                       | 568 |
| 10.8.3.5 | Types of Clinical Studies                        | 568 |
| 10.8.3.6 | Random Allocation                                | 569 |
| 10.8.4   | Methodological Assessment for Clinical Trials    | 570 |
| 10.8.4.1 | Significance of Clinical Trials                  | 570 |
| 10.8.4.2 | Critical Evaluation of RCTs                      | 570 |
| 10.8.5   | Preparing Necessary Tools for Clinical Trials    | 571 |
| 10.8.5.1 | Developing Protocol                              | 571 |
| 10.8.5.2 | Patients                                         | 571 |
| 10.8.5.3 | Interventions                                    | 573 |
| 10.8.5.4 | Outcomes                                         | 574 |
| 10.8.5.5 | How to Choose Right Outcome Measures             | 574 |
| 10.8.5.6 | Questionnaires                                   | 575 |

|          |                                          |     |
|----------|------------------------------------------|-----|
| 10.8.6   | Data Analysis and Interpretation         | 576 |
| 10.8.6.1 | Sample Size                              | 576 |
| 10.8.6.2 | Intention to Treat                       | 577 |
| 10.8.6.3 | Multiple Comparisons                     | 577 |
| 10.8.6.4 | Post Hoc Analysis and Subgroup Analysis  | 577 |
| 10.8.6.5 | Data Synthesis and Combined Result       | 577 |
| 10.8.6.6 | Side Effects                             | 578 |
| 10.8.6.7 | Reporting Clinical Trials                | 578 |
| 10.8.7   | RCT and Future Directions                | 579 |
| 10.8.7.1 | Impact from RCT to Daily Practice        | 580 |
| 10.8.7.2 | Internal Validity and External Validity  | 580 |
| 10.8.7.3 | Quantitative Study and Qualitative Study | 580 |
| 10.8.7.4 | Further Suggestions                      | 582 |
|          | References                               | 582 |

### 10.8.1 INTRODUCTION

When we look at past medical history, we can only find a few definite treatments that focus directly on their etiology, such as vaccination for some infectious diseases, vitamin C for scurvy, and, recently, gene therapy. The evidence for presently accepted treatments in our practice is insufficient; it ranges from a scattered presence to total absence. Unfortunately, apart from a few dichotomous results, we cannot expect qualitatively definite results of treatment for all clinical cases. The effect of the interventions is usually assessed quantitatively, comparing before and after treatment, according to certain outcome measures. The key to solving many of the problems in medical practice is to establish valid and reliable methods of evaluating the differences in the outcomes before and after treatment.

I started my postgraduate training as an orthopedic surgeon and then shifted my career to rehabilitation medicine. I am not, by any standard, an expert in all of the subjects covered here: clinical epidemiology, biostatistics, and related fields. However, I have had chances to join teams to organize clinical trials [1, 2]. I would like to describe the necessary steps for conducting a clinical trial from the standpoint of a clinician.

The term *evidence-based medicine* (EBM), which is the integration of best research evidence with clinical expertise and patient values, has already established wide popularity, even in the musculoskeletal field. When we review the clinical trial trends in the musculoskeletal field, we notice several characteristic points, which are obviously different from other areas.

In the musculoskeletal field, there are the typical clinical trials that use placebo-controlled, double-blind methods, such as drug trials for osteoporosis or rheumatoid arthritis. However, if we wish to compare one interventional procedure with another, the double-blind design is often impossible to do methodologically. Only in an open-label trial, where patients know about their treatment, is it possible to conduct the trial under random allocation. With such limitations, we have to suppress various biases as much as possible. Another point that needs attention is the selection of outcome measures suitable for musculoskeletal disorders. Even though there are primary malignant tumors, such as osteosarcoma or a cancer lesion as secondary



metastasis, which are sometimes lethal, the majority of musculoskeletal disorders that develop clinical manifestation are not life-threatening but do impair function. We have to use outcome measures for observing the functional ability of the patients, not for the numerical laboratory data.

In this chapter I would like to discuss the following:

1. Definition of musculoskeletal disorders
2. Background of clinical trials
3. Methodological assessment for clinical trials
4. Preparation of necessary tools for clinical trials
5. Data analysis and interpretation
6. RCT and our future direction

## 10.8.2 DEFINITION OF MUSCULOSKELETAL DISORDERS

### 10.8.2.1 Musculoskeletal Disorders

The term *musculoskeletal disorders* refers to conditions that involve the supporting structures of the body, such as the trunk and the bones and joints of the extremities. In other words, musculoskeletal disorders involve locomotor functions that have suffered from trauma, diseases, anomalies, and inevitable aging and are mainly attended to by orthopedic surgeons.

Musculoskeletal disorders are the most common cause of severe, long-lasting pain and physical disability affecting hundreds of millions of people in the world. The extent of such problems and the growing burden is described as “age quake” instead of earthquake. The impact on society from these disorders, combined with the recognition that our health care resources need to be used more efficiently, has led to the organized movement of the “Bone and Joint Decade 2000–2010” [3]. This umbrella organization is attempting to raise social awareness of the suffering and pain of musculoskeletal disorders, as well as the growing burden and cost to society that will come with this age quake. We need to advance clinical research in order to reduce the burden of musculoskeletal diseases.

### 10.8.2.2 Aims of Treatment

The aims of treatment for musculoskeletal disorders are not to reduce mortality or morbidity but to increase mobility and function, to relieve pain, or to improve quality of life. We have to assess the results of our interventions with a view to the improvement in disability or functional limitation after treatment. In this field, the clinical indicators are neither a 5-year survival rate nor an infantile survival rate. Our aims are to improve the health-related quality of life of the millions of patients suffering from musculoskeletal disorders such as joint diseases, spinal disorders, severe trauma to the body and extremities, and deformity and crippling diseases in children.

When we talk about outcome measures in disability assessment, the key issue to be discussed is what “function” and “disability” mean and how to measure them.

### 10.8.2.3 Concept of “Construct”

Our function or ability is measured in accordance with conceptual “construct” [4]. We cannot measure this function or ability directly.

In the past, orthopedic surgeons mainly evaluated the amount of disease involvement in view of physiological or biochemical parameters, pathological findings, and impairments such as range of motion or muscle strength. It was thought that these indicators were objective and independent of the unreliable responses or feelings of patients. However, are these indicators truly reflective of the needs of patients and their families to know how much they can expect as a result of treatment? Their expectations and questions would be more like, “Can I walk again?” or “Will she be able to live by herself as before?”

When a certain standardized scale is not available to measure the function or ability of interests, a new construct, according to a concept for measurement, has to be made.

Conventional clinical indicators, such as a 5-year survival rate or an infantile mortality rate, have been recognized as very important health measures in assessing health conditions in society, but lately, the situation has been changing. The recent close-up of patient-based outcome measure is deeply related to this trend [5]. We have to recognize the demand and organize those types of outcome measures for the individual [6].

### 10.8.2.4 Assessment of Function and Quality of Life (QOL)

Functional assessment, from the physician’s point of view, has been mainly used to evaluate problems at the impairment level. Traditionally, orthopedic surgeons have used clinician-based outcomes such as pain score, measurement of range of motion, muscle manual test, or X-ray findings. However, due to various biases and poor judgments of accuracy levels, clinician-based outcomes often affect the clinical ratings on the functional status of patients [7].

For the past two decades, new outcome movements have placed importance on patient-based functional assessment and measurement of health-related quality of life [5, 8].

It is because the statistically significant changes of the visual analog scale (VAS) for pains, from 68 to 42, do not mean any improvement for the patient with chronic pain. The improvement of knee flexion range of 35° does not answer the questions from patients and their families.

The end result of health services should take into account the experiences, preferences, and values of the patients [9]. The discomfort described by patients should be managed properly. We have to evaluate and treat the patient not the disease.

The so-called health-related QOL became the main target of the outcome measures related to patients’ values. As well, of course, it is clear that health-related QOL assessments alone are not enough. In 2001, the World Health Organization (WHO) proposed a new “health” concept called the International Classification of Functioning, Disability and Health (ICF) [10], which describes the pluralistic conception of functional disabilities [11]. Based on the ICF concept, for example, we would assume the problems associated with patients with knee osteoarthritis were as follows:

|                           |                                                    |
|---------------------------|----------------------------------------------------|
| Health condition          | Osteoarthritis                                     |
| Impairment                | Knee pain, limited joint movement, muscle weakness |
| Activity limitation       | Difficulty in mobility-related daily activities    |
| Participation restriction | Difficulty in participating in social life         |

Many items and domains in the outcome measures should be constructed to identify all the characteristics of subjects, with a view to disability and impairment. Irrgang and Anderson [12] described the process for developing such a new measure for health-related quality of life when they designed clinical research related to the knee. In many cases, several outcome measures must be combined to cover every aspect from functional impairment to quality of life.

### 10.8.3 BACKGROUND OF CLINICAL TRIALS

#### 10.8.3.1 Meaning of Clinical Trial

Two related statistical factors have to be considered in order to evaluate the quality of clinical studies. This is essential for the understanding of the quality of studies based on the level of evidence. First, in clinical trials we deal with human beings as research objects. Each human being is different; response is always uncertain, backgrounds vary, and mental influence is inevitable. We should compare groups that are similarly distributed, except for the specific factor that is the target of assessment. Thus, observed differences at the assessments can be directly attributed to the study targets. Second, we mainly evaluate an extracted specimen of limited numbers and then extrapolate the conclusion toward the source population. We attribute the presumption from the sample population to the general population, using statistical inference techniques. Using the statistics observed in the investigated samples, the parameters of the source population are estimated. Depending upon these two premises, we use sampling, randomization, comparison, and other statistical methods. This is a statistical outlook of the world around us.

#### 10.8.3.2 Estimate and Testing

When we treat a patient, we regard him or her not just as a single case but as a representative case among many patients with the same disorder. This concept of patient treatment had already been described by Hippocrates in ancient Greece. There are two important statistical concepts we should consider, probability and confidence interval.

The first topic is the probability, that is,  $P$  values that the result would arise by chance. Depending on the choice of cut-off level ( $P < 0.05$  or  $P < 0.01$ ), we apply a different level of “statistical significance” to the trial. If we find a result in the statistically significant range, we could reject the null hypothesis that there is no real difference between two groups. But a  $P$  value in the nonsignificant range just shows that either there is no difference between the groups or there are too few samples to reveal a difference, if it actually exists.

Also, if we repeated the same trial hundreds of times, we would not get the exact same result each time. We cannot omit an incidence in each trial because we usually conduct only one trial, but we could find a particular level of difference on average

with a 95% confidence interval. If we conducted a trial with a much larger sample size, we would be more likely to show whether a significant result exists or not. The answer is in the comparison of the 95% confidence limits and the clinical significant levels. Very narrow 95% confidence limits suggest definitive results and exclude any clinical obscurity in comparisons. The confidence intervals reveal whether the trials support the result and to what extent and whether any further studies need to be done to reinforce the result.

### **10.8.3.3 Various Biases**

The results of a trial consist of three parts: (1) true value, (2) random error (imprecision) by chance, and (3) systematic error (bias) by other factors. We cannot eradicate random error, but we can eliminate the other biases by using well-designed study protocols. The goal is to suppress the biases as much as possible and improve the quality of the trial; different study designs require different steps to reduce systematic errors. In clinical epidemiology textbooks, there are various biases pointed out such as sampling bias, selection bias, and information bias.

Confounding is the distortion of an association between two factors brought about by the association of another extraneous factor; for example, the association between lung cancer and alcohol consumption is confounded by smoking. The methods of controlling confounding in the design of a study are: restrict the participants to the study, match individuals with other comparison groups, or use random allocation. In the analysis of a study, restrict once again the participants in the data to stratify the individuals into subgroups according to categories of confounding factor and use multivariate methods.

### **10.8.3.4 Methods of Clinical Trials**

To assess the effectiveness of a certain intervention, it is necessary to compare the baseline (before state) with the result (after state) of treatment. But generally speaking, the more abnormal value often shows more improvement.

Detected improvement does not always imply that the intervention was effective. Several other factors may have contributed to the same kind of improvement; such possible factors are natural recovery, “regression to the mean” phenomenon, or the psychological effect called placebo effect. We have to compare with control to get the true change rates by test intervention.

In most clinical trials called controlled trials, there is a group receiving a certain treatment and a comparative group. An essential premise of this kind of trial is that there is genuine uncertainty as to which treatments will be best for the patient. It is this uncertainty that justifies random allocation of patients after the consent forms to enter the trial have been given. Therefore, patients should have the targeted disease and satisfy all the required conditions, that is, inclusion criteria.

### **10.8.3.5 Types of Clinical Studies**

If there were a clinical question about a patient’s management, we should perform a clinical study, no matter what the study design is. The basic structures of a study design consist of a combination of some dichotomous divisions.

1. *Experimental (Interventional) Study or Observational Study* In experimental study, a population is selected for a planned trial of a regimen whose effects are measured by comparing the outcome of the regimen in the experimental group with the outcome of another regimen in the control group. Observational study is a nonexperimental study that does not involve any intervention.
2. *Comparative or Noncomparative Design* In a comparative design, the study design has a control group to compare with the active group. In a non-comparative design, there is no control group to compare with the active group.
3. *Intergroup or Intragroup Design* This is comparison of interventional effects among the groups, or comparison in the same single group, before and after the intervention.
4. *Cross-Sectional or Longitudinal Design* This is the directionality of the study. Data are collected at a single time point or collected at two or more time points and followed up later. But some cross-sectional studies refer retrospectively to experiences in the past.
5. *Retrospective or Prospective Design* This is a part of longitudinal studies. Data are collected at baseline points and other points in the past or future. The timing of data collection is also called concurrent, historical, or mixed.

Considering these dichotomous divisions, a widely used basic study design is a parallel design, which compares the groups, or a cross-over design, which compares the before and after. In the parallel design, subjects are divided into two or more groups, each group receives each treatment concurrently, and the results of treatment among the groups are compared. In the cross-over design, subjects receive two or more treatments at stated intervals and the effect of treatment, within the same subject, is evaluated. The cross-over design has the advantage of a smaller sample size than the parallel design but requires caution to minimize the fluctuating effect of disease severity (“order effect”) and maintain the stability of patients’ symptoms using a “wash-out period.”

The randomized controlled trial (RCT) provides the most powerful proof of treatment efficacy:

*Randomized Controlled Trials* Subjects are randomly allocated into two or more groups. The groups are followed up for a specified time period and assessed in terms of outcomes defined at the start of trials. As the groups, on average, are identical except for the intervention as study target, any differences in outcomes are attributed, in theory, to the interventions.

*Open label* Subjects know which treatment they are receiving.

*More Sophisticated Randomized Controlled Trials*

*Single blind* Subjects do not know which treatment they are receiving.

*Double blind* Neither subjects nor investigators know who is receiving which treatment.

### 10.8.3.6 Random Allocation

Blinding patients to which treatments they are receiving is thought to be essentially important because the effect of the psychological expectations of the patients are

suppressed. In drug therapy, blinding is achieved with a placebo as a control. However, in certain situations, such as a comparative trial between surgery and nonsurgery, blinding is not easily possible. Double-blind techniques for surgical treatments are tentatively proposed; the anesthesiologists and surgeons in charge will maintain silence about treatment choices and patients and the persons assessing the outcomes will not be aware of the treatment choices. Because of the technical difficulty of being unaware of allocation, we sometimes have to accept that patients would know the treatment they received.

Random allocation is done immediately after completing the original registration for a trial with informed consent. However, patients often hesitate to enter a trial because, at the time of registration, they do not know which treatment is indicated for them. Zelen proposed a new method called “randomized consent design” in which random allocation is already performed at the candidate stage of being involved in a trial. Explanation of informed consent is done only about the allocated treatment [13, 14].

#### **10.8.4 METHODOLOGICAL ASSESSMENT FOR CLINICAL TRIALS**

Because of the explosive increase in health care information, all clinicians must be careful of scientific evidence regarding efficacy of treatment methods based on available medical literature. How are we to distinguish the good from the bad among the flood of information? We need a systematic way to select medical information including RCTs.

##### **10.8.4.1 Significance of Clinical Trials**

The practice of EBM is a process of self-directed, life-long, never-ending learning for our patients. Clinicians need to have the most up-to-date clinical information. Our goal is to continuously strive to improve our medical knowledge by scientific methodology. The introduction of newer information technology and medical knowledge has improved the practice of medicine. If we can properly manage to get all of the clinical information from contemporary medical practice, we will be able to master all of the medical experiences of that same period. The collected results could be the most verified medical knowledge that we have ever had.

In the musculoskeletal field, in-patients are, in most cases, operated on, but the majority of out-patients and some in-patients are treated with a combination of conservative therapeutic methods. We need to develop a systematic strategy for the indication and selection of various treatments. The accumulated results of RCTs could be the essential medical information for the most up-to-date treatments.

##### **10.8.4.2 Critical Evaluation of RCTs**

Randomized controlled trials are now regarded as having superior clinical significance, that is, the gold standard for proper treatment, and are recognized as the primary way to provide rigorous proof of efficacy. However, considering the mixture of facts and errors found in the results of studies, we have to check each step of the trial process, even in the case of RCTs. All trials can be scored on methodological

quality according to a few criteria. Basically, these criteria consist of four main categories [15–17]:

1. Design and study population
2. Description of intervention
3. Measurement of outcomes and follow-up
4. Data presentation and analysis

In the recent analyses, these 4 categories were usually further subdivided into a set of nearly 20 items, each with a given weight [18]. Various scales for quality assessment are now available and after some modifications, those for musculoskeletal disorders have been employed in several systematic reviews and meta-analyses [19, 20]. Introduced here is one of the sophisticated examples of a criteria list for assessing quality of trials in a published article (Table 1).

### **10.8.5 PREPARING NECESSARY TOOLS FOR CLINICAL TRIALS**

On the basis of my personal experience, the next few sections describe the minimum essential steps in conducting a clinical trial (Table 2); the reader should refer to appropriate textbooks for further reading [21–26]. Recently, this kind of information has also been available on the Internet; the Resource Center for Randomized Trials [27] is an example. The Resource Center has various activities, and one of the useful services is a Web-based library for accessing information about trials; there are checklists for trials, consent forms for participants, and patient information leaflets. Other public information on clinical trials is also provided by the U.S. government [28].

#### **10.8.5.1 Developing Protocol**

The first step in conducting a trial is developing the protocol. When the details of how the trial is to be conducted are determined, it is essential to reduce various relating biases. Farrell and Spark [29] described a protocol checklist as a detailed method to use when conducting a trial (Table 3). Developing a protocol takes a long time, sometimes up to a few years to get funding, secure ethical approval, and organize a trial collaboration team. I think the best thing for anyone planning a trial for the first time to do is to refer to some successful protocols. Many protocols are now available on the Internet. A protocol should clearly show how outcomes will be measured, data collected, and the analysis conducted.

#### **10.8.5.2 Patients**

Use of inclusion/exclusion criteria is important to make it possible for the trial results to be more reliable. Inclusion criteria are used to identify appropriate subjects, keep them safe, and able to answer the research questions. Exclusion criteria are also very important in participant recruitment to avoid unnecessary involvement of inappropriate subjects in the study.



**TABLE 1 Criteria List for Assessing Methodological Quality of Trials**

| Heading      | Subheading                     | Descriptor                                                                                                                                                                                                      |
|--------------|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title        |                                | Identify the study as a randomized trial.                                                                                                                                                                       |
| Abstract     |                                | Use a structured format.                                                                                                                                                                                        |
| Introduction |                                | State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses.                                                                                                        |
| Methods      | Protocol                       | Describe                                                                                                                                                                                                        |
|              |                                | Planned study population with inclusion or exclusion criteria.                                                                                                                                                  |
|              |                                | Planned interventions: their nature, content, and timing.                                                                                                                                                       |
|              |                                | Primary and secondary outcome measure(s) and the minimum important difference(s), and indicate how the target sample size was estimated.                                                                        |
|              |                                | Reasons for statistical analyses chosen, and whether these were completed on an intention-to-treat basis.                                                                                                       |
|              | Assignment                     | Mechanisms for maintaining intervention quality, adherence to protocol, and assessment of fidelity.                                                                                                             |
|              |                                | Prospectively defined stopping rules (if warranted).                                                                                                                                                            |
|              |                                | Describe                                                                                                                                                                                                        |
|              | Masking (blinding)             | Randomization (e.g., individual, cluster, geographic).                                                                                                                                                          |
|              |                                | Allocation schedule method.                                                                                                                                                                                     |
| Results      | Participant flow and follow-up | Method of allocation concealment.                                                                                                                                                                               |
|              |                                | Describe                                                                                                                                                                                                        |
|              |                                | Mechanism for maintaining blind and allocation schedule control.                                                                                                                                                |
|              |                                | Evidence for successful blinding.                                                                                                                                                                               |
|              |                                | Provide a trial profile summarizing participant flow, numbers and timing of randomization assignment, interventions, and measurements for each randomized group.                                                |
|              | Analysis                       | State estimated effect of intervention on primary and secondary outcome measures, including a point estimate and measure of precision (confidence interval).                                                    |
|              |                                | State results in absolute numbers when feasible (e.g., 10/20, not 50%).                                                                                                                                         |
|              |                                | Present summary data and appropriate descriptive and inferential statistics in sufficient detail to permit alternative analyses and replication.                                                                |
|              |                                | Describe prognostic variables by treatment group and any attempt to adjust.                                                                                                                                     |
|              |                                | Describe protocol deviations.                                                                                                                                                                                   |
| Discussion   | Discussion                     | State specific interpretations of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity, including appropriate quantitative measures when possible. |
|              |                                | State general interpretation of the data in light of the available evidence.                                                                                                                                    |

Source: Modified from Machin et al. [21], with permission.

In spite of application of well-defined inclusion/exclusion criteria, it is commonly thought that about half of all clinical trials do not achieve their planned sample size [30]. As a recruitment strategy, necessary sample size is calculated to cover the expected number of dropout cases. It seems to me that the calculated number of the sample size is just for reference as an appeal to the cautious attitude of the researcher when conducting the trial.

**TABLE 2 Necessary Steps to Conduct Trials in Musculoskeletal Disorders**

|                          |                                                                                                                                                    |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Developing protocol      | Aim and goals with using strategies<br>Simple pragmatic design<br>Minimal amount of data correction<br>Random allocation<br>Ethical considerations |
| Patients                 | Clear inclusion/exclusion criteria<br>Minimal burden for patients<br>Recruitment strategies                                                        |
| Interventions            | Drug, therapeutic exercise, physical agents, orthosis,<br>Education or care package                                                                |
| Outcomes                 | Assessment of functions and QOL, good questionnaire<br>Optimal combining outcome measures<br>Data management procedures                            |
| Time constrained         | Limited core activity                                                                                                                              |
| Staff and their training | Minimal demand<br>Sufficient preparation and supporting                                                                                            |

**TABLE 3 Developing a Protocol for Clinical Trial**

|                                                                        |
|------------------------------------------------------------------------|
| Title                                                                  |
| Summary                                                                |
| Background and rationale for the trials                                |
| Hypothesis to be tested                                                |
| Primary outcome(s)                                                     |
| Secondary outcomes                                                     |
| Inclusion and exclusion criteria                                       |
| Interventions to be tested                                             |
| Estimated sample size                                                  |
| Information for patients and consent                                   |
| Analyses plan, concluding dummy table                                  |
| How patients will be entered into the study, concealment of allocation |
| Duration and methods for follow-up                                     |
| Data collection, including questionnaires                              |
| Trial management                                                       |
| Trial supervision                                                      |
| Publication policy                                                     |
| Reference                                                              |

*Source:* Modified from Duley and Farrell [23], with permission.

### 10.8.5.3 Interventions

Treatments in musculoskeletal disorders, apart from surgical interventions, are classified into “drug therapy,” “therapeutic exercise,” “physical agents,” “orthotics and devices,” “education or care packages,” and “others.” When these intervention methods are indicated in the experimental group, it is important how comparative methods are selected in the control group. It is impossible to use the double-blind technique with placebo in musculoskeletal disorders because the comparative control is obviously different from the experimental content for the trial participants. The patients already know which treatments have been allocated and only

the assessors are unaware of the allocation results. This “open-label method” is often indicated in this field.

#### 10.8.5.4 Outcomes

The content of outcomes used in clinical trials should be important to patients, usually the five D’s: death, disease in clinical course, discomfort in symptoms, disability in activities of daily living (ADL), and dissatisfaction in QOL. There are many de facto standards for outcome measures in musculoskeletal disorders (Table 4).

Another important point in the use of outcome measures is the timing of application. Follow-up periods for outcome measures are categorized as short term (less than 6 weeks), intermediate term (6 weeks to 1 year), or long term (more than 1 year follow-up).

In an RCT, the results basically underestimate the difference between the comparative groups because the analysis is conducted with a more conservative standpoint for assessment (an intention-to-treat principle), and that will tend to dilute the estimate of the true difference. The content of interventions often fluctuates depending upon its nature, application timing, and patients’ adherence. When the patients receive the strict treatment, they can expect more obvious efficacy than that reported.

#### 10.8.5.5 How to Choose Right Outcome Measures

As an example, there is a quick-reference book that summarizes and evaluates more than 150 outcome measures for each joint of the extremities [39]. As Suk et al.’s handbook does not include the outcomes for spinal disorders, we have to have proper outcome measures for those, including neck and back problems [40, 41] (Table 4).

The necessary information for selecting outcome measures is as follows:

1. Goal of measurement
2. Nature of measurement: questionnaire, performance rating, physical properties
3. Specific population for which the instrument was developed
4. Format of measurement: number of items, response options, minimum and maximum score
5. Issues related to feasibility: time needed to perform the measurement, required equipment, and training

For clinical researchers, outcome measurements are essential for the advancement of their studies. To assess the overall quality of an outcome measure, three major elements should be considered: content of construct, psychometric evaluation, and clinical utility.

The conventional methods of studying the dimensional structure of measures are principal component analysis and factor analysis; Cronbach’s  $\alpha$  is calculated to determine the internal consistency of the dimensions. Outcome measures usually consist of one or more domains that reflect the concept of supposed construct.

**TABLE 4 Widely Used or Highly Qualified Outcome Measures for Musculoskeletal Disorders**

---

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Low back pain                             <ul style="list-style-type: none"> <li>Oswestry low back pain disability questionnaire [31]</li> <li>Roland–Morris disability questionnaire (RDQ) [32, 33]</li> </ul> </li> <li>• Rheumatoid arthritis                             <ul style="list-style-type: none"> <li>Health assessment questionnaire (HAQ) [34]</li> <li>Arthritis impact measurement scale (AIMS) [35–37]</li> </ul> </li> <li>• Osteoarthritis (hip and knee)                             <ul style="list-style-type: none"> <li>Western Ontario McMaster University (WOMAC) [38]</li> </ul> </li> </ul> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| (In this section of joint functions, the outcome measures that scored higher than 8 are shown)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| • Joint functions [39]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Shoulder                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>Disabilities of the arm, shoulder and hand (DASH)</li> <li>Flexilevel scale of shoulder function (FLEX-SF)</li> <li>Oxford shoulder score</li> <li>Shoulder instability questionnaire</li> <li>Shoulder pain and disability index (SPADI)</li> <li>Simple shoulder test (SST)</li> <li>Upper extremity function scale</li> </ul>                                                                                                                                |
| Elbow                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <ul style="list-style-type: none"> <li>Elbow functional assessment scale (EFA)</li> <li>Liverpool elbow score</li> <li>Upper extremity function scale</li> </ul>                                                                                                                                                                                                                                                                                                                                       |
| Wrist/hand                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>Boston questionnaire (also known as Brigham and Women's carpal tunnel questionnaire)</li> <li>Cochin rheumatoid hand disability scale</li> <li>Patients rated wrist evaluation (PRWE)</li> <li>Sequential occupational dexterity assessment (SODA)</li> <li>Upper extremity function scale</li> </ul>                                                                                                                                                           |
| Pelvis                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | —                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Hip                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | <ul style="list-style-type: none"> <li>AAOS hip and knee score</li> <li>Functional recovery score</li> <li>Harris hip score</li> <li>Oxford hip score</li> <li>Western Ontario and McMaster Universities OA index (WOMAC)</li> </ul>                                                                                                                                                                                                                                                                   |
| Knee                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | <ul style="list-style-type: none"> <li>AAOS hip and knee score</li> <li>Activity rating scale</li> <li>Fulkerson–Shea patellofemoral joint evaluation score</li> <li>Knee outcome survey activities of daily living scale</li> <li>Knee injury and osteoarthritis outcome score (KOOS)</li> <li>Kujala patellofemoral score (also know as the AKPS—anterior knee pain scale)</li> <li>Oxford 12-item knee questionnaire</li> <li>Western Ontario and McMaster Universities OA index (WOMAC)</li> </ul> |
| Ankle                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Foot health status questionnaire                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Calcaneus                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | —                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

---

It is very important to review, integrate, and consolidate the information with measuring instruments for cross-cultural use [6].

#### 10.8.5.6 Questionnaires

The clinical utility of an outcome measure is divided into two parts: patient friendliness (acceptability) and clinician friendliness (feasibility) [42].

If we want a questionnaire to be user friendly, for both patient and clinician, and to improve clinical utility, the questionnaire has to be refined repeatedly. The content of the questionnaire must be limited to the minimum amount of information needed and be designed to be filled out easily. Even in the case of multiple choice or fill-in-the-numeral forms, there are some problems involved in interpreting the data.

Recently, the content of questionnaires that has only been determined by so-called expert staff is no longer regarded as sufficient for assessing content validity. We have to include the opinions of patients during the development process of the questionnaires in order to check content validity and face validity (these are often used interchangeably).

## 10.8.6 DATA ANALYSIS AND INTERPRETATION

### 10.8.6.1 Sample Size

The sample size of a trial is calculated based on the estimate of primary outcome [43]. I would like to show an example of how to calculate a sample size from the difference between the two means.

We have to recognize a way of denying null hypothesis, which is always associated with false-positive and false-negative rates. The former (false positive) is known as the type I error, that is, significance level  $\alpha$ . The latter (false negative) is the type II error  $\beta$ , and  $1 - \beta$  is the power. The sample size is given here in the condition of two groups' comparison, and randomization in equal group numbers.

$$\text{Sample size } N = \frac{4(Z\alpha + Z\beta)^2}{\Delta^2}$$

where  $Z\alpha$ ,  $Z\beta$  are standardized normal deviates for given  $\alpha$ ,  $\beta$ , and  $\Delta$  is standardized effect size.

$$Z\alpha = Z_{1-\alpha/2}; \alpha = 0.05 \text{ in two-sided, then } Z_{1-\alpha/2} = Z_{0.975} = 1.96$$

$$Z\beta = Z_{1-\beta}; \text{ a power of } 1 - \beta = 0.9, \text{ then } Z_{1-\beta} = Z_{0.9} = 1.645$$

$$\Delta = \mu A - \mu B / \sigma; (\mu A - \mu B) \text{ is difference between the two means, and } \sigma \text{ is standard deviation of the endpoint}$$

If you set  $\Delta$  less than 0.1, approximately 0.5, or more than 1, it means a small, moderate, or large standardized effect, respectively.

Sample size is noted in the majority of articles on clinical trials. Sample size depends on whether we want much more power to test the difference among the groups or we want to estimate the precision of the confidence interval. If we have an available working hypothesis, we can calculate the necessary sample size according to other similar studies. The result of the calculations only indicates the number required for the data analysis. Usually, we have to add a few more cases, expecting that there will be some dropouts.

### 10.8.6.2 Intention to Treat

Intention to treat (ITT) is a conceptual principle and is not defined as a specific technical procedure. One widely used procedure is to review the data in detail when the treatment and follow-up have been completed and all the patient information has been collected, that is, at the time that all the data has been frozen. Once allocation has been done, it is the ITT principle to analyze as if the original allocation was continued, even after the intervention itself was changed. The “last observation carry-forward” method is able to cover the deficit data when the trial has been in progress for some time (more than two check points).

However, even such principles cannot be maintained at analysis in the following cases:

1. Criteria unfit for inclusion, even after allocation
2. Complete loss to follow-up

### 10.8.6.3 Multiple Comparisons

We must cut down the working hypotheses to one main one. In a usual trial comparing two groups, a statistical test provides a figure of  $P$  value as the boundary line for the verdict of hypothesis. If the calculated  $P$  value for the primary endpoint is less than the predefined line (usually  $P < 0.05$ ), we will deny the null hypothesis of no difference between the groups. However, if we use more than one endpoint, there are more than two comparisons arising (sometimes this situation becomes much more complex with multiple comparisons). In such a situation, a false-positive rate is no longer  $\alpha \times 100\%$  and increases according to the number of multiple comparisons.

### 10.8.6.4 Post Hoc Analysis and Subgroup Analysis

It is the nature of things that the therapy-responder group has a better prognosis and the nonresponder group has a worse result. That is a typical post hoc analysis. Grouping allocation should be provided at the beginning and after disclosing the result, retrograde grouping is not permitted.

At times, clinicians like to know whether certain types of patients show more obvious benefits from interventions than others, knowing that a sample group of interest is a mixture of various types of patients. This is the subgroup analysis, for example, mild or severe cases, young or old patients. Some subgroups could easily reveal positive effects by chance, even if the overall trial results are negative. A common mistake is to mix the  $P$  value of more than 0.05 in all data-combined analysis and less than 0.05 in one or more subgroup(s). This may lead to the false conclusion that a certain subgroup has a favored result by chance, in spite of the fact that the true result is negative. When you plan a subgroup analysis at the design stage, advance adjustment in sample size should be considered.

### 10.8.6.5 Data Synthesis and Combined Result

If there are several numbers of articles available for review, systematic review, is regarded as an “infrastructure” of the information system supporting EBM practice.

In regards to treatment, the pathway of systematic review is very logical; so the results provide the most accurate and authoritative guidelines for therapy. To avoid potential problems in interventional studies (clinical trials), only RCTs should be included, with complete follow-up information, blinded outcome assessment, and analysis based on intention to treat. This is applicable to an epidemiological research in observational studies, but it is still unable to solve all the problems of biases.

Meta-analysis is defined as a statistical synthesis of the numerical results of some trials (quantitative systematic review) that all examined the same question. It is a type of research that attempts to reanalyze and combine the results already reported, mainly as RCTs.

The assessment of methodological quality for meta-analysis has been recognized to be very important because even the conduct of an RCT is no guarantee of unbiased outcomes from such a study. Even meta-analysis and systematic reviews should be scrutinized carefully and analyses based on small studies should especially be treated with caution.

#### **10.8.6.6 Side Effects**

We have to pay sufficient attention to avoid adverse effects. In the case of surgery, examples of adverse side effects are nerve injury, infection, bleeding, secondary osteoarthritis, and even death. Some side effects occur even in conservative therapy.

We cannot say that treatment side effects are out of the question, even if there is no obvious difference in the frequency and degree of adverse effects between the treatment group and the control group in a certain RCT. An RCT is not an appropriate design to assess side effects because an RCT basically suppresses the number of patients to a minimum and the incidence of side effects is usually low. The role of the trial steering committee and the data-monitoring committee should be established and the range of their responsibilities defined. These committees have to take necessary actions such as early stopping, considering the results of the interim analysis, or the frequency and content of side effects. I realize that relatively small size trials do not often have the side effects and other problems that would stop a trial, but large multicentered trials have the latent possibility of such troubles.

#### **10.8.6.7 Reporting Clinical Trials**

After completing a clinical trial, an important step is to report the results, whether positive, negative, or equivocal. Selective reporting, in which only positive studies are published, often distorts a true situation; this practice is called “publication bias.” When meta-analysis is performed as a clinical trial overview, it is definitely important to include all relevant unpublished trials to gain overall results.

The International Committee of Medical Journal Editors, which consists of several leading medical core journals, decided that clinical trials should be registered in advance [44, 45]. If clinical trials are not registered in advance, journals might refuse to publish a report of a trial. This is an important step in avoiding publication bias and other inappropriate analyses.

Article writing is regarded as a form of communication or “dialog,” even in the case of a scientific journal; that is the object of our communication. The writer of an article must be aware of the target audience. Consolidated Standards of



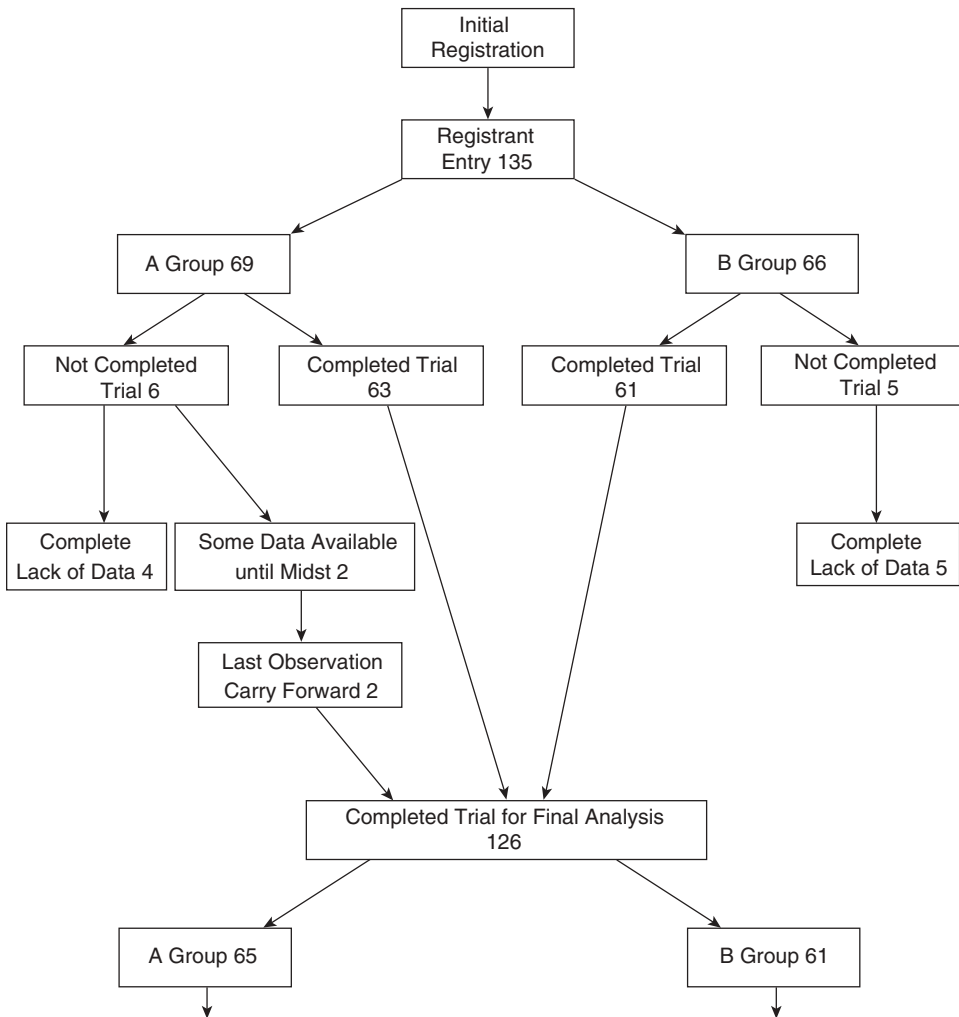


**FIGURE 1** Example flowchart of trial according to CONSORT statement.

Reporting Trials (CONSORT) is one of the examples of such a checklist or flowchart to be used, as well as the instructions for improving the quality of reporting randomized controlled trials when a paper is to be submitted [46, 47] (Fig. 1). Also suggested through discussions at a conference was a guideline for reporting systematic reviews or meta-analysis referred to as the Quality of Reporting of Meta-analysis (QUOROM). The discussions resulted in the creation of the QUOROM Statement, which consists of a checklist and flowchart [48].

### 10.8.7 RCT AND FUTURE DIRECTIONS

At the close of this section, I would like to talk about how to apply the results of an RCT to daily practice, as well as mention the potential problems when reaching the limit of a statistical outlook.

### 10.8.7.1 Impact from RCT to Daily Practice

Medical practice is a life-long, continuous process of self-learning, and, as clinicians, we are required to keep up-to-date on various medical developments. Evidence-based medicine is our way of integrating individual clinical expertise with the best available evidence to assist us in making decisions about each patient's care. Reported results of RCTs make it possible to cover the majority of our activities systematically, from daily practice for patient care to writing and reading scientific papers.

Evidence-based medicine is regarded as a new paradigm in medical practice, equal to the Human Genome Project. However, there is still much confusion and misunderstanding about the concept and content of EBM; it is often limited to searching the literature and reading articles, and to serving cost cutters and suppressing clinical freedom. The use of clinical guidelines or the managed care system is seen as intimidating the discretion of doctors in clinical practice. EBM is also seen as the fashionable trend of a group of medical academics armed with epidemiological and statistical jargon, but there are more serious basic problems behind such a statistical paradigm.

### 10.8.7.2 Internal Validity and External Validity

In clinical practice we are always surrounded by a condition called the "gray zone" [49]. Of course, there are many factors involved in this obscurity. I would like to focus closer attention on a statistical outlook of the world around us, that is, internal validity and external validity.

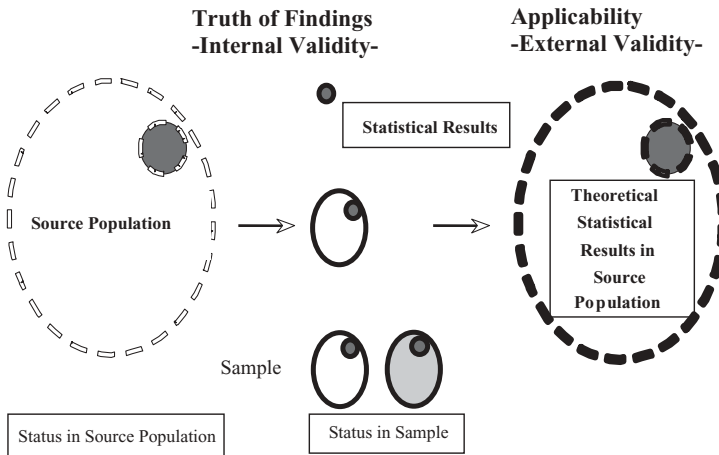
These are concepts relating to the extrapolation of the relationship between sample and source populations. In an experimental trial like an RCT, internal validity is high because the difference between the experimental group and the control group is thoroughly controlled only for the target issue, and by chance. However, external validity is low when the results from the sample with several specific premises are brought into a general finding in the mother population. On the contrary, in an observational study, such as a cohort study, there are many confounding factors involved and internal validity is low in surveyed samples. External validity in an observational study is high because a real situation in a resource population reflects just what it is (Fig. 2).

### 10.8.7.3 Quantitative Study and Qualitative Study

Clinical trials cover a wide range of medical fields and use very diverse methods. The data used is basically divided into quantitative, such as laboratory data, and qualitative, based on verbal information collected (mainly used in psychology or nursing science at present). Behind the methods of investigation and analysis in these quantitative and qualitative approaches, there is an obvious difference in concepts [50].

The general concept for study design consists of the following characteristics (Table 5):

1. Aims of study
2. Methods of outcome measurements

**Observational Study**

Cross-sectional Study  
or Cohort Study

Internal Validity: low  
External Validity: high

**Interventional Study**

Randomized Controlled Trial  
Blind & Placebo-controlled

Internal Validity: high  
External Validity: low

**FIGURE 2** Relationship between sample and mother population.

**TABLE 5** Comparison Between Quantitative Research and Qualitative Research

|                                | Quantitative Research | Qualitative Research    |
|--------------------------------|-----------------------|-------------------------|
| Study design                   | Intervention          | Observation             |
| Methods                        | Survey, experiment    | Interview               |
| Question setting               | Enumeration           | Classification          |
| Reasoning                      | Deductive             | Inductive               |
| Sampling method                | Statistical model     | Theoretical model       |
| Closeness to the truth         | Internal validity     | Credibility             |
| Applicability/generalizability | External validity     | Concept transferability |
| Consistency                    | Reliability           | Dependability           |
| Neutrality                     | Objectivity           | "Confirmability"        |

3. Selection of subjects

4. Analytic methods and provided results

In a conventional quantitative study, the main strategy is to make up a working hypothesis and test it. An RCT is a typical example of this type of strategy and is inseparable from biostatistics for data analysis. To generalize the results from samples, such as the rejection of null hypothesis in the majority of confirmatory studies, requires estimating and testing as a basic procedure. The advance of computer technology makes possible the use of multivariate analysis and the ability to search and verify potential factorial structure, even without a working hypothesis.

This technique makes it easy to analyze latent factorial structure, which occurs in some observational studies, from a calculated relationship of how the weight distribution among explanatory variables accounts for the tendency of objective variables. Therefore, this exploratory study makes it possible to investigate and verify risk factors or to produce a model of a certain phenomenon. The distinction between the term “quantitative” study and “qualitative” study will be more obscure once the quantitative studies move toward an exploratory nature from a confirmatory one for testing hypothesis.

On the contrary, a qualitative study does not always follow the deductive logic of applying the result from a sample to the mother population. Researchers try to categorize the verbal information received from subjects to establish a coding system for the data and generalize a new concept in order to build a new theory or model. It is good enough to select the subjects that fit the aim of a study as one likes. It would be possible to produce an epochal qualitative study that could produce a unique, clinically significant theory that would never have been yielded by a confirmatory study using hypotheses testing. A qualitative study is able to make progress in building up a new and significant theoretical system [51]. However, good or bad, theory is totally dependent upon the capability of the researcher.

#### 10.8.7.4 Further Suggestions

If a certain intervention shows obviously successful results for an otherwise lethal condition, we do not require RCTs and do not wait for more studies to be conducted. We can get the evidence from the basic sciences, properly designed follow-up studies, or proper cross-sectional trials.

We need to be careful not to become so skeptical as to assume that if a study is not an RCT it would have no value or use. We also need to recognize the limitations of a statistical approach through RCTs. At each step, we continue our efforts to track down the best evidence to answer our clinical questions and promote cost-effective prevention and treatment for musculoskeletal disorders.

## REFERENCES

1. Akai, M. (2002), Evidence-based medicine for orthopedic practice, *J. Orthop. Sci.*, 7, 731–742.
2. Akai, M., Doi, T., Fujino, K., et al. (2005), An outcome measure for Japanese people with knee osteoarthritis, *J. Rheumatol.*, 32, 1524–1532.
3. BJD on line (Bone and Joint Decade’s musculoskeletal Portal); <http://www.boneandjoint-decade.org/>.
4. McDowell, I., and Newell, C. (1996), *The Theoretical and Technical Foundations of Health Measurement. Measuring Health; A Guide to Rating Scales and Questionnaires*, 2nd ed., Oxford University Press, New York, pp. 10–46.
5. Guyatt, G. H., Kirshner, B., and Jaeschke, R. (1992), Measuring health status; What are the necessary measurement properties? *J. Clin. Epidemiol.*, 45, 1341–1345.
6. Dekker, J., Dallmeijer, A. J., and Lankhorst, G. J. (2005), Clinimetrics in rehabilitation medicine; current issues in developing and applying measurement instruments, *J. Rehabil. Med.*, 37, 193–201.

7. Wolfson, A. M., Doctor, J. N., and Burns, S. P. (2000), Clinician judgments of functional outcomes; how bias and perceived accuracy affect rating, *Arch. Phys. Med. Rehabil.*, 81, 1567–1574.
8. Guyatt, G. H., Feeny, D. H., and Patrick, D.L. (1993), Measuring health-related quality of life, *Ann. Int. Med.*, 118, 622–629.
9. Clancy, C. M., and Eisenberg, J. M. (1998), Outcomes research; measuring the end results of health care, *Science*, 282, 245–246.
10. World Health Organization (2001), *International Classification of Functioning, Disability and Health*, WHO, Geneva.
11. Stucki, G., and Grimby, G. (2004), Applying the ICF in medicine, *J. Rehabil. Med.*, 36 (suppl 44), 5–7.
12. Irrgang, J. J., and Anderson, A. F. (2002), Development and validation of health-related quality of life measures for the knee, *Clin. Orthop.*, 402, 95–109.
13. Zelen, M. (1979), A new design for randomized clinical trials, *N. Engl. J. Med.*, 300, 1242–1245.
14. Zelen, M. (1990), A new design for randomized consent designs for clinical trials; an update, *Stat. Med.*, 9, 645–656.
15. ter Riet, G., Kleijnen, J., and Knipschild, P. (1990), Acupuncture and chronic pain; A criteria-based meta-analysis, *J. Clin. Epidemiol.*, 43, 1191–1199.
16. Koes, B. W., Bouter, L. M., Beckerman, H., et al. (1991), Physiotherapy exercises and back pain; A blinded review, *BMJ*, 302, 1572–1576.
17. Koes, B. W., Assendelft, W. J. J., van der Heijden, G. J. M. G., et al. (1991), Spinal manipulation and mobilisation for back and neck pain; A blinded review, *BMJ*, 303, 1298–1303.
18. Koes, B. W., Bouter, L. M., and van der Heijden, G. J. M. G. (1995), Methodological quality of randomized clinical trials on treatment efficacy in low back pain, *Spine*, 20, 228–235.
19. Akai, M., and Hayashi, K. (2002), Effect of electrical stimulation on musculoskeletal systems; A meta-analysis of controlled clinical trials, *Bioelectromagnetics*, 23, 132–143.
20. Akai, M., Kawashima, N., Kimura, T., et al. (2002), Electrical stimulation as an adjunct to spinal fusions; A meta-analysis of clinical controlled trials, *Bioelectromagnetics*, 23, 496–504.
21. Machin, D., Day, S., and Green, S. Eds. (2004), *Textbook of Clinical Trials*, Wiley, West Sussex, UK.
22. Bowling, A. (2001), *Measuring Disease; A Review of Disease-Specific Quality of Life Measurement Scales*, 2nd ed., Open University Press, Buckingham, UK.
23. Duley, L., and Farrell, B. Eds. (2002), *Clinical Trials*, BMJ Books, London.
24. McDowell, I., and Newell, C. (1996), *The Theoretical and Technical Foundations of Health Measurement. Measuring Health; A Guide to Rating Scales and Questionnaires*, 2nd ed., Oxford University Press, New York.
25. Riegelman, R. K. (2005), *Studying a Study and Testing a Test; How to Read the Medical Evidence*, 5th ed., Lippincott Williams and Wilkins, Philadelphia.
26. Suk, M., Hanson, B. P., Norvell, D. C., et al. (2005), *AO Handbook; Musculoskeletal Outcomes Measures and Instruments; 155 Instruments Evaluated and Assessed*, AO Publishers, Thieme.
27. Resource Center for Randomized Trials; <http://www.rcrt.ox.ac.uk/>.
28. An introduction to clinical trials; <http://www.clinicaltrials.gov/>.
29. Farrell, B., and Spark, P. (2002), Building resources for randomized trials, in Duley, L., and Farrell, B., Eds., *Clinical Trials*, BMJ Books, London, p. 84.

30. Fayers, P. M., and Machin, D. (1995), Sample size; how many patients are necessary? (editorial review), *Br. J. Cancer*, 72, 1–9.
31. Fairbank, J. C. T., Couper, J., Davies, J. B., et al. (1980), The Oswestry low back pain disability questionnaire, *Physiotherapy*, 66, 271–273.
32. Roland, M., and Morris, A. (1983), A study of the natural history of back pain. Part II. Development of a reliable and sensitive measure of disability in low-back pain, *Spine*, 8, 141–144.
33. Roland, M., and Fairbank, J. (2000), The Roland-Morris disability questionnaire and the Oswestry disability questionnaire, *Spine*, 25, 3115–3124.
34. Fries, J. F., Spitz, P., Kraines, R. G., et al. (1980), Measurement of patient outcome in arthritis, *Arthritis Rheum.*, 23, 137–145.
35. Meenan, R. F., Gertman, P. M., and Mason, J. H. (1980), Measuring health status in arthritis; The arthritis impact measurement scales, *Arthritis Rheum.*, 23, 146–152.
36. Meenan, R. F., Gertman, P. M., Mason, J. H., et al. (1982), The arthritis impact measurement scales; Further investigations of a health status measure, *Arthritis Rheum.*, 25, 1048–1053.
37. Meenan, R. F., Mason, J. H., Anderson, J. J., et al. (1992), AIMS2; The content and properties of a revised and expanded arthritis impact measurement scales health status questionnaire, *Arthritis Rheum.*, 35, 1–10.
38. Bellamy, N., Buchanan, W. W., Goldsmith, C. H., et al. (1988), Validation study of WOMAC; A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee, *J. Rheumatol.* 15, 1833–1840.
39. Suk, M., Hanson, B. P., Norvell, D. C., et al. (2005), *AO Handbook; Musculoskeletal Outcomes Measures and Instruments; 155 Instruments Evaluated and Assessed*, AO Publishers, Thieme, pp. 37–40, 47–405.
40. Bombardier, C. (2000), Outcome assessments in the evaluation of treatment of spinal disorders; summary and general recommendations, *Spine*, 25, 3100–3103.
41. Stratford, P. W., Binkley, J., Solomon, P., et al. (1994), Assessing change over time in patients with low back pain, *Phys. Ther.*, 74, 528–533.
42. Suk, M., Hanson, B. P., Norvell, D. C., et al. (2005), *AO Handbook; Musculoskeletal Outcomes Measures and Instruments; 155 Instruments Evaluated and Assessed*, AO Publishers, Thieme, pp. 20–24.
43. Machin, D., Campbell, M. J., Fayers, P. M., et al. (1997), *Sample Size Tables for Clinical Studies*, 2nd ed., Blackwell Science, Oxford, pp. 40–78.
44. DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., et al. (2004), Clinical trial registration; a statement from the international committee of medical journal editors. *JAMA*, 292, 1363–1364.
45. DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., et al. (2005), Is this clinical trial fully registered? A statement from the international committee of medical journal editors, *JAMA*, 293, 2927–2929.
46. Begg, C., Cho, M., Eastwood, S., et al. (1996), Improving the quality of reporting of randomized controlled trials; The CONSORT statement, *JAMA*, 276, 637–639.
47. Moher, D., Schulz, K. F., Altman, D. G., for the CONSORT group (2001), The CONSORT statement; revised recommendations for improving the quality of reports of parallel-group randomized trials, *Lancet*, 357, 191–194.
48. Moher, D., Cook, D. J., Eastwood, S., et al. (1999), Improving the quality of reporting of meta-analyses of randomised controlled trials; The QUOROM statement, *Lancet*, 354, 1896–1900.

49. Naylor, C. D. (1995), Grey zones of clinical practice; Some limits to evidence-based medicine, *Lancet*, 345, 840–842.
50. Greenhalgh, T., and Taylor, R. (1997), Papers that go beyond numbers (qualitative research), *BMJ*, 315, 740–743.
51. Öhman, A. (2005), Qualitative methodology for rehabilitation research, *J. Rehabil. Med.*, 37, 273–280.





# 10.9

---

## Oncology

**Matjaz Zwitter**

*Institute of Oncology, Ljubljana, Slovenia, and Department of Medical Ethics,  
Medical School, University of Maribor, Slovenia*

### Contents

|          |                                                                                    |     |
|----------|------------------------------------------------------------------------------------|-----|
| 10.9.1   | Introduction                                                                       | 588 |
| 10.9.2   | General Comments on Ethical Issues                                                 | 589 |
| 10.9.2.1 | Ethical Demand for Research                                                        | 589 |
| 10.9.2.2 | Ethical Obligation of Patients                                                     | 590 |
| 10.9.2.3 | ALARA Principle and Positive Ethical Balance                                       | 590 |
| 10.9.2.4 | Patient with Cancer with Comments on Informed Consent and<br>Real Patient Autonomy | 590 |
| 10.9.3   | Different Medical Problems in Clinical Trials                                      | 591 |
| 10.9.3.1 | Prevention of Cancer                                                               | 591 |
| 10.9.3.2 | Induction (Neoadjuvant) Therapy                                                    | 592 |
| 10.9.3.3 | Adjuvant Therapy                                                                   | 593 |
| 10.9.3.4 | Trials of Concomittant Radiochemotherapy                                           | 593 |
| 10.9.3.5 | Systemic Treatment of Advanced Cancer                                              | 595 |
| 10.9.3.6 | Palliative Treatment                                                               | 596 |
| 10.9.4   | Types of Clinical Trials                                                           | 596 |
| 10.9.4.1 | Phase I Trials                                                                     | 596 |
| 10.9.4.2 | Phase II Trials                                                                    | 597 |
| 10.9.4.3 | Phase III Trials                                                                   | 598 |
| 10.9.5   | Endpoints                                                                          | 598 |
| 10.9.5.1 | Response to Treatment                                                              | 599 |
| 10.9.5.2 | Time to Progression                                                                | 600 |
| 10.9.5.3 | Survival                                                                           | 600 |
| 10.9.5.4 | Quality of Life                                                                    | 601 |
| 10.9.6   | Conclusion                                                                         | 602 |
|          | References                                                                         | 603 |

### 10.9.1 INTRODUCTION

Patients with cancer have their diagnostics and treatment in different departments of virtually every hospital. Still, when it comes to research, most of it would be conducted either in specialized cancer centers or in oncology departments of larger hospitals. We can say that oncology as a distinct branch of medicine indeed developed due to clinical research. Without clinical research, oncology as a discipline would not exist. Without clinical research, our only guidelines would be statements like “I think ...” or “I remember two similar patients ...”.

Many oncologists are sceptical in their attitude toward new developments. This is not unexpected. The nature of the disease itself with its unpredictable course and many false hopes do not support overt optimism at every new idea. Nevertheless, a healthy degree of scepticism should not be understood as pessimism. Looking from the perspective of one or two years, no dramatic progress may be apparent. Still, in the long run, the battle against cancer is a story of hundreds of small victories. Fifty years ago, only a tiny minority of cancer patients with truly limited disease could hope for a successful cure by surgery. For the rest, cancer stood as an invincible stone wall. Systematic clinical research led to erosion of this wall, piece by piece. Hematological malignancies and pediatric tumors, testicular cancer, breast, prostate, gastrointestinal, lung cancer—the prognosis of each of these tumors in terms of cure, prolonged survival, or quality of life has improved significantly. Thousands of clinical trials have contributed to this gradual progress.

Our understanding of how the treatment works and why it is effective in one patient but fails in another apparently similar case lags behind clinical experience. In the past, we used to apply treatments and hope for some 30% chance of remission, at the same time often ignoring the fact that the majority of patients did not benefit from the treatment and experienced only its side effects. In the past, we used to say that cancer was not one disease but about 200 different diseases. We then gave up hope for a single miracle drug against cancer and accepted the fact that each of these diseases demands a different treatment. In the future, we may need to go further in our individualized approach. Relatively common diseases such as breast cancer or non-small-cell lung cancer are now subdivided into separate entities, depending on phenotypic and genetic characteristics of tumor cells. Clinical research in oncology is turning into a truly team venture: pathology, molecular biology, genetics, and immunology provide essential information for individual classification of a tumor. So deep is the change in methodology that we can speak of a new era of clinical research in oncology [1]. Better understanding of tumor biology and of pharmacogenomics will lead to much more individualized assignment of treatment and, consequently, to more predictable outcome.

Final introductory remarks deal with professionalization of clinical research. What used to be an amateur activity several decades ago is now almost entirely in the hands of professionals supported by pharmaceutical companies or (rarely) public funds. Among clinicians, only a tiny minority of senior “opinion leaders” will be asked to participate in the design of clinical trials; for the rest, clinical research has become synonymous with recruiting patients and filling clinical record forms (CRFs). While increased funding from private funds and from for-profit companies led to important new discoveries, professionalization of clinical research has also

its dark side. It typically produces large multicenter trials focusing on common cancers where substantial profits are expected and often ignores rare tumors; yet, taken together, these rare tumors comprise more than half of all patients with cancer. Commercial interests often prevail over honest scientific evaluation of research proposals, and many promising initiatives and ideas remain unexplored. In the phase of designing a new trial, sponsors do not listen to creative young physicians. While we need large multicenter trials, we should also support simple, innovative, small trials not necessarily sponsored by industry. In a world where every new idea is tested and retested, size of a trial is of minor importance. What matters are innovative potential, reliability of research methodology, quality of the data, and fair reporting [2].

### **10.9.2 GENERAL COMMENTS ON ETHICAL ISSUES**

In any branch of medicine, ethics is at the beginning and at the end of every discussion on clinical research. Ethical issues are crucial when planning a clinical trial, throughout its actual performance, and when formulating experience and implications for routine treatment and for further research. Oncology is not an exception to this general rule.

We here offer four general statements on ethics of clinical cancer research. Some specific comments and recommendations will be made also in other parts of the text.

#### **10.9.2.1 Ethical Demand for Research**

Considering the enormous burden imposed by cancer on every patient and on society as a whole, it would be unethical not to search for new effective and safer approaches to treatment. During the past five decades, clinical research led to an improved outlook for virtually every cancer. Systematic clinical research is not only ethically acceptable but indeed an obligation for everybody in the field of oncology.

It is estimated that only between 3 and 10% of cancer patients participate in clinical trials [3], with even lower representation of elderly patients and minorities [4]. Since it is clear that our progress in the field of oncology critically depends upon the proportion of patients enrolled in clinical research, low recruitment of patients into clinical trials is not only a problem for the researchers but indeed for society as a whole. While public understanding and support for medical research generally increases, refusal of an eligible patient to participate in a clinical trial is not uncommon [5]. Time-consuming procedures for actual performance of a clinical trial with physician's time as a diminishing resource, and the impact of greater demands in a climate of decreasing health care resources have a negative effect on accrual [6, 7]. To increase trial participation, there is a critical need for infrastructure to support trials, especially additional support staff and research nurses [8]. Other obstacles to clinical research include legal restrictions, as seen in the European Union Clinical Trials Directive [9] or in the Declaration of Helsinki [10]. These documents should be critically discussed and, hopefully, amended. Please see the excellent discussion on this matter by Harris [11].

### 10.9.2.2 Ethical Obligation of Patients

From the legal standpoint, a patient with cancer is absolutely free to accept or to refuse participation in clinical research. From the ethical standpoint, however, a patient has a moral duty to participate in a clinical trial that has passed an independent scientific and ethical review and that does not present an unproportionate burden or risk. This moral duty is based on the fact that patients of today expect to be treated according to the best current knowledge. This knowledge increases continuously and is derived from the past clinical research. Patients of today, therefore, benefit from the fact that patients of yesterday participated in clinical trials—and have a moral duty to contribute to the medical knowledge that will also benefit them, future generations of patients, and society as a whole [11].

### 10.9.2.3 ALARA Principle and Positive Ethical Balance

In the practice of clinical research, ethical puritanism is rarely achievable. In spite of meticulous plan and execution of a trial, some “ethical costs” are inevitable. Instead of speaking of zero ethical costs, we follow the philosophy of ALARA—*as low as reasonably achievable* [12]. We also speak about a positive balance between the benefits of a trial (as applicable to patients in the trial and to society as a whole) and ethical costs.

Practical possibilities for lowering ethical costs and for a positive balance between benefits and ethical costs in clinical research will be presented in Sections 10.9.4 and 10.9.5.

### 10.9.2.4 Patient with Cancer with Comments on Informed Consent and Real Patient Autonomy

Among patients with cancer, we can rarely speak about full autonomy. Patients are not only medically uneducated but also under serious physical and emotional burden of the disease.

In spite of this well-recognized burden of malignant disease, information offered to cancer patients during recruitment for clinical trials is often very extensive, written in a complicated medical and legal jargon, and with detailed description of all possible and rare complications of the treatment. Such extensive and frightening information does not add to patient’s autonomy. Rather, many patients sign consent forms without an attempt read such an extensive text, leading to a diminished autonomy. According to a survey among physicians recruiting patients for randomized clinical trials, 42% of patients were not fully informed about the trial [13]. In another survey, the majority of patients who recently signed consent for participation in a trial did not recognize the nonstandard treatment nature of the trial (74%), the potential for incremental risk from participation (63%), or the unproven nature of the treatment (70%) [14]. An additional argument against extensive text of informed consent comes from Eastern Cooperative Oncology Group, which randomized patients between the standard consent procedure to an easy-to-read consent statement. A simplified text did not affect patient comprehension and resulted in significantly lower consent anxiety and higher satisfaction [15].

Even from a purely legal standpoint, it should not be too difficult to prove that many patients with cancer sign consent without being fully competent and/or do not have a free choice.

If we accept limited patient's autonomy as a reality, then we have to accept also the fact that many patients do not rely on rational considerations but simply trust their doctor. Thus, while we need patient's consent, the responsibility stays with the physician. Physician's paternalism, whether we agree with it or not, is often a reality. A physician who is also a clinical researcher should be aware of the delicacy of his or her role. He should be aware that his relation to the patient does not derive from a contract, but rather from a patient's trust. As he cannot get rid of a certain degree of paternalism, he has to strive constantly to push his role of a physician ahead of his role of a researcher.

In concluding this section, we recommend that the information for patient is adjusted to his or her real condition: brief, understandable, with realistic presentation of the problem and also offering some hope. While patients have the right to full information, they should not be excluded from research in case they opt for a simplified procedure. Rather than continuing with what is increasingly obvious to be a fiction—that fully informed consent has been obtained—we should perhaps move to a consent protocol that is basically simple consent in the presence of full information. A summary paragraph to each information sheet would allow the patients to focus on that alone if they choose to do so [16].

### 10.9.3 DIFFERENT MEDICAL PROBLEMS IN CLINICAL TRIALS

Understanding the medical problem is essential for choosing appropriate methodology for a clinical trial. We will briefly review some specific medical problems in oncology and offer comments on the research methodology.

#### 10.9.3.1 Prevention of Cancer

Various food supplements, hormones, hormone antagonists, antiviral vaccines, other substances, and even surgical procedures have been studied with the aim of cancer prevention [17–21]. While a critical coverage of this area would be clearly out of the scope of our discussion, we here offer some specific recommendations.

- Before initiating a trial, its scientific background should be thoroughly discussed. A clear distinction should be made between the two possibilities: to delay or to prevent occurrence of a particular cancer. Cancerogenesis is a long process. From the experience on secondary cancers induced by radiation or anticancer drugs, we know that there is an interval of at least 5 years between the carcinogenic mutation and clinical detection of cancer. On short terms of a few years, we may expect only a delay of occurrence of a cancer that was already in its subclinical form. If we speak about real prevention, we cannot expect a detectable effect on clinical incidence of a particular cancer in less than 8–10 years.
- The population of persons at an increased risk for a particular cancer should be clearly defined. Even so, we are dealing with healthy persons and not with

patients. The intervention studied should impose a negligible risk and minimal discomfort to participants.

- The control groups are either “pure” controls or are given placebo. Compliance is often poor: Subjects in the intervention group may not follow the procedure, while subjects in the control group may get the substance (such as a food supplement) from other sources. Besides, it is unrealistic to expect good compliance in a trial that continues over many years. Suboptimal compliance leads to dilution of the results and demands a greater number of subjects involved.
- To prove that an intervention lowers the incidence of a relatively rare event, many thousands of individuals have to be followed over a long period.
- Trials of cancer prevention are commonly linked to a program for early detection of a particular cancer. Impact of the intervention upon the incidence of a particular cancer is the obvious objective of such a trial. Still, the real value of an intervention among healthy individuals would be reduction in mortality from any cause, coupled to uncompromised quality of life. As an example, antiestrogens given to healthy women at an increased risk for breast cancer led to reduction in the incidence of breast cancer; yet, most of the prevented tumors were highly curable hormone-dependent cancers. The incidence of the more malignant hormone-resistant cancers remained almost the same, and there was no reduction in mortality. Furthermore, an increase in endometrial cancer and more frequent thrombo-embolic events were observed; the impact of antiestrogens upon subtle differences in the quality of life of healthy women remains unknown. Therefore, incidence of a particular cancer should not be the principal objective of a trial and should be regarded as a temporary surrogate endpoint.

### 10.9.3.2 Induction (Neoadjuvant) Therapy

Induction systemic therapy prior to definite local treatment with surgery and/or radiotherapy is a very promising approach that is increasingly used in apparently localized tumors [22–28]. Compared to adjuvant (postoperative) systemic treatment, induction therapy has several advantages: Treatment of micrometastatic disease begins immediately; the tumor may shrink and become operable in a higher proportion of patients; clinical, radiologic, and pathologic evaluation of response to systemic therapy will be important for further treatment strategy. The potential dangers of induction therapy in operable disease are resistance to chemotherapy, in which case a tumor may actually grow during induction treatment and become inoperable, and the risk of an increase in the incidence of perioperative complications. Prior to randomized trials, solid data from phase II trials should provide a reliable estimate of these risks.

It is essential that induction therapy and local treatment with surgery and/or radiotherapy are part of the same treatment protocol with precisely defined timing and as short a gap as possible between systemic and local treatment. Tumors that shrink during induction therapy will often regrow rapidly. If the interval between induction therapy and local treatment is too long, accelerated repopulation of the tumor may quickly annihilate all the benefits of combined treatment [29].



### 10.9.3.3 Adjuvant Therapy

Patients at an increased risk for relapse after local treatment are often given adjuvant chemotherapy, aimed at reducing this risk. Such a strategy has been widely accepted in breast cancer [30], lung cancer [23], and several gastrointestinal cancers [22]. Many new trials are in progress to improve the ratio between efficacy and side effects for the existing indications and to study this approach for new indications.

A substantial proportion of patients offered adjuvant treatment might be cured of cancer with local treatment alone. As an example, a recent trial among 1146 patients with early breast cancer has shown a 65% 10-year disease-free survival after surgery and adjuvant chemotherapy, compared to 60% after surgery alone ( $p < 0.01$ ) [31]. This experience has been presented as evidence for the indication of adjuvant therapy. Yet, these figures can be interpreted also from a different angle. We can say that 60% of patients were cured by surgery alone; that 35% had a recurrence in spite of adjuvant therapy; and that only 5% actually benefited from adjuvant treatment. It is clear, then, that the benefit of a prolonged disease-free interval and of improved survival should be weighted against side effects of the adjuvant therapy, and it is therefore essential that studies of adjuvant therapy include assessment of quality of life.

Planning a trial of adjuvant therapy has several elements in common with trials of cancer prevention. As there is no possibility for individual assessment of response, we have to enroll a large population of healthy individuals or patients after successful local treatment, and we usually measure the success with time to event (diagnosis of cancer in the case of prevention trials, time to progression in adjuvant treatment). In both instances, we should understand that time to event is only a surrogate endpoint: It is only an improvement in overall survival that validates the results. While some cancers or some recurrences of cancer that we wish to prevent are still curable, it is also important that the prognosis of what we wish to prevent is generally dismal. If this were not the case, it would make no sense to engage in prevention of a highly curable situation.

### 10.9.3.4 Trials of Concomitant Radiochemotherapy

In this section, we are dealing with two distinct categories of patients. In the first are patients with truly localized cancer such as early laryngeal or cervical cancer. A high proportion of these patients are cured either by surgery, by surgery and postoperative radiotherapy, by radiotherapy alone, or by combined irradiation and chemotherapy. When the prospects with each particular treatment are good, the decision regarding the treatment depends on patient's preferences, on side effects of each of these treatments, and on the expertise and influence of particular oncologists. The second category are patients with inoperable locally advanced disease without overt distant metastases for whom surgery is not an option.

During the past decade, there was an explosion of trials of concomitant radiochemotherapy. When nonsurgical treatment with curative intent is applied, head and neck cancers, lung, esophageal, gastric, colorectal, pancreatic, cervical and bladder carcinomas, many brain tumors, sarcomas, and most pediatric tumors are now all treated with concomitant radiotherapy and chemotherapy [32–41].

While a certain degree of uncertainty regarding the true advantage of combined-modality treatment persists, we can be confident that the trend of combined-

modality treatment will continue. Irradiation, probably to higher total doses and possibly with different schedules of fractionation will be combined with new drugs, all with the aim of achieving a better local and systemic control of the disease at the expense of acceptable toxicity. It is more than clear that both the effect against the tumor and the pattern of acute and late toxicity of combined-modality treatment are unpredictable. "Wishful thinking" should be avoided, and well-designed phase I and II clinical trials are needed [42]. In this section, we will present a selection of recent accomplishments and of basic knowledge important for the design of combined-modality trials.

Radiotherapy of today is virtually incomparable to what this treatment used to be just two decades ago. Modern technology has allowed very precise targeting of the tumor and of critical healthy tissues. Since the coverage of the tumor by radiation is much more precise and the margin of healthy tissues is getting smaller, we see fewer side effects along with higher efficacy of radiotherapy in eradication of a tumor. Instances of "individual hypersensitivity to radiation" (which in fact often represented cases with grossly imprecise dosimetry) are now rare and treatment with radiotherapy can proceed to higher total doses with predictable and acceptable toxicity.

The technological advances of radiotherapy are important not only for trials of radiation alone but also for the design of combined-modality trials. It is extremely important to understand that radiotherapy with the total dose and the fractionation that used to be standard radical radiotherapy one or two decades ago is now a sub-optimal treatment. As an example, 60 Gy in 30 fractions over 6 weeks used to be a standard for radiotherapy of lung cancer in the 1980s and early 1990s; today, the optimal dose is from 64 to 74 Gy, and in some trials even higher [43].

A reader familiar with radiobiology may skip this paragraph. For the rest who are just interested in combining radiation with drugs, it is essential to understand the importance of fractionation upon tumor control and upon the pattern of acute and late side effects of irradiation. Regarding *tumor control*, the classical and still valid advice is: "Give as many fractions as possible in as short a total time as possible." Hyperfractionated irradiation and hyperfractionated accelerated irradiation with more than one small fraction per day were tested and proven to be superior to standard radiotherapy for head and neck and lung cancers [44, 45]. Hyperfractionation produced more *acute side effects*, which were usually transient and rarely extended into chronic toxicity. On the other side of the spectrum of fractionation, we have hypofractionation with fewer large fractions of radiotherapy. While this approach is useful for palliation where the total dose is low, hypofractionation cannot be recommended for treatment with curative intent. Severe *late and permanent toxicity* was observed in trials of hypofractionated radiotherapy.

It is clear that even at very high doses, radiotherapy alone cannot cure many tumors. The concept of synergistic effect of radiotherapy and chemotherapy has been tested in vitro and confirmed in dozens of clinical studies. It was shown that most cytotoxic drugs enhance the effects of radiotherapy [46]. The broad term "radiosensitization" encompasses several phenomena: additive effect (most cytotoxic drugs) [47], true radiosensitization (gemcitabine, cisplatin when applied less than 30 minutes prior to irradiation) [48, 49], and selective sensitization of hypoxic tumor cells, which are generally resistant to irradiation (mitomycin C) [50].

Drugs that enhance the effect of irradiation on tumor also contribute toward greater side effects upon normal tissues. In radiobiological terms, adding drugs to irradiation means that each fraction of irradiation is “worth more”. At the same total dose, not only can we expect more control of the tumor but also more pronounced acute and late side effects. The simplest, but possibly not the optimal solution, is lowering of the total dose of irradiation. From the logistical perspective, adhering to the standard fractionation is the easiest way; proper attention to late side effects is essential. Hyperfractionation is another approach [51]. In addition to possible better tumor control, its clear advantage is that the maximal tolerated dose would be defined by acute and transient, rather than late and permanent, toxicity.

A detailed discussion of each particular clinical situation would be clearly beyond the scope of this chapter. In concluding this section, we wish to emphasize that the design of combined-modality treatment is a true team venture. Only optimal collaboration of radiation oncologists and medical oncologists will result in a trial that will improve the prospects of this category of patients.

### 10.9.3.5 Systemic Treatment of Advanced Cancer

Systemic treatment of advanced cancer is a rapidly expanding and exciting area of medical research. Experience on systemic treatment on advanced cancer is indeed the rostrum of cancer research. Virtually all anticancer drugs and their combinations are first tested in patients with advanced cancer.

Clinical trials of advanced cancer differ from the other previously mentioned topics in three fundamental aspects. As already mentioned, the first difference is their innovative potential. The second one is willingness of patients to participate and to agree to virtually every proposal that might prolong, improve, or save life. When suffering from advanced cancer, many patients do not wish to exert their full autonomy. A researcher in charge of the trial and a physician in charge of recruitment and care for patients in a trial should be aware of limited patient’s autonomy and, consequently, of physician’s great responsibility.

The third difference is in the choice of endpoints. For all other previously mentioned areas, survival is (or at least should be) the main endpoint. In clinical trials of advanced cancer, we have a choice of several endpoints. While survival is important, it is not always the main endpoint: With incurable disease, other factors such as quality of life come into play. A detailed discussion on endpoints follows in the next section.

Since thousands of clinical trials on treatment of advanced cancer are reported each year, an attempt to present a list of references would be clearly inappropriate. What needs to be mentioned, however, is a recent trend for biological or targeted treatments tailored to specific phenotypic or genotypic characteristics of a tumor. In common and in rare cancers, molecules that exert very specific inhibition of proliferation of tumor cells on a variety of levels from the cell membrane to the nucleus have found rapid application in clinical trials and in regular medical practice [52–57]. Clinical research with these drugs often requires additional information from pathologists; additional consent from the patient to study the tissue blocks is also required. While collection of tissue samples during the trial is a significant burden for all participants, it is virtually impossible once the trial has been

completed. Hence, it is important that these elements of research are provided during the initial version of the protocol.

Eligibility of a particular patient for the new treatment often depends upon results of detailed characterization of the tumor. Many patients will therefore sign informed consent, only to find out that they are not eligible for the expected treatment. It is highly desirable that these patients are offered an alternative treatment within the same treatment protocol. Such a provision in the treatment protocol is not only desirable from the ethical standpoint but will also enable the researchers to follow all subjects in the group, regardless of the characteristics of the tumor. In this way, the results may be compared to trials without such characterization.

#### **10.9.3.6 Palliative Treatment**

This section would be incomplete without a word on the need for clinical research in palliative treatment. For obvious reasons, informed consent should be simplified. Challenges include identifying a target population, avoiding selection bias in the face of clinician and patient denial of serious illness, developing eligibility criteria for a seriously ill population, minimizing high patient refusals due to illness, and accurate reporting of all screened and eligible participants [58].

### **10.9.4 TYPES OF CLINICAL TRIALS**

The main mission of clinical research is testing ideas for improved diagnostics and treatment under controlled setting and then offering the experience either for further research or for clinical practice. If this is to be accomplished, the experience from clinical research should be relevant for its further application. At the same time, we should not forget that patients participating in clinical research have their individual priorities, which may change with time. Design and conduct of a clinical trial, therefore, involves a delicate balance between patient's interests and objectives of research.

#### **10.9.4.1 Phase I Trials**

For obvious reasons of acute toxicity and long-term mutagenic effects of anticancer drugs, phase I trials in cancer research are not performed on healthy volunteers. The research subjects in phase I trials in oncology are patients with cancer who have progressed after all conventional therapies. In addition, patients participating in phase I clinical trials are selected for absence of any significant comorbidity. This assures that dose-limiting toxicity and maximal tolerated dose are established with a proper degree of certainty. Since the response to treatment is not among the main objectives, patients in phase I research often suffer from a spectrum of different cancers.

Phase I clinical trials test new drugs on patients resistant to one or several standard combinations of drugs. While the possibility of life-threatening toxicity is not negligible, the chances for an objective response to therapy with a new drug are below 5% [59, 60]. These chances are even smaller for the initial subgroups of patients who are offered only very small doses of the new drug. A reasonable compromise should seek a balance between patient's safety, which demands only very

gradual increases of the dose of the drug, and patient's expectations to derive benefit from the new drug, a possibility that can only be expected with higher doses of the drug [61].

In addition to the commonly discussed situation of testing a new drug as a single agent, phase I trials are done (or at least should be done) also for new combinations of drugs that have already been approved for use and for new combinations of drugs and concomitant irradiation. In testing such a treatment, one drug, an existing combination of drugs, or irradiation are applied in the standard dosing and intervals, while a new drug is added upon this treatment in progressive doses or at progressively shorter intervals. Patients in this type of trial usually suffer from a well-defined diagnosis and stage of disease. In addition to assessment of toxicity, objectives of a trial include response rate, time to progression, and survival. For this reason, the term phase I–II trial is often used for this type of research. While terminology is of minor importance, it is very important that this step is not omitted. Testing a new idea straight in a phase II or even phase III trial often leads to premature closure of the trial due to unexpected toxicity.

#### 10.9.4.2 Phase II Trials

Phase II clinical trials stand between the very early experience of phase I research and eventual testing of this experience in a phase III trial. Early promising experience in phase I, information on pharmacokinetics and metabolism of drugs, and (to be sincere) commercial interests lead to definition of patients in whom a new drug, or a new combination of drugs, will be tested. Eligibility criteria for phase II trials include a precise diagnosis and stage of disease. Precise evaluation of toxicity remains among the main endpoints. For this reason, patients with severe comorbidities are usually excluded from early phase II trials.

When compared to phase I or phase III trials, phase II trials are easier to organize and cheaper. Number of patients is much smaller than in randomized trials. Many phase II trials are designed by individual researchers in academic institutions and may be conducted even without substantial support from industry. Flexibility in design and short time between idea and actual activation of a trial are the main advantages of such trials.

Experience from phase II trials is a most valuable basis for phase III trials and for preparing meta-analyses. However, two remarks are appropriate. The first one goes to honesty in performing small or single-institution trials. The key factor is unbiased registration of patients into the trial. In final reports, many single-institution trials include only patients who have completed a certain number of treatment cycles and exclude those with early withdrawal due to complications, disease progression, or patient's refusal. Biased registration of patients in single-institution trials is probably the most common form of fraud in clinical research and is the easiest way of improving the results. Unless we are strict in implication of a rigid formal procedure for registration of patients, this form of scientific dishonesty is virtually impossible to discover: In an institution treating thousands of patients, an outside observer cannot identify additional patients who also started the same treatment and were later excluded from the report. To prevent this, patient's entry into a clinical trial should be formal and in written form registered with a person who is not under the physician's authority. From the moment of registration, the patient is included in the trial, regardless of the actual course of disease or treatment.

The second remark is on application of experience from phase II trials for clinical practice. Experience on selected groups of a few dozens or a few hundred patients is far too scarce to be taken as “pure gold.” Let us repeat that promising experience from phase II trials gives support for further phase II research (possibly with wider eligibility criteria) and for phase III trials but should not be uncritically applied in clinical practice.

#### 10.9.4.3 Phase III Trials

Phase III clinical trials present a bridge between clinical research and the application of its experience in clinical practice. It is, therefore, essential that the experience from a phase III trial is applicable to a broad population of patients. Elderly patients and those with common comorbidity (hypertension, diabetes) should not be excluded from research if they are later to be treated with the same drugs. Rather, they may be offered an adjusted treatment schedule [62]. If this is not feasible, it is essential that the authors make a clear statement on limited applicability of the experience.

The ethics of randomized clinical trials have their basis in the uncertainty principle [63]. To maintain the uncertainty throughout the trial, it is essential to keep the recruitment period as short as possible. We have to avoid a possibility that an interim analysis would show a clear, yet statistically nonsignificant, superiority of one treatment over the other. As a rule of the thumb, we propose that the recruitment period should not be longer than double the expected median time to event (e.g., with 12 months as expected median survival, recruitment period should not extend beyond 24 months).

Many randomized clinical trials face the problem of slow recruitment. The most common reaction to slow recruitment is to seek participation of additional cancer centers. Sometimes this works, sometimes it does not; and it is not rare that a trial has to close due to poor recruitment. A general rule that “prevention is better than treatment” may be applied also to the problem of slow recruitment. Quite often, the problem of slow recruitment is not due to a small number of centers but due to a small proportion of eligible patients in each center who actually participate in a trial. Even in lung cancer, the most common cancer worldwide, it is quite common that a single institution recruits fewer than 10 patients within, say, 3 years, out of 100 or more eligible patients. The burden of complicated and time-consuming CRFs has an important negative impact upon recruitment of patients into clinical trials. A busy physician treating several hundred new patients per year does not dare to register more than a dozen patients for multi-institutional clinical trial. The narrow point is not the number of eligible patients, nor registration, nor treatment: It is the fear of CRFs with thousands of unnecessary data and hundreds of replies to queries. Developing a simple CRF is therefore essential for quick recruitment and, indeed, for success (or failure) of a trial. Whenever feasible, a randomized trial should follow the principles of a “large simple trial” and collect only the essential information [64].

#### 10.9.5 ENDPOINTS

Endpoints in clinical trials in oncology are an interesting topic, yet they are rarely discussed. Few physicians really understand all biases behind simple words such as



response, time to progression, survival, and quality of life [65]. Let us briefly discuss these biases.

### 10.9.5.1 Response to Treatment

Response to treatment is now understood as objective evidence of regression of the tumor. Other indications of response such as symptomatic improvement are assessed under quality of life.

In objective evaluation of response, most clinical trials nowadays follow the RECIST guidelines (Response Evaluation Criteria In Solid Tumors) [66]. The procedure involves three steps:

1. Definition of target lesions. At least one and not more than five lesions in any organ affected by cancer should be defined as measurable. Other lesions are declared as nonmeasurable.
2. Measurement of target lesions during or after treatment. The same method (e.g., CT scan) should be used prior to treatment and at evaluation of response. Nonmeasurable lesions are also evaluated for their complete disappearance, persistence, or obvious progression.
3. Confirmation of response after a minimum of 1 month since establishing a response.

By definition, *complete response* means complete resolution of all measurable and nonmeasurable disease for more than 1 month. *Partial response* is defined as >30% reduction in the sum of all measurable lesions for more than 1 month, along with no evidence of progression of nonmeasurable disease. *Progression* is either >20% increase (in comparison with the smallest measurement) in the sum of all measurable lesions, or obvious progression of nonmeasurable disease, or any appearance of a few lesions. *Stable disease* is what remains between partial response and stable disease.

While these three steps are straightforward and should be easy to follow, evaluation of response in the practice of clinical trials is far from devoid of any bias. *Definition of target lesions* is often not done during registration of a patient for a clinical trial. Researchers are allowed to enter patients on the basis of meeting the eligibility criteria, while the target lesions are defined later in the course of treatment. This would be acceptable if all lesions of a tumor would follow the same curve of regression or progression. Still, this is not the case. The tumor population on different sites of disease is heterogeneous. In addition, there is a variable proportion of the accompanying inflammatory reaction that contributes to the bulk of a tumor, as measured radiographically. Chemotherapy itself also does not reach the same concentration in all tissues. For these reasons, we often see considerable heterogeneity in response among different organs affected by the disease, and even within the same organ. A posteriori definition of target lesions opens a possibility that those responding better will be measured, leading to an increase in response rate.

*Measurement of target lesions* is also not free of a bias. Most tumors are not round and do not shrink or progress evenly in all directions. The longest diameter is often not relevant: A lung tumor may extend into an interlobar fissure, and its longest diameter may remain unchanged even if the tumor shrinks considerably. Measuring



the tumor in other directions allows a bias similar to the one described in the previous paragraph: A posteriori definition of the direction in which response to treatment is most clearly seen leads to an increase in response rate.

*Confirmation of response* is a demand that is not always strictly followed. Even if they declare adherence to the RECIST criteria, not all researchers confirm partial remission after another month. From the clinical point of view, this seems logical: After an objective response has been documented, there is no urgent need for early repeat examinations; reevaluation every 2 or 3 months is sufficient. Finally, confirmation of response is not feasible in certain situations such as induction chemotherapy prior to surgery or irradiation where local treatment immediately follows induction chemotherapy.

### ***Recommendations and Personal View on Evaluation for Response to Treatment***

1. Target lesions for measurable disease should be defined at the time of registration of the patient for the trial.
2. Unidimensional measurement of measurable disease is not free of a bias. Precise volumetric analysis (which is quite feasible with modern computerized radiology) of predefined lesions might offer more information than unidimensional measurements and would better reflect true regression or progression of the disease.
3. Early confirmation of response does not contribute to objective evaluation and adds the burden of unnecessary diagnostics.

#### **10.9.5.2 Time to Progression**

Time to progression is defined as the interval from start of treatment until progression (see previous section for definition of progression). On the first glance, time to progression is clearly defined. Still, at least two comments may be added—one from statistical point of view and the other from the clinical standpoint.

Quite often, we read about very small differences in time to progression of 1 or 2 weeks; yet, the average interval at which the tumor is evaluated may be once every 2 or 3 months. Looking at such data with the eyes of a statistician, the situation is similar to measuring centimeters with a 1-meter scale. Unless the number of patients is really very large, differences in time to progression smaller than half of the interval between measurements should be taken with great caution.

From the clinical standpoint, very frequent exams for eventual progression may not be justified. Also, a clinician may feel that a certain treatment is still considered beneficial, in spite of minor radiological progression. A useful addition to time of radiologic progression might be *time to clinically meaningful progression*, defined as a moment when the treatment has to be changed or a new treatment modality introduced.

#### **10.9.5.3 Survival**

Survival appears as the most clear endpoint. Yet, even survival is not free of a bias.

Survival is a function of prognostic factors (such as stage, histology, age, gender, and performance status), of a specific treatment under consideration in a particular

trial, and of the treatment after progression (sometimes called “salvage treatment,” a term that is often too ambitious or misleading). Although second-line treatment rarely cures patients, it may considerably prolong survival. A trial protocol should therefore define the general guidelines for treatment at the moment of progression. Type of second-line treatment and proportion of patients who actually received such treatment should be included in the report. This will ensure that eventual difference in survival is not due to unbalanced second-line treatment.

When speaking about survival as “the ultimate endpoint,” we most often focus on statistically significant difference and do not consider the fact that a difference in survival also has to be clinically meaningful. Provided the trial is large enough, even a 2% difference in survival can become statistically significant. It is important to note that although the difference between two groups may be statistically significant to a very small probability value, the difference may be of no clinical significance [67].

In general, advantage of one treatment over another will shrink when the treatment is taken from the research setting into general use: Broader selection of patients and physicians with varying degree of expertise in the treatment and in management of complications contribute to inferior results. When improvement in survival is small, statistical significance is not the only criterion for accepting a new treatment. In such instances, the concept of clinically meaningful difference should be applied. A judgment on the applicability and relevance of the research data for the general population, on the burden of the new treatment for patients, and on the costs should be thoroughly discussed before the new treatment is accepted as a routine.

#### 10.9.5.4 Quality of Life

When it comes to incurable disease—and advanced cancer still most often falls under this category—quality of life is of at least equal importance as the other endpoints we just discussed. While we all recognize the importance of quality of life, we are far from agreement on how to approach this issue in clinical trials.

Assessment of quality of life is now regularly included in most treatment protocols [68–70]. Still, the majority of published reports do not present data on quality of life. Most often, we are left with the data on toxicity, which partially reflect quality of life but cannot offer a comprehensive picture. In a survey of randomized clinical trials for advanced breast cancer, assessment of quality of life added relatively little value to other endpoints in helping select the best treatment option, apparently largely because of suboptimal methodological standards [71].

Two fundamental reasons are behind this lack on information on quality of life. The first one is that in any particular trial, data on quality of life are virtually always incomplete. They are most often missing for those patients who do not do well—and for whom evaluation of quality of life would be most important. Patients with progression and/or severe toxicity fail to return for follow-up examinations and do not respond to the questionnaire. In such instances, one cannot avoid a bias in an analysis of an incomplete series of questionnaires [72].

The second reason for rare inclusion of quality of life issues in reports of clinical trials is the difficulty in analyzing the data. Instruments for assessment of quality of life include from 10 to more than 30 questions; a protocol often includes two

instruments (such as observer's and patient's scale). If presenting only one or two items from the lengthy quality of life questionnaire, the author could be blamed for a bias. If all the data are presented and analyzed, the volume of information is such that a separate paper might be needed.

Unlike other endpoints, quality of life data are rarely—if ever—presented in meta-analyses. The reason is inconsistency in instruments and in reporting.

The current trend in assessing quality of life of cancer patients is to use increasingly complex instruments. I am not in favor of this approach. That the mental, physical, and social domains, each containing many dimensions and items, all contribute to quality of life is uncontroversial. What is controversial is the weight of the different dimensions in overall quality of life. It has been shown to be very different between different patient populations. For individuals, assuredly complex systems, the many dimensions and items of quality of life interact, probably sometimes in chaotic ways. In these conditions, the weights of isolated items in individuals become for all practical purposes meaningless. The classical endpoints of discrete health-related functions and duration of survival are increasingly perceived as unacceptably reductionistic [73].

In our single-institutional clinical trials, we use our own simplified scale for assessment of quality of life: How do you feel, in comparison with your feeling prior to treatment?

1. Much worse
2. Worse
3. About the same
4. Better
5. Much better

Such a simple scale follows the idea that it is the patient who can best describe his or her quality of life; It is of lesser importance what precisely this means for an individual patient. The approach may be unscientific, but it is reliable, easy to use, and easy to analyze.

### 10.9.6 CONCLUSION

Most methodological issues of clinical research are similar in every field of medicine. An attempt to present a comprehensive overview of the methods of clinical research in oncology would inevitably lead to overlapping with other chapters. General questions of design of a clinical trial, its organization, statistics, and regulatory issues are to be found elsewhere in this volume. In this chapter, we focused on specific questions of design and conduct of clinical research in oncology. The choice of issues was admittedly personal, as were the views and proposals. Some dilemmas of a clinical oncologist who is also involved in research were discussed, all with the aim of facilitating research.

Our progress against cancer critically depends upon the quality and quantity of clinical research. A positive attitude of patients toward participation in clinical research, removing the obstacles that restrain physicians to enter patients into

clinical trials, and the relevance of research for regular medical practice are the three crucial points. We hope that this chapter will contribute toward a better performance on all these three points.

## REFERENCES

1. Sargent, D. J., Conley, B. A., Allegra, C., et al. (2005), Clinical trial designs for predictive marker validation in cancer treatment trials, *J. Clin. Oncol.*, 23, 2020–2027.
2. Meyerson, L. J., Wiens, B. L., LaVange, L. M., et al. (2000), Quality control of oncology clinical trials, *Hematol. Oncol. Clin. North. Am.*, 14, 953–971.
3. Go, R. S., Frisby, K. A., Lee, J. A., et al. (2006), Clinical trial accrual among new cancer patients at a community-based cancer center, *Cancer*, 106, 426–433.
4. Murthy, V. H., Krumholz, H. M., and Gross, C. P. (2004), Participation in cancer clinical trials: Race-, sex-, and age-based disparities, *JAMA*, 291, 2720–2726.
5. Lara, P. N. Jr, Higdon, R., Lim, N., et al. (2001), Prospective evaluation of cancer clinical trial accrual patterns: Identifying potential barriers to enrollment, *J. Clin. Oncol.*, 19, 1728–1733.
6. Sateren, W. B., Trimble, E. L., Abrams, J., et al. (2002), How sociodemographics, presence of oncology specialists, and hospital cancer programs affect accrual to cancer treatment trials, *J. Clin. Oncol.*, 20, 2109–2117.
7. Grunfeld, E., Zitzelsberger, L., Coristine, M., et al. (2002), Barriers and facilitators to enrollment in cancer clinical trials: Qualitative study of the perspectives of clinical research associates, *Cancer*, 95, 1577–1583.
8. Somkin, C. P., Altschuler, A., Ackerson, L., et al. (2005), Organizational barriers to physician participation in cancer clinical trials, *Am. J. Manag. Care*, 11, 413–421.
9. European Union Clinical Trials Directive; available at [http://www.wctn.org.uk/downloads/EU\\_Directive/Directive.pdf](http://www.wctn.org.uk/downloads/EU_Directive/Directive.pdf).
10. Declaration of Helsinki; available at <http://www.wma.net/e/policy/b3.htm>.
11. Harris, J. (2005), Scientific research is a moral duty, *J. Med. Ethics*, 31, 242–248.
12. Zwitter, M. (1999), Ethics of randomized clinical trials and the “ALARA” approach, *Acta Oncol.*, 38, 99–105.
13. Williams, C. J., and Zwitter, M. (1994), Informed consent in European multicentre randomised clinical trials. Are patients really informed? *Eur. J. Cancer*, 30A, 907–910.
14. Joffe, S., Cook, E. F., Cleary, P. D., et al. (2001), Quality of informed consent in cancer clinical trials: A cross-sectional survey, *Lancet*, 358, 1772–1777.
15. Coyne, C. A., Xu, R., Raich, P., et al. (2003), Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: A study of the Eastern Cooperative Oncology Group, *J. Clin. Oncol.*, 21, 836–842.
16. Jayson, G., and Harris, J. (2006), How participants in cancer trials are chosen: Ethics and conflicting interests, *Nat. Rev. Cancer*, 6(4), 330–336.
17. Lippman, S. M., and Lee J. J. (2006), Reducing, the “risk” of chemoprevention: Defining and targeting high risk—2005 AACR Cancer Research and Prevention Foundation Award Lecture, *Cancer Res.*, 66, 2893–2903.
18. Klein, E. A. (2006), Chemoprevention of prostate cancer, *Annu. Rev. Med.*, 57, 49–63.
19. Demierre, M. F., Higgins, P. D., Gruber, S. B., et al. (2005), Statins and cancer prevention, *Nat. Rev. Cancer*, 5, 930–942.

20. Kahn, J. A. (2005), Vaccination as a prevention strategy for human papillomavirus-related diseases, *J. Adolesc. Hlth.*, 37, S10–16.
21. Villa, L. L., Costa, R. L., Petta, C. A., et al. (2005), Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: A randomised double-blind placebo-controlled multicentre phase II efficacy trial, *Lancet Oncol.*, 6(5), 271–278.
22. Arnold, D., and Schmoll, H. J. (2005), (Neo-)adjuvant treatments in colorectal cancer, *Ann. Oncol.*, 16(Suppl 2), 133–140.
23. Betticher, D. C. (2005), Adjuvant and neoadjuvant chemotherapy in NSCLC: A paradigm shift, *Lung Cancer*, 50(Suppl 2), S9–16.
24. Glynne-Jones, R., Grainger, J., Harrison, M., et al. (2006), Neoadjuvant chemotherapy prior to preoperative chemoradiation or radiation in rectal cancer: Should we be more cautious? *Br. J. Cancer*, 94, 363–371.
25. Amiel, G. E., and Lerner, S. P. (2006), Combining surgery and chemotherapy for invasive bladder cancer: Current and future directions, *Expert Rev. Anticancer Ther.*, 6, 281–291.
26. Smith, I., and Chua, S. (2006), Medical treatment of early breast cancer. IV: Neoadjuvant treatment, *BMJ*, 332, 223–224.
27. Evans, D. B. (2005), Preoperative chemoradiation for pancreatic cancer, *Semin. Oncol.*, 32(6 Suppl 9), S25–29.
28. Gallo, A., and Frigerio, L. (2003), Neoadjuvant chemotherapy and surgical considerations in ovarian cancer, *Curr. Opin. Obstet. Gynecol.*, 15, 25–31.
29. El Sharouni, S. Y., Kal, H. B., and Battermann, J. J. (2003), Accelerated regrowth of non-small-cell lung tumours after induction chemotherapy, *Br. J. Cancer*, 89, 2184–2189.
30. Carlson, R. W., Brown, E., Burstein, H. J., et al. (2006), National Comprehensive Cancer Network. NCCN Task Force Report: Adjuvant therapy for breast cancer, *J. Natl. Compr. Cancer Net.*, 4(Suppl 1), S1–26.
31. Arriagada, R., Spielmann, M., Koscielny, S., et al. (2005), Results of two randomized trials evaluating adjuvant anthracycline-based chemotherapy in 1146 patients with early breast cancer, *Acta Oncol.*, 44(5), 458–466.
32. Merlano, M., and Mattiot, V. P. (2006), Future chemotherapy and radiotherapy options in head and neck cancer, *Expert Rev. Anticancer Ther.*, 6, 395–403.
33. Leonard, G. D., McCaffrey, J. A., and Maher, M. (2003), Optimal therapy for oesophageal cancer, *Cancer Treat. Rev.*, 29, 275–282.
34. Psyrri, A., and Fountzilas, G. (2006), Advances in the treatment of locally advanced non-nasopharyngeal squamous cell carcinoma of the head and neck region, *Med. Oncol.*, 23, 1–15.
35. Rigas, J. R., and Lara, P. N. Jr. (2005), Current perspectives on treatment strategies for locally advanced, unresectable stage III non-small cell lung cancer, *Lung Cancer*, 50(Suppl 2), S17–24.
36. Henson, J. W. (2006), Treatment of glioblastoma multiforme: A new standard, *Arch. Neurol.*, 63, 337–341.
37. Oehler, C., and Ciernik, I. F. (2006), Radiation therapy and combined modality treatment of gastrointestinal carcinomas, *Cancer Treat. Rev.*, 32, 119–138.
38. Sastre, J., Garcia-Saenz, J. A., and Diaz-Rubio, E. (2006), Chemotherapy for gastric cancer, *World J. Gastroenterol.*, 12, 204–213.
39. Gillespie, M. B., Marshall, D. T., Day, T. A., et al. (2006), Pediatric rhabdomyosarcoma of the head and neck, *Curr. Treat. Options Oncol.*, 7, 13–22.

40. Bosset, J. F., Lorchel, F., Manton, G., et al. (2005), Radiation and chemoradiation therapy for esophageal adenocarcinoma, *J. Surg. Oncol.*, 92, 239–245.
41. Roukos, D. H., and Kappas, A. M. (2005), Perspectives in the treatment of gastric cancer, *Natl. Clin. Pract. Oncol.*, 2, 98–107.
42. Deutsch, E., Soria, J. C., and Armand, J. P. (2005), New concepts for phase I trials: Evaluating new drugs combined with radiation therapy, *Natl. Clin. Pract. Oncol.*, 2, 456–465.
43. Bradley, J. (2005), A review of radiation dose escalation trials for non-small cell lung cancer within the Radiation Therapy Oncology Group, *Semin. Oncol.*, 32(2 Suppl 3), S111–113.
44. Bernier, J., and Bentzen, S. M. (2003), Altered fractionation and combined radio-chemotherapy approaches: Pioneering new opportunities in head and neck oncology, *Eur. J. Cancer*, 39, 560–571.
45. Baumann, M., Appold, S., Petersen, C., et al. (2001), Dose and fractionation concepts in the primary radiotherapy of non-small cell lung cancer, *Lung Cancer*, 33(Suppl 1), S35–45.
46. Wilson, G. D., Bentzen, S. M., and Harari, P. M. (2006), Biologic basis for combining drugs with radiation, *Semin. Radiat. Oncol.*, 16, 2–9.
47. Lawrence, T. S., Blackstock, A. W., and McGinn, C. (2003), The mechanism of action of radiosensitization of conventional chemotherapeutic agents, *Semin. Radiat. Oncol.*, 13, 13–21.
48. Pauwels, B., Korst, A. E., Pattyn, G. G., et al. (2003), Cell cycle effect of gemcitabine and its role in the radiosensitizing mechanism in vitro, *Int. J. Radiat. Oncol. Biol. Phys.*, 57, 1075–1083.
49. Zwitter, M., Kovac, V., Smrdel, U., et al. (2006), Gemcitabine, cisplatin and hyperfractionated accelerated radiotherapy for locally advanced non-small cell lung cancer, *J. Thorac. Oncol.*, 1, 662–666.
50. Budihna, M., Soba, E., Smid, L., et al. (2005), Inoperable oropharyngeal carcinoma treated with concomitant irradiation, mitomycin C and bleomycin—long term results, *Neoplasma*, 52(2), 165–174.
51. Bernier, J. (2005), Alteration of radiotherapy fractionation and concurrent chemotherapy: A new frontier in head and neck oncology? *Nat. Clin. Pract. Oncol.*, 2, 305–314.
52. Lynch, T. Jr, and Kim, E. (2005), Optimizing chemotherapy and targeted agent combinations in NSCLC, *Lung Cancer*, 50(Suppl 2), S25–32.
53. Nahta, R., and Esteva, F. J. (2006), Herceptin: Mechanisms of action and resistance, *Cancer Lett.*, 232, 123–138.
54. Tarn, C., and Godwin, A. K. (2005), Molecular research directions in the management of gastrointestinal stromal tumors, *Curr. Treat. Options Oncol.*, 6, 473–486.
55. Sanborn, R. E., and Blanke, C. D. (2005), Gastrointestinal stromal tumors and the evolution of targeted therapy, *Clin. Adv. Hematol. Oncol.*, 3, 647–657.
56. Cortes, J., and Kantarjian, H. (2005), New targeted approaches in chronic myeloid leukemia, *J. Clin. Oncol.*, 23, 6316–6324.
57. Boehrer, S., Nowak, D., Hoelzer, D., et al. (2006), Novel agents aiming at specific molecular targets increase chemosensitivity and overcome chemoresistance in hematopoietic malignancies, *Curr. Pharm. Des.*, 12, 111–128.
58. Bakitas, M. A., Lyons, K. D., Dixon, J., et al. (2006), Palliative care program effectiveness research: Developing rigor in sampling design, conduct, and reporting, *J. Pain Symptom Manage.*, 31, 270–284.

59. Rosa, D. D., Harris, J., and Jayson, G. C. (2006), The best guess approach to phase I trial design, *J. Clin. Oncol.*, 24, 206–208.
60. Horstmann, E., McCabe, M. S., Grochow, L., et al. (2005), Risks and benefits of phase 1 oncology trials, 1991 through 2002, *N. Engl. J. Med.*, 352, 895–904.
61. Rogatko, A., Babb, J. S., Tighiouart, M., et al. (2005), New paradigm in dose-finding trials: Patient-specific dosing and beyond phase I, *Clin. Cancer Res.*, 11, 5342–5346.
62. Townsley, C. A., Selby, R., and Siu, L. L. (2005), Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials, *J. Clin. Oncol.*, 23, 3112–3124.
63. Hansson, S. O. (2006), Uncertainty and the ethics of clinical trials, *Theor. Med. Bioeth.*, 27, 149–167.
64. Thom, E. A., and Klebanoff, M. A. (2005), Issues in clinical trial design: Stopping a trial early and the large and simple trial, *Am. J. Obstet. Gynecol.*, 193, 619–625.
65. Sakamoto, J., and Teramukai, S. (2002), Data handling in cancer clinical trials-how we can minimize potential biases, *Jpn. J. Clin. Oncol.*, 32, 1–2.
66. Therasse, P., Arbuck, S. G., Eisenhauer, E. A., et al. (2000), New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada, *J. Natl. Cancer Inst.*, 92, 205–216.
67. Lader, E. W., Cannon, C. P., Ohman, E. M., et al.; the American Heart Association (2004), The clinician as investigator: Participating in clinical trials in the practice setting: Appendix 2: statistical concepts in study design and analysis, *Circulation* 109, e305–307.
68. Kirkova, J., Davis, M. P., Walsh, D., et al. (2006), Cancer symptom assessment instruments: A systematic review, *J. Clin. Oncol.*, 24, 1459–1473.
69. Gunnars, B., Nygren, P., Glimelius, B., et al. (2001), Swedish Council of Technology Assessment in Health Care. Assessment of quality of life during chemotherapy, *Acta Oncol.*, 40, 175–184.
70. Kiebert, G. M., Curran, D., and Aaronson, N. K. (1998), Quality of life as an endpoint in EORTC clinical trials. European Organization for Research and Treatment for Cancer, *Stat. Med.*, 17, 561–569.
71. Fossati, R., Confalonieri, C., Mosconi, P., et al. (2004), Quality of life in randomized trials of cytotoxic or hormonal treatment of advanced breast cancer. Is there added value? *Breast Cancer Res. Treat.*, 87, 233–243.
72. Fayes, P. M., Hopwood, P., Harvey, A., et al. (1997), Quality of life assessment in clinical trials—guidelines and a checklist for protocol writers: The U.K. Medical Research Council experience. MRC Cancer Trials Office, *Eur. J. Cancer*, 33, 20–28.
73. Bernheim, J. L. (1999), How to get serious answers to the serious question: “How have you been?” Subjective quality of life (QOL) as an individual experiential emergent construct, *Bioethics*, 13, 272–287.



# *10.10*

---

## *Pharmacological Treatment Options for Nonexudative and Exudative Age-Related Macular Degeneration*

**Alejandro Oliver, Thomas A. Ciulla, and Alon Harris**

*Department of Ophthalmology, Indiana University, Indianapolis, Indiana*

### **Contents**

|           |                                               |     |
|-----------|-----------------------------------------------|-----|
| 10.10.1   | Introduction                                  | 608 |
| 10.10.2   | Diagnosis                                     | 609 |
| 10.10.3   | Nonexudative Age-Related Macular Degeneration | 609 |
| 10.10.3.1 | Antioxidants                                  | 609 |
| 10.10.3.2 | Drusen Ablation                               | 610 |
| 10.10.3.3 | Rheopheresis                                  | 610 |
| 10.10.4   | Exudative Age-Related Macular Degeneration    | 611 |
| 10.10.4.1 | Thermal Laser Photocoagulation                | 611 |
| 10.10.4.2 | Transpupillary Thermotherapy                  | 612 |
| 10.10.4.3 | Photodynamic Therapy                          | 612 |
| 10.10.4.4 | Radiation Therapy                             | 614 |
| 10.10.4.5 | Surgical Therapy                              | 615 |
| 10.10.4.6 | Antiangiogenic Therapy                        | 615 |
|           | References                                    | 620 |

### 10.10.1 INTRODUCTION

Age-related macular degeneration (AMD) is the leading cause of vision loss in the developed world, and it is estimated that approximately 1.2 million Americans currently suffer from severe vision loss caused by this disease [1–4]. This number is likely to increase considerably as the population in developed countries ages.

Two main types of macular degeneration have been traditionally recognized: the nonexudative (dry), composed of atrophic changes of the retinal pigment epithelium as well as deposits beneath it, and the exudative (wet) in which abnormal blood vessels develop under the retina causing blood and fluid leakage. Approximately 10–20% of nonexudative AMD will eventually progress to the exudative form, which is responsible for the majority of cases of severe visual loss from AMD [4, 5].

The traditional AMD classification criteria were revised in 1995 by the Age-Related Maculopathy Epidemiological Study Group, and the criteria for diagnosis of AMD became stricter. Patients with minimal or moderate nonexudative age-related changes in the macula were reclassified as having age-related maculopathy (ARM). By definition, advanced retinal pigment epithelium (RPE) atrophy and clumping is now required for nonexudative AMD, and the presence of choroidal neovascularization (CNV) is a requisite for the diagnosis of exudative age-related macular degeneration [6]. Currently, an estimated 85–90% of patients with age-related macular changes are ARM patients who exhibit drusen and only mild to moderate RPE changes and are typically minimally symptomatic with mild blurred central vision, difficulty reading, color and contrast disturbances, and metamorphopsia. The remaining 10–15% of patients with macular changes, who fall under the modern definition of AMD, tend to describe painless, progressive, moderate to severe blurring of central vision and moderate to severe metamorphopsia, which can be acute or insidious in onset [6].

Although some subtypes of exudative AMD are potentially treatable, the currently available treatment modalities offer limited efficacy; therefore, great interest exists in delaying or ceasing the progression of ARM or more effectively treating the factors leading to vision loss once it becomes AMD. At present, the only widely accepted method of intervention for ARM is the use of high-dose antioxidants; however, this only slows progression in some patients and does not reverse any damage already present. Once AMD becomes exudative, the treatment scheme offered to patients varies largely among physicians, as no standard therapy has been established and approved. Options currently available include laser photocoagulation, photodynamic therapy (PDT) with verteporfin, and intravitreal pegaptanib sodium. Only a minority of patients with exudative AMD shows well-demarcated “classic” CNV amenable to laser treatment, and at least half of the patients undergoing thermal laser photocoagulation suffer persistent or recurrent CNV formation within 2 years. In addition, since the treatment itself causes a blinding central scotoma when the CNV is located subfoveally, many clinicians do not treat subfoveal CNV with thermal laser. In 2000, PDT was approved by the U.S. Food and Drug Administration (FDA) as treatment for subfoveal CNV; however, it only limits vision loss and often requires multiple retreatments. Pegaptanib sodium, a vascular endothelial growth factor (VEGF) inhibitor, was approved by the FDA on December 17, 2004, and was available to physicians in January of 2005; however, its administration requires intravitreal injections every 6 weeks. Because of these treat-

ment limitations, alternative therapies for exudative AMD are being developed and include new types of photodynamic therapy, transpupillary thermotherapy, growth factor modulators, radiation, and surgical therapy. A major limitation for the design of effective treatment is still our lack of true understanding of the underlying etiology of the disease.

### 10.10.2 DIAGNOSIS

Any patient who exhibits signs and symptoms consistent with exudative AMD undergoes a thorough dilated fundus exam, stereo color fundus photography, and rapid sequence fluorescein angiography (RSFA). RSFA, which is usually the initial angiographic study, will reveal leakage—the hallmark of CNV. According to the distance from the foveal avascular zone, the leakage is classified as subfoveal, juxtafoveal (1–199  $\mu\text{m}$ ) or extrafoveal (200–2500  $\mu\text{m}$ ). In addition, indocyanine green (ICG) angiography is performed as an adjunctive study in patients with poorly delineated CNV. ICG can better delineate choroidal circulation because the near-infrared light (795–810 nm) absorbed by ICG tends to penetrate the retinal pigment epithelium better than can the shorter wavelength of light absorbed by fluorescein. Also, unlike fluorescein, ICG is strongly bound to plasma proteins, which prevents diffusion of the compound through the fenestrated choroidal capillaries, and permits better delineation of choroidal details. When a CNV is suspected, angiography is customarily performed within 72 hours of any planned treatment since CNV morphology and resulting treatment parameters can evolve rapidly.

The Macular Photocoagulation Study (MPS) defined two RSFA leakage patterns for CNV [7]. “Classic” CNV presents as discrete early hyperfluorescence with late leakage of dye into an overlying neurosensory retinal detachment. “Occult” CNV is categorized into two basic forms: late leakage of undetermined source, or fibrovascular pigment epithelial detachments. (PEDs). Late leakage of undetermined source manifests as regions of stippled leakage into an overlying neurosensory retinal detachment, without a distinct source identified on the early frames of the angiogram. Fibrovascular PEDs present as irregular elevations of the RPE associated with stippled leakage into an overlying neurosensory retinal detachment in the early and late frames of the angiogram.

### 10.10.3 NONEXUDATIVE AGE-RELATED MACULAR DEGENERATION

#### 10.10.3.1 Antioxidants

The treatment options for ARM and nonexudative AMD are limited; therefore, prevention of disease progression is viewed as critical. Currently, the only evidence-based intervention comes from the Age-Related Eye Disease Study (AREDS) [8]. This multicenter, U.S. National Institutes of Health (NIH)–supported investigation was a double-masked, randomized, prospective, clinical trial that enrolled 4357 subjects into 1 of 4 treatment groups: placebo, antioxidants (vitamins C and E plus  $\beta$ -carotene), zinc/copper, and antioxidants plus zinc/copper. The study was based on the theory that oxidative damage to the retina may contribute to the development of AMD [9–12] and a smaller, randomized, placebo-controlled clinical trial

that suggested zinc might provide protection from vision loss due to AMD [13]. AREDS concluded that “patients with extensive intermediate sized drusen, at least 1 large drusen or noncentral geographic atrophy in one or both eyes, or advanced AMD or vision loss due to AMD in one eye should consider taking a supplement of antioxidants plus zinc” [8]. The formulation (Ocuvite PreserVision, Bausch & Lomb, Rochester, New York, and ICaps, Alcon, Inc., Fort Worth, Texas) containing high doses of vitamin C, vitamin E,  $\beta$ -carotene, copper, and zinc lowered the risk of developing advanced AMD by 25% in the study [8]. Importantly, another study demonstrated that  $\beta$ -carotene supplementation increased the risk of developing lung cancer in smokers; therefore, the antioxidant formulation should exclude  $\beta$ -carotene when taken by tobacco users [14].

### 10.10.3.2 Drusen Ablation

It has been proposed that the aging RPE accumulates remnants of incomplete degradation of phagocytosed rod and cone membranes, and such accumulation results in the presence of metabolic debris and, over time, drusen formation [15, 16]. Laser treatment at or near a large drusen will often lead to resolution of the drusen, and sometimes this will be accompanied by an improvement of visual acuity. Based on the knowledge that drusen constitute a risk factor for exudative AMD, it has been proposed that laser treatment to promote clearing of drusen may result in a reduction of the risk of CNV formation. A randomized multicenter clinical trial, known as CNVPT (Choroidal Neovascularization Prevention Trial) evaluated 432 eyes treated with argon green laser. The study showed significant drusen reduction in the treated group; however, it also suggested that argon green laser treatment might increase the risk of CNV development [17]. Another study also showed that treatment with laser photocoagulation results in significant drusen reduction compared with observation at 2 years, but no differences were observed in CNV occurrence between groups [18]. In addition, two prospective randomized clinical trials are currently in progress to evaluate the effects of laser-induced drusen reduction on AMD. The NIH has supported the CAPT (Complications of AMD Prevention Trial) study, which is evaluating low-intensity argon green laser photocoagulation in patients with bilateral drusen. The study completed enrollment of 1052 patients in March 2001, and the 5-year results should be available this year [19]. The PTAMD (Prophylactic Treatment of AMD) study, a multicenter randomized prospective placebo-controlled clinical trial sponsored by Iridex Corp (Mountain View, CA) has completed enrollment and is currently in progress [20]. It is designed to evaluate the effects of extrafoveal subthreshold infrared diode treatment on stopping the progression of exudative AMD.

### 10.10.3.3 Rheopheresis

Some researchers believe that lipid deposition in sclera and Bruch's membrane leads to scleral stiffening and impaired choroidal perfusion, which in turn could adversely affect the metabolic transport function of RPE. The affected RPE would in consequence be incapable of efficiently metabolizing and removing material shed from the photoreceptors, leading to its accumulation and drusen formation [21, 22]. Based on this theory, it has been speculated that apheresis could be of benefit to

AMD patients by improving ocular blood flow and by limiting the concentration of circulating macromolecules involved in drusen formation, Bruch's membrane degradation, and retinal pigment epithelial cell dysfunction. OccuLogix LP (a joint venture between TLC Vision Co, Mississauga, Canada, and Vascular Sciences CO, Tampa, Florida) developed the Rheofilter membrane differential filter (MDF) system based on the concept of apheresis, to filter high-molecular-weight proteins and lipoproteins from the blood. A randomized prospective, double-masked pilot study involving 30 patients showed vision improvement in a significant number of patients undergoing rheophoresis. Currently, in the United States the MIRA-1 (Multicenter Investigation of Rheophoresis for AMD) is a multicenter, randomized, placebo-controlled trial of patients with large soft drusen without advanced AMD in at least one eye, and elevated serum cholesterol, IgA or fibrinogen at screening. This phase III trial was interrupted after about one-half of the planned 180 patients had been recruited due to loss of capitalization by the sponsor. An interim analysis performed on 43 patients who completed the one-year visit revealed an improvement in mean acuity of greater than one line in treated patients, compared to a mean loss of almost two lines in controls. The study has now been resumed and 185 patients have been enrolled. On November 17, 2005, OccuLogix announced that all final study visits had been completed for the MIRA-1 trial, and results would be analyzed by the end of December 2005. They were planning to file for FDA approval in the first quarter of 2006 [23].

#### 10.10.4 EXUDATIVE AGE-RELATED MACULAR DEGENERATION

##### 10.10.4.1 Thermal Laser Photocoagulation

Traditionally, ophthalmologists have used thermal laser destruction of CNV as the primary treatment of exudative AMD based on the results of the Macular Photocoagulation Study (MPS), a large, randomized, multicenter, prospective set of clinical trials comparing laser photocoagulation to observation. These studies, which were initiated in the 1980s and supported by the NIH, demonstrated that laser photocoagulation of certain types of CNV lowered the risk of large reductions in visual acuity compared to observation alone. In these studies, patients were deemed eligible for laser photocoagulation if they manifested classic CNV as determined by RSFA.

Unfortunately, only 13–26% of patients with exudative AMD presented with classic CNV eligible for laser treatment, and it became unclear whether laser photocoagulation was beneficial in a majority of patients, as they were not eligible for laser therapy in the MPS [7, 24]. Moreover, at least half of the enrolled subjects suffered from persistent or recurrent CNV formation within 2 years of treatment [24–26]. Although the arm of the MPS exploring treatment of CNV under the fovea suggested that laser photocoagulation is better than observation, treating subfoveal CNV with thermal photocoagulation is not a common practice because of the immediate central scotoma from the collateral retinal destruction. Given the large number of limitations posed by thermal laser photocoagulation, researches have searched for alternative means of subfoveal CNV treatment using a variety of laser derivatives [27, 28].

#### 10.10.4.2 Transpupillary Thermotherapy

Transpupillary thermotherapy (TTT) occludes the CNV by slowly heating the subfoveal choroidal neovascular complex with infrared (810 nm) diode laser light. The infrared wavelength is thought to traverse the retina and RPE to maximally affect the CNV, while minimizing thermal injury to the neurosensory retina. Laser application covers the entire CNV complex with a single large spot. Although the precise mechanism of CNV destruction is unclear, one study, using color Doppler imaging, suggested TTT leads to alterations in choroidal blood flow [29]. An uncontrolled phase I/II safety and efficacy study involving 113 patients showed that patients with occult CNV receiving TTT compared similarly to the verteporfin-treated patients in the verteporfin in photodynamic therapy (VIP) trial at 6 and 12 months [30]. Similarly, another uncontrolled trial with 69 patients found that TTT use compared favorably to the natural history of occult CNV [31]. The Transpupillary Thermotherapy of Occult Subfoveal Choroidal Neovascular Membranes in Patients with Age-Related Macular Degeneration Trial (TTT4CNV), the first randomized, prospective, double-blind, placebo-controlled study evaluating the effectiveness of TTT for occult CNV (or up to 10% classic CNV), enrolled 303 patients between March 2000 and March 2003 and is now following subjects [32].

#### 10.10.4.3 Photodynamic Therapy

Photodynamic therapy (PDT) utilizes laser light and intravascular dyes (i.e., photosensitizers). After intravenous injection, once sufficient time passes to concentrate the photosensitizer in neovascular tissue, the CNV is stimulated with a specific wavelength of light to react with water and create oxygen and hydroxyl free radicals [33]. These free radicals, in turn, react with cell membranes of the pathologic endothelium to induce occlusion by massive platelet activation and thrombosis, while still preserving the normal choroidal vasculature and nonvascular tissue [34, 35]. Ideally, the intensity of the exciting wavelength is low enough to spare the non-neovascular irradiated tissues from thermal damage. Important variables in this reaction include the intravascular concentration of dye, the photochemical behavior of the dye, the interval between the injection and the onset of irradiation, the intensity and specificity of the exciting irradiation, and the duration of irradiation [36–38].

**Verteporfin** The U.S. FDA approved verteporfin (Visudyne; QLT Therapeutics, Inc., Vancouver, Canada, and Novartis Ophthalmics, Bulach, Switzerland) in April 2000 for patients with “predominantly classic” subfoveal CNV caused by AMD, which demonstrates a characteristic early and well-defined RSEA stain to over 50% of the CNV complex. Similarly, marketing approval was granted in Europe in July 2000, and it is currently commercially available in over 70 countries for predominantly classic CNV.

In 1999 and 2001, the 1- and 2-year results of the Treatment of AMD with PDT (TAP) study were published. TAP consisted of two randomized, prospective, double-blind, placebo-controlled phase III trials with 609 subjects. First-year data reported the proportion of eyes with less than 15 letters of visual acuity loss on a standardized eye chart was 67% in the treated group versus 39% in control group



( $p < 0.001$ ) when the CNV was predominantly classic; however, no significant differences in visual acuity were demonstrated when the area of classic CNV was less than 50% of the entire complex. In addition, it was noted that 90% of the subjects required retreatment at 3 months, and an average of more than three retreatments over the first year [39]. Second-year follow-up data reported that 59% of treated eyes had a favorable visual outcome versus 31% in the control group when the lesion was predominantly classic [40]. The TAP trial was unmasked at 2 years of follow-up. An open-label extension to 36 months of 124 of the 159 original TAP participants with predominantly classic CNV revealed that visual acuity remained nearly constant and required fewer retreatments [41]. Because of the success of the TAP trial, the verteporfin in photodynamic therapy (VIP) trial, another randomized, prospective, double-blind, placebo-controlled clinical trial, was developed to examine many of the patients who fell outside of the inclusion guidelines set by TAP.

The VIP trial was designed to evaluate the treatment efficacy of PDT in 339 subjects with total occult subfoveal CNV, classic CNV with a visual acuity better than 20/40, or CNV secondary to pathological myopia. One-year results of the occult AMD arm showed no significant difference between visual acuity outcomes in exudative AMD patients treated with verteporfin and placebo (51% PDT and 54% placebo treated had unfavorable visual outcomes, respectively). However, 2-year follow-up data revealed that 55% of the treated subjects with occult CNV had an unfavorable outcome versus 68% in the placebo group ( $p = 0.023$ ). On average, the verteporfin-treated patients received five treatments over 24 months of follow-up. Based on this data, the study group recommended verteporfin for purely occult subfoveal CNV that demonstrated recent disease progression in all patients except those with large lesions with good visual acuity [42]. Because the FDA desired additional data before approving verteporfin for occult CNV, the Visudyne in Occult (VIO) trial was developed as a 24-month study to analyze patients with only occult CNV. Enrollment of 364 subjects was completed and is currently in the second year of follow-up as per the recommendation of the Data and Safety Monitoring Committee [43].

Several other trials have evaluated the efficacy of verteporfin in a variety of clinical situations previously lacking sufficient data. Retrospective TAP and VIP data suggested some treatment benefit for smaller minimally classic lesions. The Visudyne in Minimally (VIM) classic trial was thus initiated as a randomized, prospective, double-blind, placebo-controlled clinical trial designed to study the use of verteporfin in patients with minimally classic CNV. Phase II data on 117 patients suggests that small, recently progressive, minimally classic CNV might benefit from verteporfin therapy [44]. Two-year follow-up data revealed that fewer verteporfin-treated eyes lost three or more lines of vision on a standard visual acuity chart or converted to a predominantly classic lesion as compared to placebo [45]. Consequently, a phase III study [the visudyne minimally classic (VMC) trial] was started in late 2003 to further evaluate verteporfin in minimally classic CNV. On April 1, 2004, the U.S. Centers for Medicare and Medicaid Services (CMS) agreed to reimburse physicians for PDT of occult and minimally classic subfoveal CNV (less than 50% of CNV complex with early well-defined hyperfluorescence on RSFA) from AMD, provided that the lesion is four disk areas or less in size at least three months prior to initial treatment and evidence of progression (i.e., loss of five or more letters



on standard visual acuity charts, increase of at least one disk diameter, or appearance of blood) within 3 months of treatment.

Because 80% of vision loss in verteporfin-treated patients occurs within 6 months of developing CNV, the Verteporfin Early Retreatment (VER) trial was designed as a phase III study of 323 patients to compare the benefit of retreatment in 6-week intervals versus the standard 3 months. Twelve-month interim results of the 2-year trial did not show improved outcomes when compared to the standard treatment [46]. Additionally, the Verteporfin with Altered (Delayed) Light in Occult (VALIO) study was developed to evaluate whether delaying the light application to 30 minutes after the initiation of verteporfin infusion (versus the standard 15 minutes) would improve outcomes in occult CNV. Phase II data at 6 months of follow-up show the group treated at 30 minutes postinfusion lost 1.3 lines of vision while the standard 15 minute postinfusion treatment group lost 2–3 lines, which was not statistically significant. One-year data substantiated the 6-month findings [47, 48].

At the moment verteporfin is the only approved PDT agent, but additional photosensitizing products are under study and development.

**Rostaporfin** Rostaporfin (Photrex; Miravant Medical Technologies, Santa Barbara, California) is a purpurin with a structure similar to chlorophyll that absorbs maximally at 664 nm. Like verteporfin, the preconstituted solution is intravenously infused over 10–20 minutes. In December 2001, enrollment for a phase III placebo-controlled, double-masked clinical trial involving 920 patients was completed. Two-year follow-up data found that 58% of patients receiving a 0.5 mg/kg dose of SnET2 lost less than 15 letters compared to 42% of placebo patients ( $p = 0.0045$ ). Rostaporfin was well tolerated and demonstrated an acceptable safety profile [49]. On September 30, 2004, the FDA requested an additional confirmatory clinical trial before final marketing approval.

#### 10.10.4.4 Radiation Therapy

Since choroidal neovascular membranes are composed of rapidly proliferating pathologic endothelial cells, they may be sensitive to agents that inhibit cell division. Consequently, radiation therapy has been suggested as a treatment for subfoveal CNV. Given an apparent dose–response effect, some groups have delivered ionizing radiation to the macula using modalities that may limit the exposure of ionizing radiation to normal radiosensitive structures of the eye, such as the optic nerve or lens. These methods have included stereotactic external photon beam irradiation of the posterior pole, brachytherapy, in which radioactive plaques are sutured to the posterior pole of the eye and explanted several days later, and proton beam irradiation, which deposits almost all of its energy at the desired depth in the eye at a point called the Bragg peak and undergoes little scattering [50, 51].

Although some of the early pilot studies suggested a possible benefit, conflicting reports regarding the efficacy of radiation therapy for exudative AMD have since been published. Two prospective controlled studies, using a relatively large number of low ionizing radiation fractions, failed to show a treatment benefit for external beam radiation [52, 53]. However, two smaller prospective controlled studies using a smaller number of higher radiation fractions demonstrated a statistically significant vision benefit over controls [54, 55]. Because of the positive outcomes, a

prospective, controlled, pilot study to evaluate external beam radiation on CNV in a small number of high-energy fractions was sponsored by the National Eye Institute. An interim analysis of this study, known as the AMD Radiotherapy Trial (AMDRT), found that at 12 months follow-up 43% of radiated eyes and 50% of nonradiated eyes demonstrated a moderate visual loss ( $p = 0.60$ ) [56].

#### 10.10.4.5 Surgical Therapy

Some vitreoretinal surgeons have attempted to remove CNV with direct surgical excision, which can yield impressive results in CNV secondary to histoplasmosis and multifocal choroiditis. However, the results were disappointing for exudative age-related macular degeneration. Researchers speculate the CNV of AMD has a different morphology and grows both anterior and posterior to the RPE. The damaged RPE that remains after CNV removal causes atrophy of the underlying choriocapillaris leading to retinal disorganization [57–60]. In 1998, the National Eye Institute of the National Institutes of Health awarded funding to the submacular surgery trial (SST). This study was designed as a randomized, multicenter, prospective clinical trial comparing surgery with observation to specifically evaluate patients with large or poorly demarcated new subfoveal CNV, submacular hemorrhage from CNV associated with exudative AMD, or subfoveal CNV due to presumed ocular histoplasmosis (POHS) or idiopathic causes.

Patients were followed for 2 years and assessed for stabilization or deterioration of visual acuity (VA), change in contrast sensitivity, cataract development, surgical complications, and quality of life. Of 454 patients with subfoveal choroidal neovascularization enrolled, 228 study eyes were assigned to observation and 226 to surgery. Median VA losses from baseline to the 24-month examination were 2.1 lines (10.5 letters) in the observation arm and 2.0 lines (10 letters) in the surgery arm. Median VA declined from 20/100 at baseline to 20/400 at 24 months in both arms. Moreover, rhegmatogenous retinal detachment occurred in 12 surgery eyes (5%) and 1 observation eye. In conclusion, it was determined that submacular surgery does not improve or preserve VA for 24 months better than observation, and it is therefore not recommended for patients with subfoveal choroidal neovascularization caused by AMD or with submacular hemorrhage from CNV [61, 62].

#### 10.10.4.6 Antiangiogenic Therapy

**Vascular Endothelial Growth Factor Inhibitors** Animal and clinical studies have identified vascular endothelial growth factor (VEGF) as a key mediator of ocular angiogenesis [63]. Upregulation of VEGF expression has been reported in experimentally induced CNV in rats, and it has also been shown that VEGF is capable of inducing intraretinal and subretinal neovascularization [64]. In human clinical trials, particular attention has focused on the development of pharmaceutical agents to block VEGF expression or neutralize it once expressed. Investigators have inhibited preretinal neovascularization in experimental models with antibodies against VEGF. Others have shown similar effects using VEGF-neutralizing chimeric proteins, which were constructed by joining the extracellular domain of high-affinity VEGF receptors with IgG [65].

**Pegaptanib Sodium** The anti-VEGF pegylated aptamer, pegaptanib sodium (Macugen; Eyetech Pharmaceuticals, Inc., New York, and Pfizer, Inc., New York), demonstrated both safety and efficacy in clinical trials. This intravitreally administered polyethylene glycol (PEG)-conjugated oligonucleotide was specifically designed to bind and neutralize VEGF<sub>165</sub>, hypothesized to be the predominant VEGF isomer associated with CNV in humans. A phase I trial, involving 15 subjects receiving a single injection of pegaptanib sodium, demonstrated 80% with stable or improved vision at 3 months. More impressively, 27% of the patients had significantly improved vision: A finding missing from many of the other standard AMD treatment modalities [66]. Although small, a phase II trial involving 21 patients supported the phase I data, and, when pegaptanib sodium injections were combined with PDT, 6 of 10 (60%) patients had significantly improved vision versus 2.2% treated with PDT alone [67].

The VEGF Inhibition Study in Ocular Neovascularization (VISION), two phase II/III, multicenter, randomized, placebo-controlled studies, completed enrollment of 1186 subjects in July 2002. The 2-year follow-up data revealed less vision loss for subjects maintained on pegaptanib sodium than those who only received the medication during one year ( $p < 0.05$ ). In the group given pegaptanib at 0.3 mg, 70% of patients lost fewer than 15 letters of visual acuity, as compared with 55% among the controls ( $p < 0.001$ ). The risk of severe loss of visual acuity (loss of 30 letters or more) was reduced from 22% in the sham-injection group to 10% in the group receiving 0.3 mg of pegaptanib ( $p < 0.001$ ). More patients receiving pegaptanib (0.3 mg), as compared with sham injection, maintained their visual acuity or gained acuity (33% vs. 23%;  $p = 0.003$ ). As early as 6 weeks after beginning therapy with the study drug, and at all subsequent points, the mean visual acuity among patients receiving 0.3 mg of pegaptanib was better than in those receiving sham injections ( $p < 0.002$ ). Among the adverse events that occurred, endophthalmitis (in 1.3% of patients), traumatic injury to the lens (in 0.7%), and retinal detachment (in 0.6%) were the most serious and required vigilance. These events were associated with a severe loss of visual acuity in 0.1% of patients [68]. Based on these results, the FDA approved the use of pegaptanib sodium to slow vision loss in people with neovascular AMD on December 20, 2004 [69].

**Ranibizumab** Ranibizumab (Lucentis, Genentech Inc., San Francisco, and Novartis Ophthalmics, Basel, Switzerland), an intravitreally injected, recombinant, humanized, monoclonal antibody Fab fragment designed to actively bind and inhibit all isoforms of VEGF, has shown promise in early human trials. A phase Ib/II randomized, single-agent study found that 94% of the 50 patients receiving ranibizumab had stable vision and 44% had significantly improved vision at 6 months [70]. Additional trials have since then been initiated to provide more definitive evaluation of the clinical benefit of ranibizumab in patients with predominantly classic or minimally classic/occult CNV.

The MARINA (Minimally Classic/Occult Trial of the Anti-VEGF Antibody Ranibizumab in the Treatment of Neovascular AMD) is a phase III randomized, prospective, double-blind, placebo-controlled trial initiated in 2003 with the objective of comparing ranibizumab against verteporfin for minimally classic or occult CNV. A total of 716 patients were enrolled in this study and randomized 1:1:1 to sham injection or to ranibizumab (0.3 or 0.5 mg) injected intravitreally monthly for

24 months. Preliminary analysis of one-year MARINA data revealed that approximately 95% of patients treated with ranibizumab lost fewer than 15 letters at one year, compared with approximately 62% in the control group ( $p < 0.0001$ ). On average, the patients treated with ranibizumab had a significant improvement in visual acuity relative to their visual acuity at study entry, whereas the control group experienced a substantial decrease from baseline in mean visual acuity [71].

The ANCHOR (Anti-VEGF Antibody for the Treatment of Predominantly Classic Choroidal Neovascularization in AMD) is a multicenter, prospective, randomized, multicenter, double-masked phase III trial designed to compare a combination ranibizumab/PDT therapy to verteporfin PDT alone in 423 subjects with predominantly classic exudative AMD. This trial is still ongoing in centers in United States, Europe, and Australia [71].

The FOCUS (RhuFab V2 Ocular Treatment Combining the Use of Visudyne to Evaluate Safety) study is a randomized, single-masked phase I/II trial investigating the safety, tolerability, and efficacy of ranibizumab in combination with verteporfin PDT versus verteporfin PDT alone in patients with subfoveal predominantly classic CNV due to AMD. Enrollment of 162 patients has been completed and a preliminary analysis of one-year data indicate that approximately 90% of patients treated with the combination of ranibizumab and PDT had stable or improved visual acuity, compared with approximately 68% of patients in the control arm of PDT alone ( $p = 0.0003$ ) [71].

The Phase IIb, multicenter, randomized, double-masked, sham Injection-controlled study of the Efficacy and safety of Ranibizumab (PIER) study started enrolling approximately 180 patients in September 2004 with the objective of comparing 3-month intravitreal dosing intervals to the standard 1-month intervals [72].

**Pigment Epithelium-Derived Factor Inducer** Researchers have attempted to stimulate intravitreal production of native pigment epithelium-derived factor (PEDF), a naturally occurring potent antiangiogenic protein deficient in eyes with CNV, using gene therapy [73]. PEDF inhibits angiogenesis by inducing apoptotic death of endothelial cells stimulated to form new vessels [74]. In a laser-induced CNV murine model, choroidal neovascularization was reduced after intravitreal PEDF was produced from an adenoviral vector [75]. One study demonstrated that increased intravitreal PEDF results in 85% inhibition of neovascularization in laser-induced CNV, transgenic VEGF, and retinopathy of prematurity models. GenVec, Inc. (Gaithersburg, Maryland) has developed a PEDF producing adenovirus vector, AdPEDF (pigment epithelium-derived factor on an adenovirus vector), and completed recruitment of 51 patients in five states for phase I human trials in August 2004. Interim 12-month results of 24 patients revealed no dose-limiting toxicities or related severe adverse events [76].

**Cyclo-Oxygenase-2 Inhibitor** It is likely that a whole cascade of soluble factors play a role in ocular angiogenesis and CNV development. Cyclooxygenase-2 (COX-2) is expressed in neovascular structures, especially human cancers. Thus, it has been proposed that a cyclooxygenase-2 inhibitor might control neovascularization associated with AMD. Orally administered celecoxib (Celebrex; Pfizer, Inc., New York), a COX-2 inhibitor, significantly reduced angiogenesis and prostaglandin production in basic fibroblast growth factor (bFGF)-induced neovascularization of

rat corneas [77]. The National Eye Institute is conducting a phase I/II safety and efficacy trial comparing the use of PDT and celecoxib to PDT alone. The double-masked, randomized, placebo-controlled prospective study completed enrollment of 60 participants.

**Squalamine** Squalamine (Genaera Co., Plymouth Meeting, PA), an antiangiogenic aminosterol originally found in the body tissues of the cancer-resistant dogfish shark, acts as an inhibitor of growth factor signaling, including VEGF, integrin expression, and cytoskeletal formation. Systemic intravenous administration has inhibited iris neovascularization in primate models, oxygen-induced retinopathy in murine models, and laser-induced CNV in a rat model [78–80].

Three phase II clinical trials are currently underway to evaluate its role in AMD treatment. The largest, MSI-1256F-209, is a 100-patient prospective, randomized, controlled trial evaluating the effects of 20 or 40 mg given intravenously every week for 4 weeks followed by maintenance every 4 weeks for 48 weeks followed by 12 months of observation for exudative AMD. The second trial, MSI-1256F-208, is a 45-patient prospective controlled trial evaluating the effects of 10, 20, or 40 mg of intravenous squalamine given initially in combination with verteporfin PDT and then alone for an additional 6 months followed by 12 months of observation for exudative AMD. Preliminary results from this trial showed that subjects treated with 40 mg of squalamine lactate and concomitant PDT gained an average of 0.4 letters in visual acuity compared to study entry, while those treated with PDT alone lost an average of 4.8 letters [81]. The last trial, MSI-1256F-207, is an 18-patient open-label, parallel group trial comparing three doses of intravenous squalamine given weekly for 4 weeks followed by 4 months of follow-up on exudative AMD. On October 4, 2004, the U.S. FDA granted fast-track designation to squalamine.

**Steroid Compounds** Corticosteroid compounds have long been known to possess angiostatic properties as they alter extracellular matrix degradation and inhibit inflammatory cells, which invariably participate in neovascular responses [82]. Intravitreal administration of corticosteroids has become very popular as the blood–ocular barrier is bypassed, more constant therapeutic steroid levels are achieved, and systemic side effects are minimized. These injections have demonstrated efficacy in subretinal and preretinal neovascularization in animal models [83, 84].

**Triamcinolone Acetonide** Uncontrolled pilot studies evaluating CNV in AMD have employed the off-label use of intravitreally administered triamcinolone acetonide (Kenalog; Bristol-Myers Squibb, New York) because of its long half-life and corticosteroid properties. One study of 30 eyes receiving a single triamcinolone acetonide injection reported that 11 eyes experienced improved or stabilized vision within 1–3 months of treatment with regression of the CNV to inactive fibrosis, 15 experienced a similar outcome, except for slow extension and exudation from recurrent CNV, while 4 experienced no obvious treatment benefit [85]. In later publications, studies reported a favorable effect on the course of the disease over 6-, 12-, and 18-month follow-up; however, the lack of controls complicated the ability to compare treatment efficacy to the natural course of the disease [86–88]. The authors proposed that intravitreal triamcinolone had a beneficial effect on AMD-related CNV through inhibition of leukocytes, including macrophages, which normally

release angiogenic factors [85–87]. A randomized, double-masked, placebo-controlled clinical trial of 151 eyes receiving a single 4-mg injection of intravitreal triamcinolone found significant antiangiogenic effects at 3 months after treatment; however, no beneficial visual acuity effect was seen at 1 year [89]. The authors speculated that triamcinolone might be efficacious at a higher or more sustained dose or in concert with other modalities.

Visagen (Regenera Limited, Nedlands, Australia) is developing a triamcinolone acetonide formulation that is being developed strictly for intraocular applications. Regenera Ltd. anticipates sponsoring clinical trials in an attempt to formally gain approval for several ophthalmic indications in the near future. A different group is developing a preservative-free formulation that theoretically decreases the 0.8% sterile endophthalmitis rate observed with traditional intravitreal triamcinolone acetonide injections [90].

**Anecortave Acetate** In 1985, a class of steroid with minimal glucocorticoid or mineralocorticoid activity was developed and is now undergoing evaluation in human trials as anecortave acetate (Retaane; Alcon Laboratories, Inc., Fort Worth, Texas) [91]. The lack of corticosteroid activity minimizes commonly encountered intraocular pressure elevation and accelerated cataract formation [92]. In addition, anecortave acetate was formulated for injection into the subtenon space with a specially designed cannula. A phase II/III randomized, prospective, placebo-controlled trial involving 128 patients designed to evaluate the clinical safety and efficacy of juxtascleral injection of anecortave acetate versus placebo for the treatment of subfoveal CNV, found that baseline vision ( $p = 0.01$ ), stabilization of vision ( $p = 0.03$ ), and prevention of severe vision loss ( $p = 0.02$ ) were statistically superior to baseline at 12 months; however, the dropout rate was nearly 50% [93].

Another phase III study designed to compare anecortave acetate 15-mg suspension to visudyne PDT in patients with exudative AMD was carried out over a 12-month period. The study enrolled 530 patients with the primary objective of demonstrating that anecortave acetate 15 mg is noninferior to PDT in patients with predominantly classic subfoveal CNV. It was shown that no difference existed between the two treatment groups ( $p = 0.4305$ ) [94].

A new study (C-02-60) evaluating anecortave acetate suspension versus a sham administration procedure for prevention of progression from dry AMD to exudative AMD is currently enrolling patients. The objective of this study is to determine the safety and efficacy of anecortave acetate suspension for treatment of patients with nonexudative AMD who are at risk of progression to exudative AMD, in an attempt to arrest the development of choroidal neovascularization. A total of 2500 patients are to be enrolled in this 4-year study [95].

**Implantable Corticosteroids** Because intraocular corticosteroids have shown antiangiogenic effects with repeated intravitreal administration, the efforts to develop sustained-release intraocular implants have received special attention, in an attempt to achieve near constant intraocular steroid concentrations without repeated injections. One study demonstrated CNV inhibition using triamcinolone acetate microimplants in a laser-induced CNV rat model [96]. Furthermore, researchers at Bausch & Lomb (Rochester, New York) and Control Delivery Systems (Watertown, MA) have developed Retisert (also known as Envision TD), a nonbiodegradable intra-



vitreal implant that releases fluocinolone acetonide for up to 3 years. Phase III studies involving diabetic macular edema were promising for resolution of retinal leakage compared to placebo. However, the study also reported that after one year 58.5% of subjects receiving the 0.5-mg implant developed serious side effects such as increased intraocular pressure, vitreal hemorrhage, and cataracts, complications that occurred in only 10.7% of the standard care group. The patients will be followed for an additional 3 years to monitor the safety of the implant. Enrollment is complete for a phase II study to evaluate the effects of Retisert on exudative AMD. However, development in this indication has been discontinued [97]. Similarly, a biodegradable dexamethasone implant (Posurdex; Allergan, Irvine, California) has shown safety and benefit in recent phase II trials for macular edema from diabetes mellitus, branch or central retinal vein occlusion, uveitis, or surgery. No trials are currently evaluating Posurdex on AMD.

Given the complexity of AMD and our limited understanding of its pathophysiology, the current design of therapeutic modalities relies largely on the evaluation of a large number of candidate treatments in order to define those who offer clinical benefit. Until our knowledge of physiological processes responsible for macular degeneration increases, treatment will likely depend on a combination of approaches to limit the underlying CNV. This combination approach is already largely reflected in the design of currently ongoing clinical trials.

## REFERENCES

1. Kahn, H. A., Leibowitz, H. M., Ganley, J. P., et al. (1977), The Framingham Eye Study. I. Outline and major prevalence findings, *Am. J. Epidemiol.*, 106, 17–32.
2. Attebo, K., Mitchell, P., and Smith, W. (1996), Visual acuity and the causes of visual loss in Australia. The Blue Mountains Eye Study, *Ophthalmology*, 103, 357–364.
3. Klaver, C. C., Wolfs, R. C., Vingerling, J. R., et al. (1998), Age-specific prevalence and causes of blindness and visual impairment in an older population: the Rotterdam Study, *Arch. Ophthalmol.*, 116, 653–658.
4. Seddon, J. M. (2001), Epidemiology of age-related macular degeneration, in Ryan, S. J., Ed, *Retina*, Mosby, St Louis, pp. 1039–1050.
5. Friedman, D. S., O'Colmain, B. J., Munoz, B., et al. (2004), Prevalence of age-related macular degeneration in the United States, *Arch. Ophthalmol.*, 122, 564–572.
6. Bird, A. C., Bressler, N. M., Bressler, S. B., et al. (1995), An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group, *Surv. Ophthalmol.*, 39, 367–374.
7. Macular Photocoagulation Study Group (1982), Argon laser photocoagulation for senile macular degeneration. Results of a randomized clinical trial, *Arch. Ophthalmol.*, 100, 912–918.
8. Age-Related Eye Disease Study Research Group (2001), A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8, *Arch. Ophthalmol.*, 119, 1417–1436.
9. Fliesler, S. J., and Anderson, R. E. (1983), Chemistry and metabolism of lipids in the vertebrate retina, *Prog. Lipid. Res.*, 22, 79–131.



10. Young, R. W. (1988), Solar radiation and age-related macular degeneration, *Surv. Ophthalmol.*, 32, 252–269.
11. Gerster, H. (1991), Review: Antioxidant protection of the ageing macula, *Age Ageing*, 20, 60–69.
12. Beatty, S., Koh, H., Phil, M., et al. (2000), The role of oxidative stress in the pathogenesis of age-related macular degeneration, *Surv. Ophthalmol.*, 45, 115–134.
13. Newsome, D. A., Swartz, M., Leone, N. C., et al. (1988), Oral zinc in macular degeneration, *Arch. Ophthalmol.*, 106, 192–198.
14. The Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group (1994), The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers, *N. Engl. J. Med.*, 330, 1029–1035.
15. Eagle, R. C. J. (1984), Mechanisms of maculopathy, *Ophthalmology*, 91, 613–625.
16. Young, R. W. (1987), Pathophysiology of age-related macular degeneration, *Surv. Ophthalmol.*, 31, 291–306.
17. Ho, A. C., Maguire, M. G., Yoken, J., et al. (1999), Laser-induced drusen reduction improves visual function at 1 year. Choroidal Neovascularization Prevention Trial Research Group, *Ophthalmology*, 106, 1367–1373.
18. Olk, R. J., Friberg, T. R., Stickney, K. L., et al. (1999), Therapeutic benefits of infrared (810-nm) diode laser macular grid photocoagulation in prophylactic treatment of nonexudative age-related macular degeneration: two-year results of a randomized pilot study, *Ophthalmology*, 106, 2082–2090.
19. National Eye Institute (2006), Complications of Age-Related Macular Degeneration Prevention Trial (CAPT); available from <http://www.nei.nih.gov/neitrials/viewStudyWeb.aspx?id=70>; accessed Jan. 14, 2006.
20. PTAMD Clinical Trial (2006); available from [http://www.irisindex.com/ophthalmology/ptamd\\_clinical\\_trial.html](http://www.irisindex.com/ophthalmology/ptamd_clinical_trial.html); accessed Jan. 15, 2006.
21. Friedman, E., Krupsky, S., Lane, A. M., et al. (1995), Ocular blood flow velocity in age-related macular degeneration, *Ophthalmology*, 102, 640–646.
22. Friedman, E. (1997), A hemodynamic model of the pathogenesis of age-related macular degeneration, *Am. J. Epidemiol.*, 124, 677–682.
23. Boyer, D., and Gallemore, R. (2006), Clinical trials assess rheophoresis; available from <http://www.mdsupport.org/library/rheotri2.html>; accessed Jan. 16, 2006.
24. Macular Photocoagulation Study Group (1986), Argon laser photocoagulation for neovascular maculopathy. Three-year results from randomized clinical trials. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 104, 503–512.
25. Macular Photocoagulation Study Group (1986), Recurrent choroidal neovascularization after argon laser photocoagulation for neovascular maculopathy. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 104, 503–512.
26. Macular Photocoagulation Study Group (1991), Argon laser photocoagulation for neovascular maculopathy. Five-year results from randomized clinical trials. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 109, 110–114.
27. Macular Photocoagulation Study Group (1991), Subfoveal neovascular lesions in age-related macular degeneration. Guidelines for evaluation and treatment in the macular photocoagulation study. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 109, 1242–1257.
28. Macular Photocoagulation Study Group (1993), Laser photocoagulation of subfoveal neovascular lesions of age-related macular degeneration. Updated findings from two clinical trials. Macular Photocoagulation Study Group, *Arch. Ophthalmol.*, 111, 1200–1209.

29. Ciulla, T. A., Harris, A., Kagemann, L., et al. (2001), Transpupillary thermotherapy for subfoveal occult choroidal neovascularization: effect on ocular perfusion, *Invest. Ophthalmol. Vis. Sci.*, 42, 3337–3340.
30. Algvere, P. V., Libert, C., Lindgarde, G., et al. (2003), Transpupillary thermotherapy of predominantly occult choroidal neovascularization in age-related macular degeneration with 12 months follow-up, *Acta. Ophthalmol. Scand.*, 81, 110–117.
31. Thach, A. B., Sipperley, J. O., Dugel, P. U., et al. (2003), Large-spot size transpupillary thermotherapy for the treatment of occult choroidal neovascularization associated with age-related macular degeneration, *Arch. Ophthalmol.*, 121, 817–820.
32. TTT4CNV Clinical Trial (2003), Iridex.
33. Aveline, B., Hasan, T., and Redmond, R. W. (1994), Photophysical and photosensitizing properties of benzoporphyrin derivative monoacid ring A (BPD-MA), *Photochem. Photobiol.*, 59, 328–335.
34. Allison, B. A., Waterfield, E., Richter, A. M., et al. (1991), The effects of plasma lipoproteins on in vitro tumor cell killing and in vivo tumor photosensitization with benzoporphyrin derivative, *Photochem. Photobiol.*, 54, 709–715.
35. Hunt, D. W., Jiang, H., Granville, D. J., et al. (1999), Consequences of the photodynamic treatment of resting and activated peripheral T lymphocytes, *Immunopharmacology*, 41, 31–44.
36. Reichel, E., Puliafito, C. A., Duker, J. S., et al. (1994), Indocyanine green dye-enhanced diode laser photocoagulation of poorly defined subfoveal choroidal neovascularization, *Ophthalmic. Surg.*, 25, 195–201.
37. Hope-Ross, M. W., Gibson, J. M., Chell, P. B., et al. (1994), Dye enhanced laser photocoagulation in the treatment of a peripapillary subretinal neovascular membrane, *Acta. Ophthalmol.*, (Copenh) 72, 134–137.
38. Moriarty, A. P. (1994), Indocyanine green enhanced diode laser photocoagulation of subretinal neovascular membranes, *Br. J. Ophthalmol.*, 78, 238–239.
39. Treatment of age-related macular degeneration with photodynamic therapy (TAP) Study Group (1999), Photodynamic therapy of subfoveal choroidal neovascularization in age-related macular degeneration with verteporfin: one-year results of 2 randomized clinical trials—TAP report, *Arch. Ophthalmol.*, 117, 101329–1345.
40. Bressler, N. M. (2001), Treatment of age-related macular degeneration with photodynamic therapy (TAP) Study Group. Photodynamic therapy of subfoveal choroidal neovascularization in age-related macular degeneration with verteporfin: two-year results of 2 randomized clinical trials—tap report 2, *Arch. Ophthalmol.*, 119, 198–207.
41. Blumenkranz, M. S., Bressler, N. M., Bressler, S. B., et al. (2002), Verteporfin therapy for subfoveal choroidal neovascularization in age-related macular degeneration. three-year results of an open-label extension of 2 randomized clinical trials—TAP Report no. 5, *Arch. Ophthalmol.*, 120, 1307–1314.
42. Verteporfin In Photodynamic Therapy Study Group (2001), Verteporfin therapy of subfoveal choroidal neovascularization in age-related macular degeneration: two-year results of a randomized clinical trial including lesions with occult with no classic choroidal neovascularization—verteporfin in photodynamic therapy report 2, *Am. J. Ophthalmol.*, 131, 541–560.
43. QLT Inc. Occult AMD (2006); available at <http://www.qltinc.com/Qtinc/main/mainpages.cfm?InternetPageID=143>; accessed Jan. 16, 2006.
44. Bressler, N., Rosenfeld, P., and Lim, J. (2003), VIM Study Group: A phase II placebo-controlled, double-masked, randomized trial—verteporfin in minimally classic CNV due to AMD (VIM), *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 1100.

45. Azab, M., Boyer, D. S., Bressler, N. M., et al. (2005), Verteporfin therapy of subfoveal minimally classic choroidal neovascularization in age-related macular degeneration: 2-year results of a randomized clinical trial, *Arch. Ophthalmol.*, 123, 448–457.
46. Stur, M. (2004), VER Study Group: Verteporfin Early Retreatment (VER)—12-month results of a phase IIIB controlled clinical trial, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2275.
47. Slakter, J., and Rosenfeld, P. (2003), VALIO Study Group: Verteporfin with altered (delayed) light in occult CNV (VALIO)—Results of a phase II controlled clinical trial, *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 1101.
48. Singerman, L., and Rosenfeld, P. (2004), VALIO Study Group: Verteporfin with altered (delayed) light in occult (VALIO)—12-month results of a phase II controlled clinical trial, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2274.
49. Thomas, E. (2004), SnET2 Study Group: SnET2 photodynamic therapy for age-related macular degeneration. visual acuity efficacy outcomes from two parallel phase III trials, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2214.
50. Yonemoto, L. T., Slater, J. D., Friedrichsen, E. J., et al. (1996), Phase I/II study of proton beam irradiation for the treatment of subfoveal choroidal neovascularization in age-related macular degeneration: treatment techniques and preliminary results, *Int. J. Radiat. Oncol. Biol. Phys.*, 36, 867–871.
51. Finger, P. T., Berson, A., Ng, T., et al. (1999), Ophthalmic plaque radiotherapy for age-related macular degeneration associated with subretinal neovascularization, *Am. J. Ophthalmol.*, 127, 170–177.
52. Radiation Therapy for Age-related Macular Degeneration (RAD) Study Group (1999), A prospective, randomized, double-masked trial on radiation therapy for neovascular age-related macular degeneration (RAD Study), *Ophthalmology*, 106, 2239–2247.
53. Marcus, D. M., Sheils, W. C., Young, J. O., et al. (2004), Radiotherapy for recurrent choroidal neovascularisation complicating age related macular degeneration, *Br. J. Ophthalmol.*, 88, 114–119.
54. Bergink, G. J., Hoyng, C. B., van der Maazen, R. W., et al. (1998), A randomized controlled clinical trial on the efficacy of radiation therapy in the control of subfoveal choroidal neovascularization in age-related macular degeneration: radiation versus observation, *Graefes Arch. Clin. Exp. Ophthalmol.*, 236, 321–325.
55. Char, D. H., Irvine, A. I., Posner, M. D., et al. (1999), Randomized trial of radiation for age-related macular degeneration, *Am. J. Epidemiol.*, 127, 574–578.
56. Marcus, D., Peskin, E., Alexander, J., et al. (2003), The age-related macular degeneration radiotherapy trial (AMDRT). 1 year results, *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 3158.
57. Lambert, H. M., Capone, A. J., Aaberg, T. M., et al. (1992), Surgical excision of subfoveal neovascular membranes in age-related macular degeneration, *Am. J. Ophthalmol.*, 113, 257–262.
58. Gass, J. (1994), Biomicroscopic and histopathologic considerations regarding the feasibility of surgical excision of subfoveal neovascular membranes, *Am. J. Ophthalmol.*, 118, 285–298.
59. Ormerod, L. D., Puklin, J. E., and Frank, R. N. (1994), Long-term outcomes after the surgical removal of advanced subfoveal neovascular membranes in age-related macular degeneration, *Ophthalmology*, 101, 1201–1210.
60. Thomas, M. A., Dickinson, J. D., Melberg, N. S., et al. (1994), Visual results after surgical removal of subfoveal choroidal neovascular membranes, *Ophthalmology*, 101, 1384–1396.

61. Bressler, N. M., Bressler, S. B., Childs, A. L., et al. (2004), Surgery for hemorrhagic choroidal neovascular lesions of age-related macular degeneration: ophthalmic findings. SST report no. 13, *Ophthalmology*, 111, 1993–2006.
62. Hawkins, B. S., Bressler, N. M., Miskala, P. H., et al. (2004), Surgery for subfoveal choroidal neovascularization in age-related macular degeneration: ophthalmic findings. SST report no. 11, *Ophthalmology*, 111, 1967–1980.
63. Yi, X., Ogata, N., Komada, M., et al. (1997), Vascular endothelial growth factor expression in choroidal neovascularization in rats, *Graefes Arch. Clin. Exp. Ophthalmol.*, 235, 313–319.
64. Okamoto, N., Tobe, T., Hackett, S. F., et al. (1997), Transgenic mice with increased expression of vascular endothelial growth factor in the retina: a new model of intraretinal and subretinal neovascularization, *Am. J. Pathol.*, 151, 281–291.
65. Aiello, L. P., Pierce, E. A., Foley, E. D., et al. (1995), Suppression of retinal neovascularization in vivo by inhibition of vascular endothelial growth factor (VEGF) using soluble VEGF-receptor chimeric proteins, *Proc. Natl. Acad. Sci. USA*, 92, 10457–10461.
66. Eyetech Study Group (2002), Preclinical and phase 1A clinical evaluation of an anti-VEGF pegylated aptamer (EYE001) for the treatment of exudative age-related macular degeneration, *Retina*, 22, 143–152.
67. Eyetech Study Group (2003), Anti-vascular endothelial growth factor therapy for subfoveal choroidal neovascularization secondary to age-related macular degeneration: phase II study results, *Ophthalmology*, 110, 979–986.
68. Gragoudas, E. S., Adamis, A. P., Cunningham, E. T. J., et al. (2004), VEGF Inhibition Study in Ocular Neovascularization Clinical Trial Group: Pegaptanib for neovascular age-related macular degeneration, *N. Engl. J. Med.*, 351, 2805–2816.
69. FDA (2006), FDA Approves New Drug Treatment for Age-Related Macular Degeneration, *FDA news*; available at <http://www.fda.gov/bbs/topics/news/2004/new01146.html>; accessed Jan. 16, 2006.
70. Heier, J., Sy, J., and McCluskey, E. (2003), RhuFab V2 Study Group: RhuFab V2 in wet AMD—6 month continued improvement following multiple intravitreal injections, *Invest. Ophthalmol. Vis. Sci.*, 44, E-abstract 972.
71. Heier, J. S. (2005), Lucentis update in *American Academy of Ophthalmology Annual Meeting 2004, Retina Specialty Day Supplement*, Chicago.
72. Heier, J. S. (2004), Anti-VEGF: Genetech Ranibizumab in *American Academy of Ophthalmology Annual Meeting 2004, Retina Specialty Day*, New Orleans.
73. Takita, H., Yoneya, S., Gehlbach, P. L., et al. (2003), Retinal neuroprotection against ischemic injury mediated by intraocular gene transfer of pigment epithelium-derived factor, *Invest. Ophthalmol. Vis. Sci.*, 44, 4497–4504.
74. Stellmach, V., Crawford, S. E., Zhou, W., et al. (2001), Prevention of ischemia-induced retinopathy by the natural ocular antiangiogenic agent pigment epithelium-derived factor, *Proc. Natl. Acad. Sci. USA*, 98, 2593–2597.
75. Mori, K., Duh, E., Gehlbach, P., et al. (2001), Pigment epithelium-derived factor inhibits retinal and choroidal neovascularization, *J. Cell. Physiol.* 188, 253–263.
76. Campochiaro, P., Klein, M., Holtz, E., et al. (2004), AdPEDF therapy for subfoveal choroidal neovascularization (CNV): preliminary phase I results, *Invest. Ophthalmol. Vis. Sci.*, 45, E-abstract 2361.
77. Leahy, K. M., Ornberg, R. L., Wang, Y., et al. (2002), Cyclooxygenase-2 inhibition by celecoxib reduces proliferation and induces apoptosis in angiogenic endothelial cells in vivo, *Cancer Res.*, 62, 625–631.

78. Genaidy, M., Kazi, A. A., Peyman, G. A., et al. (2002), Effect of squalamine on iris neovascularization in monkeys, *Retina*, 22, 772–778.
79. Higgins, R. D., Sanders, R. J., Yan, Y., et al. (2000), Squalamine improves retinal neovascularization, *Invest. Ophthalmol. Vis. Sci.*, 41, 1507–1512.
80. Ciulla, T. A., Criswell, M. H., Danis, R. P., et al. (2003), Squalamine lactate reduces choroidal neovascularization in a laser-injury model in the rat, *Retina*, 23, 808–814.
81. AAO (2006), Genaera Presents Positive Preliminary Clinical Results for EVIZON for Treatment of Age-Related Macular Degeneration at the Annual AAO Meeting; available at <http://www.genaera.com/pressreleases/October%2019,%202005.pdf>; accessed 2006 Jan. 16, 2006.
82. Folkman, J., and Ingber, D. E. (1987), Angiostatic steroids. Method of discovery and mechanism of action, *Ann. Surg.*, 206, 374–383.
83. Ishibashi, T., Miki, K., Sorgente, N., et al. (1985), Effects of intravitreal administration of steroids on experimental subretinal neovascularization in the subhuman primate, *Arch. Ophthalmol.*, 103, 708–711.
84. Ciulla, T. A., Criswell, M. H., Danis, R. P., et al. (2001), Intravitreal triamcinolone acetonide inhibits choroidal neovascularization in a laser-treated rat model, *Arch. Ophthalmol.*, 119, 399–404.
85. Penfold, P. L., Gyory, J. F., Hunyor, A. B., et al. (1995), Exudative macular degeneration and intravitreal triamcinolone. A pilot study, *Aust. N. Z. J. Ophthalmol.*, 23, 293–298.
86. Challa, J. K., Gillies, M. C., Penfold, P. L., et al. (1998), Exudative macular degeneration and intravitreal triamcinolone: 18 month follow up, *Aust. N. Z. J. Ophthalmol.*, 26, 277–281.
87. Danis, R. P., Ciulla, T. A., Pratt, L. M., et al. (2000), Intravitreal triamcinolone acetate in exudative age-related macular degeneration, *Retina*, 20, 244–250.
88. Ranson, N. T., Danis, R. P., Ciulla, T. A., et al. (2002), Intravitreal triamcinolone in subfoveal recurrence of choroidal neovascularisation after laser treatment in macular degeneration, *Br. J. Ophthalmol.*, 86, 527–529.
89. Gillies, M. C., Simpson, J. M., Luo, W., et al. (2003), A randomized clinical trial of a single dose of intravitreal triamcinolone acetate for neovascular age-related macular degeneration: one-year results, *Arch. Ophthalmol.*, 121, 667–673.
90. Heriot, W. (2004), Corticosteroids for AMD in *American Academy of Ophthalmology Annual Meeting 2004. Retina Subspecialty Day*, New Orleans.
91. Crum, R., Szabo, S., and Folkman, J. (1985), A new class of steroids inhibits angiogenesis in the presence of heparin or a heparin fragment, *Science*, 230, 1375–1378.
92. Clark, A. F. (1997), AL-3789: a novel ophthalmic angiostatic steroid, *Expert Opin. Investig. Drugs*, 6, 1867–1877.
93. D'Amico, D. J., Goldberg, M. F., Hudson, H., et al. (2003), Anecortave acetate as monotherapy for treatment of subfoveal neovascularization in age-related macular degeneration: twelve-month clinical outcomes, *Ophthalmology*, 110, 2372–2383.
94. Slakter, J. S., Bochow, T. W., D'Amico, D. J., et al. (2006), Anecortave acetate (15 milligrams) versus photodynamic therapy for treatment of subfoveal neovascularization in age-related macular degeneration, *Ophthalmology*, 113, 3–13.
95. Slakter, J. S. (2005), Retaane Update in *American Academy of Ophthalmology Annual Meeting 2005. Retina Specialty Day*, Chicago.
96. Ciulla, T. A., Criswell, M. H., Danis, R. P., et al. (2003), Choroidal neovascular membrane inhibition in a laser treated rat model with intraocular sustained release triamcinolone acetate microimplants, *Br. J. Ophthalmol.*, 87, 1032–1037.
97. Anon. (2005), Fluocinolone acetate ophthalmic—Bausch & Lomb: fluocinolone acetate Envision TD implant, *Drugs R. D.*, 6, 116–119.



# 10.11

---

## *Paediatrics*

**Anne Cusick,<sup>1</sup> Natasha Lannin,<sup>2</sup> and Iona Novak<sup>3</sup>**

<sup>1</sup>*School of Biomedical and Health Sciences, University of Western Sydney, Sydney, Australia*

<sup>2</sup>*Rehabilitation Research Studies Unit, Faculty of Medicine, University of Sydney,  
Sydney, Australia*

<sup>3</sup>*Cerebral Palsy Institute, Sydney, Australia*

### **Contents**

|           |                                                 |     |
|-----------|-------------------------------------------------|-----|
| 10.11.1   | Introduction                                    | 628 |
| 10.11.2   | Definitions                                     | 628 |
| 10.11.2.1 | Paediatric Population                           | 628 |
| 10.11.2.2 | Off-Label Use                                   | 628 |
| 10.11.2.3 | Therapeutic Orphan                              | 629 |
| 10.11.2.4 | Paediatric Investigation Plan                   | 629 |
| 10.11.2.5 | Minimal Risk                                    | 629 |
| 10.11.3   | Overview of Unique Aspects of Paediatric Trials | 629 |
| 10.11.4   | Conventions                                     | 632 |
| 10.11.5   | Benchmark Regulations                           | 634 |
| 10.11.6   | Recommendations and Guidelines                  | 636 |
| 10.11.7   | Institutional Review Boards                     | 638 |
| 10.11.8   | Trial Questions                                 | 639 |
| 10.11.9   | Participant Characteristics                     | 643 |
| 10.11.10  | Paediatric Investigation Plans                  | 645 |
| 10.11.11  | Assent and Consent                              | 649 |
| 10.11.12  | Safety and Monitoring                           | 652 |
| 10.11.13  | Conclusion                                      | 653 |
|           | Appendix: Facts Summary                         | 654 |
|           | References                                      | 654 |



### 10.11.1 INTRODUCTION

This chapter introduces then explores in detail, issues and practical aspects of the context and conduct of paediatric clinical trials. Investigators need to be alert to the unique obligations that come into play when study participants are infants, children or young people. These obligations are ethical, procedural, legal and social. Investigators also need to be aware of the multiplicity of interests that operate in paediatric research so that projects can be successfully managed to completion. Ideally, realistic and reasonable paediatric trials will cause minimal harm and will meet expectations of parents, caregivers, investigators, sponsors and the public for meaningful and clinically useful outcomes.

This chapter presents the scope of challenges and opportunities in the paediatric speciality, to help support clinical trials that are child-centred, worthwhile and rigorous. Paediatric research is an area characterised by local procedural variation and rapid change in regulation, case-law, scientific evidence, scholarly opinion, public concern and professional guidelines. Consequently, this chapter provides common signposts, rather than prescriptive maps, that can be set in place by investigators as they lead the way in particular trial journeys. By the end of this chapter, investigators should have a breadth of view on issues involved in paediatric trials to enable them to apply technical information to the context of paediatrics. Investigators should also be able to reflect on their own standpoint and responsibility as trial leaders, sponsors or team members who are knowingly putting infants, children or young people at some level of risk in order to answer a question. Investigators should be able to seek out specialized paediatric trial sources after reading this chapter, having first gained a broad understanding of issues that must be proactively managed for successful trial completion.

### 10.11.2 DEFINITIONS

#### 10.11.2.1 Paediatric Population

The United Nations Convention on the rights of the child [1] defines a minor as anyone under the age of 18. While there are local variations to the legal age of adulthood, and while the clinical responsibility of the paediatric speciality may extend to 21 [2], the UN Convention should normally be observed for the purposes of trial research. For this chapter, the previable or viable foetus is not included however these are recognised as responsibilities of paediatrics [2] and specialist research sources should be consulted commencing with policy and guideline statements of relevant bodies for example, the American Academy of Pediatrics [3].

#### 10.11.2.2 Off-Label Use

Many drugs used in paediatrics are off-label and one reason underpinning the need for clinical trials is to reduce this practice:

“new drugs and biologicals [need to] include adequate pediatric labeling for the claimed indications at the time of, or soon after, approval. However, because such labelling may

not immediately be available, off-label use (or use that is not included in the approved label) of therapeutic agents is likely to remain common in the practice of pediatrics ... The purpose of off-label use is to benefit the individual patient. Practitioners may use their professional judgment to determine these uses ... The off-label use of a drug should be based on sound scientific evidence, expert medical judgement, or published literature” [4, p. 181].

### 10.11.2.3 Therapeutic Orphan

Drugs which are not approved by the Food and Drug Administration [USA] as safe and effective in children are prescribed daily. This is due in part to the fact that many drugs released since 1962 carry an “orphaning clause” in the package insert such as, “not to be used in children, since clinical studies have been insufficient to establish recommendations for its use ... Is the physician breaking the [USA] law when he prescribes drugs ... which carry the ‘orphaning clause’?. No, he is not. The physician may exercise his professional judgement in the use of any drug. However, if he deviates from the instructions in the package insert and adverse reactions occur, he must be prepared to defend himself in court if there is a malpractice suit” [5, p. 811].

### 10.11.2.4 Paediatric Investigation Plan

This is the term used in the European Union by the European Medicines Agency (EMA).

[It is] a development plan aimed at ensuring that the necessary data are obtained through studies in children, when it is safe to do so, to support the authorisation of the medicine for children. ... The paediatric investigation plan includes a description of the studies and of the measures to adapt the way the medicine is presented (formulation) to make its use more acceptable in children ... The plan should cover the needs of all age groups of children, from birth to adolescence. The plan also defines the timing of studies in children compared to adults. In some cases, studies will be deferred until after the studies in adults have been conducted, to ensure that research with children is done only when it is safe and ethical to do so [6].

### 10.11.2.5 Minimal Risk

Minimal (the least possible) risk describes procedures such as questioning, observing and measuring children, provided that procedures are carried out in a sensitive way, respecting the child’s autonomy and that consent has been given ... It is expected that research of minimal risk would not result in more than a very slight and temporary negative impact on the health of the person concerned [7, p. 15].

## 10.11.3 OVERVIEW OF UNIQUE ASPECTS OF PEDIATRIC TRIALS

This section overviews key issues that make paediatric clinical trials unique. Paediatric clinical trials seek answers about intervention effectiveness and safety—there is nothing unique about this—but to get answers, these trials are unique in requiring infants, children and youth as participants. The sample characteristics and needs thus dominate trial decisions. Paediatric populations present design challenges for trial

investigators. As participants they have inherent and continuous change in body structures, function, activities and participation, there is constant change in their social relations, exposure to physical environments, family and community influences—this continuous change is evident even when they are in good health. The scale and variability of change increases if there is illness, disease, disability or injury. Continuous underlying change in participants must therefore be anticipated in trial question, design and protocol decisions.

Paediatric populations also bring investigator responsibilities and accountabilities that extend well beyond those normally encountered in trials with adults. Young participants are inherently *vulnerable* and those who have illness or disease appear particularly exposed to the possibility of suffering. Adults make trial decisions on behalf of youngsters and act towards them in ways that may help or harm both their daily life and their life chances. Their vulnerability presents dilemmas for all involved in clinical trials. For adults to purposefully involve youngsters in studies with potential or known risk seems incompatible with obligations to protect and nurture them. For institutions, like hospitals, established to care and help, knowingly exposing youngsters to risk seems a betrayal of civic duty. These moral contradictions are a necessary and inevitable part of paediatric trials. They underpin the heightened emotion and public interest that accompanies paediatric trials. But without sound paediatric clinical trials greater harms may be perpetrated as parents, carers, medical and health personnel use clinical interventions on children at large without adequate or sometimes any scientific evidence to inform their decisions.

Clinical trials must be conducted, but their planning and accountabilities must also anticipate and accommodate paediatric participant vulnerability. Investigators who are aware of this *moral context* will ensure that questions are not only worthwhile, but also that protocols are explicitly child-centred, of rigorous design, and well conducted. It is the heightened moral context of paediatric clinical trials that is unique as infants, children and young people, not investigators, parents or sponsors, must live with the trial experience and the short and long term consequences of participation. Trial investigators who are aware of the moral context will also anticipate potential public interest and consider in advance how the lay-person might construe issues such as recruitment, design, funding, sponsorship, personnel and study procedure. Paediatric trial project planning thus needs to include strategies for public communication, public liaison and accountability trails.

The inherent vulnerability also means that a paediatric trial is one where child participants are the focus of many interested parties. *Stakeholders* include parents and guardians, medical, legal, health, and policy personnel who act as gatekeepers to and watchdogs of paediatric population participation. The media, pharmaceutical companies, organised advocacy and lobby groups play vigilant and vigorous roles in the initiation, public presentation and interpretation of trials. Stakeholder authority and influence can make the planning and conduct of paediatric trials a complex and delicate campaign of personal, professional, industry and community politics—in addition to the usual demands of a complicated scientific project. Proactive communication with stakeholders is needed regarding the question, study rationale, study conduct, regulation adherence, avoidance of conflict of interest, transparency and accountability of processes and records, and importantly the need for trials to

prevent current and future suffering caused by the use of interventions that have no scientific support.

Paediatric trials necessitate a careful balance of common sense, technical precision, and adherence to highly prescriptive regulation. Given the sensitivity of research with children, successfully engaging in paediatric trials is as much about identifying a *moral purpose*, communicating trial values and visions, ensuring justice for participants, and scanning the strategic context for threats and opportunities as it is about the study question, procedural diligence, protocol development and project management.

These “soft” project factors can publicly or professionally derail an otherwise well intentioned and well constructed trial if not proactively managed. They can destroy an investigator’s reputation even when no malicious intent or negligence was involved. The “front page of the paper” by-line may be all a public needs to discredit a researcher, damage institutional confidence and undermine a much-needed program of paediatric health research. Careful consideration of not only the scientific merit of the question, design and regulatory obligations, but also of the moral *politics* of paediatric trials is thus required. For some investigators, the interests and authority of stakeholders and gatekeepers seem like obstacles to research, however any experienced trial researcher knows these factors are just part of the “package” that is a paediatric trial.

Paediatric practice is also inherently *multidisciplinary* and the design of clinical trials must be robust enough to anticipate an array of potentially confounding factors in a child’s life emanating from health services, schools, community and family, including parental use of off-the-shelf medications, complementary medicines and alternative therapies. A clinical trial that is child-focused rather than variable focused will take into account the multiplicity of influences that may affect body structures, ability to participate in protocol requirements, recruitment, retention, confounding factors, and the short and long term consequences of trial participation.

Finally, participation, protocol adherence, study retention and the safety of young participants ultimately relies on *parent and caregiver* expertise and their understanding of and commitment to trials. Engaging parents and caregivers as protocol partners is critical to study success. Providing appropriate learning opportunities for parents to understand not only the study purpose and participants demands but also children’s rights to assent or decline participation are needed. This understanding goes to the heart of informed consent. Consideration of parent and caregiver perspectives in protocol design is essential, particularly in relation to the logistic and time demands on parents for intervention adherence and presentation of the child for outcome data measure collection. Paediatric clinical trials work better if the protocol can be reasonably integrated into the daily life of families as part of a sustainable routine.

This section of the chapter has provided paediatric investigators with an orientation to unique issues in paediatric trials. The following sections explore issues in more depth, however the following *caveats* apply. This chapter is introductory. Specialized paediatric sources, such as the Helms and Stonier (Eds) *Pediatric Clinical Research Manual* [8] which is regularly updated and has sub-speciality supplements, or the many speciality research related web-sites of regulatory bodies or professional societies should also be consulted. Research, scholarship policy and

commentary associated with paediatric research changes quickly. Investigators must therefore ensure they inform themselves about local requirements, contemporary conventions, up-to-date regulation, current public concerns, scholarly opinion and relevant scientific evidence available at the time of study commencement. A matter of months can dramatically change the paediatric research context as case law, public interest, media “frenzies”, local administration or new evidence can change what could be reasonably expected in an investigation plan. Particular attention should also be given to the most recent scientific evidence emerging from the paediatric sub-specialty under study, be that oncology, cerebral palsy, infectious disease or whatever, including evidence relating to effective sub-speciality trial methodologies. Different sub-specialities bring unique challenges relating to measurement, recruitment, consent and retention. Successful sub-speciality precedent studies can also help inform trial plan decisions. But caution! Precedent studies may have been conducted in different regulatory and social contexts making their methodologies unsuitable for contemporary replication even though the scientific findings may be rigorous and relevant.

The remainder of the chapter begins with the policy context. Policies have been developed to protect vulnerable infants, children and young people in research and to promote ethical practice. Conventions, regulations, recommendations, guidelines, and institutional review boards are introduced. The chapter continues with an exploration of issues that are particularly challenging in paediatrics: trial questions, participant characteristics, investigation plans, consent, assent, safety and monitoring.

#### 10.11.4 CONVENTIONS

Regulatory requirements compel investigators, sponsors and research partners to scope and conduct research in certain ways. Investigators and trial partners must comply with regulations or face consequences that may include prosecution and penalties ranging from public “naming and shaming” to fines or in cases of criminal conduct imprisonment. The existence of regulations means that clinical research can not only be weak or rigorous, it can also be lawful or unlawful. Investigators and members of institutional ethical review boards therefore have a duty to keep up-to-date with regulatory requirements as ignorance is normally no excuse. Understanding legal obligations, particularly those for paediatric populations, is as important as understanding scientific methods and the needs of youngsters. Investigators cannot work on the basis of what has been done in past practice or research, nor can they apply the same level of autonomous discretion they may use in their clinical life, as research-related regulations change over time, the standards for research decision-making are more prescribed and tests of due diligence, fair hearing and procedural fairness in research may be tighter.

A good place to start in understanding the regulatory context of paediatric research is with landmark conventions. These state agreed international positions on matters of importance that relate to the human condition. The United Nations (UN) General Assembly “Convention on the Rights of the Child” 1989 is probably one of the most important foundations for the paediatric speciality [1]. Although

not all countries have ratified this convention, the influence of the convention on national standards is enormous. In summary, the convention identifies that: human rights apply to children without exception; the child's best interests are the primary consideration and highest priority; children have a right to the highest attainable level of health; and they have a right to information and respect of their opinion [1]. While clinical trial plans would almost never cite the Conventions on the Rights of the Child as a methodological source and there is no mechanism to directly register trials as convention compliant with the UN, it is principles in this convention that local regulations around the world usually aim to embed and enforce.

Practically, investigators can use this convention to reflect on whether or not their study question or plan is "just". Investigators have a duty to act justly towards participants [9] and tests of justice may go beyond local regulatory requirements. The convention can provide some insight into what might reasonably be considered just. If one accepts the principle that children everywhere in any society should have human rights, then just study protocols should seek to preserve and protect those rights. Just protocols will thus include strategies to inform, listen to the opinion of and seek assent of child participants, even when parental permission has already been granted. The principle of "best interests of children", if accepted, also means that the best interests of child participants and children in general are high priorities in study decisions. In some way children must benefit, whether that is directly through trial participation or indirectly through study outcomes that enhance the well-being of children in general. Finally, the just trial will support the principle that children have a right to the highest attainable health, particularly in relation to weighing up trial benefits and risks to individual children and children in general.

Other international agreements relating to research may also apply to paediatric investigation. The most notable is the Declaration of Helsinki [10]. Clause 25 specifically relates to child involvement in medical research and it focuses on consent and assent. The United Nations Convention on the Rights of Persons with Disabilities [11], and the Declaration on the Rights of Indigenous Peoples [12] may also be relevant for studies that have targeted or incidental recruitment of youngsters from these groups. Both these conventions are relatively new, and not all countries are signatories, but again they provide a benchmark for investigators to consider whether or not the study question and plan is "just".

Conventions are thus an important and a useful background to investigator development of an ethical standpoint towards young participants. They help identify what might be considered just treatment. But most investigators do not use them. Instead they follow local regulations and procedures that codify ethical requirements. Local regulations may or may not be adequate for the moral context of paediatric clinical trials. Here is where the utility of "benchmark" regulations comes into play for investigators around the world. Although they may not apply locally, benchmark regulations provide guidelines and procedures that usually embed principles from conventions or declarations in their construction. Benchmark regulations can thus act as a guide, along with local requirements, for investigators to consider what a reasonable person would expect in a just paediatric study. The benchmark regulations of most influence are now explored. Both relate to medicines for children however the principles and processes are useful to inform research-

ers who work with other clinical interventions as they highlight the need for clear standards, accountabilities and procedural precision.

### 10.11.5 BENCHMARK REGULATIONS

One of the most significant developments in Twenty-first century paediatric clinical research has been the release of regulations in the European Union (EU) and the United States of America (U.S.). While other countries have local standards and regulations that must be consulted by investigators and adhered to in plan development and reporting, the sheer scale of the EU and U.S. regulations impact on numbers of trials and participants makes them global benchmarks that can inform and guide clinical trial decisions anywhere. This section first overviews where regulations and related sources are located, as any paediatric investigator will need to continually update and check rulings and applications of regulations. Then the EU and USA regulations themselves will be introduced, and examples of regulations from other jurisdictions that may be helpful will be provided.

The regulations and support material of the EU are easily located. A web-search using any popular search engine and the general term “paediatric clinical trial” will reveal links to the European Agency for the Evaluation of Medicinal Products (also known as European Medicines Agency, EMEA) (<http://www.emea.eu>) in addition to independent sites that hold related articles, opinion pieces, conferences and training announcements on the general topic of paediatric clinical trials and often the EMEA initiative specifically. These related items can be EMEA sponsored, or independent and they often maintain a lively watch on the EMEA initiative from an investigator and industry perspective. Sites such as these, together with subspecialty resources in particular fields available through scholarly journals, professional societies and consumer groups, help investigators understand the motivations, agendas, obligations and impacts of the EU regulation from a broader perspective. The latter can help inform the strategic decisions that need to be made by trial leaders in project planning and management. From the home page of the EMEA Medicines for Children a wealth of official resources is also available to investigators including the regulation itself, guidance for applicants, access to scientific advice, decisions and opinions on applications and importantly, paediatric related information. The latter includes information on paediatric needs, clinical trials, priority list off-patent medicines and presentations. Importantly, the decisions and opinions on particular trial applications, including class waivers and product-specific decisions are included on this site.

The EU *Regulation (EC) No 1901/2006 amended*, the “Paediatric Regulation”, came into force in January 2007, and investigators, industry and the public are still exploring the material effect it may have on paediatric research activity. The paediatric regulation aims to increase the number and availability of medicines that can be used in the paediatric population, by providing rulings, guidance and incentives to investigators, sponsors and institutions to develop paediatric specific products and to develop paediatric prescribing information for other medicines. Mechanisms to license and provide clinician and parent information are also included. One of the features of the Paediatric Regulation is the establishment of a new Paediatric Committee of the European Medicines Agency [13]. This scientific committee has



the authority to make decisions and provide opinions on applications to do research and on outcomes of that research in relation to product use. The Committee is multi-disciplinary, brings together renowned experts in fields of general practice, paediatric medicine, pharmacy, pharmacology, research, pharmacovigilance, ethics and public health. Health care professionals and patient associations are also part of the collective expertise.

The Paediatric Committee has an onerous task to ensure paediatric medicine approval is based on rigorous quality, safety and efficacy data. Strategies used by the Committee include: requiring paediatric investigation plans (PIPs) and data to be submitted to regulatory authorities; assessing PIPs and providing decisions and opinions; monitoring of the PIP compliance; supporting a paediatric research network; implementing key public communication strategies such as the use of a common symbol for medicines that have an approved paediatric use; and training investigators. In addition the regulations provide an incentive to investigators and sponsors by providing an additional six months on the supplementary protection certificate if completed PIP information is included in the Summary of Product Characteristics. For off-patent products, there is the incentive of a paediatric use marketing authorisation which has a ten year data and market protection period. The regulation also provides for establishment of a European data base of paediatric clinical trials, part of which will be publicly available.

Access to U.S. regulations is less straightforward. There are multiple web-routes to access relevant information, and the best route will depend upon the study question. One is to go direct to the U.S. Department of Health & Human Services (HHS) (<http://www.hhs.gov>), thence to the Office for Human Research Protections (OHRP) (<http://www.hhs.gov/ohrp/>). Another is to start with the U.S. Food and Drug Administration (FDA) (<http://www.fda.gov>) and consult the various pages to do with clinical trial practice that may include adults as well as children (<http://www.fda.gov/cder/pediatric>). One page, for example, provides direct links to all FDA regulations relating to good clinical practice and clinical trials (<http://www.fda.gov/oc/gcp/regulations.html>). There is also the Office of Pediatric Therapeutics (<http://www.fda.gov/oc.opt/>).

The history and recent state of USA regulations that inform paediatric research has been reviewed by Diekema [14]. He provides a concise guide to critical incidents and key sources, most notably the *Code of Federal Regulations, 45 CFR 46, Subpart D Additional Protections for Children Involved as subjects in research* 1983. Others are the *Best Pharmaceuticals for Children Act* 2002 and 2007 [15], and the *Pediatric Research Equity Act* 2003 [16] that re-established the FDA's authority to mandate paediatric drug development [17]. Informative guides and regular updates on paediatric issues have been developed and are available through HHS sites of the OHRP and FDA. An example is the *Guidance for Clinical Investigators, Institutional Review Boards and Sponsors: Process for Handling Referrals to FDA under 21 CFR 50.54; Additional Safeguards for Children in Clinical Investigations* [18]. Apart from linked guidelines, these sites provide "current thinking" of the agencies in relation to key issues variously presented as "frequently asked questions" or "guidelines". Further they provide aide-memoirs for investigators such as the "pediatric points to consider" that include a summary of unique review concerns for paediatrics, including study justification, study design, ethical issues, and pediatric protocol checklists. Later stage trial results should be reported to the clinical trial registry

and data base in accord with Regulation S3807. There are also incentives in regulations to encourage paediatric pharmaceutical research, such as six month exclusivity on manufacturer marketing licensing if the company “fairly responds” to FDA requests.

While the OHRP site provides guidance and recommendations for paediatric study conduct, it also highlights areas where compliance is required under the U.S. HHS Regulation. Compliance is specifically required for HHS supported research, but the aspects identified may provide useful benchmarks to alert investigators outside the HHS or USA for potentially high stake paediatric trial issues. As thinking may change and as there is acknowledgement that alternative approaches may be considered by the authorities, investigators are advised to check these sites for changes, opinions and rulings as part of trial planning.

Internationally, many countries have regulations that apply to research with human subjects and specifically to paediatric populations. Most feature principles and practice requirements or recommendations that are consistent with the broad approach of the EU or U.S. and related international conventions. Australia for example, has the National Health and Medical Research Council statements including the *Australian Code for the Responsible Conduct of Research* [19] and the *National Statement on Ethical Conduct in Human Research* [20]. Canada has the Medical Research Council Tri-Council policy statement on *Ethical Conduct for Research Involving Humans* [21]. Some other countries have statements on the conduct of research in humans, but have little that specifically relates to paediatric populations. For example, India has the *Ethical Guidelines for Biomedical Research on Human Participants* [22]; and South Africa has a *Code of Research Ethics* [23]. Both have only limited clauses relating to child research.

In addition to benchmark and local regulations for paediatric research in general, investigators need to be alert to any special provisions for particular groups. In Australia, for example, there are particular guidelines and requirements for research involving indigenous people [24]. Other countries may have similar provisions that should be consulted and integrated into PIPs. There are also special provisions in Australia for research that is conducted outside the country [25] and similar approaches may be taken in other nations. The particular vulnerabilities of unaccompanied children, foster children, wards of the state, and emancipated children also need attention if they are to be target or incidental participants. A lack of attention to any of these regulations can have a significant impact on the conduct of a trial or the reputation of investigators, sponsors or institutions involved.

### 10.11.6 RECOMMENDATIONS AND GUIDELINES

In addition to regulatory requirements and recommendations from statutory or government sponsored institutions, trial investigators, coordinators and employees may also be need to be registered or accredited professionals, subject to the Codes of Conduct, recommendations, guidelines or prescriptions of their professional societies or registration Acts. Depending on local conditions these may be enforceable. Each paediatric speciality and sub-speciality may also have requirements, guides

and resources. Investigators are urged to consult these at an early stage. Investigators should also satisfy themselves that they and their team members are meeting relevant professional obligations in addition to prescribed regulations, as the conduct of some assessment, intervention or outcome measure procedures may require a registered or accredited practitioner. Examples of professional societies that have guidelines for research are: the Royal College of Paediatrics and Child Health [26]; the European Academy of Paediatrics (formerly the Confederation of European Specialists in Paediatrics) [27] that provides guidance ranging from official statements to commentaries and summary presentations, for example that provided by Kurz [28]; and the American Academy of Pediatrics [3]. Profession and sub-speciality specific guides must be sought out by investigators to inform trial decisions, and if none exist, this should be noted somewhere in trial planning records so that the due diligence of trial leaders in this regard can be noted.

There are also guidelines from esteemed practice and research institutes that could be considered and used as methodological sources in trial plans. One is the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). This organisation has produced many resources including the oft-cited, *Guideline for Good Clinical Practice E6 (R1)* [29] which applies to the conduct of clinical trials, including essential documentation and archive guidelines; and the guideline for *Clinical Investigation of Medicinal Products in the Pediatric Population E11* [30], that contains a series of guidelines for drug development and registration processes. These supplement more general ethical guidelines for biomedical research [31]. The Medical Research Council [7] provides an ethics guide for *Medical Research Involving Children* that clearly summarizes key points unique to paediatric trials. Guidelines for international research [25, 31], and research involving human participants in developing societies may also be relevant to investigators working across international boundaries.

Guidelines from esteemed bodies can relate to particular paediatric populations and sub-specialities. For example, the Society for Adolescent Medicine has issued *Guidelines for Adolescent Health Research* [32]. Others such as the National Institute for Clinical Excellence (NICE) (<http://www.nice.org.uk/>) provide specialist resources on “best practice” approaches that can be incorporated into research protocols and general care of young people. While NICE guidelines cover public health, health technologies (including medicines, treatments and procedures) and clinical practice (for specific diseases and conditions) applicable within the National Health Service of the United Kingdom, they are useful practice benchmarks. Examples of approved guidelines include: *Improving Outcomes with Children and Young People with Cancer* [33] and *Feverish Illness in Children—Assessment and Initial Management in Children Younger than 5 years* [34]. Other guideline projects are underway for children. For example, *Prevention of Unintentional Injury in Children under 15* (due April 2010) and *Guidance on Looked After Children* (due September 2010). While there are no obligations on researchers to observe these guidelines in trial plans, they are useful practice benchmarks which may be required when arguing the case for equipoise in intervention or control groups.

In addition to professional society and advisory bodies, there are industry developed guidelines and issues papers. One example is the Association of the British Pharmaceutical Industry publication *Current Issues in Paediatric Clinical Trials* [35]

that reported on a conference covering matters such as regulation, ethics, parent perspectives, national frameworks and the industry perspective.

### 10.11.7 INSTITUTIONAL REVIEW BOARDS

Institutional review boards are often referred to in regulation as the local means to assess, approve and monitor studies. New investigators generally focus their approval efforts on these local bodies, and may not initially be aware of the recommendations, guidelines, regulations and conventions referred to earlier. Paediatric trial investigators should, however, set their views beyond local board requirements. A local compliance approach that breaches standards set in conventions or benchmark regulations may not be considered “reasonable” or “just” in the mind of the public or the law if a trial goes awry.

Notwithstanding the need for a broad view by investigators, many local boards are subject to regulations and guidelines at a national level and this may make local compliance adequate. For example in Australia, Human Research Ethics Committees are established through a formal process involving registration with the Australian National Health and Medical Council. Their conduct and reporting lines are mandated. In another example, local Institutional Review Boards in the U.S. have limits to their approval authority indicated by the level of risk and benefit incurred by research participants [14, 36]. So in some areas, there may be natural links from local boards to national regulations and international conventions, but this should not be assumed.

Institutional review boards that consider and approve paediatric trial research need to ensure that they have access to appropriate expertise to make informed decisions about research with youngsters. Local boards that approve paediatric research may later be on the defensive if problems arise and they did not have adequate paediatric expertise in place to make their decisions robust. Barret [37] for example, suggests that research ethics committees should have members with practical experience in working with sick children so that they can assess whether or not risks to young participants are acceptable, protocols are workable, opportunities are provided for children to withdraw, and whether their autonomy is respected. Ethics committees also need to consider whether the research team has the paediatric capacity to do the study. Drawing on ICH [30], Kurz and Gill [38] recommend that paediatric trials should be done by “medical and scientific personnel who are familiar with GCP [good clinical practice] guidelines and are capable of a trusting relationship and communication with the child and parents ... in ... a child-friendly atmosphere with a paediatric infrastructure and personnel” (p. 43).

While ethics committees can interrogate researcher profiles as part of the approval process, local boards may make assumptions based on researcher reputation that is not backed up by documentation in the applications. This can easily happen if the researcher is a local who is well known for their expertise, or if a leading international researcher is involved and local boards do not feel comfortable interrogating his or her expertise. This can create later problems for boards if the appropriateness of the decision to approve research is challenged and the grounds for the decision about researcher capacity were scant. Paediatric investigators should therefore take some care in preparing the expertise statement that most institutional review board

applications require. It may be worthwhile to detail paediatric clinical, professional and research expertise relevant to the study topic, and to demonstrate that, between all team members, there is capacity for all study demands. In Australia, many institutions also require risk assessment prior to or following institutional review board approval for trial insurance purposes. It is not uncommon, for example, for subject and/or study-specific insurance to be required by some institutions for invasive clinical trials. The risk assessment outcome may depend upon the demonstrated paediatric research capacity of the investigation team.

### 10.11.8 TRIAL QUESTIONS

Trial questions need to be worthwhile and ethical. They need to be “honest and valid” [39, p. 836]. The technical skills involved in the development of trial research questions have been covered elsewhere in this handbook, as have the general ethical principles of beneficence, maleficance and clinical equipoise, so this discussion will explore what principles underpin a worthwhile ethical paediatric trial question.

Paediatric clinical trials ask questions about the safety and efficacy of interventions for infants, children or youth. Kurz and Gill [38] proposed that paediatric should therefore “focus on the knowledge, cure, relief, or prevention of diseases of children. Biomedical studies must be devoted to reducing suffering and improving the prognosis of diseases” (pp. 42–43). To do this ethically, questions need to meet tests of relevance, benefit, originality, achievability, timing, and minimal harm. These are now explored.

The first issue facing investigators is whether or not the question is *relevant* to children and thus whether the involvement of young participants is essential rather than desirable. A high threshold needs to be met in this regard. The Royal College of Paediatrics and Child Health [26] identifies that research should only be conducted in children if the question cannot be answered by studying adults. Further, it identifies that involvement of children in clinical research will be required when an illness or condition only occurs in children, or has features that are more pronounced or have greater impact in young people. In such instances, research will be needed if there is no treatment information available or when what is known is inadequate. Paediatric research participation will also be required if the condition is in the general population but there is no paediatric treatment or there is only adult intervention evidence available [26]. Generally, paediatric inferences should not be drawn from adult studies, although occasionally a treatment will have a long history of use in children that initially relied on adult data but has subsequently been complemented by consensus expert paediatric opinion [26]. In such instances, the weight of consensus expert opinion may mean that an exception can be made and a paediatric inference can be drawn. The Medical Research Council [7] provides five questions for researchers to help determine whether or not the involvement of children is essential or whether findings from adult studies would be adequate. These questions cover [7]: age specificity, developmental understanding, implications for pharmacokinetics, applicability of adult-style therapy and issues of later life disease prevention.

The second issue facing investigators is whether or not answers to the trial question will *benefit* children. Benefit to children is essential [30]. The Royal College of

Paediatrics and Child Health [26] identifies that paediatric research must not only be well designed and well conducted but it must have a real prospect of benefiting children. Benefits to children may be direct through trial participation in treatment or control groups or indirectly to children in general—even though there may not be benefits to individual trial participants. Benefit needs to be self-evident in the trial question and study description [40]. The Royal College of Paediatricians proposed that the following issues help make potential benefits to children clear [26]: *magnitude* of the condition including severity, how *common* it is and how findings will be used; how *probable* it is that research will achieve aims; who specifically will *benefit* from the research (whether it is child participants or children in general); whether benefit to children will be *limited* because treatment is expensive or hard to deliver; the *type* of intervention and whether a less invasive one could be used; the *timing* of benefits in terms of duration or later impact; and finally the whether the *range* of child participants is adequate in terms of potential benefits [26].

One simple way for investigators to consider benefit, is to imagine that the study is finished, the trial question has been answered, then adopt the perspective of a public health official, a general practitioner, a parent or a child and ask “so-what?” What material benefit to children would accrue from the result? Does it merely confirm something already known? Is it interesting but not essential for child health and care? Is there any reasonable likelihood of direct benefit to participants or benefit to children in general? What would the “reasonable person” think about the study if they also knew the costs, risks and demands made on health care staff, participants and families to get the answer? Would the reasonable person agree that some children need to be exposed to risk to benefit children at large? Would the lay-person feel as Smyth [41] does, that “we are all aware of the dramatic impact which results of clinical trials have had on the care and survival of children” (p. 835), and “there is an energy and dynamism which is both exciting and invigorating for paediatric clinical research” (p. 837).

While common sense, careful scholarship and the “so-what?” test suffice for many trials, some studies and teams may benefit from involving ethicists in early stages of question development to ensure that a careful and considered approach to justice and benefit is taken. This is particularly important when trial interventions or measures may involve discomfort, distress or pain; or where the illness or condition brings inherent suffering or risk of death that may be exacerbated or alleviated by trial processes. A clear and scholarly beginning position on the issue of benefit to children not only helps investigators ensure their trial question is child-centred, but it also helps sponsors and participating institutions monitor progress of the trial and develop public and stakeholder communication plans.

The third challenge for investigators is to ask a question that can actually be answered. Is it *achievable*? Is it a question that the investigation team has the capacity to answer from the point of view of expertise, population access, resources and infrastructure? A question may not be realistic if there are insufficient potential participants for the study time frame; if the testing regimen is not physically or emotionally tolerable, or fails to accommodate family routines or responses; if sufficient funding, expertise, infrastructure or consumables are not guaranteed; if a-priori protocols for intervention, child care and data analysis are not transparent or cannot be adhered to; or if the findings are not going to be available or disseminated in ways that will grow the knowledge base to benefit children in general.



One of the most important issues to consider in weighing up whether a trial question is achievable is the capacity of the team. Clinical proficiency and research interest is not enough—the team as a whole needs to have the expertise to conduct all aspects of the trial. This may mean investigators must expand team membership to fill skill, knowledge and labour gaps that range from statistical analysis, ethical question design, budget management, regulation compliance, use of clinical procedures, to interpretation of findings and writing. Clinical professionals new to research sometimes learn the hard way that a good research idea is not necessarily question that can be investigated. Alternatively these researchers don't learn, they don't use structural and capacity building strategies to get the study done, and instead they can blame anyone involved in or connected to the project! The experience can leave clinical staff, managers, families, patients and researchers themselves disillusioned about the research experience and sometimes about each other. This is a particular risk for professionals who are attempting trial research in environments where they are already overloaded with clinical responsibilities and have limited prior experience of the technical and time demands of trials.

Turning a good idea into an achievable question requires a combination of scholarly, managerial and political skill that almost always involves long term collaboration of multidisciplinary experts. It almost never involves the “deity-like-researcher” leading followers or directing an operational team from a geographical or organizational distance. Such studies have inherent structural risks that make them prone to mistakes. Successful paediatric trials are always a team effort. They always involve the building of relationships over time with trust, respect and recognition. They always take effort on the part of the research leader to build and maintain a climate of open scholarly enquiry. Without such effort, mistakes may not be reported, good staff may leave, people fear betrayal or theft of their ideas or reputations, team politics rather than child participation can absorb emotional energy and the high test of ethical practice required in the moral context of paediatric practice is undermined. For paediatric researcher leaders there is, quite simply, no way to “delegate” a paediatric trial. Senior investigators must be involved in every aspect of trial planning, conduct, interpretation and writing as directors, collaborators or hands-on players. To do anything else is to risk allegations of being “front-men” or “poster girls” for trial sponsors, research institutes, or servants of their own “brilliant” careers—even when this is not true.

In addition to issues of team capacity, achievable questions rely on practical details such as estimating whether or not required sample sizes are attainable given the incidence of the condition, the recruitment target population, recruitment methods and time available. Many paediatric clinical populations are very small, hard to access, and may have high decline rates in recruitment. Trial planners may need more than epidemiological data to estimate whether or not they have a reasonable likelihood of recruiting the needed sample—they may need local “on-the-ground” informants who can estimate the impact of recruitment methods on limited potential participants. The target sample size must not only account for the clinical effect size of the intervention in question but also appropriately deal with the confounding variable of developmental maturation, and practical issues such as likely decline rates and drop outs. Trial questions that require good luck to achieve sample sizes should be put to one side. This is hard for researchers to do, particularly if the question is their passion, however it must be done. Underpowered trial findings can



be worse than no findings at all—they give the illusion of reliable evidence. They are all too common in paediatric research. A review of trials published in the *Archives of Disease in Childhood* from 1982 to 1996, found that half the trials had 40 participants or less, which in the case of the trials reviewed meant that they were often under-powered [42]. Researchers should ask themselves “what is the point?” if an adequate sample size cannot be assured. Their attention should turn to more realistic questions or less rigorous research designs.

The fourth challenge for investigators is to ensure that the trial question is *new*. Questions should not be asked when answers are already known. The case of originality should be made clear in “study rationale” sections of institutional review board and investigation plan applications. The case for originality must be strong, scholarly and set in an international context. The strength of evidence should be compelling: systematic reviews, when rigorously conducted, provide a strong evidence base to demonstrate gaps or failings in current knowledge that can be used to justify originality.

Fifth, investigators need to be confident about the *timing* of trials that involve children—they should only be done when the knowledge base is “ready”. The nature of the trial should have a good fit with knowledge already available. ICH guidelines suggest [29, 30]: Phase 1 or 2 paediatric trials are acceptable only when diseases being targeted are entirely or predominantly found in children; phase 2 or 3 trials are acceptable in children for serious diseases where no adequate treatment exists but only after safety and tolerability information has been gained from adult studies; and, Phase 2 or 3 paediatric trials are acceptable for conditions in the general population after there has been considerable research work in adults [29, 30].

Finally, trial questions need to balance anticipated harm with expected benefit and harm should be of “*minimal risk*”. Minimal risk can denote the type of procedure for data collection or intervention where only “very slight or temporary negative impact” might occur [7, p. 15], or where the risk is about the same as that in daily life, or with comparable treatments. Under regulation, [43] potential risks to research participants must be identified and minimized and the prospect of direct benefit to research participants must be maximised. Risks in this instance refer to: “any harm including physical injury, pain, distress, psychological harm, social economic or legal harms that might occur if physical injury occurs, or the potential harms that may be caused if research related information is shared with others” [43]. This places heavy obligations on investigators to identify and describe what potential risks might be. The Royal College Paediatrics and Child Health [26] provides a guide to likely harm description by identifying five aspects to consider: the *magnitude* of severity; the *probability* of harms occurring; whether the *type* of intervention is invasive or non-invasive including psychosocial procedures; the *timing* of potential harm in terms of immediate duration or later effects; and finally issues of *equity* relating to the overuse of children who have many medical problems and are used in research because they are more easily accessible. In assessing harm control, intervention and placebo conditions need to be considered—even placebos may cause harm—antihypertensive drug studies is a good case in point as this raises ethical issues regarding consent and trial design for a condition known to cause harm if left untreated [44].

While there is a requirement to assess likely risk, there is also an obligation to minimize whatever harm must be done to answer the trial question. There needs to

be an appropriate balance between the harm done and the benefit achieved: “expected benefit must exceed recognizable risks [and] serious predictable risks should be avoided” (38, p. 43). These are subjective judgements that must be made by investigators, institutional review boards and families of participants. Some judgements are easy—effective treatment should not be withheld from child participants. Other decisions about design and procedures are harder. The National Academy of the Sciences [45] provides a guide that may be helpful: researchers should consider the potential for age-related risks of harm; whether or not children are really needed; screening for known vulnerabilities; how demanding the protocol adherence is and risks arising from this; the use of only necessary procedures to answer the question; use of rigorous research designs; use of existing knowledge to estimate likely type and magnitude of risks; inclusion of adverse event information in data collection and reports of findings; ensuring investigator research and paediatric capacity; assuring appropriateness of the research setting for children; inclusion of safety monitoring, emergency arrangements, stopping rules for discontinuation; having clear guidelines for data use; and a plan for secure archiving [45].

To estimate and minimize likely harm, investigators must describe the level of risk of an intervention or outcome measure procedure. Defining level of risk is, however, a fraught task. It is one that involves value judgements because currently, risk assessment lacks an empirical standard even though risks are supposed to be identified, quantified and compared [39]. Some paediatric leaders have identified the need for guidelines [46, 47], although the use of guidelines is not universally supported [48]. Inadequate though they are, investigators should thus consult whatever regulation guidelines underpin their institutional review board requirements and ensure they benchmark their risk rating to the relevant regulation using scientific evidence for support. If there is limited scientific evidence available, the National Academy of Sciences [45] guidelines provide a framework for researchers to describe, as best they can, their strategies to minimize harm.

Researchers should also be aware that risk perception varies from the lay-person to the expert. As Afshar et al. [39] suggest: “experts usually assess harm in terms of mortality or morbidity, while a lay person may perceive harm in terms of severity, reversibility, the effect on future generations or influence on personal life. The main consideration should be the acceptability to non-experts, which in pediatric research are parents and older children. The most direct way of determining risk acceptability is to inform participants of the probability and magnitude of harm, and ask them about their preference” (p. 837). Whether or not the risk, discomfort and suffering caused is ultimately reasonable will depend on the question, interests of the child, preferences of the parents, likely benefit and comparison to likely harm in usual clinical practice.

### 10.11.9 PARTICIPANT CHARACTERISTICS

Infants, children and youth have unique body system and social attributes that need to be considered in trial research. Probably one of the most quoted phrases in paediatric research that emphasizes their unique position comes from the Royal College of Paediatrics and Child Health [26]: children are “not small adults”. While every trial will need to investigate participant characteristics relevant to the sub-speciality,

there are some common issues relating to developmental change that should be considered in all fields. These include ages, stages, incidence and heterogeneity of paediatric disease and vulnerability. These issues can be managed as potential confounders as they are known and some aspects to consider are reviewed below. Paediatric trial investigators must also be alert to the possibility of confounders that may be unknown but somehow inherent in the fact their participants are youngsters.

Investigators need to consider how the variable of *age* will be managed in the trial. Age is a proxy-indicator of developmental change and age, as a covariate, needs to be measured and controlled for. The ICH [30] recommends age should be defined in completed days, months or years, using the following stage categories: preterm newborn infants; term new born infants (0 to 27 days); infants and toddlers (28 days to 23 months); children (2 to 11 years); and adolescents (12 to 16 or 18 years depending upon the region). These categories may be used as participant inclusion/exclusion criteria in a study to restrict age variability; alternatively study specific age limits can be set that reflect the sub-speciality study question. In some studies, age may be less of a concern, however it must always be prospectively accounted for, usually by being treated as a continuous covariant in analysis particularly if participants “move” from one age category to another in the course of the study. Different end-points can be set for different age ranges, however every age strata and end-point increases the numbers required in samples.

ICH [30] age related *stages* also provide investigators with specific guidance on factors to consider in relation to drug trials. These factors may also be useful in other types of clinical studies. Investigators are, for example, alerted to [30]: the need for assessment of causes of low-birth weight in preterm newborn infants to determine whether they are immature or growth retarded; or the increased reliability of oral absorption in infants and toddlers; or the increased drug clearance (hepatic and renal) for most pathways in children; or the possibility of hormonal change affecting results of clinical studies in adolescents [30]. Kurz and Gill [38] point out that there are many differences in “physiology, pathology, pharmacokinetics and pharmacodynamics between children and adults” (p. 42); and further, growth and development can influence side effects, dose relative to body weight or surface area, severity of disease, pathological agents and natural history [38]. Investigators should consult the ICH [30], speciality trial resources such as Helms and Stonier [8], and make specific enquiries regarding age related factors in their sub-speciality. This is particularly important when [37]: considering invasive procedures such as the use of repeat blood samples in infants and children as blood volume may limit what can be done; testing off-label or unlicensed applications; or when using surrogate markers, measures of quality of life, pain or other outcomes not validated for use in children [37].

The *incidence and heterogeneity* of disease in children has also been identified as a practical challenge in paediatric research [37]. Some diseases in children are rare which creates challenges for recruiting adequate trial sample sizes. For others, individual responses can vary with age.

The *vulnerability* of young participants is a critical aspect to consider in trial research. As the *European Union Clinical Trials Directive 2001/20/EC, Article 3* states, “children represent a vulnerable population with developmental, physiologi-

cal and psychological differences from adults”. In the distant and not-so-distant past vulnerable children were exploited for medical research purposes particularly if they were institutionalised, disadvantaged or had disabilities [39]. Today, conventions, regulations and guidelines identify these actions were unacceptable, however a child’s inherent vulnerability means the risk of exploitation is always present. It was their vulnerability and perceived inability to give informed consent that lead to the post-World War Two tradition of excluding children from medical research following the establishment of the Nuremburg Code Directives for Human Experimentation 1949. Since post-war years, there is growing recognition of the need for high quality paediatric research “so that tomorrow’s children receive new and better treatments and clinicians have real evidence on which to base their decisions” [41, p. 837]. So children are now permitted to be involved in research, but with a particularly cautious and sensitive approach to acknowledge their vulnerability—this caution and the complexity of regulatory and administrative arrangements to protect them are perceived by some as “barriers” particularly in relation to drug development, but it applies equally to other clinical interventions [49]. “Barriers” are, however, better than the alternative if they are efficiently designed and managed. Barrett [37] says children are the most vulnerable patient group in research—they have fewer rights, problems expressing themselves and the potential for lasting benefit or harm from a research experience. Notwithstanding their vulnerability, she argues that paediatric trial research is essential: “because of the failure to recognise their needs and to perform appropriate research, children are denied access to safe and effective treatment that adults would demand as a fundamental right. Children have become therapeutic orphans” [37]. These “orphans” have been identified since the mid 1960’s when use of adult drugs for children through off-label or off-licence prescription was recognised as a common practice [50] and the term has since been used to identify these medicines [5, 51, 52]. In most cases physicians have little “wiggle room” as adult drugs may be potentially or demonstrably effective for children and the alternative is to provide nothing. Professional societies have tried to grapple with this problem as off-label use can mean that third-party reimbursement or government subsidy of these drugs is not permitted, and physicians may be exposed to law suits in the event of adverse reactions. The ultimate solution to the expanding family of “orphans” is the production of research that involves child participants in rigorous trials under the “parental” scrutiny of regulation, rules, reporting and peer review.

#### 10.11.10 PAEDIATRIC INVESTIGATION PLANS

Other parts of his handbook have explored the preparation of trial protocols and plans in detail. These principles and practices apply to paediatric trials too and will not be reviewed again. This section therefore highlights the special features of paediatric investigation plans that researchers should be aware of at study commencement. These features include the need for plans to be child-centred and family focussed, publicly accountable, regulation compliant, and multi-disciplinary.

Investigation plans need to have the *needs and experience of the child* as the centre-piece of development. Only the minimum number of children required for

adequately powered study design should be used. Measures should aim to demonstrate the quality, safety and efficacy of interventions for infants, children and young people. Consequently, measures selected, data collection procedures, the timing of collection, personnel involved and the study environment should be planned in such a way that developmentally appropriate and supportive approaches are used. The ICH [29, 30] recommends the following ways to minimize discomfort and distress in paediatric participants: only use personnel knowledgeable and skilled in dealing with infants, children and youth and their age-appropriate needs including skill in performing paediatric procedures; use physical settings with furniture, play equipment, activities, and food appropriate for participant's ages; aim to have the study conducted in familiar environments that may include usual places of care; minimize pain and discomfort in procedures wherever possible for example by appropriate anaesthesia, and collect research data through routine clinical tests rather than additional procedures.

Selection of methods should be based on *research evidence* from development and sub-speciality literature, together with advice of paediatric research specialists [8]. Trial plans from adult studies should not be “adapted”, but rather a fresh approach taken with the needs and attributes of children at the core of plan decisions. All personnel involved in the trial should be trained not only for the technical requirements of their role, but also for interaction with children. The trial manual or handbook should focus on the behaviour required of study personnel towards children and their families and this should be monitored for consistency throughout the study. Special training for the identification and management of paediatric adverse events should be required in study plans and a climate of openness encouraged by trial leaders so personnel feel compelled to report mistakes or gaps and are respected and valued for doing so—even if they made the mistake. Outcome measures should consider the child as a person and the demands and influences that will be made on him or her throughout the course of the study. How will the study requirements affect their daily routine, their opportunity for play or rest, their contact with friends and family, their use of favourite toys or electrical devices, the protection of their personal privacy? Things that are of marginal or no interest to researchers may be critical for the wellbeing a child or adolescent.

Apart from the notion of “minimal harm” and “benefit” there is the practical matter of what a child can *reasonably tolerate*. There is little point, for example, in setting a battery of performance tests to measure outcomes if the child can only concentrate long enough for one of them; or setting a number of blood collections if total blood volume in the child is too small; or administering medications or conducting procedures that assume adolescents are not sexually active or that they will openly tell guardians if they are. Obvious errors can be made by even experienced investigators if they focus on outcome measures and forget that they come from a child.

Trial investigation plans should also be *family focussed*. Parents and caregivers know their infants, children and youth better than anyone else. They must feel committed to and interested in the trial to first give consent, and then to maintain involvement. For any recruitment, retention, intervention or outcome measurement strategy to be successful in a paediatric clinical trial, the commitment, expertise and adherence of parents and caregivers is critical. For children in care, consent by authorities as well as carers may also be required. Investigation plans should use

recruitment and retention strategies that have been found to be effective in other trials with participants who are minors including: contact and scheduling methods (such as the use of multiple contacts for recruitment, maintaining a log of cohort follow-ups, making phone calls to contacts to locate participants), reminders (for example letters to remind of appointments), family friendly visit arrangements (such as scheduling data collection visits at the same time as routine health care visits), reimbursements (for example of transport costs), financial incentives (reimbursing for their time), non-financial incentives (such as thankyou letters), and tracking methods to monitor participation (such as regular team meetings to follow up cohort participation) [53–55]. A concerted effort to retain participants is essential. Even though “intention to treat” principles can be used in analysis, drop-outs do threaten internal and external validity of studies and limit the level of inference that can be drawn from findings.

Practically, families may need to adapt their schedule, expend resources, provide transport or other support to enable a child to participate in a trial. The question, the demands on them, the likely benefit and harm to their child will influence their decision to be involved and stay involved. If investigation plans fit into family priorities and schedules their continued involvement is more likely as it is part of a *sustainable* routine. This is particularly the case if trials require special visits for data collection. Even if a child is continually available to researchers, for example if they are hospitalised, the family must still be the focus of plan activities. Many families, for example, would expect that any trial related activity would occur in their presence and that may mean adjusting data collection or intervention schedules for study personnel. Further, families are important sources of trial information—they are usually more alert to changes in the child, are aware of more potentially confounding factors, and if their involvement as “trial partners” [17] is acknowledged by study personnel, they are more likely to adhere to protocols and provide helpful information regarding the study conduct or the child’s response.

In multi-cultural or multi-lingual societies it is important to provide information that can be understood by parents and care-givers. A common practice is translation—this should be done by professionals with expertise in health using usual process of “back-translation” to verify accuracy. But there may be more to consider—the content, presentation, method of recruitment and parent communication may require accommodations to ensure understanding and cultural acceptability. Local advisers and community members should be involved in trial planning stage rather than making adjustments as the variability of the population presents itself.

PIPs need to be publicly accountable and they need to adhere to regulatory requirements. The EMEA, for example, requires all PIPs to be submitted for consideration and to be approved by the Paediatric Committee before study commencement. Once approved, the plan is binding on investigators and sponsors. Results need to be submitted at study series conclusion in accord with the plan specifications. Investigators conducting studies outside these EMEA requirements, are well advised to mimic the transparency and provide a publicly accountable research plan as it is probably one of the best ways to manage potential criticisms or allegations of child “exploitation”. This is particularly important if clinical research organisations are employed to conduct studies: as contractors they can benefit from a clear public plan as it holds them to account at the same time as protecting the investigators and the children.



Perceptions or actual cases of exploitation in medical research do happen. Niles [56], for example, quotes Dembner, an author from the “Boston Globe” February 18, 2001 who, under the gut-wrenching title of “Dangerous dosage to make pediatric medicine safer: Thousands of children are being used to test drugs originally designed for adults” states: the “potent combination of vulnerable children, ambitious researchers, potential profits, and weak oversight can hold great peril for these children”. While the particular facts and opinions surrounding this article are not pertinent here, it is apparent that the title and commentary are, in general terms, both true and false when it comes to pediatric medicines. Yes there are thousands if not millions of children receiving adult drugs—and yes some are involved in testing these drugs—but what is the alternative? To let children go untreated and suffer? To wait until specific paediatric drugs have been developed? Not to test? To ensure only those researchers with no ambition, or those companies with no obligation to shareholders, or those universities with no interest in private funding, engage in pediatric research? To consider that the elaborate regulatory and reporting arrangements in many places are weak and can be bundled together with those that are genuinely sloppy? No wonder we are all so cautious and careful in our PIPs and no wonder there are so many “barriers” to pediatric trial research when the newspaper by-lines such as this can come hurtling into view. The best we can do as pediatric investigators is ensure that our research does meet high standards and that there is opportunity for transparency and accountability in what we do and how. PIPs that comply with reporting requirements and are clearly child-centered can protect investigators and children, and encourage the public to engage in a more informed dialogue. If investigators do not need to register with the EMEA other strategies such as submission to “trial banks” or publication of the PIP/protocol with the trial findings, or publication in open-access or independent journals can be a way to achieve transparency. Such strategies may address some of the concerns regarding potential and actual conflict of interest that have been expressed by researchers themselves [57]. There is an urgent need for informed, realistic and open public, researcher and sponsor dialogue around complex issues of trial funding, independence and ‘cross subsidizing’ of research, researchers, clinical or administrative staff through research funding and PIP transparency could help.

Investigation plans must also be tailored to *multidisciplinary* study teams. In particular, communication strategies for multidisciplinary personnel within and outside the trial are needed. All trial personnel should have a trial manual or handbook that specifies what they need to do, that uses “quick reference” and visual cues to help guide and reinforce required behaviour. Photographs, flow-charts, check-lists and so on may be useful. Small reminder posters, cue notes on files, regular follow up, thankyou or reminder calls, occasional presentations by researchers can help keep a multi-disciplinary team consistent and equivalent in their trial behaviour. These strategies can also help people feel valued and retain their commitment and enthusiasm. For clinical staff not connected with any trial activity, there may also be a need to provide information. Many parents seek the opinion of their doctor about whether or not their child should get involved in a trial [56]. Consequently, researchers should prepare information and communication strategies. Researchers could for example, contact local doctors, teachers, case-workers, therapists etc as a matter of routine. They could provide parents with information sheets that can



be given to those people parents believe are important in helping them make decisions.

### 10.11.11 ASSENT AND CONSENT

This section of the chapter explores issues of consent by parents and adults and of assent by participating children. Not enough paediatric research is conducted as there are so many apparent barriers [49] involved—trying to avoid harm, fear of litigation by sponsors and investigators, and concerns about ethics on the part of potential investigators. While great care must be taken by investigators to protect children from unnecessary risk, children ultimately have the most to gain from well intentioned and well constructed paediatric trials. Requirements relating to consent and assent can be arduous and can vary from place to place [e.g., 6, 7, 15]. Investigators must be aware of regulations that apply in jurisdictions in which their research is being conducted, as consent has legal meanings and implications, particularly in relation to the “competence” of a person, including a child, to give consent.

One way of conducting responsible research is to ensure that those who act on the behalf of children are well informed [40, 58]. If adults who make decisions for children and act towards them are well informed about research and ethical issues, paediatric needs and the autonomy and potential apprehension of children about medical procedures, then agreeing to participate in paediatric trials is done properly. The Royal College of Paediatrics and Child health [26] provides guidelines on consent that cover strategies to ensure consent is freely given, informed and that explanation, information and where possible study findings are available. But even with these safeguards not all parents read consent documents and not all have these documents explained properly to them [59]. Parents and caregivers are vulnerable to coercion, influence and intimidation even when they are well intentioned if they are not properly informed [31]. Consequently, researchers need to take the time and consider how best to explain to parents and children what will happen and why, why it is important and why children in general will benefit [60, 61]. When parents have the study specific and context relevant information they can reflect on their child’s involvement with all relevant issues laid bare. They need to be able to understand their right to withdraw and the difference between trial interventions and usual therapy [59]. If an open approach is taken, “it is now widely accepted from an ethical perspective research can be carried out on children, when there is no expected benefit for them individually, provided there is minimum risk, strict safeguards and no objection from either the child or parents” [62, p. 202]. The Medical Research Council [7] provides a helpful flow chart of processes involved in seeking consent that may be a useful procedural guide.

Explicit informed consent to participate in a clinical trial is a standard ethical requirement. In the case of paediatric trials, parental consent or consent from a person with parental responsibility on behalf of the child is necessary [63]. As noted in the last paragraph, for parental consent to be valid it must be freely given and informed. The parent can permit, approve, or agree to anything that is clearly not against the interests of the child [63]. Parental consent to something that might harm the child is not valid. The Belmont Report, cornerstone of all international ethics guidelines [64], requires protection of children regardless of parental permission.

“Minimal risk” is acceptable for parent consent and is generally accepted to be equivalent to risk encountered in the normal course of a child’s everyday life. Parents have the responsibility and authority to choose activities that define their child’s risks and benefits [65]. thus, parents can consent on their child’s behalf when the risks are equivalent to the normal risks of childhood [65]. They can also consent when the risks are comparable to other available options—referred to as “clinical equipoise” [66]. Demonstrating clinical equipoise is particularly difficult in paediatric studies—Afshar et al. [39] recommend that “genuine uncertainty about the superiority of 1 treatment over another is essential to motivate clinicians and patients to participate” (p. 838). Parents cannot and should not consent to risks that are not comparable to other available options or that go beyond minimal risk.

The process of obtaining informed parental consent is complex. Consent to treatment is usually in the child’s best interest but consent to research may not be [37]. Research benefits may be for the paediatric population at large and may not necessarily be for the individual. Or there may be little benefit at all. Why, then, do parents consent to research? The decision to participate in a trial is known to be influenced by parental, child, trial and investigator factors [67, 68, 69]. The two main reasons parents say they consent to trial participation are—contributing to clinical research and to benefit their child [70]. Caldwell, Butow and Craig [67] found parents may perceive benefits to include the offer of hope, better care, access to new treatments, access to help and information, parent-to-parent support, and the altruistic motivation of helping others.

Even though parents may support the general notion of paediatric research trials they may not want their own child involved [39]—parents fear causing harm or hurt. Parents may object to the notion of their child being used as a “guinea pig”, especially where the trial methodology involves random assignment and placebo controls [67]. Preparation of participant information sheets can help counter fears and facilitate successful recruitment. Investigator factors such as doctor recommendations, doctor invitations and communication of trial information also affect decision-making from the parent’s perspective [67]. Parents also seek out the views of their own doctors to help inform trial participation decisions [56, 68]. Parent knowledge, beliefs and emotional responses coupled with an understanding of their child’s preferences will affect decision-making [67, 68]. Other parent factors that can influence consent rates are the parent’s socio-economic background with people from lower strata agreeing more often [71, 72], parent understanding of right to withdraw, trial versus usual therapy, and the voluntary nature of participation [59]. The context of study invitations and recruitment conditions also affect consent rates and the validity of consent given. Parents in emergency or acute medical situations may not be in a state any reasonable person would consider was acceptable for making consent decisions on matters relating to harm and short or long term potential effects. Time for reflection may affect agreement rates—the longer parents are given to reflect on risks, the less likely consent will eventuate [37]. Parents of chronically ill children or children with disabilities may also be “research savvy” and able to weigh up issues in the light of previous research experience, while others may not be able to differentiate research activity from usual treatment even when this is pointed out. There is a debate in medical practice emerging about how consent for interventions should best be given [73] and this will have implications for research consent procedures.

In paediatric research, consent of parents is not enough. Even if it is given in an informed and fair manner, the interests of the child remain paramount [1]. Since children may have only a limited understanding of what the research involves, it is hard to think of their participation as anything other than “involuntary” [61]. So how do investigators observe the same ethical principles of respect, justice and beneficence as they do in adult research? Assent provides the answer. Where positive agreement can be obtained from children capable of giving it, studies should be discussed with them in age appropriate ways such as stories or photographs, and assent sought [31]. Determining whether or not a child can decline or assent in an informed manner involves a judgement about their capacity to reason [74]. Researchers must be cognisant of the child’s developmental level both when explaining research procedures and the likely outcomes, and when judging the child’s capacity to give consent [19, 20, 25]. It is understood that “children become capable of assent when they are capable of understanding the research in question and making a prospective decision whether to participate” [77, p. 233]. If a child refuses to participate and gives a reason for this that makes sense given their age, then this may be enough to indicate “competence” and hence their ability to decline in an informed manner. Welthorn and Campbell [75] suggest the age of nine years might be a useful guide and this is the intellectual age recommended by the American Academy of Pediatrics [76]; Wendler recommends the age of 14 years as suitable because at this age they can usually understand research questions [77]; and Koren et al. [78] suggest more general indicators of “maturity”—such as when they are old enough to be a “baby-sitter”—are adequate. Regardless of competence to assent, children must be given the opportunity to object. Sustained dissent should be respected in all cases even if the child is too young or unable to give assent [79].

The process for acquiring assent involves the researcher thinking about the issue from the child’s point of view. Assent statements should be written in language that is at the comprehension level of a 6-year old child, in large font using the active form [39]. It is similar to the process of converting medical jargon into plain English for adult participants on information sheets. Study participation benefits and risks are different when viewed from the child’s perspective. A child wants to know [80]: Will it be fun? Will it hurt? What will happen to me? Do I have to? Will there be parental consequences if I say no? Will assent lead to other desirable incentives, such as time off school, undivided attention of a parent, new toys and treats as rewards for good behaviour, or perhaps even the possibility of making a worried parent happy?

Children should be asked to assent only to research procedures that they are capable of understanding [79]. A child cannot become fully informed via a standard written participant information sheet and so will rely on environmental cues to confirm their impressions as part of the assent process—Is my parent looking comfortable or anxious around the investigator? Does the setting give any clues that it will be painful despite what I am being told? Does the investigator seem trustworthy to interact with? Do they provide toys and play with me? Stop when I say no? Make me fail repeatedly? An investigator who attends to the child’s emotional and environmental context in addition to providing factual information is more likely to gain the child’s agreement. The investigator should keep a written record of the assent procedure which demonstrates that they “provided the child with all the necessary information in an age-appropriate fashion, that the child understood the informa-

tion, and that the child voluntarily agreed to participate in the research project” [80, p. S32]. The investigator records are probably more important than obtaining a child’s signature, because it is not until children are older that they understand the symbolic meaning of a signature [80].

The process of gaining assent from children and parents together gives rise to many grey areas for ethics. Children and parents do not always agree. Balancing the wishes of a child who refuses to assent to research procedures against the parent’s decision and right to enrol their child in a potentially beneficial trial is a complex ethical dilemma. Ethical issues regarding consent and confidentiality become even more complex when adolescents are participants, as they may not want their parents’ involvement, but in studies where harm is greater they may need adult assistance [32].

Other issues involved in consent relate to processes and delegations. Children may have parents or care-givers, but they may not be the people who hold legal responsibility for the child. Families may be separated and careful attention to agreement of relevant adults is needed. Agencies, government or other adults may hold the authority to consent. Once in a study, parents and the children always have a right to withdraw and this should be told at the beginning and throughout the study.

Adequate recruitment and retention of homogenous participants, to paediatric research trials, is extremely challenging. Many diseases common in adults are rare in children, so it is hard to recruit sufficient numbers for statistical significance. There are temptations to “water down” high standards and consent procedure requirements in the interests of study completion. Pressure on investigators to attain sample sizes, increases the risk of parents being inappropriately pressured to enrol children to participate in trials, because the available numbers are so small. Coercion at any level is unethical even though good numbers are needed for rigorous research. Investigators must carefully plan and design trials that are clinically feasible in terms of likely sample size. It is unprincipled to enrol child in studies that cannot ever be completed due to sampling frame limitations. Recruitment and referral strategies can influence consent rates: referral by health care professionals leads to high consent rates [55], sample size attainment is more likely if large numbers of potential candidates are approached [55]. Recruitment incentives also positively influence consent rates, but their use should be informed by appropriate ethical standards [3]. Payment and reimbursement of costs have been used to provide compensation for study participation and retention to study completion [39, 81].

### 10.11.12 SAFETY AND MONITORING

Safety in clinical trials is both a process and an outcome. Safety information is an aim of trial studies. At the same time, the duty of care for trial participants means that trial processes must plan for prompt action in response to suspected and identified adverse events. Trial plans and reports therefore need to clearly outline what strategies, decision-points and decision-makers will be involved in identifying, referring, confirming and responding to suspected adverse events.

Safety of trial participants needs to be considered before, during and after the trial. Before the trial, a careful exploration of the existing evidence base should be

made to assess potential adverse events and these should be monitored and measured. If the information base relates only to adults, then a careful watch on paediatric participants may be needed as developing systems and structures may respond differently to adults [29, 30]. Before the trial, processes to monitor protocol adherence, inspect participant files and receive reports from trial and other clinical staff about suspected adverse events need to be in place. Training should be provided to everyone involved in the trial so that they know how and when to make reports of suspected adverse events. A culture of openness and respect is needed so that mistakes or adverse events are reported promptly and the honesty and vigilance involved in making that report is valued by the team. After a trial, there may be a need to continue safety and monitoring processes through long-term follow up studies. This is particularly important for paediatric trials or trials where the condition is chronic and treatment long term, as participants are growing and adverse consequences of an intervention may not be apparent initially but may emerge later in life [30]. Sammons and co-investigators [82], identified through a literature review of therapeutic clinical trials, that only 2% of studies reported use of safety monitoring committees, and only 11% of studies reported adverse events including deaths. They came to the conclusion that every paediatric trial should have an independent safety monitoring committee to assess the likelihood of risk, monitor the progress of participants and quickly respond to observed differences [82]. Independent risk assessment and safety monitoring is helpful not only for study integrity, but also for protecting children, increasing parent confidence in the well-being of their children, and demonstrating due diligence by investigators. Safety monitoring should be done by personnel with paediatric expertise, access to expert advice regarding the subspecialty, ready access to trial material, and the authority to compel action if the safety of trial participants is in doubt.

### 10.11.13 CONCLUSION

This chapter provided an introduction to common issues that are unique or pronounced in paediatric clinical trials and an outline of those issues are critical to their initiation, conduct and success. The tone adopted in the chapter was that of the “guide on the side”. These trials, perhaps more than others, require sponsors, investigators, trial coordinators and project managers to adopt a personal moral standpoint towards participants and appreciate the social context of clinical research. This chapter provided common signposts that should be helpful in whatever unique trial journey investigators, sponsors and workers may lead, recognising that the technical and procedural issues covered in other Handbook chapters will apply, and the unique evidence of sub-specialities will need to be sought.

While morality, ethics and participant characteristics feature in every trial, in paediatrics they are critical to design integrity and the public success of a trial. A perfectly crafted paediatric trial can bring down an investigator’s lifetime reputation, a company’s global brand, or reduce the esteem of institutions if the study was not child-centred and astute to the multiplicity of factors associated with participant vulnerability and their capacity to reasonably and meaningfully participate in trial requirements. The complexity and demands of paediatric clinical trials can drive some investigators, sponsors and institutions to wonder whether or not they are

worth doing. There are many perceived and actual barriers [49]. Paediatric trials are not for the faint hearted. They demand the highest intellectual capacity and ethical stance for planning, implementing, monitoring and reporting. These trials require enormous good will and generosity from sponsors, investigators, trial employees, parents/carers, and the children themselves. They are inevitably time intensive in ways that are hard to anticipate, they can be very costly with small, hard to locate and recruit samples. The financial and reputation rewards may be few and the risks are great for all concerned. There is, however, an urgent need for the brightest and best in health care to commit to paediatric clinical trials and do more of them.

Routine and widespread use of interventions with little or often no paediatric research evidence continues apace with potentially catastrophic results. Investigators, sponsors and institutions can sometimes feel caught between a rock and a hard place, as they are castigated for off-label, unscientific practice whilst suffering sometimes justified but often unwarranted public torment regarding involvement of children in trials. The irony of public concern for paediatric trial participants is that there appears little or no concern for the millions who daily receive interventions without any scientific evidence to inform parent or professional decisions. Notwithstanding the demands, risks and rigours of paediatric trial research, there is an urgent and compelling moral and practical need to involve outstanding investigators, sponsors and institutions in rigorous, community supported and well understood paediatric trials so that we are able to benefit children.

## APPENDIX: FACTS SUMMARY

Paediatric clinical trials are urgently needed to support use of clinical interventions with infants, children and youth who regularly receive off-label, off-licence or scientifically unsupported treatment.

Paediatric clinical trials require investigators to understand the moral context of their research as participants are vulnerable.

Conventions, regulations and policies from esteemed professional societies provide guidance on principles and practices that deal with implications of the moral context.

Paediatric trial questions must be worthwhile, relevant, new, achievable, timely, pose minimal risk and must benefit children.

Paediatric investigation plans need to be rigorous, evidence based, equal to the capacity of the team and resources available and benchmarked to appropriate standards and regulation particularly in relation to consent.

Safety of participants and external monitoring of adverse events are critical study processes that underpin the quality of study findings

## REFERENCES

1. United Nations (November, 1989), Convention on the rights of the child, General Assembly resolution 44/25 of 20 November 1989 Geneva, Switzerland; available at: <http://www.unicef.org/crc/>



2. American Academy of Pediatrics Council on Child and Adolescent Health (1988), Age limits of Pediatrics, *Pediatrics*, 81(5), 736; reaffirmed 2006 in *Pediatrics*, 117, 1846–1847.
3. American Academy of Pediatrics (2008), *Policy Statements* (various). Available at: [http://aappolicy.aappublications.org/policy\\_statement/index.dtl](http://aappolicy.aappublications.org/policy_statement/index.dtl)
4. American Academy of Pediatrics Committee on Drugs (2002), Use of drugs not described in the package insert (off-label uses), *Pediatrics*, 110, 181–183.
5. American Academy of Pediatrics Committee on Drugs (1970), Policy Statement: Therapeutic orphans and the package insert, *Pediatrics*, 46, 811–813.
6. European Medicines Agency (2008), *Paediatric Investigation Plans*; available at: <http://www.emea.europa.eu/htms/human/paediatrics/pips.htm>
7. Medical Research Council (2004), *MRC Ethics Guide: Medical research involving children*, MRC Publications, London; available at: [www.mrc.ac.uk](http://www.mrc.ac.uk)
8. Helms, P., and Stonier, P., eds (2005), *Paediatric Clinical Research Manual*, Euromed Communications, Surrey, UK.
9. Kahn, J., Mastroianni, A. C., and Sugarman, J., eds (1998), *Beyond consent: Seeking justice in research*, Oxford University Press, New York.
10. World Medical Association (2004), Declaration of Helsinki, *Ethical principles for medical research involving human subjects, Document 17.C* amended. Geneva, Switzerland; available at: <http://www.wma.net/e/policy/b3.htm>
11. United Nations (2008), *Convention on Rights of Persons with Disabilities, General Assembly Resolution of 3 April, 2008*, Geneva, Switzerland; available at: <http://www.un.org/disabilities/default.asp?id=259>
12. United Nations (2007), *Declaration on the Rights of Indigenous Peoples, General Assembly Resolution 13 of September 2007*, Geneva, Switzerland; available at: <http://www2.ohchr.org/english/law/>
13. European Medicines Agency EMEA (2009), *Paediatric Committee*; available at: <http://www.emea.europa.eu/htms/human/paediatrics/pdco.htm>
14. Diekema, D. S. (2006), Conducting ethical research in pediatrics: a brief historical overview and review of pediatrics regulations. *J. Pediatr.*, 149, S3–S11.
15. USA Food and Drug Administration (2002), *Best Pharmaceuticals for Children Act*, Jan 4, 2002 and 2007 (Updated provisions 2008 and 2009), Washington, DC; available at: <http://www.fda.gov/cder/pediatric/#bpca2007>
16. USA Food and Drug Administration (2003), *Pediatric Research Equity Act of 2003*. July 23, 2003. Washington, USA; available at: <http://fda.gov/cder/pediatric/#prea>
17. Rose, K. (2005), Pediatric drug development. *Appl. Clin. Trials*, January Article 140819, available at: <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=19055>
18. USA Food and Drug Administration (2006), *Guidance for Clinical Investigators, Institutional Review Boards and Sponsors, process for handling referrals to FDA Under 21 CFR 50.54 Additional Safeguards for Children in Clinical Investigations*, FDA, Rockville, MD.
19. National Health and Medical Research Council (2007), *Council Statements Including the Australian Code for the Responsible Conduct of Research*, NHMRC Publications, Canberra.
20. National Health and Medical Research Council (2007), *National Statement on Ethical Conduct in Human Research*, NHMRC Publications, Canberra.
21. Medical Research Council of Canada, Natural Sciences and Engineering Research Council of Canada and Social Sciences and Humanities Research Council of Canada



- (2005), *Tri-Council Policy Statement. Ethical Conduct for Research Involving Humans*, Ontario Public Works and Government, Ottawa; available at <http://pre.ethics.gc.ca/english/policystatement/policystatement.cfm>
22. Indian Council of Medical Research (2006), *Ethical Guidelines for Biomedical Research on Human Participants*, ICMR, New Delhi; available at: [http://www.icmr.nic.in/ethical\\_guidelines.pdf](http://www.icmr.nic.in/ethical_guidelines.pdf)
  23. Human Sciences Research Council (2009), *Code of Research Ethics*. HSRC, Pretoria, South Africa. Available at: [http://www.hsrc.ac.za/Corporate\\_Information-6.phtml](http://www.hsrc.ac.za/Corporate_Information-6.phtml)
  24. National Health and Medical Research Council (2003), *Guidelines for ethical conduct in Aboriginal and Torres Strait Islander Health Research*, NHMRC Publications, Canberra; available at: <http://www.nhmrc.gov.au/users/indig.htm>
  25. National Health and Medical Research Council (2007), *National Statement on Ethical Conduct in Research Involving Humans: People in Other Countries*, NHMRC Publications, Canberra; available at: [http://www.nhmrc.gov.au/publications/2007\\_humans/section4.8.htm](http://www.nhmrc.gov.au/publications/2007_humans/section4.8.htm)
  26. Royal College of Paediatrics and Child Health: Ethics Advisory Committee (2002), *Guidelines for the ethical conduct of medical research involving children*, Reprinted in *Arch. Dis. Ch.*, 82(2), 177–182.
  27. European Academy of Paediatrics (2009), *What is the E.A.P.?* available at: [http://www.eapaediatrics.eu/v3/lay\\_eap.cfm](http://www.eapaediatrics.eu/v3/lay_eap.cfm)
  28. Kurz, R. (2004), *Paediatric research demands child-specific guidelines for ethics and good clinical practice*. European Academy of Paediatrics: Powerpoint document. Available at: [http://www.cesp-eap.org/\\_public/lay\\_docs.cfm](http://www.cesp-eap.org/_public/lay_docs.cfm)
  29. International Conference on Harmonisation (1996), *Guideline for Good Clinical Practice E6(R1)*; available at: <http://www.emea.europa.eu/htms/human/ich/ichefficacy>
  30. International Conference on Harmonisation (2000), *Clinical Investigation of Medicinal Products in the Paediatric Population E11*; available at: <http://emea.europa.eu/htms/human/ich/ichefficacy.htm>
  31. Council for International Organizations of Medical Sciences (2002), *International Ethical Guidelines for Biomedical Research Involving Human Subjects*; available at: [http://www.cioms.ch/frame\\_guidelines\\_nov\\_2002.htm](http://www.cioms.ch/frame_guidelines_nov_2002.htm)
  32. Santelli, J. S., Rosenfeld, W. D., DuRant, H. R., et al. (1995), Guidelines for adolescent health research: a position paper of the Society for Adolescent Medicine, *J. Adol. Health*, 17, 270–322.
  33. National Institute for Clinical Excellence (2005), *Improving outcomes with children and young people with cancer*. August 2005; Available at: <http://www.nice.org.uk/guidance/index.jsp?action=byID&o=10899>
  34. National Institute for Clinical Excellence (2007), *Feverish illness in children—assessment and initial management in children younger than 5 years*, May 2007; Available at: <http://www.nice.org.uk/guidance/index.jsp?action=byID&o=11010>
  35. Association of the British Pharmaceutical Industry (2005), *Current Issues in Paediatric Clinical Trials*, ABPI Publications, London, UK.
  36. Shah, S., Whittle, A., Wilfond, B., et al. (2004), How do institutional review boards apply the federal risk and benefit standards for pediatric research? *JAMA*, 291, 476–482.
  37. Barrett, J. (2002), Why aren't more pediatric trials performed? *Applied Clinical Trials*, July, Article 83729. Available at: <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=5016>
  38. Kurz, R., and Gill, D. (2003), Practical and ethical issues in pediatric clinical trials. *Applied Clinical Trials*, September, Article 79923; available at: <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=5124>

39. Afshar, K., Lodha, A., Costei, A., et al. (2005), Recruitment in pediatric clinical trials: an ethical perspective, *J. Urol.*, 174(3), 835–840.
40. US Food and Drug Administration (2009), *Should your child be in a clinical trial?* available at: <http://www.fda.gov/consumer/updates/pediatrictrial101507.html>
41. Smyth, R. L. (2007), Making a difference: the clinical research programme for children, *Arch. Dis. Child.*, 92, 835–837.
42. Campbell, H., Surry, S., and Royle, E. (1998), A review of randomised controlled trials published in Archives of Disease in Childhood from 1982–1996, *Arch. Dis. Child.*, 79, 192–197.
43. US Department of Health and Human Services (2001), *Protection of human subjects: additional protections for children involved as subjects in research*. Code of Federal regulations Title 45, Part 46 Subpart D as revised October 1, 2001; available at: <http://www.hhs.gov/ohrp/children>
44. Flynn, J. T. (2003), Ethics of placebo use in pediatric clinical trials: the case of antihypertensive drug studies, *Hypertension*, 42, 865–869.
45. National Academy of Sciences (2004), *Ethical Conduct of Clinical Research Involving Children*, NAS, Washington DC.
46. Nelson, R. M. (2007), Minimal risk, yet again. *J. Pediatr.*, 150, 570–572.
47. Nelson, R. M., and Ross, L. F. (2005), In defense of a single standard of research risk for all children, *J. Pediatr.*, 147, 565–566.
48. Wendler, D., and Glantz, L. (2007), A standard for assessing the risks of paediatric research: pro and con, *J. Pediatr.*, 150, 579–582.
49. Vanchieri, C., Butler, A. S., Khutsen, A. (Rapporteurs) (2008), *Addressing the barriers to pediatric drug development workshop summary*, The National Academies Press, Washington, D.C.
50. Shirkey, H. (1968), Therapeutic orphans, *J. Pediatr.*, 72, 119–120.
51. Shirkey, H. (1999), Editorial comment: therapeutic orphans, *J. Pediatr.*, 104, 583–584.
52. Wilson, J. T. (1999), An update on the therapeutic orphan, *J. Pediatr.*, 104, 585–590.
53. Meyers, K., Webb, A., Frantz, J., et al. (2003), What does it take to retain substance-abusing adolescents in research protocols? Delineation of effort required, strategies undertaken, costs incurred, and 6-month post-treatment differences by retention difficulty, *Drug Alch. Depend.*, 69, 73–85.
54. Robinson, J. L., Fuerch, J. H., Winiewicz, D. D., et al. (2007a), Cost effectiveness of recruitment methods in an obesity prevention trial for young children. *Prev. Med.*, 44, 499–503.
55. Robinson, K. A., Dennison, C. R., Wayman, D. M., et al. (2007b), Systematic review identifies number of strategies important for retaining study participants. *J. Clin. Epidemiol.*, 60, 757–765.
56. Niles, J. P. (2003), Pediatric subjects and their parents. *Appl. Clin. Tr.*, Sept, 46–48. Article 79923; available at: <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=5489>
57. Smith, R. (2005), Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Medicine*, 2(5), e138. doi: 10.1371/journal.pmed.0020138.
58. Diekema, D. S., and Stapleton, F. B. (2006), Current controversies in pediatric research ethics: Proceedings introduction. *J. Pediatr.*, 149, S1–S2.
59. Hazen, R. A., Drotar, D., and Kodish, E. (2007), The role of the consent document in informed consent for pediatric leukaemia trials, *Contemp. Clin. Trials*, 28, 401–408.

60. Brown, K. E., Barton, R. P., Short, M. A., et al. (2006). Positive approach to pediatric informed consent. *Appl. Clin. Trials*, June, Article: 334576. available at: <http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/author/authorDetail.jsp?id=33197>
61. Royal Australasian College of Physicians (2008), *The Royal Australasian College of Physicians' Paediatric Study policy on Ethics of Research in Children*, RACP, Sydney.
62. British Medical Association (2001), *Consent rights and choices in health care for children and young people*. BMJ Books: London, UK.
63. Royal College of Physicians (2007), *Guidelines on the practice of ethics committees in medical research with human participants*, Fourth Edition RCP, London, UK.
64. National Commission for the protection of human subjects of biomedical and behavioural research (1979), *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*, Department of Health Education and Welfare, U.U. Government, Washington; available at: [http://www.biomedethics.gov/reports/part\\_commissions/index.html](http://www.biomedethics.gov/reports/part_commissions/index.html)
65. National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research (1977), *Report and recommendations—research involving children*. Department of Health Education and Welfare, US Government, Washington; available at: [http://www.bioethics.gov/reports/past\\_commissions/index.html](http://www.bioethics.gov/reports/past_commissions/index.html)
66. Freedman, B. (1987), Equipoise and the ethics of clinical research. *New Eng. J. Med.*, 317, 141–145.
67. Caldwell, P. H., Butow, P. N., and Craig, J. C. (2003), Parents' attitudes to children's participation in randomized controlled trials. *J. Pediatr.*, 142, 554–559.
68. Caldwell, P. H., Murphy, S. B., Butow, P. N., et al. (2004), Clinical trials in children. *Lancet*, 364(9436), 803–811.
69. Tait, A. R., Voepel-Lewis, T., Robinson, A., et al. (2001), Priorities for disclosure of the elements of informed consent for research: a comparison between parents and investigators. *Paediatr. Anaesth.*, 12, 332–336.
70. van Stuijvenberg, M., Suur, M. H., de Vos, S., et al. (1998). Informed consent, parental awareness, and reasons for participating in a randomised controlled study, *Arch. Dis. Child.*, 79, 120–125.
71. Harth, S. C., and Thong, Y. H. (1990), Sociodemographic and motivational characteristics of parents who volunteer their children for clinical research: a controlled study. *BMJ*, 300(6736), 1372–1375.
72. Harth, S. C., Johnstone, R. R., and Thong, Y. H. (1992), The psychological profile of parents who volunteer their children for clinical research: a controlled study. *J. Med. Ethics*, 18, 86–93.
73. Elwyn, G. (2008), Patient consent-decision or assumption? *BMJ*, 336, 1259–1260.
74. Broome, M. E. (2001). *Children's assent to clinical trial participation: a unique kind of informed consent*. Available at: <http://cancertrials.nci.nih.gov/understanding/indepth/protectons/assent/index.html>
75. Welthorn, L. A., and Campbell, S. B. (1982), The competency of children and adolescents to make informed treatment decisions. *Child Dev.*, 53, 1589–1598.
76. American Academy of Pediatrics Committee on Bioethics (1995), Informed consent, parental permission, and assent in pediatric practice. *J. Pediatr.*, 95, 314.
77. Wendler, D. S. (2006). Assent in paediatric research: theoretical and practical considerations. *J. Med. Ethics*, 32, 229–234.
78. Koren, G., Carmeli, D. B., Carmeli, Y. S., et al. (1993), Maturity of children to consent to medical research: the babysitter test. *J. Med. Ethics*, 19, 142–147.

79. Wendler, D., and Jenkins, T. (2008), Children's and their parents' views on facing research risks for the benefit of others. *Arch. Pediatr. Adolesc. Med.*, 162, 9–14.
80. Ungar, D., Joffe, D., and Kodish, E. (2006), Children are not small adults: documentation of assent for research involving children. *J. Pediatr.*, 149, S31–S33.
81. Wendler, D., Rackoff, J. E., Emanuel, E. J., et al. (2002). The ethics of paying for children's participation in research. *J. Pediatr.*, 141, 166–171.
82. Sammons, HM, Gray, C, Hudson, H., et al. (2008). Safety in paediatric clinical trials—a 7 year review. *Acta. Paediatr.*, 97(4), 474–477.



# 10.12

---

## *Clinical Trials in Dementia*

**Encarnita Raya-Ampil<sup>1</sup> and Jeffrey L. Cummings<sup>2</sup>**

<sup>1</sup>*Department of Neurology and Psychiatry, University of Santo Tomas, Manila, Philippines*

<sup>2</sup>*Departments of Neurology and Psychiatry and Biobehavioral Sciences,  
David Geffen School of Medicine at UCLA, Los Angeles, California*

### **Contents**

|          |                                                         |     |
|----------|---------------------------------------------------------|-----|
| 10.12.1  | Introduction                                            | 662 |
| 10.12.2  | Defining AD, MCI, and VaD                               | 662 |
| 10.12.3  | Severity of Dementia                                    | 663 |
| 10.12.4  | Ethical Conduct of Dementia Trials and Informed Consent | 665 |
| 10.12.5  | Generalizability of Clinical Trial Results              | 666 |
| 10.12.6  | Outcome Assessments in Dementia Trials                  | 667 |
|          | 10.12.6.1 Primary Outcome Measures                      | 668 |
|          | 10.12.6.2 Secondary Outcome Measures                    | 670 |
| 10.12.7  | Clinical Trial Designs                                  | 673 |
|          | 10.12.7.1 Special Clinical Trial Design Features        | 677 |
|          | 10.12.7.2 Randomization                                 | 678 |
|          | 10.12.7.3 Length of Clinical Trials                     | 678 |
| 10.12.8  | Statistical Analyses                                    | 679 |
| 10.12.9  | Drug–Placebo Difference                                 | 681 |
| 10.12.10 | Placebo Responses                                       | 681 |
| 10.12.11 | Attrition and Adverse Effects                           | 685 |
| 10.12.12 | Presenting Clinical Trial Results                       | 687 |
| 10.12.13 | Disease-Modifying Trials                                | 687 |
|          | Acknowledgment                                          | 687 |
|          | References                                              | 690 |

### 10.12.1 INTRODUCTION

Dementia is a health problem affecting millions of people worldwide. Its prevalence increases with age and it is almost always a disease of the elderly. The impact of this disorder lies not only in the loss of patient autonomy and caregiver burden that ensue as it progresses but also in its marked economic effects. With improving health care management in both affluent and developing countries, a rise in the aging population and subsequent rise in dementia cases are anticipated.

Alzheimer's disease (AD) is the leading cause of dementia. In the United States alone, AD was estimated to affect 4 million Americans in 1990 [1]. This number is expected to rise to 8.5 million by the year (2030) [2] and 14 million by the year (2050) [1]. The overall prevalence rate of AD is 2–3% at age 65 [3]. This is estimated to double every 5 years so that almost 50% of individuals 85 years and older may be affected by this disorder [3]. The United States is spending \$100 billion per year to care for individuals with AD [4, 5].

The large population affected by dementia and the unmet need for more efficacious treatment have led to randomized controlled trials. The largest numbers of trials have been done with cholinesterase inhibitors (ChE-Is), the first class of agents approved by the U.S. Food and Drug Administration (FDA) for the treatment of AD. Recently, trials led to the approval of memantine, an *N*-methyl-*D*-aspartate (NMDA) antagonist as a treatment for patients with moderate to severe AD. These trials of approved agents are very influential in determining how future trials of antidementia agents will be conducted for AD, vascular dementia (VaD), and other entities such as mild cognitive impairment (MCI). Lessons learned from these trials will guide trial conduct not only for symptomatic agents with AD but also for compounds that may have disease-modifying effects. This chapter reviews published trials to derive guidelines on how future trials may be conducted. We concentrate on trials of AD and MCI only.

### 10.12.2 DEFINING AD, MCI, AND VaD

Precision in clinical diagnosis is essential to ensure valid outcomes in clinical trials. Randomized controlled trials in AD and MCI have utilized various diagnostic criteria to guarantee subject homogeneity.

Mild cognitive impairment had been a diagnostic dilemma from the time that the term was coined. Most regard it as a transition state between normal aging and early AD. More recently, new concepts of MCI emerged, making assessment and management more complex. Some regard MCI as incipient AD [6] (see Table 1).

Currently, MCI can be classified on the basis of the affected cognitive domain/s [7]—single memory domain, single non-memory domain, and multiple cognitive domain with or without involvement of memory. The amnesic type or the single memory domain deficit is the one that is closely correlated with AD. The rest of the MCI types may lead to either AD or other dementia syndromes, broadening the possibility of patient outcome. At the moment, therapeutic trials have limited recruitment to the amnesic type of MCI [8, 9] since this is the most certain prelude to AD. A majority of studies have adopted delay to progression to AD using a survival type of outcome as the research design approach.



**TABLE 1 Criteria for Amnesic Mild Cognitive Impairment**

- 
1. Memory complaint, preferably corroborated by informant
  2. Impaired memory function for age and education
  3. Preserved general cognitive function
  4. Intact activities of daily living
  5. Not demented
- 

**TABLE 2 DSM-IV Criteria for Diagnosis of Alzheimer's Disease**

- 
- A. Alzheimer's disease is characterized by progressive decline and ultimately loss of multiple cognitive functions, including both:
    1. Memory impairment—impaired ability to learn new information or to recall previously learned information.
    2. At least one of the following:
      - a. Loss of word comprehension ability (aphasia)
      - b. Loss of ability to perform complex tasks involving muscle coordination (apraxia)
      - c. Loss of ability to recognize and use familiar objects (agnosia)
      - d. Loss of ability to plan, organize, and execute normal activities
  - B. The problems in A represent a substantial decline from previous abilities and cause significant problems in everyday functioning.
  - C. The problems in A begin slowly and gradually become more severe.
  - D. The problems in A are not due to:
    - Other conditions that cause progressive cognitive decline, among them stroke, Parkinson's disease, Huntington's chorea, brain tumor, etc.
    - Other conditions that cause dementia, among them hypothyroidism, HIV infection, syphilis, and deficiencies in niacin, vitamin B<sub>12</sub>, and folic acid
  - E. The problems in A are not caused by episodes of delirium.
  - F. The problems in A are not caused by another mental illness: depression, schizophrenia, etc.
- 

Source: From [10].

Alzheimer's disease is a progressive degenerative disorder which leads to cognitive decline that is severe enough to cause functional deterioration. Several criteria are available to clearly define this entity: *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV) [10] (Table 2); *International Classification of Diseases*, tenth revision (ICD-10) [11]; and the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [12] criteria (Table 3). The level of diagnostic certainty is indicated in the NINCDS-ADRDA criteria. Neuroimaging has been included for the purpose of excluding other dementias in the differential diagnoses. Its postconsensus sensitivity and specificity of the diagnosis of AD range from 0.83 to 0.95 and 0.79 to 0.84 [13–15] while clinicopathological sensitivity and specificity are 0.64–0.86 to 0.89–0.91 [14, 16]. Interrater reliability is moderate [14]. The NINCDS-ADRDA criteria for AD has been regularly used in trials due to its established validity.

### 10.12.3 SEVERITY OF DEMENTIA

It is essential for dementia severity to be specified in clinical trials to ensure patient homogeneity and gauge treatment response. The spectrum of dementia depends on

**TABLE 3 NINCDS/ADRDA Criteria for Diagnosis of Probable Alzheimer's Disease**


---

|       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I.    | Dementia established by clinical examination and documented by a standard test of cognitive function (e.g., Mini-Mental State Examination, Blessed Dementia Scale) and confirmed by neuropsychological tests                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| II.   | Significant deficiencies in two or more areas of cognition, for example, word comprehension and task completion ability                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| III.  | Progressive deterioration of memory and other cognitive functions                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| IV.   | No loss of consciousness                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| V.    | Onset from age 40 to 90, typically after 65                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| VI.   | No other diseases or disorders that could account for the loss of memory and cognition                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| VII.  | Diagnosis of probable AD is supported by: <ol style="list-style-type: none"> <li>1. Progressive deterioration of specific cognitive functions: language (aphasia), motor skills (apraxia), and perception (agnosia)</li> <li>2. Impaired activities of daily living and altered patterns of behavior</li> <li>3. A family history of similar problems, particularly if confirmed by neurological testing</li> <li>4. The following laboratory results: normal cerebrospinal fluid (lumbar puncture test), normal electroencephalogram (EEG) test of brain activity, and evidence of cerebral atrophy in a series of computerized tomography (CT) scans</li> </ol> |
| VIII. | Other Features consistent with AD: <ol style="list-style-type: none"> <li>1. Plateaus in the course of illness progression</li> <li>2. CT findings normal for the person's age</li> <li>3. Associated symptoms, including depression, insomnia, incontinence, delusions, hallucinations, weight loss, sex problems, and significant verbal, emotional, and physical outbursts</li> <li>4. Other neurological abnormalities, especially in advanced disease: including increased muscle tone and a shuffling gait</li> </ol>                                                                                                                                       |
| IX.   | Features that decrease the likelihood of AD: <ol style="list-style-type: none"> <li>1. Sudden onset</li> <li>2. Such early symptoms as seizures, gait problems, and loss of vision and coordination</li> </ol>                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

---

*Source:* From [12].

its longitudinal course, which is defined by three domains: cognition, behavior, and function. Generally, dementia severity can be divided into three stages. In the early and mild stage the initial manifestations of minor memory impairment emerge with concomitant decline in complex activities. In the moderate stage, behavioral changes become more prominent, with involvement of cognitive domains other than memory, typically language and visuospatial skills. Behavioral changes become more apparent at this point and patients have more difficulty in coping with daily activities such as housework and hobbies. In the severe stage, patients are unable to live without assistance and they manifest with more disturbing behavioral symptoms. Institutionalization is common in this stage of the illness.

The Global Deterioration Scale (GDS) [17], Clinical Dementia Rating (CDR) [18], and Mini-Mental State Examination (MMSE) [19] are instruments that are frequently used for assessing dementia severity. Both the GDS and CDR are global assessment instruments which examine cognition, function, and behavior. The MMSE, however, is limited to the measurement of cognition in terms of orientation, attention, memory, language, and figure copying. It has been regularly used as an instrument for the staging of dementia since it examines the “core” manifestations of the disorder which are the main target of therapeutic trials.

The MMSE was formulated by Folstein as a practical method of grading the cognitive state [19] of psychiatric inpatients. Advantages of this instrument are its

**TABLE 4 MMSE Range of Subjects in AD and MCI Randomized Controlled Trials**

| Diagnosis                               | Agent                  | RCTs     | Duration of RCT (weeks)                      | MMSE/sMMSE Range  |
|-----------------------------------------|------------------------|----------|----------------------------------------------|-------------------|
| Alzheimer's disease<br>Mild to moderate | Metrifonate            | [21]     | 30 (12 <sup>a</sup> ), 36 (26 <sup>a</sup> ) | 10–26             |
|                                         | Tacrine                | [22–25]  | 6, 12, 30                                    | 10–26             |
|                                         | Donepezil              | [26–28]  | 12, 24, 52                                   | 10–26             |
|                                         | Galantamine            | [29]     | 20                                           | 10–22             |
|                                         |                        | [30, 31] | 12, 24                                       | 11–24             |
|                                         | Rivastigmine           | [32, 33] | 26, 26                                       | 10–26             |
|                                         | Diclofenac/Misoprostol | [34]     | 25                                           | 11–25             |
|                                         | Prednisone             | [35]     | 52                                           | 13–26             |
|                                         | Estrogen               | [36]     | 52                                           | 14–28             |
|                                         | Rofecoxib or naproxen  | [37]     | 52                                           | 13–26             |
|                                         | Rofecoxib              | [38]     | 52                                           | 14–26             |
|                                         | Acetyl-L-carnitine     | [39]     | 52                                           | 13–26             |
|                                         | Donepezil              | [40]     | 24                                           | 5–17 <sup>b</sup> |
|                                         | Memantine              | [41]     | 28                                           | 3–14              |
| Mild cognitive impairment               | Donepezil              | [8, 9]   | 24, Survival design (time to reach endpoint) | ≥24               |

Note: SMMSE: standardized Mini–Mental State Examination; RCT, randomized controlled trial.

<sup>a</sup>Duration of active treatment.

<sup>b</sup>In this trial sMMSE was used instead of MMSE.

brevity and ease of administration. It has an adequate sensitivity of 86% and a specificity of 92% when a cutoff score of 23–24/30 is used [20]. Its ceiling and floor effect, though, lessens its sensitivity in detecting mild and severe cognitive impairment, respectively. MMSE scores are influenced by age and education. The presence of within-subject and between-subject variability challenges the applicability of MMSE scores; they have wide standard errors of measurement. The natural variability of the disease also contributes to score fluctuations. Despite these caveats, the use of MMSE is widespread in the staging of dementia and is commonly used to define a restricted range of dementia severity for patients included in clinical trials (Table 4).

#### 10.12.4 ETHICAL CONDUCT OF DEMENTIA TRIALS AND INFORMED CONSENT

Research trials involve experimentation and, to protect human subjects, various guidelines have been implemented. As in all research studies involving human subjects, the principles of minimization of harm, beneficence, and veracity are maintained in dementia trials. Conduct of dementia trials is complicated by inclusion of subjects with varying severity of cognitive impairment, a condition that makes them vulnerable to exploitation.

A multitude of “ethically approved” randomized controlled trials in dementia therapy have been conducted in the past years. From these studies, acetylcholinesterase inhibitors have been approved as a class of drugs that is effective and safe in AD patients. In 2001, the American Academy of Neurology issued guidelines in the management of AD and it described acetylcholinesterase inhibitors as the standard drug for this disorder [42]. Thus, assignment of subjects to placebo in future clinical

trials will be unethical since a standard drug is already available. Administration of this standard drug to all the groups in a dementia clinical trial becomes necessary.

Determination of the capacity to consent is the most discussed aspect of dementia trials. Differing levels of disease severity produce differing levels of incapacity and challenge the ability of the patient to participate in consent discussions. The presentation of the purpose, methodology and procedures, risks, benefits, and alternatives must be very clear and simple for cognitively impaired subjects to comprehend. The extent of the subject's grasp of the details of the study and decision-making capacity should be evaluated. This involves examination of (1) understanding of the relevant facts of the trial disclosed to the patient, (2) appreciation of the research risks and potential benefits, and (3) reasoning in terms of comparing options and drawing consequences from these options [43].

If a subject is deemed incompetent to give informed consent, proxy consent is obtained from an individual who has the capacity and legal authority to give it. Typically, the caregiver is given this task since he or she is the one who addresses the daily needs of the subject and knows the latter's preferences. Competence and benevolence of the caregiver must be ensured in these circumstances.

In the mild to moderate stages of the disease process, the subject will still be able to contribute to research trial decisions. The investigator must ensure that the subject is still part of the choices made by the proxy. Later, the subject loses the ability to participate meaningfully and the caregiver becomes the sole decision maker. It is at this point that assent (or dissent) should be obtained from the subject. This is judged behaviorally based on the cooperativeness of the subject with the study procedures [43]. Consistent resistance of the subject with study procedures may be taken as dissent, a probable basis for discontinuation from the study.

### 10.12.5 GENERALIZABILITY OF CLINICAL TRIAL RESULTS

A drug's value is ultimately tested once it is marketed in the targeted population. The terms "efficacy" and "effectiveness" distinguish between how well the intervention/drug *can* work under ideal circumstances such as that in clinical trials and how the intervention/drug *does* work under "field conditions" such as that in the community [44]. The drug's applicability depends on the inherent characteristics (demographic and medical) of the cohort of subjects enrolled in the clinical trial and how representative they are of the population who will eventually use it. Subject selection by using inclusion and exclusion criteria potentially limit the generalizability of the results of the clinical trial. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (NIMH-CATIE) is an example of an effectiveness study proposal of psychotropic use in AD with a methodological design that assures applicability of the results in the community setting [45].

In randomized controlled trials of acetylcholinesterase inhibitors, recruitment bias was observed toward subjects who are healthier, better educated, younger, and of higher socioeconomic status [46, 47]. This has the effect of excluding subjects who have complicated medical histories and are taking specific medications. Caregivers may be more aggressive in seeking medical help when their relatives are in the early stage of dementia, exposing the patients to a higher probability of being recruited.

Another confounding factor on the generalizability of trials is the drug's eventual applicability in different ethnic populations. Randomized controlled trials require major funding and, for this reason, these are conducted in affluent countries. This fundamentally explains why Caucasians compose the greater part of the subjects in trials globally. However, there is a mixture of ethnic populations in the United States and Caucasians should not dominate disproportionately. Examination of the population recruited in acetylcholinesterase inhibitor trials which were conducted in the United States show that the samples were 91–99% white [46]. Problem may arise due to differences in the pharmacokinetics of the drug in various ethnic populations resulting in varied first-pass metabolism and systemic bioavailability. Activity of the cytochrome P450 enzymes differ among various ethnic groups, possibly leading to disparate side-effect profiles, dosaging, and efficacy.

Highly specialized centers conduct clinical trials. Involved physicians have expertise in handling dementia cases and are highly motivated to recruit patients. Complete laboratory and neuroimaging work-up are also on hand. Subjects in these centers are diagnosed with precision and handled differently as compared to those in the community setting. There is stronger and more accessible patient support system in trial centers. This may have great impact in the care of dementia patients since it would result in better compliance and fewer problem behaviors. This setting again distinguishes clinical trials from routine care.

### 10.12.6 OUTCOME ASSESSMENTS IN DEMENTIA TRIALS

Alzheimer's disease affects primarily the memory with subsequent involvement of other cognitive domains. This change causes concomitant deterioration in function and behavior. The efficacy of antidementia drugs therefore should be evaluated based on these domains: cognition, behavior, and function.

The efficacy of an antidementia drug is measured through various outcome assessment scales and psychometric tests. As required by the FDA, the cognitive improvement that results from drug administration should be supported by (1) positive change in a performance-based cognitive instrument and (2) clinically meaningful effect seen globally [48]. These two factors comprise the primary outcome measures in dementia clinical trials. Secondary outcome measures are included to further document the effect of a drug on the other aspects of the subject's life. These secondary outcome measures do not need to be positive for the drug to be approved and marketed. Assessment scales examining the subject's quality of life, noncognitive behavioral symptoms, and economic impact of the illness usually compose the secondary outcomes.

Outcome measures must possess certain properties before these can be employed in clinical trials. The instruments must have been proven valid, reliable, and sensitive. Validity confirms whether the tool measures what it is intended to measure. Reliability (test/retest, interrater and/or intrarater) refers to the replicability of results so that the same value will be obtained under multiple circumstances. Sensitivity is the capacity to detect change over time, especially with treatment. It is best if the instruments are effective in detecting changes even in the extremes of the spectrum of the illness, eliminating floor and ceiling effects. Other important properties are ease of use and short administration time and availability of multiple equivalent forms to avoid practice effects with repeated use. Ideally, these should

be independent of demographic and socioeconomic factors such as gender, education, and cultural background.

### 10.12.6.1 Primary Outcome Measures

**Performance-Based Cognitive Assessment** Cognition is composed of multiple domains, all of which are inevitably affected at the late stage of dementia. Consensus on cognitive domains that may be assessed include the following: memory, attention, processing speed, visuospatial function, praxis, language, executive function, and abstraction [49]. A combination of different psychometric tests can fully evaluate these domains. However, it is more appropriate for disease-specific instruments to be used in a chronically progressive illness so that it can appropriately reflect the outcome at study endpoint.

The Alzheimer's Disease Assessment Scale (ADAS) was designed to evaluate the severity of both cognitive and noncognitive manifestations of AD patients [50]. Cognitive domains include memory, orientation, language, and praxis while the noncognitive domains include mood state and behavioral changes. The cognitive portion has a maximum score of 70 with a higher score indicating more severe impairment. Its advantages include short administration time, ability to detect changes from the mild to severe stages, and being appropriate for patients in different environments [50]. There is excellent interrater and test/retest reliability of the ADAS-cog at 0.99 and 0.92, respectively. It is the most widely used objective cognitive assessment scale in dementia randomized controlled trials (RCTs). The ADAS-cog is the prevailing primary outcome measure of efficacy that evaluates changes in the core manifestations of dementia. The ADAS-cog lacks tests examining executive function, a domain that is affected frequently. The Alzheimer's Disease Cooperative Study (ADCS) has extended the ADAS-cog by adding two executive tests, a cancellation and a maze task.

**Global Measures** The overall clinical impact of an antidementia drug is evaluated using global assessment scales. These assess the multidimensional manifestations of the illness in terms of cognition, behavior, and function. There are two categories of global measures: (1) global severity scales, which ascertain the absolute severity of the patient's condition, and (2) global change scales, which determine the overall improvement or deterioration of the subject. Global scales have less structure, thus partially avoiding the influence of subject characteristics and rating variance. The available global scales were specifically designed for use in AD or primary degenerative dementia so that results are specific to the illness. These are sensitive measures for long-term assessment of efficacy since changes can be quantified when other assessments are affected by floor effects.

Global assessment scales of change were developed to measure a "clinically meaningful" treatment effect that translates into practical usefulness. All of the measures are seven-point scales which are rated as follows: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, 7 = very much worse. These assess change from a specified baseline. Unlike the symptom scales, these are relatively unstructured, relying on an experienced clinician to conduct a thorough and accurate interview on which the

**TABLE 5 Global Outcome Measures (Severity and Change Scales) Utilized in AD and MCI Trials as Either Primary or Secondary Measure of Efficacy**

| Global Outcome Measures          | RCT That Used Outcome as Primary Measure of Efficacy | RCT That Used Outcome as Secondary Measure of Efficacy |
|----------------------------------|------------------------------------------------------|--------------------------------------------------------|
| Severity                         |                                                      |                                                        |
| CDR (global and/or sum of boxes) | [39]                                                 | [8, 26, 27, 35–37, 54, 58]                             |
| GDS                              | [34]                                                 | [8, 22, 25, 28, 32, 41]                                |
| Change                           |                                                      |                                                        |
| CGI [51]                         | [23, 24, 34, 55]                                     | [39]                                                   |
| CIBI [52]                        | [25]                                                 |                                                        |
| CIBIC Plus [52]                  | [21, 22, 26, 27, 29–33, 38, 40, 41]                  |                                                        |
| ADCS-CGIC [53]                   | [36], MCI version [9]                                |                                                        |

*Note:* CIBIC-plus: Clinical Interview Based Impression of Change plus caregiver information; CGIC: Clinical Global Impression of Change.

rating will be based. Even though reliability is compromised with this format, the sensitivity to measure meaningful change is retained so that it remains as one of the primary efficacy measures.

The first global assessment scale was the Clinician Global Impression (CGI), an unstructured instrument that was widely used in neuropsychopharmacological trials [52]. The CGI was first utilized in a dementia RCT involving tacrine by Davis et al. with ADAS-cog as the other primary measure of efficacy [23]. However, significant improvement was only noted in the ADAS-cog but not in the CGI, which indicated the latter's lack of sensitivity to the treatment effect of tacrine. A guideline-based global assessment scale, the Clinical Interview Based Impression (CIBI), was then developed which was used in the 30-week tacrine trial by Knapp et al. [25]. Both ADAS-cog and CIBI yielded significantly positive results which led to the FDA approval of tacrine. Several global assessment scales emerged to further improve the instrument's reliability (Table 5). To date, CIBIC plus is the most frequently used global assessment instrument in dementia RCTs.

Global severity or staging scales such as the CDR scale [18] and the GDS [17] are frequently used as entry criteria or as secondary outcome measures. The CDR is a worksheet-based semistructured interview that evaluates six domains: three cognitive (memory, orientation, judgment, and problem solving) and three functional (community affairs, home and hobbies, and personal care). Rating is based on a five-point scale in which 0 = none, 0.5 = questionable dementia, 1 = mild impairment, 2 = moderate impairment, and 3 = severe impairment. It can be scored in two ways: (1) as a sum of boxes (SB) by obtaining the sum of the ratings of each of the six CDR domains/boxes and (2) as a global rating based on a scoring system wherein the memory box/domain is the main consideration. Preference for this instrument is due to its clinically based assessment and high interrater reliability resulting in a level of agreement of 80% [56]. The GDS is also a useful instrument in staging primary degenerative dementia. It is capable of accurately delineating stages of dementia throughout the course of AD [17]. It rates cognitive decline based on a seven-point scale with the following scoring system: 1 = none, 2 = very mild, 3 = mild, 4 = moderate, 5 = moderately severe, 6 = severe, 7 = very severe. Interrater and test/retest reliability are high, both at 0.92 [57].



### 10.12.6.2 Secondary Outcome Measures

**Mini-Mental State Examination** The MMSE may be included as an entry criterion or as a supplementary measure of cognition [8, 21, 23, 25–28, 32, 34, 36, 38, 39, 41, 54, 55, 58]. Compared to the ADAS-cog, its result is better understood by non-AD specialists and can be easily translated into more practical terms. However, its limited sensitivity makes it a poor primary outcome measure.

**Activities of Daily Living** Functional impairment is an essential component of the clinical syndrome of dementia. It is required for the clinical diagnosis of dementia and included in the NINCDS-ADRDA criteria [12] and the DSM-IV criteria [10]. The resulting dependence affects not only the patient but also the quality of life of the caregiver and becomes an important factor that leads to institutionalization. Changes in activities of daily living (ADLs) are frequently included as a secondary outcome measure. Functional deterioration is only moderately correlated with the cognitive status of patients with AD [59] and seems to be an expression of other integrative abilities of the individual. This makes functional assessment all the more important in clinical trials since cognitive tests cannot fully gauge improvement in other aspects of the patient's life.

Drug effect in terms of reversibility, stabilization, or slower deterioration of ADLs can be monitored through the use of functional assessment scales [60]. Different instruments for this purpose have been developed for AD (Table 6), incorporating either basic or instrumental/complex ADLs or both. It is necessary for complex ADLs to be incorporated in the examination since functional deterioration occurs in hierarchical order, initially affecting difficult tasks before simpler ones. Ideally, caregiver information should be obtained for data reliability since loss of insight as the disease progresses makes self-report impossible.

These functional assessment scales were used as either primary or secondary outcome measures of efficacy in different dementia RCTs (Table 7)

**Neuropsychiatric Symptoms** Neuropsychiatric manifestations are common in AD. They reflect the underlying neuropathological and neurotransmitter changes in the brain. Incidence increases with disease severity but symptoms are variable and differ among afflicted individuals. Assessment of these behavioral manifestations is essential since it can influence the individual's state, further aggravating the existing cognitive and functional impairment. Presence of disruptive behavior increases caregiver burden and is one of the determinants of eventual nursing home placement. These symptoms can also predate the onset of dementia and may present in MCI.

Several instruments that measure behavioral symptoms are available and employed in dementia clinical trials. These are used as primary outcome measures when the target symptom is behavior (e.g., with psychotropic agents) and as secondary outcome measures when the target symptom is cognition (e.g., with cholinergic agents) and behavior is evaluated as an auxiliary effect. Characteristically, these can be categorized into broad-spectrum scales when they sample a comprehensive range of symptoms and focused scales when more items are dedicated into the subtleties of a particular behavioral domain [64]. Some characterize the symptoms by their frequency and severity, which is very helpful in gauging the disruptiveness of the behavior and becomes an indirect measure of caregiver burden. Frequently used

**TABLE 6 Comparison of General Characteristics and Reliability of Different Functional Assessment Scales Utilized in Dementia RCT**

| Description of Instrument          | PSMS and IADL [61]                                                                  | PDS [62]                                                                          | DAD [59]                                                                                                                                                                                                         | ADCS-ADL [63]                           | ADFACS [54]                   |
|------------------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|-------------------------------|
| General characteristic             | Caregiver questionnaire                                                             | Caregiver questionnaire developed as measure of QOL changes; bipolar analog scale | Proxy respondent scale (completed either by caregiver as questionnaire or as structured interview)                                                                                                               | Informant-based questionnaire           | Informant-based questionnaire |
| Population/environment             | AD/community and residential care                                                   | AD/not indicated                                                                  | AD/community dwelling                                                                                                                                                                                            | AD/community dwelling                   | AD/not indicated              |
| Administration time                | Not indicated                                                                       | 10–15 min                                                                         | <15 min                                                                                                                                                                                                          | 30–45 min                               | Not indicated                 |
| Gender influence                   | IADL different for males and females (5 pts. male; 8 pts. female)                   | Not indicated                                                                     | No influence (including age and education)                                                                                                                                                                       | Gender-biased items removed             | No influence                  |
| ADL inclusion                      | Both BADL and IADL                                                                  | Both BADL and IADL                                                                | Both BADL and IADL                                                                                                                                                                                               | Both BADL and IADL                      | Both BADL and IADL            |
| Number of items (score range)      | PSMS 6 items, IADL 8 items                                                          | 29 QOL factors (0–100)                                                            | 40 items (0–100)                                                                                                                                                                                                 | 24 items (0–78)                         | 16 items (0–54)               |
| Correlation with dementia severity | Not indicated                                                                       | Yes (GDS)                                                                         | Yes (MMSE and GDS)                                                                                                                                                                                               | Yes (MMSE)                              | Not indicated                 |
| Reliability                        | Test/retest: reproducibility coefficient 0.96; interrater: 0.87–0.91 (Pearson $r$ ) | Test/retest: 0.898 (Pearson product-moment correlation)                           | Test/retest: 0.96 (intraclass correlation coefficient); interrater: 0.95 (intraclass correlation coefficient)                                                                                                    | Test/retest: 0.4–0.75 ( $k$ statistics) | Not indicated                 |
| Special qualities                  |                                                                                     |                                                                                   | Evaluates aspects of activities that are impaired (initiation, planning, organization, effective performance); nonapplicable domain does not influence scoring since total score is converted into a percentage. |                                         |                               |

*Note:* PSMS: Physical Self-Maintenance Scale; IADL: instrumental activities of daily living; PDS: Progressive Deterioration Scale; DAD: Disability Assessment in Dementia; ADCS-ADL: Alzheimer's Disease Cooperative Study—Activities of Daily Living; ADFACS: Alzheimer's Disease Functional Assessment and Change Scale; BADL: Basic Activities of Daily Living; QOL: quality of life.

**TABLE 7 Functional Assessment Scales Utilized in Different Dementia RCTs as Primary or Secondary Outcome Measures of Efficacy**

| Functional Assessment Scale | RCT That Used Scale as Primary Outcome Measure | RCT That Used Scale as Secondary Outcome Measure |
|-----------------------------|------------------------------------------------|--------------------------------------------------|
| PSMS and IADL               |                                                | [21, 23, 25, 34, 39, 40]                         |
| PDS                         | [32, 33]                                       | [23, 25, 28]                                     |
| DAD                         |                                                | [22, 30, 31, 40]                                 |
| ADCS-ADL                    | Adapted to severe state [41]                   | MCI version [8]; [29, 37, 38]                    |
| ADFACS                      | [54]                                           |                                                  |

assessment scales in dementia clinical trials are the following: Neuropsychiatric Inventory (NPI) [65], Behavioral Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD) [66], CERAD Behavior Rating Scale for Dementia (CBRSD) [67], Cohen Mansfield Agitation Inventory (CMAI) [68], Cornell Scale for Depression in Dementia (CSDD) [69], Brief Psychiatric Rating Scale [70, 71], and Positive and Negative Syndrome Scale (PANSS) [72].

The NPI, BEHAVE-AD, and CBRSD are examples of broad-spectrum scales developed primarily for the evaluation of dementia patients (Table 8). These are valid and reliable instruments that can be used in clinical and research settings. The NPI is widely used in dementia clinical trials due to its comprehensive list of symptoms that commonly occur in individuals with dementia. It also has accompanying questions for every symptom so that appropriate probing can be done and includes a frequency and severity component for every symptom category. Major dementia RCTs [22, 28–30, 37, 40, 41] have used the NPI as a standard gauge of behavior outcome. CBRSD was used in the trial by Sano et al. [58] as a secondary outcome measure. The BEHAVE-AD was more frequently employed in trials involving psychotropic agents [74–76].

The CMAI and CSDD are focused scales. These are useful when particular symptoms are targeted by the clinical trial. For example, the CMAI was used as a primary outcome measure to demonstrate the efficacy of risperidone in agitation and aggression in subjects with AD [76]. The CSDD was likewise the primary outcome measure in the sertraline RCT [77] that determined the drug's effect on depression in subjects with AD.

The BPRS and PANSS are broad-spectrum scales that were developed for use in psychiatry populations. These have been employed in several dementia trials involving psychotropic agents since it covers a variety of symptoms aside from psychosis. The BPRS was the primary measure of efficacy for the control of agitation and aggression in dementia in carbamazepine [78] and divalproex sodium RCTs [79]. The PANSS has five symptomatic dimensions (negative, positive, excitation, cognitive, and depression) and the quetiapine trial [80] used the excitation component (PANSS-EC) to evaluate the drug's efficacy in agitation that occurs in dementia.

**Other Measures of Efficacy** Aside from treatment efficacy, pharmacoeconomics, caregiver burden, and quality of life in dementia have been addressed in dementia RCTs. Although these are not required as outcome measures for drug approval,

their usefulness is based on a more practical aspect of efficacy, one that can be seen and experienced in daily life. Caregiver burden scales determine the impact of dementia on the people who are considered direct health care providers. They experience stress and burden. One way of determining burden is by recording how much time is spent attending to the patient. This can be assessed using the Caregiver Activities Time Survey, the Caregiver Activity Survey [81], or the Caregiver Time Questionnaire [82].

The economic impact of dementia is overwhelming since the burden extends to the national level. It is directly related to the number of people affected by the disorder and the duration of the disease [83]. Incurred costs include (1) medical resources used to treat the illness (medication, cost of institutionalization, etc.), (2) nonmedical resources, and (3) lost productivity caused by disability (both patient and caregiver) [83]. Since maintenance of patients with dementia eventually translates into cost of care, some RCTs include pharmacoeconomic measures. The AD2000 trial [84] determined the health care cost of dementia from the societal perspective in terms of delaying institutionalization or requiring other health services and the cost-effectiveness of the drug (donepezil). Reisberg et al. [41] similarly evaluated this effect by utilizing the Resource Utilization in Dementia (RUD) [85], which assesses the burden on the caregiver and provides AD-related health economics data, in a memantine trial.

Quality-of-life scales encompass the overall effect of the treatment on the patient in terms of social, psychological, cognitive, and functional well-being. Either the patient or the caregiver's quality of life can be assessed. However, information will be derived mainly from the caregiver once the patient's communication problems emerge.

Transcultural application of these instruments may alter their validity and reliability. Interpretation of results should therefore be done with caution when these are used outside the cultures in which they were developed.

### 10.12.7 CLINICAL TRIAL DESIGNS

The FDA requires that agents tested should be superior to placebo in terms of its cognitive and global effects. The standard format of antidementia drug RCTs is the parallel group design wherein there is direct comparison between the experimental and placebo group. In this head-to-head comparison, the active drug and placebo are allocated to different subjects randomly. The 12-week tacrine study by Farlow et al. [24] utilized this design, although randomization of subjects was stratified. Alzheimer's disease drug trials (acetylcholinesterase inhibitors, NMDA antagonists, psychotropics, SSRIs, etc.) follow this standard format due to its straightforward approach. This design was also used in the MCI trial of Salloway et al. [9] comparing donepezil with placebo. Between-patient variability can deleteriously affect the outcome of parallel group studies and cannot be eliminated with this format. The crossover design eradicates this confounding variable by applying the treatment sequentially to the same subject. Aside from reducing error variance, sample size is diminished since subjects can be used several times in the analysis while increasing the statistical power of the clinical trial. However, this strategy is rarely used primarily due to difficulty in reobtaining baseline values during the

TABLE 8 Comparison of Characteristics of NPI, BEHAVE-AD, and CERAD, CBRSD

| Instrument Characteristics | BEHAVE-AD                                                                                                                              | CBRSD                                                                                                                                                                 | NPI                                                                                                                                                                                                                                                                                               |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Overview                   | Developed through retrospective review of 57 AD outpatient charts using information from nursing staff, physicians, and family members | Items from literature review and expert panel; some items drawn from other scales; items are homogeneously scaled, anchored; emphasize frequency rather than severity | Assesses a wide range of behaviors with potential to differentiate between dementia syndromes; items suggested by expert panel by "Delphi" method; format used minimizes administration time and optimizes capture of information                                                                 |
| Population                 | AD outpatients with mild to severe cognitive loss                                                                                      | AD outpatients                                                                                                                                                        | Elderly dementia patients                                                                                                                                                                                                                                                                         |
| Informant                  | Caregiver                                                                                                                              | Caregiver                                                                                                                                                             | Caregiver                                                                                                                                                                                                                                                                                         |
| Administration             | Clinician in person or by phone or self-administered                                                                                   | Structured interview by technician                                                                                                                                    | Clinician interview                                                                                                                                                                                                                                                                               |
| Administration time        | 45min                                                                                                                                  | 30 min                                                                                                                                                                | 20 min                                                                                                                                                                                                                                                                                            |
| Time interval              | 2 weeks                                                                                                                                | 1 month (behaviors occurring prior to 1 month, since dementia onset, are also noted)                                                                                  | 1 month                                                                                                                                                                                                                                                                                           |
| Items                      | 25 plus 2 global ratings                                                                                                               | 48                                                                                                                                                                    | 10 items in 5 domains: mood changes (dysphoria, euphoria), agitation (aggression, aberrant motor behavior), personality alterations (apathy/indifference, irritability/lability, disinhibition), psychosis (delusions, hallucinations), and anxiety; sleep and appetite changes included later on |

|                  |                                                                                                                                                                                            |                                                                                                                                                                                           |                                                                                                                                                              |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Frequency        | No                                                                                                                                                                                         | 0 = not since illness began; 1 = 1–2 days/past month, 2 = 3–8 days, 3 = 9–15 days, 4 => 16 days, 8 = since illness began but not in past month                                            | 1 =< 1 time/wk, 2 = 1 time/wk, 3 = several times/week but <1 time/day, 4 ≥ 1/day                                                                             |
| Severity         | Rating differs for each specific item; includes severity estimates that indirectly consider caregiver impact: absent = 0, mild = 1, moderate = 2, severe = 3                               | No                                                                                                                                                                                        | 1 = mild, 2 = moderate, 3 = severe                                                                                                                           |
| Caregiver impact | Directly assessed by items 21 and 23 and part II                                                                                                                                           | No                                                                                                                                                                                        | No                                                                                                                                                           |
| Reliability      | Interrater: $r = 0.90$ (severity ratings for scale as a whole); reliability of symptom presence or absence = 0.62–1.0                                                                      | Interrater: $k = 0.77$ –1.0 ( $n = 104$ )                                                                                                                                                 | Interrater: $r = 0.96$ –1.0 (frequency), 0.98–1.0 (severity); Test/retest: $r = 0.79$ (frequency), $r = 0.86$ (severity); Overall (Cronbach $A$ ) $r = 0.88$ |
| Validity         | Has content and construct validity (derived from patient records)                                                                                                                          | Has construct and content validity; convergent validity with CMAI                                                                                                                         | Has construct and content validity; convergent validity with HAM-D and BEHAVE-AD                                                                             |
| Scoring          | Total score or in seven domains (paranoid/delusional ideation, hallucinations, activity disturbances, aggressive behavior, sleep disturbance, affective symptoms, anxiety/phobic symptoms) | Initial study suggested eight factors: depressive symptoms, psychotic symptoms, poor self-regulation, irritability/agitation, vegetative features, apathy, aggression, affective lability | For each of the 10 items, total score = frequency $\times$ severity                                                                                          |
| Suggested use    | Demonstration of behavioral disturbance and quantification of behavioral effects of antipsychotics or other interventions                                                                  | Detailed elicitation/quantification of broad range of psychopathology in mild to moderate AD patients                                                                                     | Assessment and quantification of symptoms                                                                                                                    |

Source: From [73].

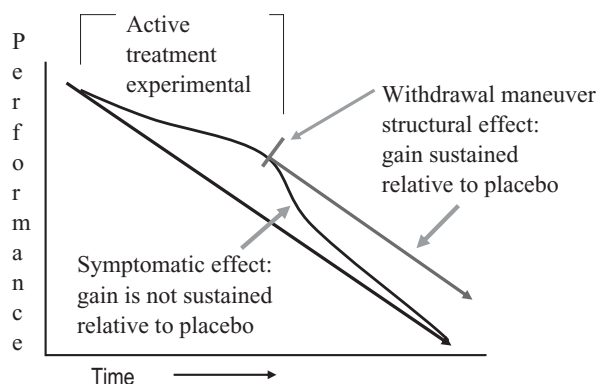


**FIGURE 1** Randomized withdrawal design. (Adapted from [86])

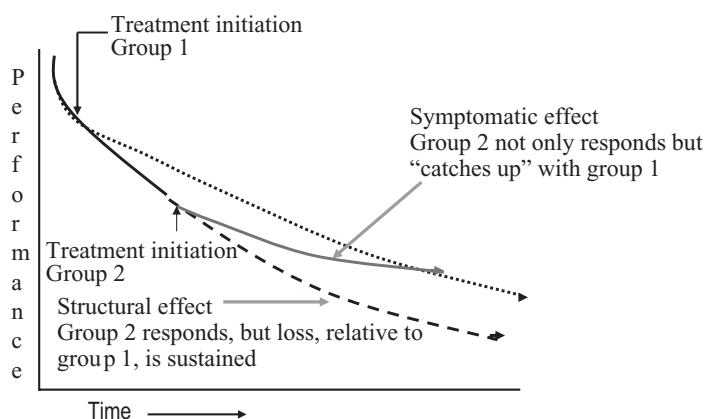


**FIGURE 2** Randomized start design. (Adapted from [86])

treatment shift. There is also uncertainty if the changes noted with the second treatment are due to the remaining or carry-over effect of the first one. The required washout periods of the drugs are unknown and may be lengthy, causing prolongation of the clinical trial.

The current trend in antidementia drug development is toward disease modification. The above-cited study designs are adequate for measuring symptomatic change (has no effect in the survival and function of degenerating neurons) and are less applicable to assessing disease modification. A variety of trial designs have been suggested to assist in establishing disease-modifying effects. Paul Leber [86] suggested the randomized withdrawal and randomized start designs to determine the disease-modifying effect of the antidementia drug (Figures 1 and 2). The randomized withdrawal design (Figure 1) is based on the hypothesis that withdrawal of a disease-modifying drug will not produce loss of the incremental changes noted compared to the withdrawal of a symptomatic drug. The major drawback of this design is the uncertainty as to how long the subjects need to be observed to determine the outcome. Consequently, it will require a considerable number of subjects since attrition is expected to be high secondary to the length of the trial and with-



drawal of a treatment that may be beneficial to the subject. Blinding is also compromised since everyone becomes aware that all patients are taking placebo at the end of the trial. These potential limitations led to the formulation of the staggered start design (Figure 2) which has two phases: The first is similar to a placebo-controlled, randomized, parallel study while the second entails all subjects to be given the active treatment. A significant drug placebo difference should be existent after the first phase before the second phase can be started. If the active drug has a disease-modifying effect, the placebo group's performance cannot catch up with that of the treatment group in the second phase of the study.

Sano et al. [87], utilized a time-oriented, valid outcome measure to assess disease progression that includes the main domains affected in AD—cognition and function. The survival design examines the longitudinal course of dementia to measure the antidementia drug effect on disease progression. Endpoints are not primarily based on neuropsychological tests, unlike in other trial designs. Instead, readily observable, clinically meaningful endpoints that represent disease progression are utilized. The primary efficacy outcome measure is the time to reach specified endpoints. Similarly, Mohs et al. [54] used a survival design to evaluate the length of time the function is preserved in AD patients with the use of donepezil.

Amnesic MCI lends itself to use of survival designs since longitudinal examination of patients can reveal either progression or nonprogression to AD. Most MCI clinical trials used this study design with progression to AD as the endpoint [8, 88]. A limitation of this scheme is the difficulty in precisely defining progression to AD.

An enrichment design was utilized in the tacrine study of Davis et al. [23]. The first or enrichment phase consisted of finding the “best dose” response of each patient to tacrine as defined by the ADAS without producing intolerable side effects. Subjects who continued in the subsequent parallel group randomized controlled trial were the “enriched population” or those who had a potential to respond to the treatment as seen in the first phase of the study. RCTs which include subjects who have a higher risk of acquiring AD such as those with positive family history, presence of ApoE4 alleles, presence of cerebrospinal fluid (CSF) AD biomarkers, hippocampal atrophy on magnetic resonance imaging/computerized tomography (MRI/CT) scan, or bilateral temporoparietal hypometabolism on positron emission tomography (PET) scan are recruiting an “enriched population.” A drawback of this study design is the potential to skew the outcome favorably toward the active drug and in the end limit the generalizability of the findings. Enrichment design for an antidementia drug is useful in proof-of-concept studies or as a means of entering the market by showing evidence of its efficacy. However, it should be followed by a trial that employs less restrictive designs so as not to limit the generalizability of findings.

#### 10.12.7.1 Special Clinical Trial Design Features

Long-term drug trials may incorporate special design features to improve accuracy of findings and to confirm additional hypotheses. Single-blind run-ins added before initiation of the double-blind phase involve administration of placebo to both treatment groups for a specified period of time without the knowledge of subjects and caregivers. Its primary aim is to minimize the “placebo effect” to which the initial changes from baseline may be contributed. By doing this, both groups equally start

at the same level. The galantamine efficacy and safety trials utilized this feature [29–31].

Single-blind washout periods are sometimes done at the end of the double-blind treatment. This entails administration of placebo to both treatment groups in an effort to determine the withdrawal effect of the study drug. In the donepezil trials [26, 27], the washout periods showed the return of the performance of the donepezil group to baseline values—an indication that the observed improvement after drug administration was a symptomatic and not a disease-modifying effect. This is also useful for the assessment of untoward reactions secondary to drug removal and to gain information on how long the clinical response is sustained after drug discontinuation.

Open-label extensions conducted after the double-blind study is a method of determining safety of prolonged treatment. Limited efficacy data are collected in open-label extensions.

### 10.12.7.2 Randomization

Randomization allocates subjects to treatment arms wherein each arm is exposed to a different condition. Usually only two arms are employed in phase III trials: experimental and control. In phase II or when several treatment strategies are available, additional treatment arms become necessary. For example, in the MCI trial by Petersen et al. [8], three arms were utilized and subjects received (1) 2000 IU of vitamin E daily, (2) 10 mg of donepezil daily, or (3) placebo. Sano et al. [58] likewise expanded the treatment allocation by means of a  $2 \times 2$  factorial design to further examine treatment combinations. Four treatment arms were used: (1) selegiline and  $\alpha$ -tocopherol, (2) placebo and  $\alpha$ -tocopherol, (3) selegiline and placebo, and (4) placebo only. Multiple treatment arms are used to compare response to different dosages of the drug [26–32, 36].

### 10.12.7.3 Length of Clinical Trials

Duration of the trial and number of recruited patients are very important factors that contribute to the validity of the outcomes of the RCTs. Duration of phases II and III should be long enough to ensure that the efficacy and safety of the drug are adequately explored. Efficacy of antidementia drugs cannot be seen before 3 months of drug exposure and may be evident only after 6 months. Although longer duration (>6 months) is ideal, the number of subjects completing the study and their compliance are concerns in longer studies.

Adequate clinical trial length is essential in demonstrating treatment effect in a progressively deteriorating disorder. Drug effect (symptomatic vs. disease modifying), target symptoms (cognitive vs. noncognitive/behavioral), and outcome measures (biological vs. assessment scales) are important factors that should be considered in determining the trial length. Trials with insufficient duration can yield inaccurate and potentially misleading results in view of the fact that small but important changes from baseline may remain undetected. Longer trials ( $\geq 1$  year) are ideal considering the chronicity of the disorder being examined, but ethical, attrition, and compliance problems may be encountered which could com-

promise study interpretation. Winblad et al. [28] conducted the first published long-term (1-year) efficacy and safety study of donepezil on AD while the AD (2000) trial [84] has extended the duration to 2 years. Most trials involving agents that may have a neuroprotective or disease-modifying effect are conducted for 1 year (Table 4) since time must be allotted for their structural effect to become apparent.

Neuropsychiatric symptoms are encountered throughout the spectrum of dementia and may even herald its onset. Numerous studies have been conducted involving psychotropics (typical and atypical), anticonvulsants (carbamazepine and valproate), and antidepressants [selective serotonin reuptake inhibitors (SSRIs)] to determine which agent alleviates these symptoms. Patterns of analysis may be either (1) reduction in emergence of behavioral symptoms wherein asymptomatic patients are followed up longitudinally to determine which regimen has fewer symptoms at endpoint or (2) comparing which regimen produced more change/reduction in the behavioral symptoms from baseline to endpoint. The length of the trials is notably shorter than those addressing cognitive symptoms, ranging from 6 to 12 weeks, since acute symptomatic improvement is the goal.

### 10.12.8 STATISTICAL ANALYSES

Subject noncompliance and dropouts always complicate clinical trials. These subjects cannot be excluded from analysis since they may have demographic or disease characteristics that are different from those who completed the trial and adhered to the treatment randomly assigned to them. It is for this reason that most clinical trials use the intention-to-treat principle, which provides unbiased and reliable interpretation of treatment effect. Intention-to-treat analyses typically include all subjects who were randomized to treatment, received at least one study drug dose, and provided a baseline assessment and at least one postbaseline assessment. It is a conservative method that makes it possible for the potential treatment benefit on patients to be evaluated regardless of whether or not they completed the study. With this, the random assignment of subjects to treatment groups is preserved during data analysis and potential bias is reduced [44]. This differs from the fully evaluable (per-protocol) population analysis wherein those who completed the entire phase of the study and remained compliant to the treatment regimen based on compliance rules set prior to study initiation are the only ones included.

Analysis of longitudinal data can be problematic since missing data are evident. Different methods of statistical analysis have been adopted to treat these data sets. The most widely used technique that addresses missing data is the last observation carried forward (LOCF), a method in which the subject's last available assessment is imputed for all remaining unobserved response measurements. It has the advantage of preserving the sample size, but it presumes that the subjects' responses have been constant from the last observed value to the trial endpoint. This unwarranted assumption about the missing data may result in either underestimating or overestimating the treatment effects. Type I errors can be generated from this so that a treatment difference can be falsely endorsed when in fact there is none. Despite these caveats, the method is frequently used due to its simplicity and ease of implementation and relatively conservative method of treating data. On the other hand, an observed cases (OC) analysis utilizes only the data of subjects remaining in the

trial at a specified point in time. In this method, a direct relationship of the data used and the obtained results is observed. However, loss of power and subsequent validity of results may occur since data of the noncompleters are unexploited. Results that are statistically significant for both types of analysis clearly support their accuracy. Conversely, cautious interpretation of results should be done when results are significant only in the OC analysis.

All statistical analysis should be planned in detail prior to the initiation of the clinical trial since changes afterward may introduce bias in the system. However, changes that are made prior to breaking of the blind still have limited implications for study interpretation. Analyses that are made afterward are less compelling. This also applies for specification of subgroup analysis, which is done on the basis of an expectation of a larger treatment effect in some subgroups than in others [44].

Clinical trials with survival data require different statistical treatment since subjects have varying endpoints producing asymmetries in data distribution. If the data from the subjects who did not make it to the endpoint are excluded, bias may be introduced into the results. Studies using the survival design have three general objectives: to estimate the time to event, to compare the time to event between/among the groups, and to determine the relationship of the covariables to the time to event. The hazard ratio and survival time are two important functions that are examined to generate the answers to these questions [89]. The Cox proportional-hazards regression and Kaplan–Meier analysis were used to evaluate these, respectively, in the trials of Mohs et al. [54], Sano et al. [58], and Petersen et al. [8]. The Cox proportional-hazards model controls for any bias in the predetermined covariates among the treatment arms since these change over time. It is used to estimate the hazard ratio, which is the risk of progression to an event over time in the treatment group versus the placebo group. The Kaplan–Meier method provides survival time estimates to clinically evident decline or to a chosen event (e.g., death, institutionalization, or loss of the ability to perform basic ADLs).

An adequate sample size ensures that the treatment effect can be reliably derived from the clinical trial at a specified endpoint. For ethical reasons, the sample size should be well justified; samples too small or too large are not warranted. Factors that affect sample size determination are the power of the study to detect a drug–placebo difference and the chosen level of significance of the statistical tests. Confounders and attrition rate should be considered in sample size determination, for which it should be adjusted appropriately.

The study should have enough “power” to accurately detect the smallest possible difference in the primary outcome measure that has clinical significance produced by the treatment. Power is usually set at 80% so that there is a 20% probability of missing the difference between the treatment and placebo group. Some clinical trials use a power of 90% to further reduce the chance of a false-negative result. The level of significance, or *p* value, is the probability of incorrectly identifying a treatment difference between the treatment and placebo arm when actually there is none (false-positive result). By convention, a value of  $\leq 0.05$  is frequently used. Sample size is inversely proportional to the chosen level of significance while it is directly proportional to the power of the study.

Sample size calculation is based on the primary outcome measure and how much change is required to produce a clinically meaningful effect. Previous phases II and

III clinical trials and longitudinal studies establish these changes. For example, in the ADAS-cog a four-point change is utilized for 6-month trials and seven-point change for 1-year trials for clinically significant change to be detected [90]. For the CIBIC-plus a 0.3–0.4 change is usually targeted [26, 40]. In the 1-year study of Mohs et al. [54] where preservation of function was determined as an effect of donepezil, power was calculated based on functional performance, that is, the 1-year value for significant functional decline in the placebo and donepezil group based on a previous study. Most base the power of the study on one of the primary outcomes (either ADAS-cog or the global assessment) while some base it on dual outcomes (both ADAS-cog and global assessment) [21, 22].

### 10.12.9 DRUG–PLACEBO DIFFERENCE

The drug–placebo difference is the discrepancy between the deterioration of the placebo group and the improvement, stabilization, or reduced deterioration in the treatment group [46]. It is derived by determining the difference between the mean change from baseline scores of the actively treated and placebo groups at a specified endpoint. The FDA requires proof of efficacy in terms of statistically significant improvement on specified outcome measures in the treatment group.

The effect size is the definitive basis of a drug's efficacy. It is determined by dividing the drug–placebo difference by the standard deviation using specific outcome measures.

A summary of the drug–placebo differences of the RCTs in MCI and AD based on outcome measures can be seen in Tables 9, 10, and 11. The drug–placebo difference varies among the studies and among the class of therapeutic agents. The range of treatment effect produced by acetylcholinesterase inhibitors in the ADAS-cog is 2–3.9 and 0.2–0.47 in the global assessment scales among subjects with mild to moderate AD. Among the other class of therapeutic agents, only Ginkgo biloba [55] produced significant drug–placebo difference in the ADAS-cog, but this was not supported by the global evaluation.

### 10.12.10 PLACEBO RESPONSES

Use of a placebo arm in a RCT is a standard procedure as long as it is ethical and feasible. Comparison of drug–placebo outcomes determines the investigational drug's efficacy. In AD, it is expected that an efficacious drug will produce improvement or stabilization in primary assessments while the placebo group continues its course of deterioration.

Irregularities in the placebo response are apparent in some clinical trials (Figure 3). Factors that contribute to this effect are fluctuation of the symptoms, methodological inconsistencies, and the beneficial effects of improved medical care provided during the study. It is hypothesized that placebo effects or trial effects can result from the attention that subjects receive from health care providers involved in the study. In some studies, the occurrence of adverse events in the placebo group can approach the level of occurrence observed in the treatment group. Placebo and trial effects may account for the initial improvement that is seen in the placebo

**TABLE 9 Drug–placebo Differences on ADAS-cog and Global Assessment Scales among Various Therapeutic RCTs in MCI and Mild to Moderate AD (Intent-to-Treat Analyses<sup>a</sup>)**

| Agent/Study                     | ADAS-cog          |                                   |                   | Global Assessment Scale<br>(CGI, CGIC, CIBIC-Plus) |
|---------------------------------|-------------------|-----------------------------------|-------------------|----------------------------------------------------|
|                                 | D/P<br>Difference | Significance<br>( <i>p</i> Value) | D/P<br>Difference | Significance<br>( <i>p</i> Value)                  |
| Tacrine                         |                   |                                   |                   |                                                    |
| Farlow et al. <sup>b</sup> [24] |                   |                                   |                   |                                                    |
| 20 mg                           | 0.9               | NS                                | 0                 | NS                                                 |
| 40 mg                           | 1.4               | NS                                | 0.1               | NS                                                 |
| 80 mg                           | 3.8               | 0.015                             | 0.5               | 0.015                                              |
| Davis et al. [23]               |                   |                                   |                   |                                                    |
| 40 or 80 mg                     | 2.4               | <0.001                            | NR                | NS                                                 |
| Knapp et al. [25]               |                   |                                   |                   |                                                    |
| 80 mg                           | 1.4               | NS                                | 0.1               | NS                                                 |
| 120 mg                          | 2                 | 0.008                             | 0.2               | 0.04                                               |
| 160 mg                          | 2.2               | 0.002                             | 0.2               | 0.04                                               |
| Donepezil                       |                   |                                   |                   |                                                    |
| Rogers et al. [27]              |                   |                                   |                   |                                                    |
| 5 mg                            | 2.49              | <0.0001                           | 0.36              | 0.0047                                             |
| 10 mg                           | 2.88              | <0.0001                           | 0.44              | 0.0001                                             |
| Rogers et al. [26]              |                   |                                   |                   |                                                    |
| 5 mg                            | 2.5               | <0.001                            | 0.3               | 0.003                                              |
| 10 mg                           | 3.1               | <0.001                            | 0.4               | 0.008                                              |
| Galantamine                     |                   |                                   |                   |                                                    |
| Tariot et al. [29]              |                   |                                   |                   |                                                    |
| 16 mg                           | 3.1               | <0.001                            | 0.41              | <0.001                                             |
| 24 mg                           | 3.1               | <0.001                            | 0.44              | <0.001                                             |
| Raskind et al. [31]             |                   |                                   |                   |                                                    |
| 24 mg                           | 3.9               | <0.001                            | 0.28              | <0.01                                              |
| 32 mg                           | 3.4               | <0.001                            | 0.29              | <0.05                                              |
| Rockwood et al. [30]            |                   |                                   |                   |                                                    |
| 24 or 32 mg                     | 1.7               | <0.01                             | NR                | 0.01                                               |
| Rivastigmine                    |                   |                                   |                   |                                                    |
| Corey-Bloom et al. [32]         |                   |                                   |                   |                                                    |
| 1–4 mg                          | 1.73              | NR                                | 0.26              | NR                                                 |
| 6–12 mg                         | 3.78              | <0.001                            | 0.29              | <0.01                                              |
| Rosler et al. [33]              |                   |                                   |                   |                                                    |
| 1–4 mg                          | 0.03              | NS                                | 0.14              | NS                                                 |
| 6–12 mg                         | 1.6               | NS                                | 0.47              | <0.001                                             |
| Metrifonate                     |                   |                                   |                   |                                                    |
| Cummings et al. [21]            |                   |                                   |                   |                                                    |
| 10–20 mg                        | 1.5               | 0.02                              | 0.04              | NS                                                 |
| 15–25 mg                        | 1.3               | NS                                | 0.29              | 0.005                                              |
| 30–60 mg                        | 2.94              | 0.0001                            | 0.35              | 0.0007                                             |
| Morris et al. [22]              |                   |                                   |                   |                                                    |
| 30–60 mg                        | 2.86              | 0.0001                            | 0.28              | 0.0071                                             |
| Mulnard et al. [36]             |                   |                                   |                   |                                                    |
| Estrogen 0.625 or 1.25 mg/d     | 2                 | NS                                | 0.1               | NS                                                 |
| Le Bars et al. [55]             |                   |                                   |                   |                                                    |
| Ginkgo biloba 120 mg/d          |                   |                                   |                   |                                                    |
| a. AD and MID                   | 1.4               | 0.04                              | 0                 | NS                                                 |
| b. AD only                      | 1.7               | 0.02                              | 0                 | NS                                                 |
| Reines et al. [38]              |                   |                                   |                   |                                                    |
| Rofecoxib 25 mg/d               | 0.6               | NS                                | 0.03              | NS                                                 |
| Scharf et al. [34]              |                   |                                   |                   |                                                    |
| Diclofenac/Misoprostol          | 1.14              | NS                                | 0.24              | NS                                                 |

Note: Only studies with complete data are included. ADAS-cog: Alzheimer's Disease Assessment Scale; D/P: drug–placebo.

<sup>a</sup>Except Farlow et al. [24].

<sup>b</sup>Evaluable population analysis.

**TABLE 10 Drug–Placebo Differences on MMSE among Various Therapeutic Randomized Controlled Trials in MCI and Mild to Moderate AD (Intent-to-Treat Analyses<sup>a</sup>)**

| Agent/Study                      | MMSE           |                                |
|----------------------------------|----------------|--------------------------------|
|                                  | D/P Difference | Significance ( <i>p</i> Value) |
| <b>Tacrine</b>                   |                |                                |
| Farlow et al. <sup>b</sup> [24]  |                |                                |
| 20 mg                            | 0.1            | NS                             |
| 40 mg                            | 0.4            | NS                             |
| 80 mg                            | 0.7            | NS                             |
| Davis et al. [23]                |                |                                |
| 40 or 80 mg                      | 0.7            | NS                             |
| Knapp et al. [25]                |                |                                |
| 80 mg                            | 0.6            | NS                             |
| 120 mg                           | 0.4            | NS                             |
| 160 mg                           | 0.9            | 0.02                           |
| <b>Donepezil</b>                 |                |                                |
| Rogers et al. [27]               |                |                                |
| 5 mg                             | 1.21           | 0.0007                         |
| 10 mg                            | 1.36           | 0.0002                         |
| Rogers et al. [26]               |                |                                |
| 5 mg                             | 0.96           | <0.004                         |
| 10 mg                            | 1.26           | <0.001                         |
| Mohs et al. [54]                 |                |                                |
| 10 mg                            | NR             | NS                             |
| Winblad et al. [28]              |                |                                |
| 10 mg                            | NR             | Significant                    |
| <b>Rivastigmine</b>              |                |                                |
| Rosler et al. [33]               |                |                                |
| 1–4 mg                           | 0.15           | NR                             |
| 6–12 mg                          | 0.68           | <0.05                          |
| <b>Metrifonate</b>               |                |                                |
| Cummings et al. [21]             |                |                                |
| 10–20 mg                         | 1.11           | 0.003                          |
| 15–25 mg                         | 0.63           | NS                             |
| 30–60 mg                         | 1.37           | 0.003                          |
| Morris et al. [22]               |                |                                |
| 30–60 mg                         | 0.43           | NS                             |
| Sano et al. [58]                 |                |                                |
| Vitamin E, selegiline, or both   | NR             | NS                             |
| Mulnard et al. [36]              |                |                                |
| Estrogen 0.625 or 1.25 mg/d      | 0.4            | NS                             |
| Thal et al. [39]                 |                |                                |
| Acetyl-L-carnitine 3 g/d         | 0.4            | NS                             |
| Reines et al. [38]               |                |                                |
| Rofecoxig 25 mg/d                | 0.44           | NS                             |
| Scharf et al. [34]               |                |                                |
| Diclofenac/misoprostol           | 1.37           | NS                             |
| Feldman et al. <sup>c</sup> [40] |                |                                |
| Donepezil 10 mg                  | 1.79           | <0.0001                        |
| Reisberg et al. [41]             |                |                                |
| Memantine 20 mg                  | 0.7            | NS                             |
| Petersen et al. [8]              |                |                                |
| Donepezil 10 mg                  | 0.44           | NS                             |

Note: Only studies with complete data are included. NS, not significant; NR, not reported.

<sup>a</sup>Except Farlow et al. [24].

<sup>b</sup>Evaluable population analysis.

<sup>c</sup>sMMSE—standardized Mini–Mental State Examination.



TABLE 11 Comparative Effects of Therapeutic Agents in AD Clinical Trials on ADAS-cog and MMSE at 6 months (Intent-to-Treat Analyses<sup>a</sup>)

| Agent/<br>RCT                  | Year | Treatment<br>Duration (weeks) | ADAS-cog Drug–<br>Placebo Difference | MMSE Drug–<br>Placebo Difference |
|--------------------------------|------|-------------------------------|--------------------------------------|----------------------------------|
| Metrifonate [22],<br>30–60 mg  | 1998 | 26                            | 2.86                                 |                                  |
| Tacrine [25]<br>80 mg          | 1994 | 30                            | 2.2                                  | 0.8                              |
| 120 mg                         |      |                               | 3.1                                  | 0.9                              |
| 160 mg                         |      |                               | 4.2                                  | 2.5                              |
| Donepezil [28]<br>10 mg        | 2001 | 52                            | —                                    | 1.4                              |
| Rivastigmine<br>[33]<br>1–4 mg | 1999 | 26                            | 0.03                                 | 0.15                             |
| 6–12 mg<br>[32]                | 1998 | 26                            | 1.6                                  | 0.68                             |
| Galantamine [31]<br>24 mg      | 2000 | 52                            | 3.9                                  |                                  |
| 32 mg                          |      |                               | 3.4                                  |                                  |
| Ginkgo biloba [55]             | 1997 | 52                            | 1.33                                 |                                  |
| Estrogen [36]                  | 2000 | 52                            | 1.3                                  | 0.7                              |

Note: Included trials have duration of ≥6 months and either have data at 6 months or have graphs from which data can be extrapolated.

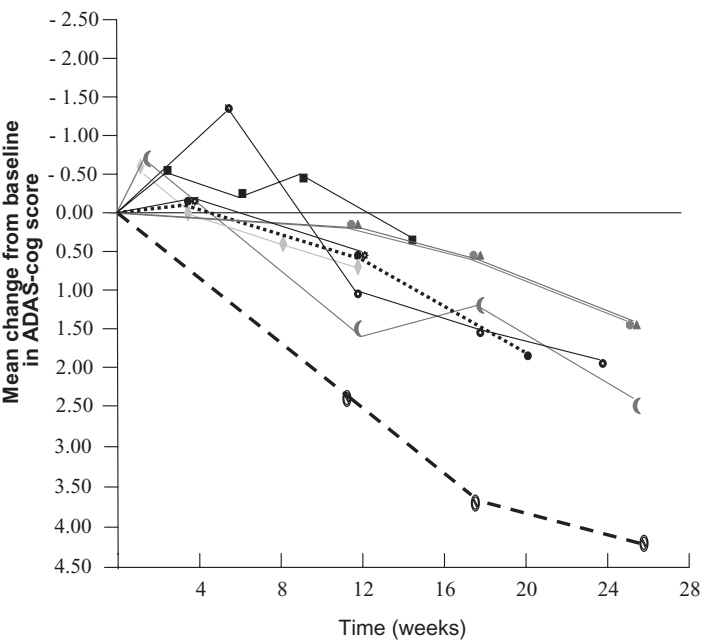


FIGURE 3 Mean change from baseline (intent-to-treat analyses) in ADAS-cog scores of placebo groups of different acetylcholinesterase inhibitor clinical trials with duration of 3–6 months in mild to moderate AD (mean change from baseline derived from available data while other values were extrapolated from graphs): ▲ [31], ■ [33], ● [22], ⋯ [29], ○ [27], ◆ [21], ✱ [30], ■ [26], ◊ [32].

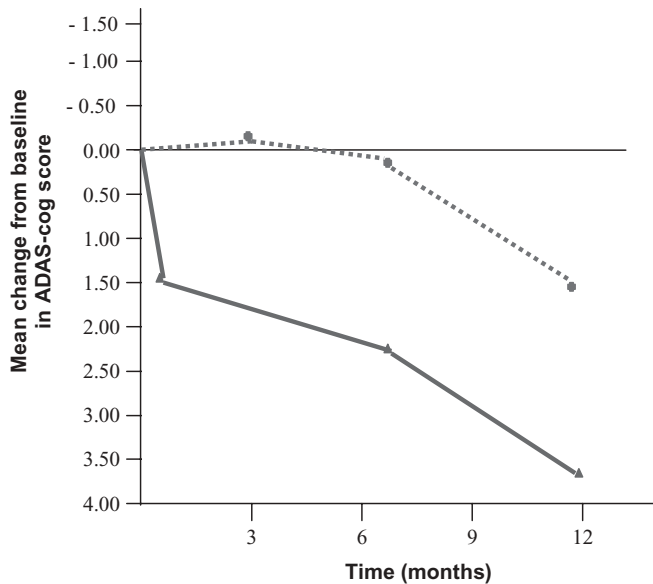


**FIGURE 4** Mean change from baseline (intent-to-treat analyses) in ADAS-cog scores of placebo groups of two 1-year clinical trials involving estrogen replacement and Ginkgo biloba (mean change from baseline derived from available data while other values were extrapolated from graphs):  $\triangle$  [36],  $\bullet$  [55]. Patients in the estrogen clinical trial had an MMSE range of 14–28 while those in the Ginkgo biloba trial had an MMSE range of 9–26. Other trials not included due to unavailable mean ADAS-cog scores of change from baseline along the course of the trial.

group in the first few weeks of some clinical trials [21, 22, 26, 27, 29, 30]. Deterioration is inevitable in trials of symptomatic agents despite these initial improvements and fluctuations.

The expected rate of change in the ADAS-cog scores annually in patients with AD is seven to eight points [91]. This varies depending on the severity of the cognitive impairment; less severe patients have a slower rate of decline [90]. This is evident in the Ginkgo biloba trial [55], for example, which included subjects with a lower range of MMSE scores and higher magnitude of decline, compared to those in the estrogen trial, which included subjects with higher MMSE scores (Figure 4).

#### 10.12.11 ATTRITION AND ADVERSE EFFECTS

Clinical trials suffer from subject dropout and the effects of attrition must be anticipated in trial design. Dropouts are more commonly due to refusal to follow up, lack of efficacy, violation of study protocol, and presence of adverse events.

Diminution of the subject pool is expected as the length of the trial is increased due to dropouts. Attrition is a potential problem because each dropout is equivalent to lost data, which may eventually lead to inadequate sample or, worse, skewing of the outcomes. Selective attrition can produce changes in the subject pool so that it inadvertently becomes unrepresentative of the subjects that originally entered into the study. Results become unreliable in this setting.

**TABLE 12** Gastrointestinal and Selected Other Adverse Events in Placebo Group (Percent of Occurrence) of Acetylcholinesterase Inhibitors in Mild to Moderate AD Clinical Trials

| Adverse Events          | Donepezil |      |      |      | Rivastigmine | Galantamine |      |      |
|-------------------------|-----------|------|------|------|--------------|-------------|------|------|
|                         | [26]      | [27] | [28] | [54] |              | [30]        | [29] | [31] |
| Anorexia                | —         | 2    | —    | 6    | 2            | 2.4         | 3.1  | 5.6  |
| Nausea                  | 8         | 4    | —    | 9    | 10           | 11.2        | 4.5  | 13.1 |
| Vomiting                | —         | 2    | —    | —    | 6            | 4           | 1.4  | 7.5  |
| Abdominal pain          | —         | —    | 5.6  | —    | 3            | 1.6         | —    | 4.2  |
| Diarrhea                | 3         | 7    | 6.9  | 17   | 9            | —           | 5.9  | 9.9  |
| Constipation            | —         | —    | 6.3  | —    | —            | —           | —    | —    |
| Dyspepsia               | —         | —    | —    | 6    | —            | —           | —    | 4.7  |
| Weight loss             | —         | —    | —    | 6    | —            | —           | —    | —    |
| Agitation               | —         | —    | —    | 13   | —            | 0.8         | 9.4  | —    |
| Confusion               | —         | —    | 6.3  | —    | —            | —           | —    | —    |
| Insomnia                | —         | —    | 6.9  | 8    | —            | —           | —    | —    |
| Somnolence              | —         | —    | —    | —    | —            | 0.8         | —    | —    |
| Dizziness               | —         | 4    | 4.2  | —    | 7            | 4           | —    | 11.3 |
| Headache                | 8         | —    | 6.3  | 9    | 8            | —           | —    | —    |
| Fatigue                 | —         | 2    | —    | —    | 3            | —           | —    | —    |
| Syncope                 | —         | —    | 2.8  | —    | —            | —           | —    | —    |
| Vertigo                 | —         | —    | 2.1  | —    | —            | —           | —    | —    |
| Anxiety                 | —         | —    | 5.6  | —    | —            | —           | —    | —    |
| Depression              | —         | —    | 7.6  | —    | —            | —           | —    | —    |
| Asthenia                | —         | —    | 3.5  | 7    | —            | —           | —    | —    |
| Urinary tract infection | 13        | —    | 6.9  | 13   | —            | —           | —    | —    |

*Note:* Included are adverse events with rate of occurrence that differed by at least 5% between the treatment group and the placebo.

Adverse events (AEs) is another cause of subject dropout. Adverse events are defined as newly emergent events/symptoms occurring after the first dose of the study drug or clinically significant worsening of a preexisting condition after the first dose of the study drug. Treatment emergent signs and symptoms (TESSs) is another term that is frequently used to indicate AEs. A serious adverse event (SAE) pertains to any event that is considered life threatening, resulting in hospitalization, permanent disability, or death. Occurrence of a serious AE that is deemed related to drug exposure may be grounds for withdrawal of a drug that is already released in the market. Clinical trials include tolerability and safety studies critical to determining any AEs that may occur with a drug.

Both the active treatment and placebo group experience AEs and all are reported whether these are deemed related or not related to the drug exposure. AEs range from incidental symptoms, part of the inherent deterioration of the disorder, secondary to existent medical disorder or related to drug exposure. Subject variability may account for the assortment of symptoms encountered as AEs and they can be observed among placebo groups as well as active intervention groups (Table 12).

Most AEs in acetylcholinesterase inhibitor trials are causally related to the drug's mechanism of action. Consistent findings are cholinergic manifestations that are usually transient and mild, although some may be moderate in severity. These are dose related and studies which utilized a forced titration rate increased cholinergic

AEs in subjects [22, 26]. In comparison, those that employed a longer dosage titration schedule had lower incidence of these AEs and better drug tolerability [27].

Deaths occur during clinical trials. The probability of its occurrence is increased in a dementia trial due to an expected lower survival rate in this population and the age of the subjects included. Deaths have been rare in most trials and were judged to be unrelated to the drug intake.

### 10.12.12 PRESENTING CLINICAL TRIAL RESULTS

The CONSORT statement was developed in (1996) [92] with the purpose of improving the accuracy of conducting and reporting RCTs. This obligates investigators to provide transparency as to the details of the trial from recruitment of subjects to its results. By following the CONSORT criteria, investigators are more likely to conduct ethical trials and produce valid, unbiased results. CONSORT criteria make it easier for readers to evaluate and understand the RCTs and there is an assurance that the results can be relied upon. Inadequate reporting of RCTs has led to biased interpretation of results with overestimation or underestimation of treatment effects. Currently, a 22-item CONSORT checklist and a flow diagram are available for use and these are constantly updated and modified [93].

Table 13 shows how CONSORT criteria can be applied to trials of AD and MCI.

### 10.12.13 DISEASE-MODIFYING TRIALS

Disease-modifying drugs have been the focus of dementia management for some time. The methodological strategies of clinical trials involving this class of therapeutic agent are apparently different from those involving symptomatic agents since the former target the neurodegenerative process of AD. Patients in the early stage of AD should ideally be included since they will receive the most benefit from disease intervention. Evidence of disease course alteration can be established in longer trials, that is,  $\geq 1$  year, since the natural progression of AD needs to be observed. This will also ensure that the treatment effect is not due to disease fluctuation that is present in AD. A larger sample population is necessary since attrition rate is expectedly higher in clinical trials with longer duration. Outcome measures other than neuropsychological tests such as biomarkers need to be employed to support the disease-modifying effect of the therapeutic agent. Use of biomarkers as surrogate endpoints can also reduce the required sample size and the duration of clinical trial, saving on time and resources. Pharmacoeconomic measures should also be integrated to justify long term societal benefits from the drug.

## ACKNOWLEDGMENT

The authors wish to acknowledge Lynn Fairbanks, Ph.D., a professor in the Department of Psychiatry and Biobehavioral Sciences, UCLA, for her special contribution in the statistical section of this chapter.

**TABLE 13 Features of Current MCI and ADRCT as These Are Adapted to CONSORT Checklist**

| Paper Section and Topic                                                    | Item | Examples of Features of MCI and Dementia Trials                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|----------------------------------------------------------------------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title and abstract                                                         | 1    | <ul style="list-style-type: none"> <li>Intervention stated (duration of trial sometimes included) in randomized, double-blind, placebo-controlled study</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Background                                                                 | 2    | <p>Introduction</p> <ul style="list-style-type: none"> <li>Basis for trial of intervention in human subjects: e.g., cholinergic hypothesis for AChEI's/glutamatergic excitotoxicity via NMDA receptor for memantine/oxidative damage and free-radical generation for antioxidants/ inflammatory response for prednisone and NSAIDS/estrogen replacement</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Participants                                                               | 3    | <p>Methods</p> <ul style="list-style-type: none"> <li>Subjects of both sexes, any race, <math>\geq 50</math> years old (or other age range), no significant medical disorder, adequate hearing and vision for testing, recruited from community/nursing home/other institutions, ambulant/or assisted</li> <li>Diagnosis of MCI, AD, VaD based on criteria ([7] for MCI; NINDS-AIREN [94] for VaD; NINCDS-ADRDA [12] for AD)</li> <li>Severity of dementia specified via MMSE range or through other measures (e.g., CDR)</li> <li>Acceptable concomitant medications</li> <li>Exclusion of those with contraindication to administration of intervention</li> <li>Thorough discussion of approval by ethical board, if ethical standards are met, and how consent is taken</li> <li>Multicenter involving tertiary centers with adequate facilities and personnel trained to diagnose, treat, and handle MCI and dementia cases</li> </ul> |
| Interventions                                                              | 4    | <ul style="list-style-type: none"> <li>Placebo and interventional drug compared via parallel-group trial; treatment arms specified</li> <li>Fixed dose/dose titration/titration sequence if dose-finding study</li> <li>Pill distribution (both placebo and interventional drug) including time</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Objectives                                                                 | 5    | <ul style="list-style-type: none"> <li>Efficacy and safety of intervention</li> <li>Time to reach an endpoint (cognitive and/or functional deterioration)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Outcomes                                                                   | 6    | <ul style="list-style-type: none"> <li>Primary and secondary outcome measures assessing cognitive, global, functional states, and/or quality of life</li> <li>Intervals of assessment stated, including endpoint</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Sample size                                                                | 7    | <ul style="list-style-type: none"> <li>Derived through power analysis (usually power of 80% with specified alpha value)</li> <li>Based on review of clinical studies of drugs of same class and results of earlier phase II studies</li> <li>Determined using primary efficacy variable to achieve certain power; estimate based on clinically relevant change in the score of said variable from baseline</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Randomization (sequence allocation/allocation concealment/ implementation) | 8–10 | <ul style="list-style-type: none"> <li>Sequence allocation by simple/blocked or stratified randomization</li> <li>Via computer-generated randomization list by pharmaceutical company</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Blinding (masking)                                                         | 11   | <ul style="list-style-type: none"> <li>Double blind with occasional inclusion of single-blind run-in or single-blind wash-out</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

TABLE 13 *Continued*

| Paper Section and Topic | Item | Examples of Features of MCI and Dementia Trials                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Statistical methods     | 12   | <p>Efficacy:</p> <ul style="list-style-type: none"> <li>• Drug–placebo difference between baseline and endpoint in specified outcome measures using least-squares (LS) mean</li> <li>• Statistical analysis used specified for both categorical and continuous variables</li> <li>• ITT or PP (LOCF and OC): population defined</li> <li>• Fully evaluable population (EP): population defined</li> <li>• Survival analysis</li> </ul> <p>Safety:</p> <ul style="list-style-type: none"> <li>• Included data defined (e.g., all patients who received at least one dose of study medications and who provided any postbaseline follow-up data)</li> <li>• Incidence of AEs or TESS compared among treatment arms</li> </ul> |
| Results                 | 13   | <p>Participant flow</p> <ul style="list-style-type: none"> <li>• Diagram of number of subjects in every stage of trial (enrollment, intervention, allocation, follow-up, and analysis)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Recruitment             | 14   | <ul style="list-style-type: none"> <li>• Recruitment period specified</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Baseline data           | 15   | <ul style="list-style-type: none"> <li>• Demographic and clinical characteristics of population allocated to placebo and intervention shown</li> <li>• Comparison made between groups to determine any significant difference in characteristics that may contribute to biased outcome</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Numbers analyzed        | 16   | <ul style="list-style-type: none"> <li>• Number of subjects who were included in each analysis (dependent on type of analysis, e.g., ITT, LOCF, and OC)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Outcomes and estimation | 17   | <ul style="list-style-type: none"> <li>• Effect size based on clinically meaningful change produced by the intervention</li> <li>• Statistically significant drug–placebo difference based on each of primary and secondary efficacy variables (level of significance specified)</li> <li>• For survival studies: statistically significant difference between placebo and intervention in delaying time to reach endpoint</li> </ul>                                                                                                                                                                                                                                                                                       |
| Ancillary analyses      | 18   | <ul style="list-style-type: none"> <li>• Subgroup analysis based on certain features of subjects where efficacy of drug may be different (e.g., early onset vs. late onset)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Adverse events          | 19   | <ul style="list-style-type: none"> <li>• Including serious AEs and any deaths</li> <li>• Determination if occurrence is beyond acceptable limit and if there is association with exposure to intervention</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Interpretation          | 20   | <p>Discussion</p> <ul style="list-style-type: none"> <li>• Interpretation and explanation of positive and/or negative results</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Generalizability        | 21   | <ul style="list-style-type: none"> <li>• Limitations of study that will limit its application to targeted population in community setting (e.g., differential recruitment in terms of race, highly specialized centers)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Overall evidence        | 22   | <ul style="list-style-type: none"> <li>• Evidence-based conclusion regarding efficacy and safety of intervention in subset of patients studied</li> <li>• Recommendations in future studies</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

ITT: Intent-to-treat analysis; PP: per protocol.

## REFERENCES

1. Evans, D. A. (1990), Estimated prevalence of Alzheimer's disease in the United States, *Milbank Q*, 68, 267–289.
2. National Institute of Health, National Institute on Aging (1999), *Progress Report on Alzheimer's Disease, 1999* (NIH Pub. No. 99-4664), U.S. Department of Health and Human Services, Bethesda, MD.
3. Evans, D. A., Funkenstein, H. H., Albert, M. S., et al. (1989), Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported, *JAMA*, 26, 2551–2556.
4. Institute on Aging (1996), *Progress Report on Alzheimer's Disease, 1996* (NIH Pub. No. 96-4137), National Institute on Aging, Bethesda, MD.
5. Ernst, R. L., and Hay, J. W. (1994), The US economic and social costs of AD revisited, *Am. J. Public Health*, 84, 1261–1264.
6. Morris, J. C., Storandt, M., Miller, P., et al. (2001), Mild cognitive impairment represents early-stage Alzheimer's disease, *Arch. Neurol.*, 58, 397–405.
7. Petersen, R. C., Doody, R., Kurz, A., et al. (2001), Current concepts in mild cognitive impairment, *Arch. Neurol.*, 58, 1985–1992.
8. Petersen, R. C., Thomas, R. G., Grundman, M., et al. (2005), Vitamin E and donepezil for the treatment of mild cognitive impairment, *N. Engl. J. Med.*, 352, 2379–2388.
9. Salloway, S., Ferris, S., Kluger, A., et al. (2004), Efficacy of donepezil in mild cognitive impairment: A randomized placebo-controlled trial, *Neurology*, 63, 651–657.
10. American Psychiatric Association (1994), *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (DSM-IV), American Psychiatric Association, Washington, DC, pp. 143–147.
11. World Health Organization. (1993), *The ICD-10 Classification of Mental and Behavioral Disorders. Diagnostic Criteria for Research*, World Health Organization, Geneva, Switzerland.
12. McKhann, G., Drachman, D., Folstein, M., et al. (1984), Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, *Neurology*, 34, 939–944.
13. Holmes, C., Cairns, N., Lantos, P., et al. (1999), Validity of current clinical criteria for Alzheimer's disease, vascular dementia and dementia with Lewy bodies, *Bri. J. Psych.*, 174, 45–50.
14. Blacker, D., Albert, M., Bassett, S., et al. (1994), Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease. The National Institute of Mental Health Genetics Initiative, *Arch. Neurol.*, 51, 1198–1204.
15. Lopez, O. L., Litvan, I., Catt, K. E., et al. (1999), Accuracy of four clinical diagnostic criteria for the diagnosis of neurodegenerative dementias, *Neurology*, 53, 1292–1299.
16. Tierney, M., Fisher, R., Anthony, J., et al. (1988), The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: A clinicopathologic study of 57 cases, *Neurology*, 38, 359–364.
17. Reisberg, B., Ferris, S., de Leon, M., et al. (1982), The Global Deterioration Scale for assessment of primary degenerative dementia, *Am. J. Psych.*, 138, 1136–1139.
18. Hughes, C., Berg, L., Denziger, W., et al. (1982), A new clinical scale for the staging of dementia, *Bri. J. Psych.*, 140, 566–572.



19. Folstein, M., Folstein, S., and McHugh, P. (1975), Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician, *J. Psychiatric Res.*, 12, 189–198.
20. O'Connor, D. W., Pollitt, P., Hyde, J. B., et al. (1989), The reliability and validity of the Mini-Mental state in a British community survey, *J. Psychiatric Res.*, 23, 87–96.
21. Cummings, J., Cyrus, P., Bieber, F., et al. (1998), Metrifonate treatment of the cognitive deficits of Alzheimer's disease, *Neurology*, 50, 1214–1221.
22. Morris, J., Cyrus, P., Orazem, J., et al. (1998), Metrifonate benefits cognitive, behavioral, and global function in patients with Alzheimer's disease, *Neurology*, 50, 1222–1230.
23. Davis, K., Thal, L., Gamzu, E., et al. (1992), A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease, *N. Engl. J. Med.*, 327, 1253–1259.
24. Farlow, M., Gracon, S., Hershey, L., et al. (1992), A controlled trial of tacrine in Alzheimer's disease, *JAMA*, 268, 2523–2529.
25. Knapp, M., Knopman, D., Solomon, P., et al. (1994), A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease, *JAMA*, 271, 985–991.
26. Rogers, S., Doody, R., Mohs, R., et al. (1998), Donepezil improves cognition and global function in Alzheimer's disease, *Arch. Intern. Med.*, 158, 1021–1031.
27. Rogers, S., Farlow, M., Doody, R., et al. (1998), A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease, *Neurology*, 50, 136–145.
28. Winblad, B., Engedal, K., Soininen, H., et al. (2001), A 1-year, randomized, placebo-controlled study of donepezil in patients with mild to moderate AD, *Neurology*, 57, 489–495.
29. Tariot, P., Solomon, P., Morris, J., et al. (2000), A 5-month, randomized, placebo-controlled trial of galantamine in AD, *Neurology*, 54, 2269–2276.
30. Rockwood, K., Mintzer, J., Truyen, L., et al. (2001), Effects of a flexible galantamine dose in Alzheimer's disease: A randomized, controlled trial, *J. Neurol. Neurosurg. Psych.*, 71, 589–595.
31. Raskind, M., Peskind, E., Wessel, T., et al. (2000), Galantamine in AD: A 6-month randomized, placebo-controlled trial with a 6-month extension, *Neurology*, 54, 2261–2268.
32. Corey-Bloom, J., Anand, R., Veach, J., et al. (1998), A randomized trial evaluating the efficacy and safety of ENA 73 (rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderately severe Alzheimer's disease, *Int. J. Ger. Psychopharm.*, 1, 55–65.
33. Rosler, M., Anand, R., Cicin-Sain, A., et al. (1999), Efficacy and safety of rivastigmine in patients with Alzheimer's disease: International randomized controlled trial, *BMJ*, 318, 633–638.
34. Scharf, S., Mander, A., Ugoni, A., et al. (1999), A double-blind, placebo-controlled trial of diclofenac/misoprostol in Alzheimer's disease, *Neurology*, 53, 197–201.
35. Aisen, P., Davis, K., Berg, J., et al. (2000), A randomized controlled trial of prednisone in Alzheimer's disease, *Neurology*, 54, 588–593.
36. Mulnard, R., Cotman, C., Kawas, C., et al. (2000), Estrogen replacement therapy for treatment of mild to moderate alzheimer disease, *JAMA*, 283, 1007–1015.
37. Aisen, P., Schafer, K., Grundman, M., et al. (2003), Effects of rofecoxib or naproxen vs. placebo on Alzheimer disease progression, *JAMA*, 289, 2819–2826.
38. Reines, S., Block, G., Morris, J., et al. (2004), Rofecoxib: No effect on Alzheimer's disease in a 1-year, randomized, blinded, controlled study, *Neurology*, 62, 66–71.

39. Thal, L., Carta, A., Clarke, W., et al. (1996), A 1-year multicenter placebo-controlled study of acetyl-L-carnitine in patients with Alzheimer's disease, *Neurology*, 47, 705–711.
40. Feldman, H., Gauthier, S., Hecker, J., et al. (2001), A 24-week, randomized, double-blind study of donepezil in moderate to severe Alzheimer's disease, *Neurology*, 57, 613–620.
41. Reisberg, B., Doody, R., Stoffler, A., et al. (2003), Memantine in moderate-to-severe Alzheimer's disease, *N. Engl. J. Med.*, 348, 1333–1341.
42. Doody, R., Stevens, J., Beck, C., et al. (2001), Practice parameter: Management of dementia (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology, *Neurology*, 56, 1154–1166.
43. Alzheimer Association (2004), Research consent for cognitively impaired adults: Recommendations for institutional review boards and investigators, *Alzheimer Dis. Assoc. Disord.*, 18, 171–175.
44. Koepsell, T., and Weiss, N. (2003), *Epidemiologic Methods*, Oxford University Press, New York, pp. 64–67, 261, 327–329.
45. Schneider, L., Tariot, P., Lyketsos, C., et al. (2001), National Institute of Mental Health Clinical Antipsychotic Trial of Intervention Effectiveness (CATIE): Alzheimer Disease Trial Methodology, *Am. J. Ger. Psych.*, 9, 346–390.
46. Cummings, J. (2003), Use of acetylcholinesterase inhibitors in clinical practice: Evidence-based recommendations, *Am. J. Ger. Psych.*, 11, 131–145.
47. Schoenmaker, N., and Van Gool, W. (2004), The age gap between patients in clinical studies and in the general population: A pitfall for dementia research, *Lancet Neurol.*, 3, 627–630.
48. Leber, P. (1990), *Guidelines for the Clinical Evaluation of Antidementia Drugs*.
49. Ferris, S., Lucca, U., Mohs, R., et al. (1997), Objective psychometric tests in clinical trials of dementia drugs, *Alzheimer Dis. Assoc. Disord.*, 11, 34–38.
50. Rosen, W., Mohs, R., Davis, K., et al. (1984), A new rating scale for Alzheimer's disease, *Am. J. Psych.*, 141, 1356–1364.
51. Guy, W., ed. (1976), Clinical Global Impression (CGI). *ECDEU Assessment Manual for Psychopharmacology*, U.S. Department of Health and Human Services, Public Health Service, Alcohol Drug Abuse and Mental Health Administration. NIMH Psychopharmacology Research Branch, Rockville, MD, pp. 218–222.
52. Knopman, D., Knapp, M., Bracon, S., et al. (1994), The Clinician Interview-Based Impression (CIBI): A clinician's global change rating scale in Alzheimer's disease, *Neurology*, 44, 2315–2321.
53. Schneider, L., Olin, J., Doody, R., et al. (1997), Validity and Reliability of the Alzheimer's Disease Cooperative Study—Clinical Global Impression of Change, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 2), S22–S32.
54. Mohs, R., Doody, R., Morris, J., et al. (2001), A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients, *Neurology*, 57, 481–488.
55. Le Bars, P., Katz, M., Berman, N., et al. (1997), A placebo-controlled, double-blind, randomized trial of an extract of ginkgo biloba for dementia, *JAMA*, 278, 1327–1332.
56. Burke, W., Miller, O., Rubin, E., et al. (1988), Reliability of the Washington University Clinical Dementia Rating, *Arch. Neurol.*, 45, 31–32.
57. Reisberg, B., Scaln, S., Franssen, E., et al. (1994), Dementia Staging in chronic care populations, *Alzheimer Dis. Assoc. Disord.*, 8 (Suppl. 1), S188–S205.
58. Sano, M., Ernesto, C., Thomas, R., et al. (1997), A Controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease, *N. Engl. J. Med.*, 336, 1216–1222.

59. Gelinas, I., Gauthier, L., McIntyre, M., et al. (1999), Development of a functional measure for persons with Alzheimer's disease: The disability assessment for dementia, *Am. J. Occup. Ther.*, 53, 471–481.
60. Gauthier, S., Bodick, N., Erzigkeit, E., et al. (1997), Activities of daily living as an outcome measure in clinical trials of dementia drugs, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 3), 6–7.
61. Lawton, M., and Brody, E. (1969), Assessment of older people: Self-maintaining and instrumental activities of daily living, *Gerontologist*, 9, 179–186.
62. DeJong, R., Osterlund, O., and Roy, G. (1989), Measurement of quality-of-life changes in patients with Alzheimer's disease, *Clin. Ther.*, 11, 545–554.
63. Galasko, D., Bennett, D., Sano, M., et al. (1997), An inventory to assess activities of daily living for clinical trials in Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 2), S33–39.
64. Patterson, M., and Bolger, J. (1994), Assessment of behavioral symptoms in Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 8 (Suppl 3), 4–20.
65. Cummings, J., Mega, M., Gray, K., et al. (1994), The Neuropsychiatric Inventory: An efficient tool for comprehensively assessing psychopathology in dementia, *Neurology*, 44, 2308–2314.
66. Reisberg, B., Auer, S. R., Monteiro, I. M., et al. (1996), Behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale, *Int. Psychogeriatr.*, 8 (Suppl 3), 301–308.
67. Tariot, P., Mack, J., Patterson, M., et al.; the CERAD Behavioral Pathology Committee (1995), The CERAD Behavior Rating Scale for Dementia (BRSD), *Am. J. Psych.*, 152, 1349–1357.
68. Cohen-Mansfield, J. (1985), Agitated behaviors in the elderly ill. Preliminary results in the cognitively deteriorated, *J. Am. Ger. Soc.*, 34, 722–727.
69. Alexopoulos, G., Abrams, R., Young, R., et al. (1988), Cornell Scale for Depression in Dementia, *Biol. Psych.*, 23, 271–284.
70. Overall, J., and Gorham, D. (1962), The Brief Psychiatric Rating Scale, *Psychol. Rep.*, 10, 799–812.
71. Overall, J., and Gorham, D. (1988), Introduction: The Brief Psychiatric Rating Scale (BPRS): Recent developments in ascertainment and scaling, *Psychopharm. Bull.*, 24, 97–99.
72. Kay, S., Opler, L., Lindenmayer, J., et al. (1989), The Positive and Negative Syndrome Scale: A rationale and standardization, *Br. J. Psych.*, 155 (Suppl 7), 59–65.
73. Weiner, M., Koss, E., Wild, K., et al. (1996), Measures of psychiatric symptoms in Alzheimer patients: A review, *Alzheimer Dis. Assoc. Disord.*, 10, 20–30.
74. De Deyn, P., Rabheru, K., Rasmussen, A., et al. (1999), A randomized trial of risperidone, placebo, and haloperidol for behavioral symptoms of dementia, *Neurology*, 53, 946–955.
75. Katz, I., Jeste, D., Mintzer, J., et al. (1999), Comparison of risperidone and placebo for psychosis and behavioral disturbances associated with dementia: A randomized, double-blind trial, *J. Clin. Psych.*, 60, 107–115.
76. Brodaty, H., Ames, D., Snowden, J., et al. (2003), A randomized placebo-controlled trial of risperidone for the treatment of aggression, agitation and psychosis of dementia, *J. Clin. Psych.*, 64, 134–143.
77. Lyketsos, C., Sheppard, J., Steele, C., et al. (2000), Randomized, placebo-controlled, double-blind clinical trial of sertraline in the treatment of depression complicating Alzheimer's disease: Initial results from the depression in Alzheimer's disease study, *Am. J. Psych.*, 157, 1686–1689.

78. Tariot, P., Erb, R., Podgorski, C., et al. (1988), Efficacy and tolerability of carbamazepine for agitation and aggression in dementia, *Am. J. Psych.*, 155, 54–61.
79. Porteinsson, A., Tariot, P., Erb, R., et al. (2001), Placebo-controlled study of divalproex sodium for agitation in dementia, *Am. J. Ger. Psych.*, 9, 58–66.
80. Zhong, K., Tariot, P., Minkwitz, M. C., et al. (2004), Quetiapine for the treatment of agitation in the elderly institutionalized patients with dementia: A randomized, double blind trial. International Conference on Alzheimer's Disease and Related Disorders meeting, poster # P2-442.
81. Davis, K., Marin, D., Kane, R., et al. (1997), The Caregiver Activity Survey (CAS): Development and validation of a new measure for caregivers of persons with Alzheimer's disease, *Int. J. Ger. Psych.*, 12, 978–988.
82. Wimo, A., Wetterholm, A., Mastey, Y., et al. (1998), Evaluation of the healthcare resource utilization and caregiver time in anti-dementia drug trials, in Wimo, A., Jonsson, B., Karlson, G., and Wilblad, B., Eds., *Health Economics of Dementia*, Wiley, Chichester, England, pp. 465–499.
83. Whitehouse, P. (1997), Pharmacoeconomics of dementia, *Alzheimer Dis. Assoc. Disord.*, 11 (Suppl 5), S22–33.
84. AD2000 Collaborative Group (2004), Long term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): Randomized double-blind trial, *Lancet*, 353, 2105–2115.
85. Winblad, B., Wimo, A., and Almkvist, O. (2000), Outcome measures in Alzheimer's disease: Do they go far enough? *Dem. Ger. Cogn. Dis.*, 11 (Suppl 1), 3–10.
86. Leber, P. (1997), Observations and suggestions on antidementia drug development, *Alzheimer Dis. Assoc. Disord.*, 10 (S1), 31–35.
87. Sano, M., Ernesto, C., Klauber, M., et al. (1996), Rationale and design of a multicenter study of selegiline and  $\alpha$ -tocopherol in the treatment of Alzheimer disease using novel clinical outcomes, *Alzheimer Dis. Assoc. Disord.*, 10, 132–140.
88. A Randomized Double-Blind, Placebo-Controlled Trial to Evaluate the Efficacy and Safety of Galantamine in Subjects with Mild Cognitive Impairment (MCI) Clinically at Risk for Development of Clinically Probably Alzheimer's Disease. Gal-INT-18 and Gal-INT-11. <http://www.alz.org/news/05q1/012405.asp>; accessed November 17, (2005).
89. Everitt, B., and Pickler, A. (2004), *Statistical Aspects of the Design and Analysis of Clinical Trials*, Imperial College Press, London, pp. 45–50, 203–204.
90. Doraiswamy, P., Kaiser, L., Bieber, F., et al. (2001), The Alzheimer's Disease Assessment Scale: Evaluation of psychometric properties and patterns of cognitive decline in multicenter clinical trials of mild-to-moderate Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 15, 174–183.
91. Morris, J., Edland, S., Clark, C., et al. (1993), The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part. IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease, *Neurology*, 43, 2457–2465.
92. Begg, C., Cho, M., Eastwood, S., et al. (1996), Improving the quality of reporting of randomized controlled trials: The CONSORT statement, *JAMA*, 276, 637–639.
93. Moher, D., Schulz, K., Altman, D., et al. (2001), The CONSORT Statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials, *JAMA*, 285, 1987–1991.
94. Roman, G. C., Tatemichi, T. K., Erkinjuntti, T., et al. (1993), Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Work Group, *Neurology*, 43, 250–260.

# 10.13

---

## *Clinical Trials in Urology*

**Geoffrey R. Wignall, Carol Wernecke, Linda Nott, and Hassan Razvi**

*Schulich School of Medicine and Dentistry, University of Western Ontario,  
London, Ontario, Canada*

### **Contents**

|           |                                             |     |
|-----------|---------------------------------------------|-----|
| 10.13.1   | Introduction                                | 695 |
| 10.13.2   | Basics                                      | 696 |
| 10.13.2.1 | Clinical Trial Objectives                   | 696 |
| 10.13.2.2 | Study Population                            | 697 |
| 10.13.2.3 | Ethical Considerations                      | 698 |
| 10.13.2.4 | Review Boards                               | 700 |
| 10.13.3   | Specific Issues with Surgical Device Trials | 701 |
| 10.13.3.1 | Regulatory Issues                           | 701 |
| 10.13.3.2 | “First-in-Man” Studies                      | 702 |
| 10.13.4   | Summary                                     | 702 |
|           | References                                  | 703 |

### **10.13.1 INTRODUCTION**

The wide scope of current urological practice has led to a proliferation of research and clinical trials in this specialty. As a surgical subspecialty devoted to the medical and surgical care of patients with genitourinary disorders, both pharmaceutical and device trials have equally important roles in the advancement of the future care of the urologic patient. Pharmaceutical trials are ongoing for conditions such as erectile dysfunction, urinary incontinence, and oncology to name a few. Significant advancements in surgical instrumentation such as robotics and prosthetic devices for the treatment of genitourinary diseases are also being witnessed and are the

subject of clinical investigation. The objective of this chapter is to provide an overview of the objectives, design, and implementation of clinical trials specifically relevant to those conducting urological research. While many of the domains covered in the chapter are subjects commonly faced by researchers in other specialties, we will attempt to highlight issues of particular importance to urological investigators.

## 10.13.2 BASICS

In this section we provide an overview of the issues related to the development of clinical trial objectives and the study populations of importance in urological research.

### 10.13.2.1 Clinical Trial Objectives

The first and most important step in designing any clinical trial is defining the study objective [1]. Careful attention to the formulation of the primary and secondary study objectives will allow investigators to design a quality clinical trial that is adequately powered to answer the study question and avoid many potential pitfalls.

This crucial first step attempts to formulate the primary question or that question that is most significant to be answered by the trial. This may seem self-evident, as this objective is likely the reason the trial was initiated in the first place. However, more often than not the initial concept is too broad or poorly defined to be useful as a primary question. Investigators must ask themselves whether this primary objective can be realistically addressed with a clinical trial. The sample size calculation is based on the primary question. In some instances, that calculation may reveal that more participants would be needed than would be expected to be available in a reasonable time frame. Reevaluation of the primary outcome measurement may provide a more realistic number of participants that would be needed for the trial.

When designing a study to evaluate a new treatment or drug to treat a disease such as cancer, the urological researcher should ensure the most clinically relevant outcomes will be assessed. Time to disease progression, disease-specific survival, and overall survival are just a few of the commonly determined measurements. The effect of the study treatment may or may not be evident for several years, depending on the disease process, and the effect of the study treatment. Therefore, the investigators will need to define how these study endpoints will be assessed over time.

In formulating the primary question, investigators may either seek to prove a hypothesis [2] or to show a lack of a difference in outcomes between participants and controls. An example of the former would be the situation where investigators hypothesize that patients receiving a study drug for the treatment of prostate cancer may have a decrease in cancer-specific mortality. At other times it is beneficial to show that there is no difference between subjects and controls such as when an intervention is deemed to be unnecessary. An example of this would be the argument against the routine placement of a ureteral stent following uncomplicated ureteroscopy where investigators demonstrated that there was no advantage in the placement of a stent [3].

While formation of a clear primary objective is of the utmost importance, secondary objectives must also be defined. These may be evident at the start of the study or may become clear through collection of data designed to address the primary objective. Using the above example of a drug for the treatment of cancer, if the primary objective is to demonstrate an improvement in cancer-specific mortality, a secondary objective may be to see which subgroups in particular benefit from this effect. Whenever possible, these secondary objectives should be outlined at the commencement of the trial. Additionally, secondary objectives may be distinctly separate from the primary question such as determining whether there are any adverse effects from the drug or a difference between subjects and controls with respect to quality of life issues. The primary objective is used for power and sample size calculation; therefore, it is not uncommon for secondary objectives to be inconclusive and lack power due to small sample size.

Urologists have several tools to evaluate outcomes in clinical trials. While questionnaires are able to evaluate subjective outcomes, the study design should also include objective measurements. Objective measurements must be relative to the disease under study and be considered to be part of the standard of care or evaluation of that same patient population.

The International Prostate Symptom Score includes symptom, bother, and quality of life scores that are useful in accessing primary and secondary outcomes in several types of urology trials [4]. The International Index of Erectile Function (IIEF) can be used to evaluate the primary outcome in erectile dysfunction trials, and secondary outcomes of other trials affecting the male population where the evaluation of sexual function is relevant [5]. Quality of life measurements are commonly utilized in urologic trials as secondary objectives. The SF-36 Health is considered the most comprehensive instrument to measure general health-related quality of life (HRQOL) [6]. Various validated instruments are also available to assess a number of specific urologic conditions. The UCLA Prostate Cancer Index was devised to assess health-related quality of life in patients with prostate cancer [7]. The Ureteral Stent Questionnaire (USSQ) uses both the visual analog scale (VAS) and quality of life questions to evaluate the patient's stent-related symptoms [8]. Specific urological questions, either as primary or secondary outcomes, can be evaluated in the trial by using the VAS to assess the effect of the study treatment on participants' perceived results, ranging from pain to treatment effectiveness or reduction of urinary symptoms [9].

#### **10.13.2.2 Study Population**

Careful selection of the study population will improve the chances of successfully answering the study question in a meaningful and convincing fashion. Well-defined inclusion and exclusion criteria must be clearly defined in the study protocol. In general, a simple approach to defining the study population begins with including the entire population at large and methodically narrowing the definition of the eligible subject. To illustrate this process we consider the example of a large prospective urological study, The Prostate Cancer Prevention Trial (PCPT) [10, 11]. The objective of PCPT was to assess the effect of the drug finasteride for the prevention of prostate cancer. The first step toward selecting the study population in this trial was to identify those individuals with or at risk for prostate cancer. In the case of



prostate cancer, young men are much less likely to manifest the disease; therefore, the designers of the PCPT trial opted to include only those men 55 years of age or older. As this study looked at the development of prostate cancer in men without the disease, it was important to include only men with a normal digital rectal exam (DRE). One must also be aware of potential confounding conditions that may detract from the ability to assess for the condition being studied. Benign prostatic hyperplasia (BPH) may result in a benign rise in prostate specific antigen that may be perceived as potential prostate cancer. To minimize the impact of BPH, men with an American Urological Association symptom score of 20 or greater were excluded from the study.

Randomized clinical trials usually have at least one control group as well as one or more treatment groups. While this tends to increase the costs of the trial due to the greater number of subjects required, it is essential to statistically answer the primary study question.

Control subjects may be assigned as either negative or positive controls. In the setting of a pharmaceutical trial, a patient receiving a placebo (i.e., sugar pill) would be a negative or passive control, meaning that they are not getting active treatment. Still, patients in placebo arms of studies often report improvement or deterioration in their condition, demonstrating the so-called placebo effect. This is a well-described phenomenon in urology, especially noteworthy in benign prostatic hyperplasia clinical trials. Improvements in symptom severity in particular may range from 5 to 15% for placebo or sham treated cohorts for both medical and minimally invasive device treatment trials of BPH patients [12].

A negative control may be used either alone (if no gold standard treatment exists) or in conjunction with a positive control arm. A positive or active control may be used when there is a current treatment that is considered the standard of care for a particular disease or condition. The control arm consists of patients receiving the standard treatment while subjects in the experiment arm receive the treatment under investigation. By actively treating both treatment and control groups, investigators hope to demonstrate superiority of the experimental treatment over the established standard. It has been suggested that active controls may protect patients and should be routinely utilized [13, 14].

Subject-to-subject variability is controlled by choosing a cross-over design. In a cross-over medical device trial, subjects can be used as their own control. Cross-over trials can be considered when the study includes two or more interventions. The interventions may be two separate treatment devices or one may be a placebo treatment. Each subject in the clinical study is assigned a randomized treatment group, allowing a period between interventions for a “wash out” of any carry-over from the previous intervention. Patients treated as part of the cross-over design are then followed from the baseline of the study’s schedule as outlined in the protocol. In studies with a placebo arm, patients who were randomized to the placebo arm are offered open-label study treatment at a predetermined time. Often this is defined as when the study treatment is unblinded and revealed to all of the participants [15].

### 10.13.2.3 Ethical Considerations

Research has been ongoing throughout human history; however, it has not always been carried out in a fashion that would be deemed ethical by today’s standards.

Modern investigators must be acutely aware of the ethical aspects of clinical trials and should strive to exceed the expectations of their research review boards. Given the nature of the urological patient population, significant ethical issues are not infrequently encountered in the course of urological clinical trials.

The formulation of a relevant and important research question helps to lay the foundation for ethically sound research [16]. Researchers must thoroughly consider the relevance of their research proposal, ensuring that it is based on solid scientific principles and that the results of the study are likely to improve patient care or enhance knowledge in the field [17]. Patients enrolled in a clinical trial should be confident that their participation should ultimately result in a valuable scientific contribution [16].

In order to best protect the rights of individual patients, investigators should adhere to the principles of autonomy, beneficence, and justice [18]. Autonomy implies that the individual should make important decisions intentionally and free of external influence. Beneficence centers on the responsibility of the investigator to maximize the positive outcome of the trial for each patient. This is a fundamental principle of the Hippocratic oath along with the responsibility to “do no harm” and thereby minimize the potential negative outcomes of the trial [19]. Justice dictates that the design and implementation of a clinical trial be fair to participants. The selection of subjects should be free of bias, and drawn from a relevant patient population.

The Council for International Organizations of Medical Sciences (CIOMS) and the World Medical Association (WMA) have both published guidelines for the ethical management of biomedical research. The WMA Declaration of Helsinki (Ethical Principles for Medical Research Involving Human Subjects) was first adopted in 1964 and has undergone several revisions with the most recent in 2000 [20]. The CIOMS International Ethical Guidelines for Biomedical Research Involving Human Subjects consists of 21 guidelines relating to ethical justification scientific validity of research, informed consent, and equity regarding burdens and benefits among other issues [21]. Each set of guidelines provides an excellent framework for the ethical conduct of biomedical research and are of benefit when preparing a research proposal for institutional review.

No clinical trial may be deemed ethical without a properly formulated and freely given informed consent. Not only must this consent be legal according to the laws of the nation where the study is to be performed but should also conform to the ethical guidelines laid out by the WMA and/or CIOMS. Additionally, in the event that pertinent information becomes available during the course of a trial, it is the investigators' responsibility to advise the participants of any information that pertains to the informed consent [1]. First, the potential study subjects should be provided with an overview of the proposed research to decide whether they would be interested in inclusion into the trial. This includes informing subjects of their responsibilities during the trial (i.e., taking of medications, follow-up appointments), as well as any procedures that will be performed. There should be an honest and open dialog between the individual obtaining the consent and the prospective subject with no attempt to mislead or deceive. The consent must be given voluntarily with participants entering the trial of their own free will with no fear of reprisal should they decline (i.e., withholding of care). Subjects must be told why they are appropriate candidates for the trial (inclusion criteria) and what would make them ineligible (exclusion criteria).

Participants must be informed of all potential risks and benefits of participation in the trial at the time of consent. Risks may include but are not limited to potential physical or mental injury as a result of participation in the study. Additionally, the investigators should provide the patient with an explanation of what will be done to minimize these risks. The benefits of inclusion in the trial may include an improvement to the patient's symptoms or condition, although the patient should be made aware that this is not guaranteed. It is essential to explain any alternate treatments that are available to the subject, including the standard drugs or procedures that would be used in his or her situation and make the patient aware if any of these will be withheld due to inclusion in the study. An example of this would be a trial studying a new form of ureteroscopic lithotripsy. The patient should be made aware of the potential risks of the procedure (i.e., ureteral injury, failure to fragment the stone) as well as possible benefits (i.e., improved stone fragmentation). Additionally, participants must be made aware of alternative procedures such as laser lithotripsy that would be withheld in order to study the new device. The described benefits should also include the potential benefits to society as a whole such as improved stone-free rates using the new technology.

Confidentiality is an important ethical consideration when carrying out a clinical trial and should be maintained from consent through to the study conclusion. Personalized information including subject name, date of birth, and ethnicity along with any other personal identifiers must remain confidential. Study participants should be notified at the time of consent if any personal identifiers will be disclosed during the course of the trial and what measures will be taken to maintain confidentiality. In the event that the trial requires the use of video or photographic data, patients should be aware that they have the right to access such files should they wish to do so.

Participants should be provided with written material summarizing the above information in lay language that is clear and understandable. This information should include appropriate contact information for the reference during the course of the study. The informed consent document must be signed and dated by both the participant and the individual obtaining consent, although verbal consent may be allowed at the discretion of the review board in some instances.

#### **10.13.2.4 Review Boards**

While the actual requirements to conduct research vary based on the country in which the trial is to be conducted, most nations have adopted a model that incorporates an institutional review board (IRB) or a research ethics board (REB). The general mandate of the IRB/REB is to objectively review research proposals involving the use of human subjects and to ensure that trials are ethically sound according to the previously described principles. The board protects the rights of individual participants in research trials through external review to assess the relevance and ethics of proposals and to make certain that confidentiality and safety are maintained by investigators. In general, ethics boards are comprised of multiple individuals of both genders, at least one of whom should be from a nonscientific discipline. Additionally, one or more members may be appointed from outside of the institution at which the research is to be conducted [16].

Researchers embarking on a trial should be aware of the principles of ethical research and identify any potential pitfalls that may delay approval of the ethics application. The primary research question must be clear and relevant, and all secondary objectives should be stated. Relevant statistics such as the sample size calculation must be performed ahead of time to justify the required number of participants. Any potential risks to participants should be addressed in the application as well as measures through which the investigators will minimize these risks and deal with adverse events should they arise. While guidelines are, for the most part, consistent between institutions, researchers should make themselves familiar with the requirement at their institution. This preparation will help to avoid any unnecessary delays.

### **10.13.3 SPECIFIC ISSUES WITH SURGICAL DEVICE TRIALS**

#### **10.13.3.1 Regulatory Issues**

Urologists employ a wide array of surgical devices ranging in complexity from simple guidewires to robotic operating systems. Each of these devices must pass through regulatory safeguards prior to introduction into clinical use. Issues relating to drug trials are covered elsewhere in this text; however, we feel it necessary to briefly touch on matters relating to device trials and how this may differ from pharmaceutical trials.

The U.S. Food and Drug Administration (FDA) Center for Devices and Radiological Health oversees the development and marketing of new devices in the United States. If the investigational device study presents a significant risk to the subjects, an investigational device exemption (IDE) application must be completed [22]. The IDE must be approved by the FDA, and the individual site(s) institutional review board prior to study initiation. The process of IDE approval is often lengthy, and once approved the clinical study must be conducted according to the IDE regulations (CFR 812). If an investigator or a sponsor proposes that the device under study does not pose a significant risk (NSR) to the participants, and the ethics board agrees and approves the study, no FDA/IDE application is needed.

Medical devices are classified by risk, with class I devices being lowest risk, and class III devices being the highest risk as defined by FDA standards [22]. FDA class I devices (i.e., handheld surgical instruments) present minimal potential harm to the user and require the least amount of regulatory control. Class II devices include instruments such as uroflowmeters, monitors, pump systems, and X-ray devices. These devices require additional special controls including special labeling requirements, mandatory performance standards, and postmarket surveillance. Class III devices (class III/IV in Canada) have the strictest regulatory controls and include devices that support or sustain human life or lack sufficient safety information to assure safety and efficacy (e.g., ureteral stents, penile prostheses, microwave thermotherapy units).

The process in Canada is similar. Medical devices are regulated by Health Canada's Therapeutic Products Directorate under the Food and Drugs Act [23]. The goal of this regulation is to ensure that medical devices available in Canada are safe and effective and of sufficient quality for general distribution. Once Health Canada

approval has been secured, the local institutional review boards are then able to proceed with their review of the research protocol. Health Canada is entitled to ask the sponsor for any information it deems necessary to make a sufficiently informed decision on the safety of the product (additional information request). The cycle of Health Canada issuing the request, the time for the sponsor to compose its response, and a new review time of 30 days from Health Canada's receipt of the response may further prolong the review cycle. Health Canada differs slightly by stratifying devices from class I to class IV but follows the same general principles as the FDA.

If an investigator or a sponsor proposes a study using a device previously approved for the same application, a Health Canada Clinical Trial Application (CTA) is not required. However, if the study proposes to use an approved device for another indication or in association with a drug not previously approved for the same indication, a CTA must be submitted to Health Canada [24].

For specific details on the regulatory process in countries outside North America, interested researchers should contact the relevant national health department's regulatory branch.

### 10.13.3.2 "First-in-Man" Studies

First-in-man trials present a particular concern for drug studies, especially since March of 2006 when six healthy volunteers were administered the study drug TGN1412 and subsequently developed multiorgan failure requiring intensive care [25]. Fortunately, there were no deaths involved with this trial. However, this brought to light particular issues regarding first-in-man trials. In 2007, the Royal Statistical Society published a study focusing on statistical consideration for first-in-man studies [26]. They recognized that first-in-man studies are vital in the development of new drugs and devices and such trials must have a high level of safety. Recommendations were made for improvement in the areas of preparatory clinical work, experimental design (particularly with respect to drug dosing), risk assessment, and communication.

With respect to device trials, first-in-man studies may be slightly less cumbersome to undertake. Regulatory bodies require evidence that the safety and effectiveness of the device has been established as far as possible prior to human use, that the study will accomplish its objectives, and that trial subjects are protected to every possible extent. For first-in-man studies, it is generally easier to justify small subject numbers. However, the "pilot" study can often be omitted through the appropriate use of preclinical and bench-top studies where the risks to patients can be shown to be minimal.

### 10.13.4 SUMMARY

The specialty of urology is a fertile ground for the astute scientist with many opportunities for cutting-edge drug and device research. Only through the conduct of properly constructed clinical trials and the proper interpretation of study findings, however, will our patients truly reap the benefits of this activity. It is hoped the

information provided in this chapter will aid those considering urological investigation and inspire a commitment to the highest quality research endeavor.

## REFERENCES

1. Friedman, L. M., Furberg, C., and DeMets, D. L. (1998), *Fundamentals of Clinical Trials*, 3rd ed., Springer, New York, pp. 41–78.
2. Cutler, S. J., Greenhouse, S. W., Cornfield, J., et al. (1996), The role of hypothesis testing in clinical trials. Biometrics seminar, *J. Chronic Dis.*, 19, 857–882.
3. Denstedt, J. D., Wollin, T. A., Sofer, M., et al. (2001), A prospective randomized controlled trial comparing nonstented versus stented ureteroscopic lithotripsy. *J. Urol.*, 165, 1419–1422.
4. Anon. (1992), American Urological Association symptom index for benign prostatic hyperplasia, *J. Urol.*, 148, 1549–1557.
5. Rosen, R., Riley, A., Wagner, G., et al. (1997), The international index of erectile dysfunction (IIEF): A multidimensional scale for assessment of erectile dysfunction, *Urology*, 49, 822–830.
6. Health Assessment Lab, Medical Outcomes Trust and QualityMetric Incorporated, 1992–2002; available at <http://www.outcomes-trust.org/instruments.htm#SF-36>.
7. Litwin, M. S., Hays, R. D., Fink, A., et al. (1998), The UCLA Prostate Cancer Index: Development, reliability and validity of a health-related quality of life measure, *Med. Care*, 36, 1002–1112; available at <http://roadrunner.cancer.med.umich.edu/epic/epicmain.html>.
8. Joshi, H. B., Newns, N., Stainthorpe, A., et al. (2003), Ureteral stent symptom questionnaire: Development and validation of a multidimensional quality of life measure, *J. Urol.*, 169, 1060–1064.
9. Beiko, D. T., Watterson, J. D., Knudsen, et al. (2004), Double-blind randomized collected trial assessing the safety and efficacy of intravesical agents for ureteral stent symptoms after extracorporeal shockwave lithotripsy, *J. Endourol.*, 18, 723–730.
10. Higgins, B., and Thompson, I. M. (2004), The prostate cancer prevention trial: Current status. *J. Urol.*, 171, S15.
11. Thompson, I. M., Goodman, P. J., Tangen, C. M., et al. (2003), The influence of finasteride on the development of prostate cancer. *N. Engl. J. Med.*, 349, 215–224.
12. Roehrborn, C. G. (1996), The placebo effect in the treatment of benign prostatic hyperplasia, in Kirby, R. S., McConnell, J. D., Fitzpatrick, J. M., Roehrborn, C. G., and Boyle, P., Eds., *Textbook of Benign Prostatic Hyperplasia*, Oxford, UK, Isis Medical Media, pp. 239–253.
13. Good, P. I. (2006), *A Manager's Guide to the Design and Conduct of Clinical Trials*, 2nd ed., Wiley-Liss, Hoboken, NJ, pp. 23–36.
14. Barbui, C., Violante, A., and Garattini, S. (2000), Does placebo help establish equivalence in trials of new antidepressants? *Eur. Psychiatry*, 15, 268–273.
15. Tan, A. H., Nott, L., Hardie, R., et al. (2005), Long-term results of microwave thermotherapy for symptomatic benign prostatic hyperplasia, *J. Endourol.*, 19, 1191–1195.
16. Portney, L. G., and Watkins, M. P. (2000), *Foundations of Clinical Research: Applications to Practice*, 2nd ed., Prentice Hall Health, Upper Saddle River, NJ, pp. 33–133.
17. Weijer, C., Dickens, B., and Meslin, E. M. (1997), Bioethics for clinicians: 10. research ethics, *Can. Med. Assoc. J.*, 156, 1153–1157.

18. Levine, R. J. (1986), *Ethics and Regulation of Clinical Research*, 2nd ed., Urban & Schwarzenberg, Baltimore, pp. 87–98.
19. The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research; available at <http://www.impactcg.com/docs/BelmontReport.pdf>.
20. Anon. (2000), World Medical Association Declaration of Helsinki: Ethical principals for medical research involving human subjects, *JAMA*, 284, 3043–2045.
21. CIOMS (2002), Council for International Organizations of Medical Sciences (CIOMS). International ethical guidelines for biomedical research involving human subjects, *Bull. Med. Ethics*, 182, 17–23.
22. U.S. Food and Drug Administration. Center for Devices and Radiographic Health; available At: <http://www.fda.gov/CDRH/>.
23. Health Canada; available at: <http://www.hc-sc.gc.ca>.
24. Beiko, D. T., Watterson, J. D., Knudsen, B. E., et al. (2004), Double-blinded prospective randomized controlled trial assessing the safety and efficacy of intravesical agents for ureteral stent symptoms following extracorporeal shock wave lithotripsy, *J. Endourol.*, 18, 723–730.
25. Hemelaar, J. (2007), Minimising risk in first-in-man trials, *Lancet*, 369, 1496–1497.
26. Royal Statistical Society's working party on statistical issues in first-in-man studies; available at: <http://www.rss.org.uk/>.



# 10.14

---

## *Clinical Trials on Cognitive Drugs*

**Elisabetta Farina and Francesca Baglio**

*Neurorehabilitation Unit, Don Carlo Gnocchi Foundation, Scientific Institute and University,  
IRCCS, Milan, Italy*

### **Contents**

|           |                                                                            |     |
|-----------|----------------------------------------------------------------------------|-----|
| 10.14.1   | Kinds of Diseases Representing Target in This Field                        | 706 |
| 10.14.1.1 | Dementia                                                                   | 706 |
| 10.14.1.2 | Trauma                                                                     | 707 |
| 10.14.1.3 | Stroke                                                                     | 707 |
| 10.14.2   | Selecting Population for Study                                             | 708 |
| 10.14.2.1 | Common Criteria                                                            | 708 |
| 10.14.2.2 | Particular Criteria                                                        | 710 |
| 10.14.3   | Endpoint selection and Assessment Tools for Efficacy and Safety Evaluation | 713 |
| 10.14.3.1 | Efficacy Evaluation                                                        | 713 |
| 10.14.3.2 | Cognitive Tools                                                            | 713 |
| 10.14.3.3 | Outcome Measures of Function                                               | 715 |
| 10.14.3.4 | Global and Quality-of-Life Measures                                        | 717 |
| 10.14.3.5 | Measures of Neuropsychiatric and Behavioural Changes                       | 719 |
| 10.14.3.6 | Other Measures: Outcomes Specific To Caregivers                            | 720 |
| 10.14.3.7 | New and Future Measures                                                    | 721 |
| 10.14.3.8 | Safety Evaluation                                                          | 723 |
| 10.14.4   | Duration of study and Possible bias in Trials on Cognitive Impairment      | 723 |
| 10.14.4.1 | Duration                                                                   | 723 |
| 10.14.4.2 | Possible Bias                                                              | 724 |
| 10.14.5   | Conclusion                                                                 | 725 |
|           | References                                                                 | 725 |

10.14.1 KINDS OF DISEASES REPRESENTING TARGET IN THIS FIELD

Cognitive disabilities, especially memory disturbances and cognitive deficits, posttraumatic brain injury (TBI), and poststroke are important targets for clinical pharmacological trials given their important epidemiological and social impact (see Table 1).

Until a few years ago, patients with dementia, stroke, or TBI faced a bleak future with a diminished quality of life, reduced productivity in terms of their capacity to work and contribute to family life, and dependence on caregivers. Without a cure or any means of alleviating the patient’s symptoms, for example, by prescribing drugs that slowed the progression of a degenerative disease or neurorehabilitation to restore some of the brain function lost through stroke or traumatic brain injury, the doctor’s role was reduced to advising the patient on coping strategies. However, over recent years and thanks to new scientific data, the treatment of brain diseases has undergone a revolution.

10.14.1.1 Dementia

Brain changes leading to cognitive decline and dementia, and their relation to disability, are key topics requiring investigation at the population level. Global aging is a recent phenomenon. Its potential impact on social and economic aspects of more developed countries highlighted aging as an important issue toward which resources must be directed. Dementia is an increasingly common diagnosis in our aging population, and the numbers are expected to rise exponentially in coming years. Alzheimer’s disease (AD) alone actually affects 4.5 million people in the United States [2], while millions more are currently affected by vascular dementia (17.6% of all incident dementia [3]), Lewy body disease, and frontotemporal dementia. Each of these is a distinct entity, though overlapping symptoms and comorbidities occur frequently. Within the past two decades research has progressed rapidly on multiple fronts, including epidemiology, etiology, pathology, diagnosis, and treatment. It is important for clinicians to recognize early signs and symptoms of dementia and to note critical differences among them. Dementia research has moved beyond description of symptoms and clinicopathological correlation to the elucidation of risk factors, the pathobiology of the disease process, and, most important, to the first generation of dementia treatments. Furthermore, the question of how to divide aging phenomena into categories of healthy, normal, or disease-related remains a difficult but important goal in medicine: We are entering an era of dementia care that will be based upon the identification of potentially modifiable risk factors and early disease markers and the application of new disease-specific diagnostic tools and treatment modalities. Literature data showed, for example, that older persons

TABLE 1 Epidemiological and Social Impact of Dementia, Stroke, and TBI

| Disease                | Prevalence       | Annual Cost (euro)/Year |
|------------------------|------------------|-------------------------|
| Dementia               | 5 million people | 55 billion              |
| Traumatic brain injury | 700,000          | 3 billion               |
| Stroke                 | 1 million people | 22 billion              |

Source: Data from Olsen et al. [1].

can develop demonstrable cognitive impairment (“mild cognitive impairment,” MCI; [4]), especially memory deficits without crossing the threshold for dementia, but that these patients have an increasing risk of developing dementia, especially Alzheimer disease [5]. We expect that pharmacological approaches will become increasingly integrated with behavioral, genetic, and neuroimaging methods to investigate not only disease processes but also the normal and preclinical individual differences that underlie successful—or unsuccessful—aging. These future directions share an important feature with the current focus of the field: a shift from the dismal characterization of aging as an inevitable process of brain damage and decline to the concept (emerging from cognitive neuroscience) that aging can be successful associated with gains and not only with the losses. It is not necessarily a unidirectional process but rather a complex phenomenon characterized by reorganization, optimization, and enduring functional plasticity that can enable the maintenance of a productive and happy life.

#### 10.14.1.2 Trauma

Traumatic brain injury is another leading cause of disability, and survivors often suffer cognitive, mood, and behavioral disorders. TBI is an insult to the brain that leads to temporary or permanent impairments of cognitive abilities and physical functioning. TBI results principally from vehicular incidents, falls, acts of violence, and sports injuries and is more than twice as likely to occur in men as in women. The estimated incidence in U.S. rate is 100 per 100,000 persons, with 52,000 annual deaths. The highest incidence is among persons aged 15–24 years and 75 years or older, with a less striking peak in incidence in children aged 5 years or younger [6]. Since TBI may result in lifelong impairment of physical, cognitive, and psychosocial functioning and prevalence is estimated at 2.5–6.5 million individuals, it is a disorder of major public health significance. Mild TBI is significantly underdiagnosed, and the likely social burden is therefore even greater. Given the large toll of TBI and absence of a cure, prevention is of paramount importance. More recently, a hospitalized incidence rate of 229 per 100,000 was calculated for England for TBI in 2001–2002 [7]. Again, significant local variation was noted (91–419 per 100,000 across health authorities). In the United Kingdom, “serious” TBI is estimated to have an incidence rate of 52 per 100,000, while an incidence of 12 per 100,000 was reported in Australia for “severe” brain injury [8].

Research in traumatic brain injury has shown that the primary traumatic insult sets in training a cascade of events in the brain that leads to secondary injury, and that this secondary injury can be exacerbated by systemic insults such as hypoxia and shock [9]. Strategies to minimize both secondary insults and the evolution of secondary damage in the acute phase, while maximizing plasticity in the following course, are essential goals for those patients, but pharmacological studies in this field are lacking. The same can be said for long-term cognitive impairment deriving from a previous TBI, a condition affecting a significant percentage of TBI survivors.

#### 10.14.1.3 Stroke

Stroke is a vascular disorder characterized by the sudden death of brain cells due to a reduced or blocked supply of blood. The symptoms depend on the site of the

stroke, but they can include motor, speech deficits (aphasia), visuospatial deficits (including neglect), cognitive decline, and dementia. The incidence of stroke is reaching pandemic proportions: Stroke is the third leading cause of death and one of the leading causes of adult disability in North America, Europe, and Asia [10–12]. Each year across the world, 5.5 million people die as result of a stroke. A further 15 million people survive but are disabled and face the prospect of a later, even more serious stroke [13]. Costs of care, which have to be met by either the individual or the state, can also prove crippling [14].

A number of well-designed randomized stroke trials and case series have now been reported in the literature to evaluate the safety and efficacy of pharmacologic therapy for the treatment of acute ischemic stroke, but there are few data on long-term cognitive outcomes and possible treatments after first-ever stroke. Literature data showed that recurrent stroke significantly contributes to global cognitive decline after a first-ever stroke. Furthermore, studies explained that stroke prevention would be important in reducing dementia related to stroke (e.g., hypertension treatments; [15]). There is a growing interest in the latter because patients with only mild vascular cognitive impairment and no dementia are at increased risk of cognitive deterioration and should have more opportunities for treatment and prevention [16]. Further investigation of these issues may help to understand mechanisms underlying stroke-related cognitive decline and assist in planning interventions to prevent dementia related to stroke, and to ameliorate existing cognitive deficits.

## 10.14.2 SELECTING POPULATION FOR STUDY

### 10.14.2.1 Common Criteria

**Diagnostic Accuracy** Before starting a clinical pharmacological trial, clear classification of the diseases and well-recognized diagnostic criteria are necessary to realize the selection of the study population. This is relevant not only for scientific reasons (standardization and comparability between data collected by several scientific groups) but also for practical aspects (e.g., without a clear stadiation of disease, it is unthinkable to lay down different therapeutic options). A detailed anamnesis as well as physical examination at presentation are necessary for the overall evaluation of the single subject; the employment of consensus criteria for the diagnosis is essential for the group analysis comprehension. Actually, consensus criteria are well defined in dementia (Table 2 cites the most recent clinical criteria). Unlike dementia, cognitive and behavioral sequelae in stroke and TBI recognize only clinical classifications of the specific symptoms. Nevertheless, attributing single individuals to the classic syndrome categories is not easy, due to the complexity of the underlying brain damage and the existence of clinical variants. For example, aphasia, the loss or impairment of language caused by damage to several discrete cognitive systems, is one of the most devastating cognitive sequelae of stroke. Well-defined classic aphasic profiles (anomia, Broca's, conduction, Wernicke's, and transcortical) are frequent when patients with only single lesions are considered [24], whereas global and nonclassified aphasias account for 50% of cases admitted to acute stroke units, especially among patients with previous strokes [25]. A difficult to ascertain proportion of aphasic individuals cannot easily be assigned to the classic

**TABLE 2 Recent Clinical Criteria in Dementia**

| Type of Dementia        | Criteria                                                                                                                                                                                                                                                                       |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Alzheimer's disease     | Criteria of the Diagnostic and Statistical Manual of Mental Disorders [17]<br>The National Institute of Neurologic, Communicative Disorders and Stroke—Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) criteria [18];<br>NINCDS-ADRDA research criteria [5] |
| Dementia with Lewy body | The Consortium for DLB diagnostic criteria [19, 20]                                                                                                                                                                                                                            |
| Frontotemporal dementia | Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria [21]<br>Report of the Work Group on Frontotemporal Dementia and Pick's Disease [22]                                                                                                             |
| Vascular dementia       | The National Institute of Neurologic Disorders and Stroke and the Association Internationale pour la Recherche et l'Enseignement en Neuroscience (NINDS-AIREN) diagnostic criteria [23]                                                                                        |

*Note:* The clinical diagnosis should rely on criteria that have been proposed to increase the reliability and accuracy of the diagnosis. The accuracy of these diagnostic criteria varies as a function of the dementia.

syndrome categories, whereas others display atypical clinicoradiological correlations (e.g., Wernicke's aphasia associated with frontal lobe lesions), in part, a result of the idiosyncratic brain organization of language networks [26]. Indeed, other right-handed individuals show no aphasia despite having large lesions in the left hemisphere because in such cases language is innately lateralized to the right hemisphere [27]. However, functional imaging [functional magnetic resonance imaging (MRI)], mapping language brain areas, may currently theoretically provide means to improve the diagnostic accuracy by clinicians [28, 29].

In any case, it is now recommended submitting patient to baseline imaging evaluation before beginning trial [computed tomography (CT) or magnetic resonance imaging (MRI), within 6–12 months]. This is important to confirm the original diagnosis and order to exclude other concomitant pathologies (e.g., vascular impairment for degenerative dementia, brain tumor).

**Severity Criteria** Another relevant aspect to consider when planning a trial is the stadiation of disease that includes both temporal dimension and severity. In pharmacological trials stadiation is important because different phases of illness require different pharmacological approaches. For example, this is true for the treatment of AD: Literature data showed that cholinesterase inhibitors for AD treatment are useful in the mild-to-moderate phase and ineffective in the severe stage. On the contrary, memantine, an *N*-methyl-D-aspartate (NMDA) antagonist, is approved for the treatment of moderate to severe AD. Predefined severity criteria should be made explicit in future trials to ensure transparency with regard to comparability and general applicability of results.

**Cultural and Educational Background** Cultural and educational background of the population under study is another relevant characteristic to hold in consideration. Really, cognitive and behavioral measures are particularly vulnerable to the effect of culture, language, and education. The relevance of some domains may

vary between cultures; functional disability, for example, may be less important in cultural contexts where independence and autonomy in activities of daily living are not part of the older person's role [30]. Similarly, memory impairment is not regarded as important in all cultures [31]. Moreover, the recruited subjects must have the same range of years in age and in school attendance. The age and school attendance setting is an essential requisite for making a comparison between different group performances on assessment tools.

**Baseline Safety Evaluation** Regardless of the drug that we want to try, it is also essential to assess adverse effects of treatment in order to determine the risk–benefit ratio. For this reason, it is well advised to have a safety assessment that at least includes laboratory examinations (hematology and urinalysis) and an electrocardiogram (ECG). Further clinical or instrumental evaluations before starting and during trial are established in accordance with the preliminary safety data of the drug to try.

#### 10.14.2.2 Particular Criteria

**Population with Dementia** In patients with dementia some aspects are typical, and they must be kept in mind before starting the clinical trial. The power of clinical trials in dementia is significantly reduced by diagnostic inaccuracy; therefore, there is a tension between requirements for diagnostic accuracy and the necessity for clinical trial results to be generally applicable to the clinical population that might subsequently receive the intervention.

First, it is well recognized that diagnostic accuracy and the detection of treatment effects may be easier to attain in moderate than in mild dementia, but the potential benefits of disease modification are likely to be greatest in the early or prodromal stages. Earlier intervention with disease-modifying therapies is likely to be more effective when there is a lower burden of amyloid and hyperphosphorylated tau and may truncate the devastating effects of secondary events due to inflammatory, oxidation, excitotoxicity, and apoptosis. Also for this reason revised research criteria for AD [National Institute of Neurologic, Communicative Disorders and Stroke–Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) research criteria; [5]] are now available. These new criteria would allow diagnosis when symptoms first appear, before full-blown dementia, thus supporting earlier intervention at the prodromal stage. They are centered on a clinical core of early and significant episodic memory impairment, and they stipulate that there must also be at least one or more abnormal biomarkers [among structural neuroimaging with MRI, molecular neuroimaging with positron emission tomography (PET), and cerebrospinal fluid analysis of amyloid  $\beta$  or tau proteins]. However, these criteria require significant expertise, technical skills, and financial resources to allow the comprehensive assessment of MRI, PET, and cerebrospinal fluid, and this diagnostic framework may not yet be feasible in all memory clinics and certainly not in most epidemiological studies and all clinical trials: The research criteria will need to be adapted for use in standard clinical settings.

The appearance of these criteria, however, could overcome problems encountered in clinical trials targeting MCI population, where the clinical heterogeneity of patients with this diagnosis makes the demonstration of clinical effects of a drug very difficult.

A second aspect to consider is the coexistence of several pathologies in these patients given their old age. Indeed, dementia onset is most often situated in old age (subjects who are 65 years old or more), and it is common knowledge that age is an independent risk factor for other pathologies (hypertension, diabetes, depression, etc.). Therefore, it would be necessary to exclude those conditions that may adversely affect the cognitive domain and/or giving safety problems. However, it is important to avoid a hyperselection of the study population in order to obtain a group of subjects similar to the general patient population, that is, patients seen in everyday clinical practice. This is also important to avoid insurmountable difficulties in recruitment.

Another point to consider is the common presence in dementia of behavioral and psychological symptoms [Behavioral and Psychological Symptoms of Dementia (BPSD) or neuropsychiatric symptoms; [32]]. These disturbances are distressing, present major difficulties for caregivers, and accelerate nursing home placement [33]. For these reasons antipsychotic drugs are commonly prescribed to many people with dementia (up to 45%) in residential or nursing homes but also at home, often for prolonged periods. However, studies reported that antipsychotics have a negative impact on cognition in patients with dementia [34, 35]. Even though it would be appropriate to exclude patients treated with these drugs from clinical trials, also in this case it is important to avoid a population hyperselection in order to reproduce as much as possible the real clinical situation in everyday practice. Therefore, you can evaluate whether to include patients in some specific cases: when the patient assumes a low dose of benzodiazepines with hypnotic purpose or when he has been in steady therapy with antidepressant for at least 6 months. Furthermore, it is possible to consider in some specific trials the sporadic use of low doses of neuroleptics (e.g., haloperidol).

Approaching dementia, another critical aspect must be well known: the mixed dementia. Mixed dementia, a combination of definite AD and vascular encephalopathy, is a controversial but important issue to consider given the frequency of this condition in the oldest patients. In particular, it is unclear how many vascular lacunar lesions are compatible with a diagnosis of AD (generally it is considered only one), taking into account that isolated leukoariosis is not considered an exclusion criteria for the diagnosis of neurodegenerative diseases. In the future, it could be useful to include into pharmacological trials also patients with mixed dementia.

### ***Population with Cognitive and Behavioral Deficits after TBI and Stroke***

Different than in dementia, several cognitive and behavioral clinical situations can occur following TBI or stroke, depending on the degree and type of brain injury. The variability of pretraumatic and prestroke brain function further adds elements to the clinical complexity. For example, although some brain structures are more vulnerable in TBI than others, the vast heterogeneity of the TBI patients is likely



to cause high variability in the therapeutic response to any agent. Moreover, a certain cognitive symptom may have a variable biochemical background. For example, apathy and poor initiation may arise from deficits in the dopaminergic, serotonergic, noradrenergic, or cholinergic systems or from an imbalance between these systems. At the clinical level, however, the manifestations may be very similar.

Therefore, in order to evaluate the efficacy of a pharmacological treatment in these patients, it is important to choose the population using stringent and accurate multicriteria. These criteria can be both clinical (e.g., clinical presentation at the neurological examination: neglect, aphasia, etc.) and founded on neuroimaging findings (injury level established by CT or MRI examination). Functional imaging, in particular, may theoretically provide means to evaluate the basis for treatment interventions [36], but for practical and economical reasons it seems unrealistic that, at an individual level and at larger scale, we could effectively use such tools for treatment decisions in the foreseeable future.

Another important point is that, since the damage can occur at any age, age is an important prognostic factor, and the recovery can be different between several patients, it is necessary that patients recruited for the trial are similar not only in the demographic characteristics (age, sex, and education) but also as far as premorbid conditions are concerned (e.g., it is necessary to exclude preexisting psychiatric or substance abuse problems, poor general health). The same can be said for comorbidities (such as chronic pain, depression, substance abuse, life stress, unemployment, and protracted litigation). The use of poorly matched experimental and control groups increases the likelihood that neuropsychological group differences will reflect premorbid group differences rather than the effect of damage postinjury. Moreover, it is of primary importance the good knowledge of TBI clinical details (e.g., duration of posttraumatic amnesia, coma duration, patient level at the Glasgow Coma Scale). Clearly, according to previous cited common criteria, another relevant aspect to consider is the stadiation of disease: temporal dimension postinjury and severity of the pathology at the time of the enrollment. The temporal dimension postinjury is relevant: The overall effect size for the influence of mild TBI on neuropsychological functioning, for example, in the *postacute* phase (more than 3 months postinjury) is very different than in the *acute* phase. Indeed, in a meta-analysis on neuropsychological outcome following TBI, Schretlen and Shapiro [37] concluded that mild TBI results in clear decrements in cognitive functioning that resolve in 1–3 months. In contrast, they illustrated that patients with moderate or severe brain injuries show clear evidence of early recovery but they are at risk for frank deficits more than 2 years after injury. Moreover, studies of spontaneous recovery in poststroke aphasia have shown that the greatest improvement occurs in the first 2 or 3 months, with a discernible but less evident improvement in the following months, and most patients reaching a plateau after 1 year [38]. If the stability of cognitive sequelae is ascertained, TBI and stroke survivors can be candidates to cross-over studies and not only to classical prospective clinical trial. This is not possible in dementia due to the progressive nature of the disease.

To evaluate the severity of TBI pathology at the time of the enrollment, specific scales are actually available to establish a global cognitive evaluation of the patients. They are also useful to provide follow-up and comparison of the patients [e.g., Rancho Los Amigos (RLA) Levels of Cognitive Functional Scale for TBI].

### 10.14.3 ENDPOINT SELECTION AND ASSESSMENT TOOLS FOR EFFICACY AND SAFETY EVALUATION

#### 10.14.3.1 Efficacy Evaluation

Dementia, stroke, and TBI are multifactor disorders that negatively impact cognitive, behavioral, and functional abilities. Complete and accurate evaluation of the efficacy of treatments is crucial, but the heterogeneity of symptoms makes uniform assessment difficult. Due to the complex nature of these diseases, it is crucial to design assessment procedures that accurately track disease progression across a wide variety of symptom domains. Although a number of outcome measures have been developed for this purpose, it is suitable to use the diagnostic tools that have sensibility, specificity, and relevancy enough to test modifications of estimated characteristics (primary and secondary endpoints). Generally, primary outcome measures assess key pathology domains (such as cognition, function, global impression, and quality of life in dementia), while secondary endpoint measures assess other meaningful and measurable domains, which may constitute appropriate targets for interventions (e.g., caregiver outcomes, behavioral measures, activity of daily living).

These assessment tools must be based on reliable and validated scales and scores and must be usable in many settings. Ideally, the administration of the instrument should be brief; however, many of the measurement tools used in clinical research take too much time to administer to be employed in clinical settings.

**Dementia** The goals of treatments in this field can be to prevent dementia emerging in patients with MCI, or, in those with established dementia, to evaluate improvement and stability versus deterioration. The outcome measures used to evaluate efficacy of dementia treatments should be sensitive to both improvement and decline and have minimal floor and ceiling effect [39]. The primary outcomes include key domains of dementia: cognition, function, global impression, and/or quality of life. Secondary measures assess behavioral ratings and other outcomes such as nursing home placement, caregiver burden, and others.

Table 3 shows the principal primary and secondary outcome measures used to evaluate efficacy in dementia treatments. For MCI patients, one of the primary outcome measures should be the conversion to dementia, that is, how much a drug is able to reduce the absolute and relative risk of conversion. However, this is a critical primary endpoint to evaluate, due to difficulties in establishing when the individual patient exactly passes the threshold of dementia (it basically depends from loss of independence in everyday life), and to the necessity of planning a long-term study to catch enough cases of conversion to dementia (at least 2 years).

#### 10.14.3.2 Cognitive Tools

The cognitive tools should provide an assessment of multiple cognitive functions that are affected by dementia (memory, orientation, language, frontal abilities, praxis, and others), plus a measure of global cognitive decline. The Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog) is the de facto standard primary outcome neuropsychological measure for AD trials [40]. It measures several

**TABLE 3 Primary and Secondary Outcomes Used to Evaluate Efficacy in Dementia Treatments (see text for further details)**

| Outcome                               | Scale                                                                                                                                                                                                                                                                                                                                    |
|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cognition                             | Alzheimer's Disease Assessment Scale—cognitive subscale (ADAS-Cog)<br>Mini-Mental State Examination (MMSE)<br>Severe Impairment Battery (SIB)                                                                                                                                                                                            |
| Function                              | Activities of daily living (ADL): basic ADL (BADL) and instrumental ADL (IADL)<br>The Alzheimer's Disease Activities of Daily Living International Scale (ADL-IS)<br>Performance measures:<br>The Physical Performance Test (PPT)<br>The Direct Assessment of Functional Status (DAFS)<br>The Functional Living Skills Assessment (FLSA) |
| Global impression                     | Clinical Interview-Based Impression of Change (CIBIC, CIBIC-plus)                                                                                                                                                                                                                                                                        |
| Quality of life                       | Quality of Life—Alzheimer Disease Scale (QOL-AD)<br>The Late-Stage Dementia Scale (QUALID)<br>The Alzheimer Disease-Related Quality of Life Scale (ADRQL)<br>The Dementia Quality of Life Scale (D-QoL)                                                                                                                                  |
| Behavioral and psychological symptoms | Neuropsychiatric Inventory (NPI)<br>ADAS non-Cog<br>The Behavioral Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD)<br>Cornell Geriatric Scale                                                                                                                                                                                  |
| Other                                 | Global measures of burden                                                                                                                                                                                                                                                                                                                |

cognitive domains, including memory, language, and praxis. Total scores range from 0 to 70, with scores greater than 40 indicating severe impairment and a score of 10–25 indicating mild impairment. Many regulatory authorities recognize a four-point change on the ADAS-Cog at 6 months as indicating a clinically important difference [41–44]. Interestingly, the ADAS-Cog scale has the advantage of having parallel forms that avoid learning as a confounding factor. Unfortunately, it has limitations that are worth noting. The first of these is important when comparing results in different kinds of dementia. Since the ADAS-Cog is an assessment designed for AD, other dementias require the use of further neuropsychological instruments to assess the cognitive domain particularly affected in that specific type of disease. In patients with frontotemporal dementia, for example, frontal abilities that are specifically involved should be investigated (e.g., with the Frontal Assessment Battery [45]); in patients with Lewy body dementia, visuospatial deficits and cognitive fluctuations should be taken into consideration (for visuospatial deficits, several tests exist to evaluate visuoperceptive or visuospatial abilities, for example, some subtests of the Birmingham Object Recognition Test and the Benton Line Orientation Test, respectively; For the latter, short forms are also available [46]; cognitive fluctuations should be evaluated with specific questionnaires, such as the DLB fluctuation scale by Ferman et al. [47]).

The ADAS-Cog takes 30–45 minutes to administer; therefore it is widely used only by researchers. On the contrary, Mini-Mental State Examination (MMSE; [48]) is a brief cognitive screening tool that only takes 5–15 minute to administer. Therefore, it is used both in research and clinical settings.

As previously noted, the measures should be sensitive to disease progression over time and useful for a broad range of severity. Another limitation of ADAS-Cog is that there is a floor effect in severe AD patients [49]. For this reason, this scale is

not generally used to evaluate the cognitive changes in advanced patients, and the Severe Impairment Battery (SIB; [50–52]) is currently the most accredited instrument for tracking progression in severely demented patients [53]. The SIB includes six subscales (attention, orientation, language, memory, visual perception, and construction) and a brief assessment of social skills, praxis, and response to name. Total score ranges from 1 to 100, with a score of 63 or less being considered “very severely impaired”; this battery provides accurate assessment in patients with MMSE scores of 13 or less [54]. Like the ADAS-Cog, the original SIB is a long tool to be administered to the patient (it takes approximately 30 minutes). For this reason, a short form of the Severe Impairment Battery (SIB-S) has been developed that only takes 10–15 minutes to administer, making it more appropriate for the use in clinical pharmacological trials in patients with very severe dementia, while it maintains the attributes of the original SIB [55].

### 10.14.3.3 Outcome Measures of Function

Outcome measures of function are other important data to collect in order to demonstrate the efficacy of a drug in clinical pharmacological trials. Loss of functioning in complex tasks of everyday life is a hallmark feature of dementing illness: In fact, current clinical diagnostic criteria for dementia [17, 18] require documentation of cognitive decline as well as loss of competence in either social or occupational domains. Furthermore, measures of function allow clinicians to assess the need for personal and institutional care, and to design individually tailored interventions, so that a demented individual receives sufficient help while avoiding unnecessary assistance that may lessen patient self-esteem, increase caregiver burden, and accelerate functional deterioration. The ability to perform activities of daily living (ADLs) has a significant impact on the patient, caregiver, and society. Among the functional disability scales based on patient or informant response, we can distinguish generic tools.

Generic scales are designed to measure function regardless of the person’s specific medical history: With these scales, the disability of a patient with dementia can be compared with the disability of a patient with arthritis, for example. This type of tool evaluates basic ADLs (BADLs) and/or instrumental ADLs (IADLs). IADLs include activities that concern the ability to adapt to the environment: Some of these activities (such as telephone usage, financial management, and shopping) are particularly dependent on cognitive functioning, and often serve as earlier predictors of poor outcome; this subset of IADLs is known as advanced cognitive IADL [56]. The most popular generic BADL scale is the Katz ADL scale [57]: This measure of basic functioning requires an observer to rate the patient’s ability in bathing, dressing, going to the toilet, transferring, maintaining continence, and feeding. The most commonly used IADL scale is the Lawton IADL scale [58].

This last one includes eight items: telephoning, shopping, meal preparation, housekeeping, laundry, transportation, medications, and handling finances. A problem with the Katz IADL is that the original version of this scale uses all items for women and only five tasks for men (three items are removed for men to avoid tasks that men did not usually perform because of societal norms associated with the division of labor in household management: But these norms are changing in the last decades). A low performance in IADL scales is associated with mild

dementia, while ADL deficits characterize more severe dementia [59]. Although, measuring ability to perform ADLs and IADLs is important when assessing the therapeutic efficacy of a drug, these scales are very poorly sensitive to change, and these generic scales have been criticized for a relative insensitivity to functional losses due to cognitive impairment [60]. Therefore, alternative tools should be identified, such as more detailed functional scales or other performance instruments.

The most popular specific informant-based scales are the Blessed scale [61], the Nurses' Observation Scale for Geriatric Patients [62] (it includes a subscale specifically designed to assess IADLs, but this last one contains only five items), and the Alzheimer's disease activities of daily living international scale (ADL-IS) [63]. The ADL-IS was launched to develop an ADL scale aimed to measure pharmacological response in AD patients [63]; it was carefully developed in a 10-year period and covers a large sample of ADLs. It shows high correlations with measures of cognitive functions and stage of dementia, thus appearing a promising tool also for clinical setting. However, clinical experience with this scale is still limited. It must be added that the use of informant reports is appealing because it can allow comprehensive and longitudinal evaluations but has potential disadvantages: The source of information is filtered through an observer, and it may contain distortions based on the personal needs and opinions of the informant [64]. For this reason, in the last years, some performance measures of everyday functioning have been proposed to assess demented patients.

Performance measures are tools in which an individual is asked to perform an activity and is evaluated in a formal manner on that performance with standardized criteria (counting, timing, observation of the need for assistance, and appropriateness or completeness of the task). These measures have been used in medical rehabilitation for diagnosis and to demonstrate therapeutic progress, and have shown to predict survival, hospitalization, use of assistance, long-term-care, and nursing home placement [65–69]. The main advantage to this approach is that it directly measures ability, thus eliminating faulty perceptions. Moreover, with performance measures it is possible to take into account a large variability of the ability to perform a task, thus increasing the sensitivity of the tool [70]. Other claimed advantages are good patient acceptability and good interrater reliability [68, 71–73]. Examples are the ADL Situational Test (AST; [74]), the Physical Performance test (PPT; [72]), the Structured Assessment of Independent Living Skills (SAILS; [73]), the Direct Assessment of Functional Abilities (DAFA; [75]), the Everyday Problems Test for the Cognitively Challenged Elderly (EPCCE; [76]), and the Direct Assessment of Functional Status (DAFS; [71]). In the AST [74], evaluation is limited to one BADL (dressing) and three IADL (meal preparation, telephoning, and purchasing), each ADL item broken down into subtasks. For each item there are two scores: the time required to complete each ADL task and a performance score derived from the level of assistance needed by the subject to complete the task.

As far as the PPT [72] is concerned, this tool has been proposed to evaluate a general geriatric population rather than patients affected by dementia. The test includes some motor activities, three BADLs (eating, dressing, and walking), and writing a sentence. Therefore, IADLs are not evaluated in the PPT.

The DAFS [71] features mainly IADL skills (communication abilities, transportation, financial skills, shopping skills), along with two BADL skills (eating and

dressing/grooming skills), and time orientation. Performance is rated as correct or incorrect. The DAFS, which represents an interesting tool, but it explores only four IADL skills and does not define the degree of dependence of the patient.

The SAILS [73] includes a subset of BADL (dressing, eating) and IADL (meal preparation, telephoning, and money-related skills), as well as fine and gross motor tasks, and high-order abilities (communication, social interaction, and orientation). Each item includes five subtasks. Each subtask is scored from “unable” to “normal,” with an intermediate level of disability, with both time (on timed tasks) and accuracy (in all tasks) contributing to the score.

The DAFA [75] is a direct measure of IADLs conceived to be compared with the Pfeffer Functional Activities Questionnaire (PFAQ; [77]). Ten items are adapted to explore seven functional domains: money management, shopping, hobbies, meal preparation, awareness, reading, and transportation. Items on the DAFA and PFAQ are both evaluated by using an integer score from 0 (independent functioning) to 3 (dependent functioning).

Finally, the EPCCE [76] has been specifically designed to provide an objective measure of problem solving with respect to cognitively demanding tasks encountered in daily livings. The patient is shown 16 stimuli and asked to solve two problems related to each stimulus. Each item is scored right or wrong. The authors propose the EPCCE as a complement to other performance measures (e.g., the DAFS). These performance instruments, for example, the DAFS, represent interesting tools, but it must be recognized that these instruments could show some limitation in terms of applicability because of different sociocultural backgrounds and temporal bias (the amount of time required to administer the scales).

Our group has developed a new direct performance measure for very mild to moderate patients with dementia: the Functional Living Skills Assessment (FLSA; [78]). Eight areas of interest are evaluated (resources, consumer skills, public transportation, time management, leisure, telephone skills, self-care, and health). Performance is scored according to completeness and level of assistance. We propose its use when a high sensibility to different levels of functional impairment is needed, as evaluation of treatment efficacy, diagnosis of dementia in the very initial phase, and identification of relatively intact functional areas to plan cognitive rehabilitation.

#### 10.14.3.4 Global and Quality-of-Life Measures

Global measures are one of the most important data used in clinical pharmacological trials. Global assessment scales assess whether treatment effects are clinically meaningful and whether a given agent leads to a change, and they appear to be one of the principal modality to detect changes in mild cognitive impairment patients. The global scales encompass two distinct categories: first, the clinician’s interview-based global severity scales, and, second, the clinician’s interview-based global change scales [79]. The Clinical Dementia Rating (CDR; [80]) and the related CDR-sum of boxes (CDR-SB), the Global Deterioration Scale (GDS; [81]), and the Functional Assessment Staging (FASTM; [82]) procedure are the global severity scales most often used in clinical trials. These measures are clearly useful in subject categorization in treatment trials (they are relatively free of many of the sociocultural biases inherent in psychometric descriptors and psychobehavioral measures).



They can also be used to demonstrate therapeutic efficacy in terms of the general progression of the dementia process and have also proven to be useful in sensitively assessing pharmacotherapeutic effects in AD treatment trials.

On the other hand, the most popular clinician's interview-based global change scales are the Clinician's Interview-Based Impression of Change (CIBIC; [83]) and the Clinical Global Impressions of Change (CGIC; [84]). With the CIBIC the clinician assesses four areas of patient functioning (general, cognitive, behavioral, and ADLs) by interviewing the patient. Another version of this scale is also available: the Clinician's Interview-Based Impression of Change Plus Caregiver Input (CIBIC-Plus; [85]), which is much more elaborate and includes information derived from caregiver input. In both versions, disease severity is assessed at baseline, and the rate of change from baseline is assessed at each follow-up examination. Score ranges from 1 to 7, where a score of 1 indicates marked improvement, a score of 4 indicates no change, and a score of 7 indicates marked worsening. The CIBIC-Plus procedures require an independent clinician assessment and can provide independent, comprehensive evidence of therapeutic efficacy. The CIBIC-Plus procedures may also be useful in sensitively assessing efficacy in prevention trials (e.g., MCI therapeutic trials, and perhaps in future trials with people referring subjective cognitive complaints).

The CGIC assesses cognitive, behavioral, social, and daily functioning domains, allows for caregiver input, and provides a minimal number of sample probes to guide clinicians in their assessments. Although CIBIC and CGIC are useful global measures, the lack of structure of global assessment reduces the interrater reliability, and it can be difficult to compare these measures across drug studies [84]. Assessment terms, such as "minimal improvement," are not well defined, and the rating of change by the clinician is a subjective measure that can be prone to biases (e.g., the clinician's perception of what constitutes change; sources of information about the patient; how information is recorded; the tests used to assess cognitive functioning; how the change is rated) [83, 86].

Dementia is one of the main causes of decreased quality of life (the degree of need satisfaction within the physical, psychological, social, activity, material, and structural area) [87] among older adults. Quality of life is recognized as an abstract and broad concept encompassing physical well-being, perceptions of well-being, satisfaction, and sense of self-worth. Actually, a substantial number of researchers agree that measuring quality of life is just as important as measuring cognition, disease severity, symptom response, behavioral disturbance, functional abilities, caregiver burden, and resource utilization [88]. Currently, the consensus group for measuring treatment benefits in dementia [89] have indicated that an improvement in quality of life could be an acceptable alternative to global improvement and have identified a number of other meaningful and measurable domains that may constitute appropriate targets for interventions (caregiver outcomes, including psychological consequences and time spent in delivering care, BPSD, quality of life, activities of daily living, global assessment, health utility, and health care costs). Furthermore, quality of life is mentioned as one of the primary outcomes of interest in dementia drug trials in a recent Cochrane update on cholinesterase inhibitors for AD [90].

However, quality-of-life scales are used as outcome measure in only 4.4% of all dementia/MCI-related randomized controlled trials [91]. Quality-of-life measures



cover a range of domains: physical status, functional ability including role functioning, social and community interactions, economic status, psychological status, well-being, somatic sensation, and life satisfaction [92, 93]. The most frequent quality-of-life measures applied in pharmacological trials in dementia or MCI are the Quality-of-Life–Alzheimer Disease Scale (QOL-AD), the patient-rated scale according to Blau (PRB), the Late-Stage Dementia scale (QUALID), and the Goal Attainment Scaling (GAS). Other currently used scales are the Dementia Care Mapping (DCM), the Alzheimer Disease Related Quality-of-Life scale (ADRQL), the Dementia Quality-of-Life scale (D-QoL), the psychosocial domain of the Functional Limitation Profile (FLP), and the European Quality-of-Life scale (Euro-QoL) (for the review of these tools, see [94]). Rating of quality of life in dementia trials, three methods are generally used: self-rating (patient-reported outcomes), proxy rating, and proxy observation scales. Patient-reported outcomes (including measures of subjective symptom report, health-related quality of life, and treatment satisfaction) are unique indicators of disease activity, complementary to established outcome measures [95]. Persons with MCI and mild to moderately severe dementia can be considered good informants of their own subjective states. At more severe stages proxy measures or direct observation may be preferred. The disadvantage of proxy ratings is that they represent a subjective measure, passed throughout the filter of the caregiver's own expectations, mood, burden of care, and the specific relationship with the person being rated [96–98]. In case a patient-reported outcome is impossible, observational evaluation by an uninvolved professional caregiver is the best alternative [96].

#### 10.14.3.5 Measures of Neuropsychiatric and Behavioral Changes

Clinicians are now aware that neuropsychiatric symptoms (also known as BPSD) are very common and diverse, and that scores on specific symptom clusters, such as measures of affect, apathy, and psychosis, may be more meaningful than an overall score [99]. BPSDs are distressing, present major difficulties for caregivers, and accelerate nursing home placement. These concepts need to be further developed to facilitate valid measurement of treatment effects on specific symptoms or domains. The measure of neuropsychiatric symptoms and behavioral changes most commonly used in dementia trials is the Neuropsychiatric Inventory (NPI; [100]), which assesses 10 common BPSD areas (e.g., delusions, irritability, disinhibition, agitation, sleep disorders) through an interview with the patient's caregiver. The frequency and severity of each behavior is measured, as well as the resulting caregiver distress. Each behavioral domain score ranges from 0 to 12, with a total NPI range of 0–120 (a score of less than 20 indicates mild disturbance; a score ranging from 20 to 50 indicates moderate disturbance; and a score superior to 50 indicates severe disturbance). The NPI takes approximately 10–15 minutes to administer; therefore, it is used both in research and clinical settings. A subscale of the Alzheimer's Disease Assessment Scale (ADAS-non-Cog; [40]) and the behavioral pathology in Alzheimer's disease rating scale (BEHAVE-AD; [101]) are other global neuropsychiatric measures used in AD trials. The ADAS-non-Cog assesses 10 domains (e.g., hallucinations, delusions, depression, power of concentration) and the score range is from 0 to 50 (maximum of deficits). The BEHAVE-AD is a good measure in clinical pharmacological trials to follow up the efficacy of behavioral

drug treatments, taking into account behavioral changes independent from cognitive deficit.

Depression is a relevant psychiatric symptom in dementia, and it may deserve a specific evaluation: the Cornell Geriatric Scale, for example, is a useful scale specifically conceived to detect depression in patients affected by dementia [102].

Because specific neuropsychiatric symptoms vary in prevalence between the different forms of dementia, in clinical trials recruiting non-AD dementia patients, other instruments conceived to evaluate particular neuropsychiatric and behavioral areas according to the dementia type can be selected between the different options offered by the dementia literature.

#### 10.14.3.6 Other Measures: Outcomes Specific to Caregivers

Measures of caregiver-related outcomes are various and are useful measures as endpoints for AD and related disorders studies. In clinical trials burden and time used to assist the patient represent the most extensively examined outcomes specific to caregivers, followed by other caregiver-specific variables (such as psychological well-being, health care costs, and satisfaction with treatment) and other caregiver-dependent outcomes (institutionalization, proxy survey of interview).

Burden outcome is generally reported using the Neuropsychiatric Inventory Distress rating (NPI-D; [103]) or more global measures of burden included in the Caregiver Burden Inventory (CBI; [104]), the Screen for Caregiver Burden (SCB; [105]), and the Caregivers Stress Scale (CSS; [106]). The NPI-D measures distress in response to 10 behaviors (e.g., delusions, agitation, apathy, disinhibition, sleep disorders) exhibited by the patient.

The CBI, SCB, and CSS are measures based on a conceptual model of stress of caregivers of patients with AD. The CBI is a 24-item, multidimensional 5-subscale (e.g., it measures time, dependence burden, and developmental burden) instrument to evaluate the impact of providing care to AD patients on family members' life. The SCB is a 25-item assessment of subjective and objective burdens that has been validated against a theoretical model of caregiver stress. The CSS is a 15-domain assessment that targets caregiver's primary stressors (e.g., cognitive and behavioral symptoms), secondary role strains (e.g., family conflict), secondary intrapsychic strains (e.g., role captivity, loss of self), and mediators (e.g., management of situation, expressive support). Other burden instruments include the Relatives' Stress Scale [107], which measures caregivers' personal distress, life upset, and negative feelings associated with the caregiving role, and the Cognitive Subscale of the Poulshock and Deimling tool [108], which assesses distress responses to care recipient cognitive symptoms.

Another outcome specific to caregivers is the "time use": caregivers' time spent assisting patients to perform basic or instrumental activities of daily living. Generally, the active time used by informal caregivers is examined, and the estimations of passive care time is reported (time providing supervision but not engaging in a task per se), and of paid care time. Measures of time use includes the Allocation of Caregiver Time Survey [109] and the Caregiver Activity Time Survey [110]. Both measures assess the amount of time spent assisting the patient in activities of daily living. Similarly, the Instrumental Activities of Daily Living Scale-Plus [57], the Physical Self-Maintenance Scale Plus [111], and the Resource Utilization in Demen-

tia questionnaire [112] generate caregiver estimates of the time spent providing such instrumental activities of daily living and self-care assistance.

Although caregiver outcomes generally represent secondary endpoints in AD and related disorders, drug development, with little time and effort information derived from such instruments, could be considerably improved. In order to implement and enrich the quality and interpretability of dementia clinical research with caregiver-specific outcomes, some recommendations are useful to remember. According to Lingler et al. [113], we suggest that investigators should specify entry criteria for caregiver study participants and should collect basic caregiver sociodemographic data (caregiver sociodemographic characteristics—sex, relationship to patient, and age, years caring for the patient, primary or secondary caregiver status, caregiver marital status if a child, co-residence with the patient and other relatives, caregiver ethnicity, income, physical, and mental health status, level of social support). This will aid in the interpretation of data about drug effects on caregiver measures and data about proxy-dependent patient effects. It will also facilitate the identification of variables influencing such effects, and it will allow researchers to investigate not only whether the benefits of antidementia drugs extend to caregivers, but also to what type of caregiver (e.g., spouses or children) and under what circumstances.

#### 10.14.3.7 New and Future Measures

In the past two decades much progress has been made in refining our understanding of the neurobiology and clinical phenomenology of dementia, particularly in Alzheimer's dementia (see new research criteria for Alzheimer's disease [5]. Actually, several "candidate" biomarkers for diagnosis, severity, progression, or prediction of response are recognized (e.g., structural brain changes visible on MRI with early and extensive involvement of the medial temporal lobe in AD, molecular neuroimaging changes seen with PET with hypometabolism or hypoperfusion in specific areas, and changes in cerebrospinal fluid biomarkers). Volumetric neuroimaging, which seems particularly promising as a measure of progression, and markers of oxidative stress may also become useful. Biomarkers may be informative indicators of the mechanism of action of novel treatments or may help identify potential responders (possible strategies to facilitate work in this area in the future include stratification of study populations to identify predictors of response such as known or putative biomarkers, e.g., Apolipoprotein E—APOE—isoform ApoE- $\epsilon$ 4), but it would be considered as supplementary data to collect and not a proxy for clinically significant measures of treatment response.

***Patients with Cognitive and Behavioral Deficits after TBI and Stroke*** Currently, poor level of scientific evidence in treating the cognitive and behavioral deficits due to TBI is available. Future studies are necessary, and, in order to gather scientific evidence, selecting appropriate and reliable outcome measures is a critical issue. An important obstacle in scientific studies is the problem of how to measure cognition and behavior. It is well known that victims of TBI may show fairly good test performance, yet their everyday life is severely restricted or vice versa. In addition, pronounced fluctuation in daily functioning is a common sequel of TBI, reflecting the low cognitive reserve of the injured brain. The Functional Independence

Measurement (FIM) [114], a widely used index of rehabilitation outcome, measures the level of assistance that an individual requires to perform basic life activities. It is an 18-item, 7-level scale that rates the ability of a person to perform independently in self-care, sphincter control, transfers, locomotion, communication, and social activity. The total score is obtained by summing the partial scores, and it ranges from 18 (maximally dependent) to 126 (maximally independent). Two motor and cognitive subscales can be obtained by summing the 13 motor items (range 13–91) and the 5 cognitive items (range 5–35). The Rancho Los Amigos (RLA) Levels of Cognitive Functional Scale is another valid tool to measure global outcome in cognitive sequelae of TBI [115]. This scale was developed to use in the planning of treatment, tracking of recovery, and classifying outcome levels in TBI. There are 8 classification levels, ranging from no response (level I) to confused and agitated (level IV) to purposeful and appropriate (level VIII). It is an easy and simple to perform tool for the global cognitive evaluation of the TBI patients and an appropriate and valid scale as it can both provide follow-up and comparison of the patients.

Several cognitive and behavioral clinical situations can occur following stroke, depending on the degree and type of brain injury. We will consider in the text only poststroke aphasia as an example of cognitive sequela. While speech-language therapy actually remains the mainstay treatment of aphasia, there have been some attempts to introduce the use of pharmacological agents (piracetam, drugs acting on catecholamine systems—bromocriptine, dexamfetamine—and on the cholinergic system—donepezil—as therapeutic strategies in aphasia and other stroke sequelae (e.g., spatial neglect).

Currently, there is unanimous consensus that evaluation of aphasic deficits should be comprehensive to allow the planning of rational therapies [116]. However, there is less agreement regarding the language assessment methodology with debated problems about the different levels of analysis that should be implemented to formally assess language impairment. Some researchers [116, 117] recommend evaluation of groups of aphasic patients using standardized batteries. These batteries examine oral language, reading, and writing and combine the information gathered from different language subtests that assess spontaneous speech, comprehension, repetition, and naming to obtain a graphic profile of performance (e.g., Boston Diagnostic Aphasia Examination; [118]), overall scores of aphasia severity (e.g., Western Aphasia Battery, WAB; [119]), or measures of communicative ability (e.g., Porch Index of Communicative Abilities, PICA; [120]). The estimation of the initial aphasia severity and its clinical profile using standardized assessment batteries usually couple with mapping of lesion size. The overall score of these batteries are increasingly being used as primary outcome measures to estimate changes in the global level of performance after pharmacological treatments [121, 122].

The alternative position to evaluate aphasia is the neuropsychological approach. This approach examines the nature of both normal and abnormal language functioning in terms of current information-processing models. One of the assessment tools based on neuropsychological point of view is the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA; [123]). The PALPA is useful for identifying the different clusters of linguistic impairment as well as the residual areas of strength in the aphasic patient. However, administration of the PALPA or

similar batteries is not always feasible because they are time consuming and not applicable to patients with severe aphasia. In recent years, there has been a growing convergence on the usefulness of both these two assessment approaches: The surface symptoms of aphasia and its global severity are better recognized using standardized aphasia batteries such as the WAB or PICA, whereas evaluations using the cognitive neuropsychological approach and tools such as the PALPA more appropriately assess the nature of language deficits [124].

Like in dementia, given that aphasia and cognitive-behavioral sequelae of TBI negatively affect quality of life, there has been increasing interest in developing reliable and valid assessment instruments in this field (e.g., Stroke and Aphasia Quality of Life Scale-39; [125]). Future studies also need to include psychiatric comorbidity as a predictive factor of outcome in these diseases because depression, anxiety, and social withdrawal can have a negative impact on rehabilitation, potential benefits of drugs treatment, and psychosocial functioning.

#### **10.14.3.8 Safety Evaluation**

Safety evaluation does not depart from the normal pharmacological trial procedures: Assessment of safety included reports of all possible adverse events, laboratory examinations (chemistry, hematology, and urinalysis), and vital sign measurements (electrocardiogram). Especially in the population affected by dementia, given the old age, it is important to monitor laboratory findings (renal and hepatic function, etc.) and possible cardiac adverse events by repeating electrocardiogram.

### **10.14.4 DURATION OF STUDY AND POSSIBLE BIAS IN TRIALS ON COGNITIVE IMPAIRMENT**

#### **10.14.4.1 Duration**

When performing pharmacological trials for cognitive disturbances, it is necessary to study the patient for a long time in order to detect changes. Recently, Cortes et al. [125], trying to determine the best duration of follow-up necessary to demonstrate the impact of new drugs in dementia, showed that changes undergone by AD patients under cholinesterase inhibitors after 6 months are not adequate to demonstrate the effect of a new treatment. The authors propose that an 18-month trial appear to have the potential to demonstrate clearly the effect of a new drug.

Even if previous randomized controlled trials were of limited duration (often 24 weeks) in order to contain costs and to not loose too many patients, and changes were statistically significant in some cases at 6 months, it is natural to think such a time span is far shorter than the duration of the disease. Therefore, the duration of a randomized controlled study should be at least of 9 or 12 months, with planned evaluations, in our opinion, every 3 months. In addition, it is possible to consider from the beginning an open extension of the randomized trial to obtain further data on patient evolution. Sure, a careful planning of the power of the study (i.e., of the number of patients necessary to demonstrate a treatment effect, calculated by taking into account previous data about the amplitude of the effect of existing drugs for cognitive deficits) is necessary, also taking into consideration the possibility to loose patients at the last follow-up evaluations.

#### 10.14.4.2 Possible Bias

A bias is a systematic error that alters the results of a study. We will have a bias when the population of treated patients is different from the control population, according to a characteristic that can change results in primary or secondary efficacy measures. To correctly evaluate drug efficacy, the two populations (treated patients and controls) must be identical in all their characteristics, with the exception of the experimental treatment. Possible confounders in studies on cognitive impairment may be, for example, age, education, severity of the disease, initial severity of trauma for TBI, lesion location for cognitive sequelae of stroke, comorbidity, or treatment with other drugs. The best way to avoid systematic bias is to perform an accurate randomization of subjects included in the study; randomization must follow the general rules, particularly it must be performed by a center that is external to centers involved in the study. Good randomization avoids selection bias. Another important system to avoid bias is to maintain the double-blind condition; this is not always possible, however, particularly when the drug has frequent and easily detected side effects (this was the case for interferon in multiple sclerosis trials). Sometimes it can be appropriate to perform stratification of the studied population according to main possible confounders before randomization: This is important also to identify subpopulations of patients who better respond to therapy. At the statistical level, it is indicated to perform analysis by covariating for different factors (e.g., treatment with other drugs to avoid execution bias, comorbidity, age, sex, education, center, and severity).

As in all drug studies, the intention to treat analysis allows maintaining advantages assured by randomization, and it avoids the introduction of systematic errors due to reduced compliance (which can be particularly problematic in cognitively impaired patients), that is, exclusion bias.

To avoid bias, it is also necessary to diversify the sources to recruit patients and controls. Actually, all pharmacological trials are multicentric; generally, centers of different countries are involved. This also allows to recruit a sufficient number of subjects for the study: Calculation of the study power is very important to avoid underscoring drug efficacy due to recruitment of an insufficient patient number. On the other hand, when considering standard statistical significance ( $p = 0.05$ ), we will find a casual association (and not an association due to treatment) once every 20 times. Therefore, it is necessary to limit the number of comparisons, which is the number of primary and secondary endpoints.

Other bias can derive from some deficiency of instruments used to evaluate efficacy. For example, measures of caregiver burden do not take into account some important variables, such as caregiver education, social status, relationship with the patient. (Is he/she a family member or a professional caregiver? Does he/she live with the patient?) Functional scales based on informant reports may be influenced by the same variables. Some cognitive instruments may not have parallel forms to avoid learning bias. Another bias can be represented by the interpretation bias: This arises when different stakeholders assign their individual values to the interpretation of the final results of randomized controlled trials. This can be a frequent bias in the particular domain of trials on diseases causing cognitive impairment, where really effective and satisfying treatments are lacking.



### 10.14.5 CONCLUSION

The field of drug trials in medical situations leading to cognitive impairment is a fascinating one. The population of patients who can potentially benefit from therapy is enormous. Actually, treatments available for AD are of limited efficacy, and they do not succeed in stopping cognitive deterioration and loss of function, or to significantly reduce caregiver and social burden. In other forms of dementia (with the exception of dementia with Lewy bodies, for which limited data indicate some efficacy of cholinesterase inhibitors) there is no treatment. Nearly the same can be said for long-term cognitive sequelae of TBI and stroke.

We need the pharmaceutical industry to invest its energy and financial resources in trials aimed at improving this sad situation. However, in order not to waste resources, it is necessary to plan very careful drug trials.

### REFERENCES

1. Olsen, J., Baker, M., Freund, T., et al. (2006), Consensus document on European brain research, *J. Neurol. Neurosurg. Psychiatry*, 77(Suppl I), i1–I49.
2. Hebert, L. E., Scherr, P. A., Bienias, J. L., et al. (2003), Alzheimer disease in the U.S. population: Prevalence estimates using the 2000 Census, *Arch. Neurol.*, 60, 1119–1122.
3. Fratiglioni, L., Launer, L. J., Andersen, K., et al. (2000), Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group, *Neurology*, 54, S10–S15.
4. Petersen, R. C., Smith, G. E., Waring, S. C., et al. (1999), Mild cognitive impairment: Clinical characterization and outcome, *Arch. Neurol.*, 56(3), 303–308.
5. Dubois, B., Feldman, H. H., Jacova, C., et al. (2007), Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria, *Lancet Neurol.*, 6(8), 734–746.
6. Consensus Conference (1999), Rehabilitation of persons with traumatic brain injury. NIH Consensus Development Panel on Rehabilitation of Persons with Traumatic Brain Injury, *JAMA*, 282(10), 974–983.
7. Tennant, A. (2005), Admission to hospital following head injury in England: Incidence and socio-economic associations, *BMC Public Health*, 5, 21.
8. Wenden, F. J., Crawford, S., and Wade, D. T. (1998), Assault, post-traumatic amnesia and other variables related to outcome following head injury, *Clin. Rehabil.*, 12, 53–63.
9. Bullock, R., Chesnut, R. M., Clifton, G., et al. (2002), Guidelines for the management of severe head injury. BrainTrauma Foundation, American Association of Neurological Surgeons, Joint Section on Neurotrauma and Critical Care, *J. Neurotrauma*, 17, 449–627.
10. Adelman, S. M. (1981), The national survey of stroke: Economic impact, *Stroke*, 12, 169–187.
11. Taylor, T. N., Davis, P. H., Torner, J. C., et al. (1996), Lifetime cost of stroke in the United States, *Stroke*, 27, 1459–1466.
12. Bergen, D. C., and Silberberg, D. (2002), Nervous system disorders: A global epidemic, *Arch. Neurol.*, 59, 1194–1196.
13. MacKay, J., and Mensah, G. A., eds. (2004), *The Atlas of Heart Disease and Stroke*, World Health Organisation, Geneva.



14. Palmer, A. J., Valentine, W. J., Roze, S., et al. (2005), Overview of costs of stroke from published, incidence-based studies spanning 16 industrialized countries, *Curr. Med. Res. Opin.*, 21, 19–26.
15. Srikanth, V. K., Quinn, S. J., Donnan, G. A., et al. (2006), Long-term cognitive transitions, rates of cognitive change, and predictors of incident dementia in a population-based first-ever stroke cohort, *Stroke*, 37, 2479–2483.
16. Serrano, S., Domingo, J., Rodriguez-Garcia, E., et al. (2007), Frequency of cognitive impairment without dementia in patients with stroke: A two-year follow-up study, *Stroke*, 38, 105–110.
17. American Psychiatric Association (APA) (2000), *Diagnostic and Statistical Manual of Mental Disorders*, DSM-IV, 4th ed., APA, Washington, DC.
18. McKhann, G., Drachman, D., Folstein, M., et al. (1984), Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, *Neurology*, 34, 939–944.
19. McKeith, I. G., Galasko, D., Kosaka, K., et al. (1996), Consensus guidelines for the clinical and pathological diagnosis of dementia with Lewy bodies (DLB): Report of the consortium on DLB international workshop, *Neurology*, 47, 1113–1124.
20. McKeith, I. G., Dickson, D. W., Lowe, J., et al. (2005), Diagnosis and management of dementia with Lewy bodies: Third report of the DLB Consortium, *Neurology*, 65(12), 1863–1872.
21. Neary, D., Snowden, J. S., Gustafson, L., et al. (1998), Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria, *Neurology*, 51, 1546–1554.
22. McKhann, G. M., Albert, M. S., Grossman, M., et al. (2001), Clinical and pathological diagnosis of frontotemporal dementia: Report of the Work Group on Frontotemporal Dementia and Pick's Disease, *Arch. Neurol.*, 58, 1803–1809.
23. Roman, G. C., Tatemichi, T. K., Erkinjuntti, T., et al. (1993), Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop, *Neurology*, 43, 250–260.
24. Pedersen, P. M., Vinter, K., and Olsen, T. S. (2004), Aphasia after stroke: Type, severity and prognosis. The Copenhagen aphasia study, *Cerebrovasc. Dis.* 17(1): 35–43.
25. Godefroy, O., Dubois, C., Debachy, B., et al. (2002), Vascular aphasia: Main characteristics of patients hospitalized in acute stroke units, *Stroke*, 33(3), 702–705.
26. Basso, A., Lecours, A. R., Moraschini, S., et al. (1985), Anatomoclinical correlations of the aphasias as defined through computerized tomography: Exceptions, *Brain Lang.*, 26, 201–229.
27. Ferro, J. M., and Madureira, S. (1997), Aphasia type, age and cerebral infarct localisation, *J. Neurol.*, 244, 505–509.
28. Binder, J. (1997), Functional magnetic resonance imaging. Language mapping, *Neurosurg. Clin. N. Am.*, 8(3), 383–392.
29. Drobyshevsky, A., Baumann, S. B., and Schneide, W. (2006), A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function, *Neuroimage*, 31(2), 732–744.
30. Gureje, O., Ogunniyi, A., Kola, L., et al. (2006), Functional disability in elderly Nigerians: Results from the Ibadan Study of Aging, *J. Am. Geriatr. Soc.*, 54(11), 1784–1789.
31. Chiu, H. F., and Zhang, M. (2000), Dementia research in China, *Int. J. Geriatr. Psychiatry*, 15(10), 947–953.

32. Ballard, C., Holmes, C., McKeith, I., et al. (1999), Psychiatric morbidity in dementia with Lewy bodies: A prospective clinical and neuropathological comparative study with Alzheimer's disease, *Am. J. Psychiatry*, 156, 1039–1045.
33. Ryu, S. H., Katona, C., Rive, B., et al. (2005), Persistence of and changes in neuropsychiatric symptoms in Alzheimer disease over 6 months: The LASER-AD study, *Am. J. Geriatr. Psychiatry*, 13, 976–983.
34. McShane, R., Keene, J., Gedling, K., et al. (1997), Do neuroleptic drugs hasten cognitive decline in dementia? Prospective study with necropsy follow up, *BMJ*, 314, 266–270.
35. Ballard, C., et al. (2005), Quetiapine and rivastigmine and cognitive decline in Alzheimer's disease: Randomised double blind placebo controlled trial, *BMJ*, 330(7496), 874.
36. Strangman, G., O'Neil-Pirozzi, T. M., Burke, D., et al. (2005), Functional neuroimaging and cognitive rehabilitation for people with traumatic brain injury, *Am. J. Phys. Med. Rehabil.*, 84, 62–75.
37. Schretlen, D. J., and Shapiro, A. M. (2003), A quantitative review of the effects of traumatic brain injury on cognitive functioning, *Int. Rev. Psychiatry*, 15, 341–349.
38. Lendrem, W., and Lincoln, N. B. (1985), Spontaneous recovery of language in patients with aphasia between 4 and 34 weeks after stroke, *J. Neurol. Neurosurg. Psychiatry*, 48, 743–748.
39. Jarvik, L. F., Berg, L., Bartus, R., et al. (1990), Clinical drug trials in Alzheimer disease. What are some of the issues? *Alzheimer Dis. Assoc. Disord.*, 4, 193–202.
40. Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984), A new rating scale for Alzheimer's disease, *Am. J. Psychiatry*, 141, 1356–1364.
41. Matthews, H. P., Korbey, J., Wilkinson, D. G., et al. (2000), Donepezil in Alzheimer's disease: Eighteen month results from Southampton Memory Clinic, *Int. J. Geriatr. Psychiatry*, 5, 713–720.
42. Le Bars, P. L., Kieser, M., and Itil, K. Z. (2000), A 26-week analysis of a doubleblind, placebo-controlled trial of the ginkgo biloba extract EGb 761 in dementia, *Dement. Geriatr. Cogn. Disord.*, 11, 230–237.
43. Farlow, M., Potkin, S., Koumaras, B., et al. (2003), Analysis of outcome in retrieved dropout patients in a rivastigmine vs placebo, 26-week, Alzheimer disease trial, *Arch Neurol.*, 60, 843–848.
44. Aisen, P. S., Schafer, K. A., Grundman, M., et al. (2003), Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: A randomized controlled trial, *JAMA*, 289, 2819–2826.
45. Dubois, B., Slachevsky, A., Litvan, I., et al. (2000), The FAB: A Frontal Assessment Battery at bedside, *Neurology*, 55(11), 1621–1626.
46. Qualls, C. E., Bliwise, N. G., and Stringer, A. Y. (2000), Short forms of the Benton Judgment of Line Orientation Test: Development and psychometric properties, *Arch. Clin. Neuropsychol.*, 15(2), 159–163.
47. Ferman, T. J., Smith, G. E., Boeve, B. F., et al. (2004), DLB fluctuations: Specific features that reliably differentiate DLB from AD and normal aging, *Neurology*, 62, 181–187.
48. Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975), 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician, *J. Psychiatr. Res.*, 12, 189–198.
49. Caban-Holt, A., Bottiggi, K., and Schmitt, F. A. (2005), Measuring treatment response in Alzheimer's disease clinical trials, *Geriatrics*, Jun (Suppl):3–8.

50. Saxton, J., McGonigle Gibson, K. L., et al. (1990), Assessment of the severely impaired patient: Description and validation of a new neuropsychological test battery, *Psychol. Assess.*, 2, 298–303.
51. Panisset, M., Roudier, M., Saxton, J., et al. (1994), Severe impairment battery: A neuropsychological test for severely demented patients, *Arch. Neurol.*, 51, 41–45.
52. Schmitt, F. A., Ashford, W., Ernesto, C., et al. (1997), The severe impairment battery: Concurrent validity and the assessment of longitudinal change in Alzheimer's disease, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 2), S51–S56.
53. Reisberg, B., Doody, R., Stöffler, A., et al. (2003), Memantine Study Group. Memantine in moderate-to-severe Alzheimer's disease, *N. Engl. J. Med.*, 348(14), 1333–1341.
54. Schmitt, F. A., Cragar, D., Ashford, J. W., et al. (2002), Measuring cognition in advanced Alzheimer's disease for clinical trials, *J. Neural. Transm. Suppl.*, 62, 135–148.
55. Saxton, J., Kastango, K. B., Hugonot-Diener, L., et al. (2005), Development of a short form of the Severe Impairment Battery, *Am. J. Geriatr. Psychiatry*, 13(11), 999–1005.
56. Wolinsky, F. D., Callahan, C. M., Fitzgerald, J. F., et al. (1992), The risk of nursing home placement and subsequent death among older adults, *J. Gerontol. Social Sci.*, 47, S173–S182.
57. Katz, S., Ford, A. B., Moskowitz, R. W., et al. (1963), The index of ADL: A standardized measure of biological and psychosocial function, *JAMA*, 185, 914–919.
58. Lawton, M. P., and Brody, E. M. (1969), Assessment of older people: Self-maintaining and instrumental activities of daily living, *Gerontologist*, 9, 179–186.
59. Stern, Y., Hesdorffer, D., Sano, M., et al. (1990), Measurement and prediction of functional capacity in Alzheimer's disease, *Neurology*, 40, 8–14.
60. Spector, W. D. (1997), Measuring functioning in daily activities for persons with dementia, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 6), 81–90.
61. Blessed, G., Tomlinson, B. E., and Roth, M. (1968), The association between quantitative measures of dementia and of senile change in the cerebral gray matter of elderly subjects, *Br. J. Psychiatry*, 114, 797–811.
62. Spiegel, R., Brunner, C., Ermini-Fünfschilling, D., et al. (1991), A new behavioral assessment scale for geriatric out and in-patients the NOSGER (Nurses' Observation Scale for Geriatric Patients), *J. Am. Geriatr. Soc.*, 39, 339–347.
63. Reisberg, B., Finkel, S., Overall, J., et al. (2001), The Alzheimer's Disease Activities of Daily Living International Scale (ADL-IS), *Int. Psychogeriatr.*, 13, 163–181.
64. Rubenstein, L. Z., Schairer, C., Wieland, G. D., et al. (1984), Systematic biases in functional status assessment of elderly adults: Effects of different data sources, *J. Gerontol.*, 39, 686–691.
65. Kuriansky, J., and Gurland, B. (1976), The performance test of activities of daily living, *Int. J. Aging Hum. Dev.*, 7, 343–352.
66. Jette, A. M., and Branch, L. G. (1985), Impairment and disability in the aged, *J. Chronic Dis.*, 38, 59–65.
67. Williams, M. E. (1987), Identifying the older person likely to require long-term care services, *J. Am. Geriatr. Soc.*, 35, 761–766.
68. Guralnik, J. M., Branch, L. G., Cummings, S. R., et al. (1989), Physical performance measures in aging research, *J. Gerontol.*, 44, 141–146.
69. Reuben, D. B., Siu, A. L., and Kimpau, S. (1992), The predictive validity of self-report and performance-based measures of function and health, *J. Gerontol.*, 47, M106–110.
70. Zimmerman, S. I., and Magaziner, J. (1994), Methodological issues in measuring functional status of cognitively impaired nursing home residents: The use of proxies

- and performance based measures, *Alzheimer Dis. Assoc. Disord.*, 8(Suppl 1), S281–S290.
71. Lowenstein, D. A., Amigo, E., Duara, R., et al. (1989), A new scale for the assessment of functional status in Alzheimer's disease and related disorders, *J. Gerontol.*, 44, 114–121.
  72. Reuben, D. B., and Siu, A. L. (1990), An objective measure of physical function of elderly outpatients. The Physical Performance Test, *J. Am. Geriatr. Soc.*, 38, 1105–1112.
  73. Mahurin, R. K., De Bettignies, B. H., and Pirozzolo, F. J. (1991), Structured assessment of independent living skills: preliminary report of a performance measure of functional abilities in dementia, *J. Gerontol.*, 46, 58–66.
  74. Skurla, E., Rogers, J. C., and Sunderland, T. (1988), Direct assessment of activities of daily living in Alzheimer's disease, *J. Am. Geriatr. Soc.*, 36, 97–103.
  75. Karagiozis, H., Gray, S., Sacco, J., et al. (1998), The Direct Assessment of Functional Abilities (DAFA): A comparison to an indirect measure of instrumental activities of daily living, *Gerontologist*, 38, 113–121.
  76. Willis, S. L., Allen-Burge, R., Dolan, M. M., et al. (1998), Everyday problem solving among individuals with Alzheimer's disease, *Gerontologist*, 38, 569–577.
  77. Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., et al. (1982), Measurement of functional activities in older adults in the community, *J. Gerontol.* 37, 323–329.
  78. Farina, E., Fioravanti, R., Chiavari, L., et al. (1999), Functional Living Skills Assessment: A standardized instrument built to monitor activities of daily living in patients with dementia: Preliminary data, *Journal of Neurology*, 246(suppl. 1), 1/101.
  79. Reisberg, B. (2007), Global measures: utility in defining and measuring treatment response in dementia, *Int. Psychogeriatr.*, 19, 421–456.
  80. Hughes, C. P., and Berg, L. (1982), A new clinical scale for the staging of dementia, *Br. J. Psychiatry*, 140, 566–572.
  81. Reisberg, B., Ferris, S. H., deLeon, M. J., et al. (1982), The Global Deterioration Scale for assessment of primary degenerative dementia, *Am. J. Psychiatry*, 139, 1136–1139.
  82. Sclan, S. G., and Reisberg, B. (1992), Functional assessments staging (FAST) in Alzheimer's disease: Reliability, validity, and ordinality, *Int. Psychogeriatr.*, 4(Suppl. 1), 55–69.
  83. Knopman, D. S., Knapp, M. J., Gracon, S. I., et al. (1994), The Clinician Interview-Based Impression (CIBI): A clinician's global change rating scale in Alzheimer's disease, *Neurology*, 44, 2315–2321.
  84. Schneider, L. S., Olin, J. T., Doody, R. S., et al. (1997), Validity and reliability of the Alzheimer's Disease Cooperative Study—Clinical Global Impression of Change. The Alzheimer's Disease Cooperative Study, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 2), S22–S32.
  85. Reisberg, B., Schneider, L., Doody, R., et al. (1997), Clinical global measures of dementia. Position paper of the International Working Group on Harmonization of Dementia Drug Guidelines, *Alzheimer Dis. Assoc. Disord.*, 11(Suppl 3), 8–18.
  86. Quinn, J., Moore, M., Benson, D. F., et al. (2002), A videotaped CIBIC for dementia patients: Validity and reliability in a simulated clinical trial, *Neurology*, 58, 433–437.
  87. Joyce, C. R. B., O'Boyle, C. A., and McGee, H. (1999), *Individual Quality of Life. Approaches to Conceptualisation and Assessment*, Harwood Academic, Singapore.
  88. Rockwood, K., and Gauthier, S., eds. (2006), *Trial Designs and Outcomes in Dementia Therapeutic Research*, Taylor & Francis, London and New York.

89. Katona, C., Livingston, G., Cooper, C., et al. (2007), International Psychogeriatric Association consensus statement on defining and measuring treatment benefits in dementia, *Int. Psychogeriatr.*, 19(3), 345–354.
90. Birks, J. (2006), Cholinesterase inhibitors for Alzheimer's disease, *Cochrane Database Syst. Rev.*, CD005593.
91. Scholzel-Dorenbos, C., Van der Steen, M., Engels, L., et al. (2007), Assessment of quality of life as outcome in dementia and MCI intervention trials, *Alzheimer Dis. Assoc. Disord.*, 21, 172–178.
92. Bowling, A. (1993), in Keynes, M., Ed., *Measuring Health: A Review of Quality of Life Measurement Scales*, Open University Press, Philadelphia.
93. Spilker, B. (1990), *Quality of Life Assessment in Clinical Trials*, Raven Press, New York.
94. Naglie, G. (2007), Quality of life in dementia, *Can. J. Neurol. Sci.*, 34(Suppl 1), S57–61.
95. Wiklund, I. (2004), Assessment of patient-reported outcomes in clinical trials: The example of health-related quality of life, *Fundam. Clin. Pharmacol.*, 18, 351–363.
96. Brod, M., Stewart, A. L., Sands, L., et al. (1999), Conceptualization and measurement of quality of life in dementia: The dementia quality of life instrument, *Gerontologist*, 39, 25–35.
97. Logsdon, R. G., Gibbons, L. E., McCurry, S. M., et al. (2002), Assessing quality of life in older adults with cognitive impairment, *Psychosom. Med.*, 64, 510–519.
98. Scocco, P., Fantoni, G., and Caon, F. (2006), Role of depressive and cognitive status in self-reported evaluation of quality of life in older people: Comparing proxy and physician perspectives, *Age Ageing*, 35, 166–171.
99. Lyketsos, C. G., Breitner, S., and Rabins, P. V. (2001), An evidence-based proposal for the classification of neuropsychiatric disturbance in Alzheimer's disease, *Int. J. Geriatr. Psychiatry*, 16, 1037–1042.
100. Cummings, J. L., Mega, M., Gray, K., et al. (1994), The Neuropsychiatric Inventory: Comprehensive assesment of psychopathology in dementia, *Neurology*, 44, 2308–2314.
101. Reisberg, B., et al. (1988), BEHAVE-AD. A clinical rating scale for the assessment of pharmacologically remediable behavioral symptomatology in Alzheimer's disease, in Altman, H., (Ed.), *Alzheimer's Disease and Dementia: Problems, Prospects, and Perspectives*, Plenum, New York.
102. Conn, T. (2007), Assessment of behavioral and psychological symptoms associated with dementia, *Can. J. Neurol. Sci.*, 34(Suppl 1), S67–71. Review.
103. Kaufer, D. I., Cummings, J. L., Christine, D., et al. (1998), Assessing the impact of neuropsychiatric symptoms in Alzheimer's disease: The Neuropsychiatric Inventory Caregiver Distress Scale, *J. Am. Geriatr. Soc.*, 46, 210–215.
104. Novak, M., and Guest, C. (1980), Application of a multidimensional caregiver burden inventory, *Gerontologist*, 29, 798–780.
105. Vitaliano, P. P., Russo, J., and Young, H. M. (1991), The screen for caregiver burden, *Gerontologist*, 31, 76–83.
106. Pearlin, L. I., Mullan, J. T., Semple, S. J., et al. (1990), Caregiving and the stress process: An overview of concepts and their measures, *Gerontologist*, 30, 583–594.
107. Greene, J. G., Smith, R., and Timbury, G. C. (1982), Measuring behavioral disturbances of elderly demented patients in the community and its effects on relatives: A factor analytic study, *Age Aging*, 11, 121–126.
108. Poulshock, S. W., and Deimling, G. T. (1984), Families caring for elders in residence. Issues in the measurement of burden, *J. Gerontol.* 39, 230–239.

109. Blesa, R. (2000), Galantamine: Therapeutic effects beyond cognition, *Dement. Geriatr. Cogn. Disord.*, 11(Suppl 1), 28–34.
110. Clipp, E. C., and Moore, M. J. (1995), Caregiver time use. An outcome measure in clinical trial research on Alzheimer's disease, *Clin. Pharmacol. Ther.*, 58, 228–236.
111. Mohs, R. C., Doody, R. S., Morris, J. L., et al. (2001), A 1-year placebo-controlled preservation of function survival study of donepezil in AD patients, *Neurology*, 57, 481–488.
112. Wimo, A., Wetterholm, A. L., Mastey, V., et al. (1988), Evaluation of the resource utilization and caregiver time in anti-dementia drug trials: A quantitative battery, in Wimo, A., Jonsson, B., Karlsson, G., Eds., *The Health Economics of Dementia*, Wiley, London. pp. 465–499.
113. Lingler, J. H., Martire, L. M., and Schulz, R. (2005), Caregiver-specific outcomes in antidementia clinical drug trials: A systematic review and meta-analysis, *J. Am. Geriatr. Soc.*, 53(6), 983–990.
114. Wright, J. (2000), The Functional Assessment Measure. The Center for Outcome Measurement in Brain Injury.
115. Hagen, C., Malkmus, D., and Durham, P. (1972), Rancho Los Amigos cognitive scale. Rancho Los Amigos Hospital.
116. Blumstein, S. E. (1997), A perspective on the neurobiology of language, *Brain Lang.*, 60, 335–346.
117. Grodzinsky, Y., Pinango, M. M., Zurif, E., et al. (1999), The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia, *Brain Lang.*, 67, 134–147.
118. Goodglass, H., and Kaplan, E. (1972), *Assessment of Aphasia and Related Disorders*, Lea and Febiger, Philadelphia.
119. Kertesz, A. (1982), *The Western Aphasia Battery*, Grune and Stratton, New York.
120. Porch, B. (1982), *The Porch Index of Communicative Abilities*, Consulting Psychologists Press, Palo Alto, CA.
121. Walker-Batson, D., Curtis, S., Natarajan, R., et al. (2001), A double-blind, placebo-controlled study of the use of amphetamine in the treatment of aphasia, *Stroke*, 32, 2093–2098.
122. Berthier, M. L., Hinojosa, J., Martin, M. C., et al. (2003), Open-label study of donepezil in chronic poststroke aphasia, *Neurology*, 1218–1219.
123. Kay, J., Lesser, R., and Coltheart, M. (1992), *Psycholinguistic Assessments of Language Processing in Aphasia (PALPA)*, Lawrence Erlbaum, Hove.
124. Nickels, L. (2002), Theoretical and methodological issues in the cognitive neuropsychology of spoken word production, *Aphasiology*, 16, 3–19.
125. Hilari, K., Byng, S., Lamping, D. L., et al. (2003), Stroke and aphasia quality of life scale-39 (SAQL-39): evaluation of acceptability, reliability and validity, *Stroke* 34, 1944–1950.
126. Cortes, F., Portet, F., Touchon, J., et al. (2007), Six and 18-month changes in mild to moderate Alzheimer's patients treated with acetylcholinesterase inhibitors: What can we learn for clinical outcomes of therapeutic trials? *J. Nutr. Health Aging*, 11(4), 330–337.





# *10.15*

---

## *Bridging Studies in Pharmaceutical Safety Assessment*

**Jon Ruckle**

*Covance Clinical Research Unit, Honolulu, Hawaii*

### **Contents**

|           |                                                                 |     |
|-----------|-----------------------------------------------------------------|-----|
| 10.15.1   | Introduction                                                    | 734 |
| 10.15.1.1 | Are Ethnic Factors Real?                                        | 734 |
| 10.15.1.2 | What Are “Bridging Studies”?                                    | 735 |
| 10.15.2   | Why Do Bridging Studies?                                        | 736 |
| 10.15.2.1 | Traditional Practice of Medicine in Japan and the United States | 736 |
| 10.15.2.2 | Pharmaceutical Development in Japan                             | 737 |
| 10.15.2.3 | Safety versus Efficacy                                          | 738 |
| 10.15.2.4 | Market Factors                                                  | 738 |
| 10.15.2.5 | Social and Political Factors                                    | 739 |
| 10.15.3   | Regulatory Guidance: ICH and E5                                 | 740 |
| 10.15.3.1 | ICH History                                                     | 740 |
| 10.15.3.2 | E5 Guidelines                                                   | 741 |
| 10.15.3.3 | Implications for Non-ICH Jurisdictions                          | 744 |
| 10.15.4   | Ethnic Pharmaceutical Safety Issues: Pharmacokinetics           | 745 |
| 10.15.4.1 | ADME: Absorption                                                | 745 |
| 10.15.4.2 | ADME: Distribution                                              | 746 |
| 10.15.4.3 | ADME: Metabolism                                                | 746 |
| 10.15.4.4 | ADME: Elimination                                               | 749 |
| 10.15.4.5 | ADME: Dietary Factors                                           | 749 |
| 10.15.4.6 | ADME: Multiple Mechanisms                                       | 751 |

|           |                                                                    |     |
|-----------|--------------------------------------------------------------------|-----|
| 10.15.5   | Ethnic Pharmaceutical Safety and Efficacy Issues: Pharmacodynamics | 752 |
| 10.15.6   | Bridging Study Design                                              | 753 |
| 10.15.6.1 | Candidate Compound                                                 | 754 |
| 10.15.6.2 | Basic Study Design                                                 | 754 |
| 10.15.6.3 | Subject Population                                                 | 755 |
| 10.15.6.4 | Diet and Food Effects                                              | 758 |
| 10.15.6.5 | Restrictions                                                       | 759 |
| 10.15.7   | Bridging Study Conduct                                             | 761 |
| 10.15.7.1 | Recruitment                                                        | 761 |
| 10.15.7.2 | Informed Consent                                                   | 761 |
| 10.15.7.3 | GCP Practices                                                      | 762 |
| 10.15.7.4 | Adverse Events                                                     | 762 |
| 10.15.7.5 | Documentation                                                      | 763 |
| 10.15.7.6 | Monitoring and Audits                                              | 763 |
| 10.15.8   | Bridging Study Experience                                          | 763 |
| 10.15.8.1 | Examples of Success                                                | 763 |
| 10.15.8.2 | Challenges                                                         | 764 |
| 10.15.9   | Summary and Future Directions                                      | 765 |
|           | References                                                         | 765 |

## 10.15.1 INTRODUCTION

Globalization increasingly characterizes the pharmaceutical industry. We see mergers and acquisitions within industry, rapid international information sharing, and emerging centers for drug discovery. With this, we see more urgency for international regulatory approval and marketing of pharmaceutical compounds. However, regulatory approval in a new region is generally not a simple matter of reciprocity. Practitioners and regulatory agencies have long been concerned with ethnic differences that may affect the optimal use of a medication in a new region, even if it is well understood in the original region. Agencies have understandable concerns about defining the indication, dose selection, product labeling, safety, tolerability, and monitoring of the medication in the new region. “Bridging studies” are one important mechanism for addressing these questions.

### 10.15.1.1 Are Ethnic Factors Real?

As an American working in Honolulu, Hawaii, arguably one of the world’s major cross roads for Eastern and Western cultural exchange, I have had the opportunity to participate in discussions regarding clinical pharmacology development programs with people worldwide. In these discussions, I hear two distinct schools of thought regarding “ethnic differences.”

One school maintains that ethnic differences are real and important considerations in evaluating new medications. For example, some see individuals composing Japanese society as unique and different from those in other societies, thus foreign data must be viewed with great caution. Support for this view comes from review of the historic and cultural differences between Japanese and Western societies, including differences in diet, average weight, expected life span, social homogeneity, and epidemiological patterns of disease. In addition, Japanese practitioners typically

utilized considerably lower doses of medications than their Western colleagues, even when using the same medication for the same indication. Data from controlled studies illustrating substantially different pharmacokinetic profiles in Japanese versus Caucasians for certain compounds, and differences in adverse event profiles in Japanese versus Caucasians for certain medications are cited to add scientific support to this perspective.

The other school asserts that there is no such thing as “ethnic differences.” This school maintains that “under the skin” all the world’s peoples are essentially the same, and that apparent ethnic differences can be explained on the basis of identifiable variables or individual differences. For example, Japanese have traditionally been smaller and lighter than Westerners and many apparent pharmacokinetic ethnic differences disappear once the profiles are adjusted for weight or percent body fat. Other differences can be largely explained by identifiable factors such as heterogeneity in hepatic cytochromes or other metabolizing enzymes, or known food effects and nutritional/dietary differences. This view is supported by data from controlled trials showing that individual intraethnic differences in pharmacokinetic profiles for many compounds are often greater than interethnic differences.

Which school is correct? In this chapter, we will discuss these issues and then revisit this question.

#### **10.15.1.2 What Are “Bridging Studies”?**

When considering two regions with ethnic differences, bridging studies are performed to facilitate approval of a medication in the new region after it has already been studied or approved in another region. A bridging study is defined as: “A supplemental study performed in the new region to provide pharmacodynamic or clinical data on efficacy, safety, dosage and dose regimen in the new region that will allow extrapolation of the foreign clinical data to the new region. Such studies could include additional pharmacokinetic information” [1].

The concept of a “bridge” is a useful metaphor. Physical bridges have long enabled people to travel and share goods and services across rivers and other natural barriers. Similarly, a bridging study provides data to members of the new jurisdiction, enabling them to interpret and apply data from the original jurisdiction. Especially if the results allow the new region to believe the drug will behave similarly in both regions, or at least reasonably similarly with a dose adjustment, the study creates a bridge allowing data from the first region to be used to support approval and labeling in the new region.

In their simplest form, bridging studies are clinical pharmacology studies done in healthy individuals or patients to compare the pharmacokinetic (PK) profile, safety and tolerability of single or multiple doses in subjects representing the new region, compared with subjects representing the original region. Such studies may be performed in either or both regions and are usually conducted in a relatively small number of participants, who may be “healthy” and “normal.” These studies are generally modeled after phase I studies performed in the United States and primarily determine if a dose adjustment is appropriate in the new region.

In the larger context, additional studies performed in the new region confirming pharmacokinetics and/or the efficacy, safety, and tolerability are also considered bridging studies if their design is based on data from foreign studies and their

purpose is to confirm or clarify the drug's behavior in the new population compared with the reference population. These studies are generally performed within the new region and enroll the target patient population. Their design is similar to phase II or phase III studies performed in the United States.

### 10.15.2 WHY DO BRIDGING STUDIES?

From a business standpoint, it is about saving time and money. The bridge allows a large body of existing safety and efficacy data to be used that would otherwise have to be re-created, allowing pivotal efficacy and safety studies in the new region to be smaller and shorter than they would be otherwise.

The key variable is the dose. Should the dose in the new region be the same, higher, or lower? Coupled closely with identifying the appropriate dose, product labeling and safety monitoring may be different in a new region, reflecting both the pharmacodynamic (PD) effects, definition of disease and identification of the target population, and the priorities and medical practices of the new region.

In this section, we will be mainly discussing the United States, the European Union (EU), and Japan, with primary attention to medications studied or approved in the United States and the European Union being considered for approval in Japan. To better understand the background for bridging studies, it may be helpful to review the cultural context for medical practice in Japan and the traditional Japanese approach to pharmacological research.

#### 10.15.2.1 Traditional Practice of Medicine in Japan and the United States

The practice of medicine in Japan has traditionally been quite different from that in the United States. Medical practice in the United States has long incorporated a fee for service system with multiple third-party payers, a competitive marketplace for drug pricing, and a physician–patient relationship based on considerable patient autonomy and expectation of information, second opinions, and informed consent.

Not so in Japan. Rihito Kimura, a Japanese bioethicist, summarizes as follows:

In the long tradition of Japanese medical practice, the Confucian notion of *jin* (benevolence) has been one of the most important ethical elements; medicine itself is known as *jinjyutsu* (the art of *jin*). Physicians, as conduits of *jin*, were required to act with benevolence toward their patients, and were responsible for the welfare of patients in a fiduciary (trust) relationship (Kimura, 1991a). It was obligatory to use medicine, a gift of benevolence, for the good of others even without payment. Physicians fulfilled their responsibility toward their patients and the patients' family members by acting in a paternalistic and authoritative way; the Japanese, nurtured in the Confucian ethos to respect law, order, authority, and social status, acquiesced without murmur to the superior knowledge of the physician. [2]

Kimura further notes:

It is important to note that owing to the character of Japanese society and its distinctive historical understanding of medicine and the role and responsibilities of the physician, it was not until the 1960s that the bioethical and sociolegal concerns about the practice

of medicine began to be deliberately reflected in Japanese society, and only during the 1980s that the notions of autonomy and rights in medicine, and of bioethics in general, became gradually influential. [2]

In addition to these cultural factors and relationship patterns, Japanese medical practice also often incorporated acupuncture and herbal medications. A universal health insurance covered most medical services and medication prices were standardized centrally.

#### 10.15.2.2 Pharmaceutical Development in Japan

The traditional approach to pharmaceutical development in Japan reflected their cultural and dynamics of health care delivery in general. Trials were generally performed under the direction of a chief investigator who wrote the protocol and reviewed the data. It was not until 1980 that the Pharmaceutical Affairs Law (PAL) included an explicit clause regulating clinical trials, including the sponsor's obligation to submit a clinical trial plan, and the minister's authority to instruct sponsors to modify the trials to avoid possible health hazards.

Even still, investigators were allowed to enroll subjects with oral rather than written consent, the discussion about possible adverse events was generally abbreviated (by Western standards), protocols were not routinely reviewed by independent ethics review boards, sponsors and auditors were not allowed access to the original source documents to establish data validity, accountability for investigational products at the trial site was informal, and reporting of adverse events was essentially up to the investigator. Studies were usually done without much if any support from clinical research coordinators, research nurses, and data management staff.

These practices changed in 1998 when the Japanese regulations were revised to meet the standards of the International Conference on Harmonisation (ICH) on good clinical practices (GCP) regulations, reflecting pharmaceutical development practices in Europe and the United States.

For cultural reasons, incorporation of the new regulations was particularly difficult regarding obtaining written informed consent, and enrollment of subjects decreased dramatically for several years. Japanese people were not accustomed to making a written contract for their daily activities, especially regarding health care. Physicians were accustomed to making patient care decisions, and many patients were more comfortable letting physicians do whatever they thought was right rather than being informed fully and being asked to make a decision themselves. In particular, Japanese people are sensitive about possible side effects of drugs. A full description of the major undesirable possibilities is not usual in daily medical practice in Japan, thus the tendency is to refrain from study participation after seeing these in written form, even though the risk is small.

Obtaining informed consent presented a dilemma for Japanese researchers. Although they recognized the importance, obtaining consent is not common in daily medical practice and contradicts the traditional doctor–patient relationships deeply rooted in the culture.

Furthermore, study participation often lacked incentives. With universal health insurance coverage, study participation did not necessarily provide services otherwise unavailable and instead required more frequent visits, procedures, and

observations than normal treatment. Stipend payments for study participation have only been recently introduced. In general, there were negative perceptions regarding study participation, rather than seeing study participation as an opportunity to contribute to society and thus an honorable duty, such as donating blood.

#### 10.15.2.3 Safety Versus Efficacy

Both safety and efficacy are important everywhere. But traditionally, there is a different cultural emphasis on safety versus efficacy in Japan compared with the United States. While Americans want safe medications, the cultural bias also insists “give me something that works,” with high interest in efficacy. If there are potential adverse effects, most individuals are willing to weigh these risks (once properly advised) against the expected benefit.

In Japan, it is almost the opposite. While they want effective medications, both providers and patients highly desire to avoid anything that could make them worse. As a cultural practice, they are generally willing to use low doses or less efficacious medications if such will also offer a safer approach.

#### 10.15.2.4 Market Factors

Of necessity, economic considerations are major factors in drug development, particularly regarding bridging studies. The largest market for pharmaceutical products is the United States, followed by the European Union (when considered collectively), then Japan, with Japan as second largest when considering individual nations. To illustrate, in 2002, the global market for sales of pharmaceutical products was approximately \$401 billion. Sales are markedly disproportional to population. U.S. sales accounted for over 50% of the world's pharmaceutical sales, of about \$204 billion, and continues to grow at 12% per year. The European Union followed the United States with combined sales of about \$91 billion, with Japan at \$47 billion. Thus these three jurisdictions account for less than 15% of the world's population, yet about 85% of the world's pharmaceutical sales.

Active drug discovery is ongoing in all three jurisdictions, and all three have and continue to produce important compounds. Historically, the European Union invested the most in pharmaceutical research and development, a leadership role that continued through 1990. Growth in the U.S. pharmaceutical industry was such that U.S. investment exceeded Europe in the early 1990s and has continued to grow rapidly. In 2004 research input by U.S. pharmaceuticals nearly matched the investments of Europe and Japan combined.

Each jurisdiction has its own approach to patent protection, marketing, distribution, and pricing. For example, U.S. pharmaceutical companies generally base pricing on market considerations. In Japan, prices have been established under the oversight of the Ministry of Health, Labor, and Welfare (MHLW), with downward adjustments applied, on the average, every 2 years. Even considering these differences, if a pharmaceutical company has a product that is commercially successful in one jurisdiction, there is usually sufficient potential for financial gain to interest industry in offering the compound in the other jurisdictions.

Market factors interface with political structures and regulatory practices and affect product development. During the 1980s and 1990s, it took a U.S.

pharmaceutical company 10 years, on average, to develop and launch a new product, whereas it required 15–17 years for the Japanese counterparts. Partially for this reason, of the 178 new chemical entities launched worldwide between 1999 and 2003, just 47 were launched in Japan. On the other hand, drugs discovered in Japan tended to be rapidly acquired and successfully launched by Western companies. Examples include Pravachol, Crestor, and Aricept, discovered, respectively, by Sankyo, Shionogi, and Eisai laboratories yet marketed outside of Japan by Bristol Meyers Squibb, AstraZeneca, and Pfizer.

Market factors are mentioned here because they are such powerful considerations, although an adequate discussion of this topic is beyond the scope of this chapter.

#### 10.15.2.5 Social and Political Factors

Collective social opinion is a significant factor in the pharmaceutical industry, influencing regulation, pricing strategies, and standards for determining what is “generally safe and well tolerated.” Social influences are particularly important in bridging studies and reflect a continuation of the usual perspectives of one’s own culture and the natural comfort of seeing and doing things in ways that are familiar. It takes substantial effort to understand perspectives and traditions that are different from your own. As will be discussed in more detail below, cultural factors are significant considerations in the practice of medicine in general and the role of pharmaceuticals in particular. Thus it is understandable that the regulatory agencies in each jurisdiction will operate with primary concern for their own population, in ways which generally represent local medical and business practices. These concerns and practices vary from one country to another. Successful global pharmaceutical development strategies require appreciation for these political factors.

The agencies in each jurisdiction and the approval process for new compounds operate within this milieu and change over time. The milestones for the U.S. Food and Drug Administration (FDA) include a long list of laws and regulations. Historically, most were “political” responses to problems within the industry [3]. Recently, the emphasis has been on increased efficiency.

Historical development in Japan was somewhat similar, with a recent emphasis on greater efficiency. Japan introduced a revised Pharmaceutical Affairs Law (PAL) in 2002, with compliance effective by 2005, with organizational and regulatory changes in product registration, industry standards, bridging studies, clinical trials labeling, advertising, product classification, and intellectual property. Restructuring under the revised PAL includes the creation of the Pharmaceutical and Medical Devices Agency (PMDA), combining the Organization for Pharmaceutical Safety and Research (OPSR), the Pharmaceutical and Medical Device Center (PMDEC), and the Japan Association for the Advancement of Medical Equipment (JAAME). The purpose of the PMDA is to expedite what had historically been a slow review process. The PMDA revised systems and regulations involving the import, marketing, and contract manufacturing requirements of non-Japanese pharmaceutical companies and signaled an intent to gradually increase the acceptance of new drug application (NDA) submissions featuring bridging studies.

Political factors and the structure of each jurisdiction’s regulatory agencies are also topics beyond the scope of this chapter. They are introduced here out of



recognition for their significance and as a transition to the discussion of regulatory guidance.

### 10.15.3 REGULATORY GUIDANCE: ICH AND E5

Bridging studies would not be possible without consensus regulatory guidance to allow data from studies done in one jurisdiction to be accepted in another. Consensus at an appropriate level of detail on issues this important is not easily achieved. Despite recognition of the substantial benefits to all concerned, consensus guidelines required diligent effort for nearly a decade before they were agreed, adopted, and utilized.

The regulatory guidance currently utilized is the effort of the ICH. The history is briefly summarized below, with attention to the topic (E5) describing use of foreign data.

#### 10.15.3.1 ICH History

The importance of independent evaluation of medicinal products before they are allowed on the market was realized at different times in different regions. In the United States, the role of the FDA in reviewing medications was markedly increased following a tragic mistake in the formulation of a sulfanilamide syrup in 1937. Japan began requiring all medicinal products to be registered for sale starting in the 1950s. Response to the thalidomide tragedy of the 1960s expanded regulatory oversight in Europe.

The different regulatory systems were based on the same fundamental obligations to evaluate the quality, safety, and efficacy of the pharmaceutical products, yet the detailed technical requirements diverged over time to such an extent that industry found it necessary to duplicate many time-consuming and expensive test procedures in order to market new products internationally.

By the late 1980s, the urgent need to rationalize and harmonize regulation in an increasingly global marketplace was impelled by concerns over the rising costs of health care, including pharmaceutical R&D, in juxtaposition to the need to meet the public expectation that there should be a minimum of delay in making safe and efficacious new treatments available to patients in need.

Harmonization of regulatory requirements was pioneered by the European Community, in the 1980s, as the European Community (now the European Union) moved toward the development of a single market for pharmaceuticals. The success achieved in Europe demonstrated that harmonization was feasible. At the same time there were bilateral discussions between Europe, Japan, and the United States on possibilities for harmonization.

The birth of the ICH took place at a meeting in April 1990. Representatives of the regulatory agencies and industry associations of Europe, Japan, and the United States met, primarily, to plan an international conference. The ICH Steering Committee, which was established at that meeting, has since met at least twice a year, with the location rotating between the three regions. ICH work groups have met frequently, and six ICH meetings have been held, most recently in November 2003 in Osaka, Japan. Although the scope of the ICH topics has expanded over time, the

fundamental goals are still generally consistent with the early goals, as expressed in a statement by the ICH Steering Committee from October, 1990:

The Parties cosponsoring this Conference, represented at the 2nd Steering Committee Meeting in Tokyo, 23–24 October 1990 re-affirmed their commitment to increased international harmonization, aimed at ensuring that good quality, safe and effective medicines are developed and registered in the most efficient and cost-effective manner. These activities are pursued in the interest of the consumer and public health, to prevent unnecessary duplication of clinical trials in humans and to minimize the use of animal testing without compromising the regulatory obligations of safety and effectiveness.

ICH Guidelines cover a variety of topics, generally arranged as:

- Quality Topics (e.g., Q7, on Good Manufacturing Practices and Q8 on Pharmaceutical Development)
- Safety Topics (e.g., S2 discussing genotoxicity studies, and S5 addressing reproductive toxicity)
- Efficacy Topics relating to studies in human subjects (e.g., E6 on good clinical practices)
- Multidisciplinary Topics, which do not fit uniquely into the above categories (e.g., M1: Medical Terminology).

### 10.15.3.2 E5 Guidelines

Of special relevance for bridging studies, ICH Efficacy Topic E5 discusses Ethnic Factors in the Acceptability of Foreign Clinical Data. The current document was recommended for adoption by the ICH Steering Committee on February 5, 1998, approved by the EU in March, 1998, published in the U.S. *Federal Register* in June 1998, and adopted by the MHLW in Japan in August 1998 [1].

Early experience in all ICH regions indicated the need for some clarification, and a Questions and Answers document was issued in November 2003. This was adopted by the European Union in November 2003, by Japan in February 2004, and by the United States in June 2004. The bridging studies discussed in this chapter were conducted within the regulatory environmental matrix described in these two documents. Although classified by ICH as an “efficacy” topic, the E5 guidelines are equally concerned with issues of patient safety.

The perspective of this author is that E5 comprises a particularly key component of ICH, providing the essential mechanism for the success of the ICH effort in achieving its goals. All topics are important. However, if data cannot be shared from one jurisdiction to the other, the pharmaceutical sponsors are essentially back where they started: duplicating the investigational program in a new region.

**E5 Objectives** The objectives of E5 are:

- To describe the characteristics of foreign clinical data that will facilitate their extrapolation to different populations and support their acceptance as a basis for registration of a medicine in a new region.

- To describe regulatory strategies that minimize duplication of clinical data and facilitate acceptance of foreign clinical data in the new region.
- To describe the use of bridging studies, when necessary, to allow extrapolation of foreign clinical data to a new region.
- To describe development strategies capable of characterizing ethnic factor influences on safety, efficacy, dosage, and dose regimen.

The guidelines further state “All regions acknowledge the desirability of utilizing foreign clinical data. . . . However, concern that ethnic differences may affect the medication’s safety, efficacy, dosage and dose regimen in the new region has limited the willingness to rely on foreign clinical data,” and thus explains the reason why there was so much duplication of the research program. The guidelines are based on the assumption that many medicines will have comparable characteristics and effects across regions, although there is acknowledgment that ethnic differences may affect the safety, efficacy, dosage, or dose regimen.

E5 “is based on the premise that it is not necessary to repeat the entire clinical drug development program in the new region.” It is also “not intended to alter the data requirements for registration in the new region; it seeks to recommend when these data requirements may be satisfied with foreign clinical data,” noting that “additional studies conducted in any region may be required by the new region to complete the clinical data package.”

**Regional Requirements** The three ICH regions have developed a basic consensus on the fundamentals of appropriate study conduct, summarized as good clinical practices (E6). Still, the regions vary in what constitutes an approvable clinical data package. At a minimum, studies need to provide data that contain adequate characterization of pharmacokinetics, pharmacodynamics, dose response, efficacy, and safety. Once these are established, further clinical trials should be performed according to GCP with appropriate choice of controls and endpoints using medical and diagnostic definitions acceptable to the new region. In particular, the pharmacodynamic endpoints for safety, tolerability, and efficacy must be performed either in the new foreign region or in a population representative of the new region.

Several ICH guidelines address these aspects of program design and provide guidance for a complete clinical data package. These include GCPs (E6), evaluation of dose–response (E4), adequacy of safety data (E1 and E2), conduct of studies in the elderly (E7), conduct of studies in a pediatric population (E11), reporting of study results (E3), general considerations for clinical trials (E8), statistical considerations (E9), and choice of control groups (E10). Guidelines are also emerging regarding specific therapeutic areas (e.g., hypertension, E12A) and clinical safety issues (e.g., QT/QTc interval prolongation for non-antiarrhythmic drugs, E14).

**Ethnic Factors** Within E5, “ethnic factors are factors relating to races or large populations grouped according to common traits and customs.” The guidance specifically notes that ethnic factors go beyond racial factors per se. The focus of attention for ethnic factors includes diet and lifestyle considerations with implications for the expected safety and efficacy of the medication in the region. For example, the Japanese authorities are more concerned with the effects of a medication within

Japan than the effects of the medication when taken by Japanese individuals within the United States.

Ethnic factors are broadly classified as Extrinsic and Intrinsic.

**Extrinsic Factors** Extrinsic factors are associated with the environment and culture in which a person resides and are more behaviorally than genetically derived. Examples include diet, use of tobacco and alcohol, exposure to pollution and sun-shine, socioeconomic status, medical practice, compliance with prescribed medica-tions, and practices in clinical trials design and conduct.

**Intrinsic Factors** Intrinsic factors apply more to the genetic features of the indi-vidual and the collective subpopulation of individuals within a region, including the physical aspects of genetics plus the lifestyle of the region. Examples include genetic polymorphism, age, gender, height, weight, lean body mass, body composition, and organ dysfunction. Table 1 illustrates intrinsic and extrinsic ethnic factors as defined in E5.

**Sensitivity to Ethnic Factors** Ethnic sensitivity is compound specific, depending on its pharmacologic class, indication, and the usual age and gender of the target patient population. No single factor is predictive of the relative sensitivity, but the following properties as a whole are generally indicative.

Examples of properties making a compound less sensitive to ethnic factors include:

- Linear pharmacokinetics
- A flat pharmacodynamic curve for both efficacy and safety

TABLE 1 Classification of Intrinsic and Extrinsic Ethnic Factors

| INTRINSIC                                   |                                           | EXTRINSIC                        |
|---------------------------------------------|-------------------------------------------|----------------------------------|
| Genetic                                     | Physiological and pathological conditions | Environmental                    |
| Gender                                      | Age (children-elderly)                    | Climate<br>Sunlight<br>Pollution |
| Height                                      | Liver                                     | <b>Culture</b>                   |
| Bodyweight                                  | Kidney                                    | Socioeconomic factors            |
|                                             | Cardiovascular functions                  | Educational status               |
| ADME                                        |                                           | Language                         |
| Receptor sensitivity                        |                                           | Medical practice                 |
| Race                                        |                                           | Disease definition/Diagnostic    |
| Genetic polymorphism of the drug metabolism |                                           | Therapeutic approach             |
|                                             |                                           | Drug compliance                  |
| Genetic diseases                            | Diseases                                  | Smoking                          |
|                                             |                                           | Alcohol                          |
|                                             |                                           | Food habits                      |
|                                             |                                           | Stress                           |
|                                             |                                           | Regulatory practice/GCP          |
|                                             |                                           | Methodology/Endpoints            |

- A wide therapeutic dose range
- Minimal metabolism or metabolism distributed among multiple pathways
- High bioavailability, that is, less susceptibility to dietary absorption effects
- Low potential for protein binding
- Little potential for drug–drug, drug–diet, and drug–disease interactions
- Nonsystemic mode of action
- Little potential for inappropriate use and abuse

The opposite properties make a compound more sensitive to ethnic factors, especially:

- Metabolism by enzymes known to show genetic polymorphism
- Administration as a prodrug, with potential for ethnically variable enzymatic conversion
- Low bioavailability, with more susceptibility to dietary absorption effects

***Bridging Data Package*** E5 specifies that the complete clinical data package intended for registration contain clinical data that fulfill the regulatory requirements of the new region and contain pharmacokinetic data relevant to the population of the new region. Use of a bridging study allows inclusion of selected information from the original region, including pharmacokinetic, dose–response, and pharmacodynamic data, with appropriate extrapolation to the population of the new region based on the results of the bridging study.

***Global Development Strategies*** Sponsors are increasingly interested in concurrent international development, and use of ICH guidelines allow data to be used across the three major jurisdictions. Thus we are seeing Japanese–Caucasian bridging studies with a clinical pharmacology focus performed concurrently with phase II studies in the United States and European Union, with the expectation that phases II and III studies will be performed in Japan concurrently with the phase III program in the United States.

### 10.15.3.3 Implications for Non-ICH Jurisdictions

This may be all well and fine for Japan, the European Union, and the United States. How about the rest of the world? The ICH regions account for approximately 85% of global pharmaceutical sales, but less than 15% of the global population. The non-ICH regions contain some of the most dynamic and changing pharmaceutical markets. These regions are also seeking innovative and better quality medicines, are also interested in benefiting from the medicines developed within the ICH regions, and are also attempting to develop world class drug/biological regulatory systems.

Long recognizing these global implications, the ICH has made efforts to involve stakeholders with similar needs and related interests from other regions. For example, the ICH Steering Committee statement of July 1997 included language that “the widespread adoption and use of ICH guidelines ... is essential if the long-term

benefit of international harmonization in terms of quicker access to effective new medicines, is to be available to patients throughout the world.”

In 1999, the ICH steering committee created a Global Cooperation Group (GCG) subcommittee to share information and harmonization practices with other nations. Participants include Canada and the World Health Organization (WHO), representing 180 countries.

Harmonization initiatives are also being coordinated on a regional basis by the following organizations:

- Asia-Pacific Economic Cooperation (APEC)
- Association of Southeast Asian Nations (ASEAN)
- Gulf Cooperation Countries (GCC)
- Pan American Network on Drug Regulatory Harmonization (PANDRH)
- South African Development Community (SADC)

These agencies are participants or interested observers in the ICH process and adopt or incorporate ICH guidelines in their regions where applicable. Thus ICH guidelines, or guidelines similar to ICH, may facilitate global availability of new compounds, global development of new compounds, and globalization of clinical research.

#### **10.15.4 ETHNIC PHARMACEUTICAL SAFETY ISSUES: PHARMACOKINETICS**

Pharmacology is a science. The ethnic, historical, and regulatory issues primarily provide a context for understanding protocol design and the regulatory approval process. The heart and soul of the clinical trials is to provide scientific data according to recognized practices that answer the most important questions regarding the behavior and effects of a specific compound following human exposure.

Sections 4 and 5 provide an overview of pharmacokinetics (PK), that is, the absorption, distribution, metabolism, and elimination (ADME) of the subject compound and the pharmacodynamic (PD) considerations and experience of bridging studies. Only selected examples are cited here, in the interest of illustrating ethnic issues relevant to planning bridging studies.

##### **10.15.4.1 ADME: Absorption**

This discussion of absorption will focus primarily on the gastrointestinal tract. For our purposes, intravenous, intramuscular, and subcutaneous routes largely bypass the absorption process. There may be ethnic differences in absorption of transdermal medications, as these may vary with skin type, including racial differences in skin thickness, barrier function, and transepidermal water loss. In terms of ethnic differences, intraocular, intravaginal, intrarectal, and other routes of administration have not received extensive investigation.

Absorption from the gut is a passive process for many medications. In this case, ethnic differences in absorption would not be expected, and none have been demonstrated. However, a number of medications are actively metabolized within

the intestinal membranes or absorbed via transport proteins. At last three major enzymes systems are involved with drug absorption across the intestinal membrane, CYP (especially CYP 3A4), the multidrug resistance proteins (MDR1-5) including P-glycoprotein (MDR-1), and the organic anion transporting polypeptide (OATP) family, which transport a large array of structurally divergent drugs [4]. While their relevance for ethnic differences at the population level have yet to be defined, individual heterogeneity and single nucleotide polymorphism (SNP) variation have been identified for a number of these transport proteins. Undoubtedly, other transport proteins have yet to be defined.

The bridging studies done with Japanese to date have shown several compounds, such as risedronate and rosuvastatin, with a higher time to maximum concentration ( $T_{\max}$ ) and area under the curve (AUC) yet a similar half-life ( $T_{1/2}$ ). These observations remain after adjustment for weight, gender, percent body fat, and related individual factors, suggesting the ADME difference is primarily one of increased absorption rather than differences in distribution, metabolism, or elimination. Bridging studies have documented these observations, even though the precise mechanism remains to be elucidated.

#### 10.15.4.2 ADME: Distribution

Insofar as a medication's distribution varies throughout body compartments, generally lipophilic medications would be expected to show a difference in the volume of distribution, with ethnic effects reflecting the population's typical percent body fat versus lean body mass. In Japan, obesity is defined as a body mass index (BMI) of greater than 25, whereas in the United States an individual is considered "overweight" with BMI 27–30 and "obese" when the BMI is over 30. This may partially account for the greater effect of diazepam in Asians than in Caucasians.

Some ethnic differences in protein binding have been described. The two major drug-binding proteins in blood plasma are albumin and  $\alpha_1$ -acid glycoprotein (AGP). Albumin has a relatively low drug affinity, but a high binding capacity with three major binding sites and the ability to bind acids, bases, and neutral compounds. Conversely, AGP has a high drug affinity but low capacity. AGP binds many basic psychotropic compounds, including haloperidol, chlorpromazine, and fluvoxamine. The structures of albumin and AGP are genetically determined, with polymorphic variations across racial groups in several studies. In general, evidence indicates that Caucasians (especially those of European origin) have higher levels of AGP than Asians and blacks, although albumin levels are similar.

Finally, ethnic differences in transport proteins such as P-glycoprotein and the OATP enzyme systems may account for differences in intracellular concentration within the liver, kidney, brain, and other tissues.

#### 10.15.4.3 ADME: Metabolism

A substantial body of literature exists regarding heterogeneity of metabolizing enzymes, particularly the hepatic cytochrome P450 (CYP) enzyme families. Considering the wide range of intraethnic mutant alleles, some authorities may argue that these approach "individual" factors. Nonetheless, the incidence of certain genotypes shows ethnic patterns and must be considered when discussing ethnic differences.



Wider application of genotyping has clarified our understanding while adding considerable complexity, as we have identified multiple mutant genotypes with similar phenotypical function. As the CYP system is also involved with detoxification or metabolism of a wide range of dietary components and environmental toxins, implications for risk of certain cancers in various ethnic populations may be as important clinically as the implications for drug metabolism and remains an area of active research.

Overall, enzymes in the CYP 2 superfamily show far greater heterogeneity than CYP 1 and CYP 3. In addition to the allele heterogeneity, some studies also show considerable individual and ethnic differences in enzyme content and activity [5–7]. Selected CYP enzyme systems with ethnic relevance for drug metabolism are considered next.

**CYP 1A1, 1A2, and 1B1** CYP 1A2 substrates include caffeine, theophylline, and tacrine. Four types of CYP 1A1 and three types of CYP 1B1 polymorphisms have been described, with distinct incidence patterns in Japanese compared with Caucasians [8], and may play a role in cancer risk. CYP 1A2 polymorphisms vary in the Chinese population, with some more similar to Japanese and some more similar to Caucasians [9]. Additional CYP 1A2 polymorphisms have been identified and shown to have variable incidence in Taiwanese, Caucasian, and African American populations. However, the functional differences of these variants is fairly small, and relevance for dose selection has not been established [10, 11].

**CYP 2C9** CYP 2C9 is a major enzyme, accounting for about 20% of the total P450 enzyme pool in both Japanese and Caucasians. Substrates include many nonsteroid anti-inflammatory medications, glipizide, losartan, and S-warfarin. There is large interindividual variability in 2C9 content and considerable intraethnic variation in both population. Some mutant alleles are shared in both populations and some are unique to each population [12].

CYP 2C9 is polymorphic, with six major allele variants. The different alleles vary in ethnic distribution between Asian, Caucasian, and African populations, and different alleles vary in their metabolic activity for common substrates.

**CYP 2C19** CYP 2C19 substrates include omeprazole and phenytoin. Along with CYP 2D6, CYP 2C19 contributes to the metabolism of selective serotonin reuptake inhibitors, tricyclic antidepressants, and a variety of other psychotropic medications.

**CYP 2D6** CYP 2D6 substrates include codeine and other opiate medications,  $\beta$  blockers, and antidysrhythmic medications such as flecainide and propafenone. CYP 2D6 also contributes significantly to the metabolism of tricyclic antidepressants, selective serotonin reuptake inhibitors, and other psychotropic medications such as haloperidol and thioridazine.

**CYP 2E1** The classic pharmacological substrate for CYP 2E1 is chlorzoxazone, although 2E1 is also a metabolic pathway for *p*-nitrophenol, ethanol, and dimethylnitrosamine. Unlike most of the other CYP enzymes, 2E1 metabolism has a tendency to activate substrates into more toxic metabolites, such as acetaminophen,

ethanol, carbon tetrachloride, thioacetamide, and the N-nitroso compounds found in red meat and processed meats. Thus a higher level of 2E1 activity may increase the risk of alcoholic liver injury and various malignancies [13–15].

A comparison of 39 Japanese and 45 Caucasians showed three types of polymorphisms, with seven genotypes found in the Japanese and four in the Caucasians [7a]. Although little difference in metabolic activity was found between the ethnic groups with one probe substrate despite the genotype heterogeneity, there is a typical difference between Japanese and Caucasians in the metabolism of chlorzoxazone, with lower clearance for the Japanese even after adjusting for differences in body weight, suggesting a lower level of catalytic activity in Japanese [4].

**CYP 3A4 and 3A5** CYP 3A4 is present in the intestinal brush border as well as hepatocytes and is the major “workhorse” for drug metabolism, that is, the major metabolic pathway for as many as 60% of our currently known pharmaceutical compounds. This lack of substrate specificity also lends itself to a large number of drug and diet interactions [16]. Individual 3A4 content and activity varies widely, with up to 40-fold individual content variation and up to 11-fold individual variability in midazolam oral clearance [5, 17]. More than 30 SNP variations have been identified, which vary in ethnic distribution between Japanese, Caucasian, and African American subjects. Despite this variety, allele frequency is low (<5%) and appears as heterozygous with the wild type. None of the alleles have been associated with clinically significant differences in drug metabolism, especially regarding interethnic comparisons [18].

CYP 3A5 is a coenzyme for many 3A4 substrates. CYP 3A5 is also polymorphically expressed with varying ethnic distribution, but the functional significance has not been established [19].

As we learn more about individual intraethnic differences in content and activity, allele function, and diet interactions, ethnic differences previously reported in CYP 3A4 activity may be partly due to individual differences in studies with small sample sizes and/or dietary factors [6, 20].

**Non-CYP Metabolic Enzymes** A number of enzymes other than the CYP system are also important in drug metabolism and demonstrate ethnic heterogeneity. The intra- and interethnic genotype variation is not always linked with phenotype expression, and this remains an area of active inquiry. Selected examples include:

- Aldehyde dehydrogenase (ALDH) is primarily significant for metabolism of ethanol, especially ALDH2. An atypical allele, ALDH2(2) results in accumulation of acetaldehyde with histamine release and flushing. Approximately 30% of Asians are ALDH deficient [21].
- *N*-acetyltransferases (NAT) demonstrate polymorphism and ethnic variation and are involved with metabolism of environmental carcinogens and a number of arylamine and hydrazine drugs [22]. As expected, SNP and allele analysis show substantial heterogeneity, both within Japanese and within Caucasians [23].
- Thiopurine *S*-methyltransferase (TPMT) is responsible for the metabolism of thiopurine medications such as azathioprine. Patients with deficient TPMT

activity are at high risk for toxicity and can be successfully treated with 10- to 15-fold lower dose. Mutant alleles vary in their ethnic distribution [24].

- Glucuronosyltransferase (UGT) enzymes catalyze the transfer of glucuronic acid to alcohols, carboxylic acids, amines, and free sulfhydryl groups. Twenty-four distinct UGT genes have been identified, with significant ethnic variation in Caucasian, Asian, and African populations [25].

#### 10.15.4.4 ADME: Elimination

The major routes of elimination are hepatic and renal. Ethnic differences in hepatic function are primarily metabolic. To my knowledge, no ethnic differences in post-hepatocyte biliary elimination have been described.

Renal clearance is the net effect of glomerular filtration, tubular secretion, and tubular reabsorption. Glomerular filtration and tubular reabsorption are largely passive processes, where ethnic differences have not been described. However, tubular secretion is an energy-requiring active process, and ethnic differences could possibly result from genetic heterogeneity in this process although they are particularly compound specific. For example, renal clearance for morphine and metabolites morphine-3-glucoronide and morphine-6-glucoronide are higher in Chinese than Caucasians, primarily due to tubular secretion. However, other medications that undergo tubular secretion have not shown ethnic differences, such as furosemide and procainamide.

#### 10.15.4.5 ADME: Dietary Factors

Most of the ADME factors described above are primarily genetically determined. But diet can have a significant influence on how these enzymes function, which may exaggerate or introduce new ethnic differences.

As a context, the “traditional diet” in Japan is quite different from most regions within the United States. Despite globalization, certain generalities still apply:

- Source of complex carbohydrates: Japanese consume more rice, and Americans more wheat, corn, and potatoes.
- Source of protein: Japanese consume more fish and soy products, Americans more beef and dairy products.
- Sources of fat: Japanese consume less fat overall, and the fat consumed is more likely to be from fish and vegetable sources. Americans consume more fat overall, and a higher percent is saturated fat from red meat and other animal sources.
- Typical vegetables and fruits vary between Japan and the United States, in addition to the seasonal variations in both regions.
- Common food preparation techniques and preferred sauces, flavorings, dressings, and spices are typically quite different.

Macronutrient differences may have some affect on drug absorption, as will physiological changes in the fed versus the fasted state such as gastric pH, gastric and intestinal motility, pancreatic and biliary secretion, drug adherence to dietary

fiber, passive drug absorption along with meal uptake, and the like. However, we should draw particular attention to the vegetables, fruits, herbs, and spices. Consumption of these items often varies by culture or region, and many fruits, vegetables, and herbs have been shown to have isoflavones or other compounds that can affect drug absorption, metabolism, or both.

One well-known example is grapefruit and grapefruit juice, which contain isoflavones that significantly inhibit CYP 3A4 in both the intestinal wall and the liver. For medications metabolized by 3A4, grapefruit intake reduces metabolism in the gut, thus increasing absorption and the  $C_{\max}$ . Grapefruit also prolongs the  $T_{1/2}$  by inhibiting hepatic metabolism. Both mechanisms contribute to a higher AUC. These effects persist at least 24 hours after dietary exposure. This grapefruit effect has presumably been present throughout human biological history, even though it was only defined in the 1990s [26]. This effect apparently applies across all ethnic groups, yet some would classify this as an “ethnic factor” as grapefruit consumption varies by region and subpopulation.

Herbal products can interact with prescription medications. St. John’s wort (hyperforin and quercetin) is commonly used as an antidepressant and has been shown to be a potent inducer of CYP 3A4 and P-glycoprotein, although it may inhibit other CYP enzymes [27]. St. John’s wort has been shown to lower the plasma concentration (and/or the pharmacological effect) of a number of drugs including alprazolam, amitriptyline, cyclosporine, digoxin, fexofenadine, indinavir, irinotecan, methadone, nevirapine, simvastatin, tacrolimus, theophylline, warfarin, phenprocoumon, and oral contraceptives, with interactions that may cause a variety of adverse clinical effects [28].

Other herbs, spices, and “natural medications” have significant effects on CYP enzymes. Ginkgo biloba (Ginkgo) can exert potent inhibition on 1A2, 2C9, and 2C19, with less inhibition of 2D6 [29]. A variety of spices, herbal and black teas, and soybean products were tested for their in vitro inhibition of CYP enzymes. All were found to inhibit 3A4 mechanisms to a variable extent, and 26 showed significant inhibition of 2C9, 2C19, or 2D6. Examples of products tested include black tea, herbal tea mixtures, sage, thyme, cloves, goldenseal, and seven soybean varieties [30]. Similar inhibition of CYP 1A2, 2C9, 2C19, 2D6, and 3A4 was noted by herbal extracts of devil’s claw root (*Harpagophytum procumbens*), feverfew herb (*Tanacetum parthenium*), fo-ti root (*Polygonum multiflorum*), kava-kava root (*Piper methysticum*), peppermint oil (*Mentha piperita*), eucalyptus oil (*Eucalyptus globules*), red clover blossom (*Trifolium pretense*), garlic (allicin), and evening primrose oil (cis-linoleic acid) [29, 31]. Gallic acid, a plant phenol found in wines and teas, can inhibit CYP 3A4 under certain conditions [32].

In general, a wide variety of clinically significant adverse events have been documented from herb–drug interactions, through a combination of pharmacokinetic and pharmacodynamic mechanisms [33]. To the extent that herbal use characterizes an ethnic group in a particular region, herbal use can be considered an ethnic factor.

Diet components may also affect transport proteins. Grapefruit, orange, and apple juice have been shown to inhibit human OATP activity, with reduced oral drug bioavailability [34]. Another study of 20 naturally occurring flavonoids and some of their corresponding glycosides demonstrated significant inhibition of OATP drug uptake, indicating the potential for drug interactions [35].

P-glycoprotein is also inhibited by a wide variety of dietary and herbal compounds, including capsaicin (in chili peppers), curcumin (in turmeric), [6]-gingerol (in ginger), and resveratrol (in grapes) [36]. Eight dietary flavonoids commonly present in fruits, vegetables, and plant-derived beverages have been shown to inhibit P-glycoprotein [37, 38]. Citrus (rutaceous) herbs are often used in traditional Japanese medicine and cuisine and contain compounds known as monoterpenoids. A number of these have been shown to be potent P-gp inhibitors [39].

Even food preparation preferences can affect drug pharmacokinetics. Charcoal-broiled and smoked foods contain polycyclic aromatic hydrocarbons formed during their preparation, which can induce intestinal CYP 1A [40].

Much of the research in this area is fairly recent and more remains. We are likely to define new transport proteins, new receptors, and new dietary influences on the pharmacokinetic or pharmacodynamic effects of drugs.

#### 10.15.4.6 ADME: Multiple Mechanisms

Certain compounds will be influenced by combinations of the above factors. Before we complete our discussion about ethnic factors in pharmacokinetics, let us use warfarin, one of the world's most widely prescribed anticoagulant medications, as an example. Although effective at reducing the incidence of thromboembolic disease in common clinical settings, there is a greater than 10-fold interpatient variability in the dose required, necessitating careful monitoring and dose adjustment for each patient.

Warfarin exists as optical S and R isomers. S-warfarin has three to five times more anticoagulant potency than its optical congener. Warfarin is well absorbed orally, although delayed in the presence of food. Absorption through the bowel as well as target intracellular concentrations may be affected by P-glycoprotein. Once absorbed, warfarin is highly protein bound, mainly to albumin. Warfarin is primarily metabolized by CYP 2C9, with minor activity by CYP 1A2 and CYP 3A, and elimination of inactive metabolites in the urine and the stool.

As discussed previously, CYP 2C9 enzymes show marked phenotypical and genotypical variability in various ethnic groups worldwide, with multiple alleles demonstrating considerable differences in rates of S-warfarin metabolism. Multiple variants have also been described in the promoter, exonic, intronic, and 3'-untranslated regions of CYP 2C9.

A wide variety and a large number of medications have reported interactions with warfarin, including azole, macrolide and quinolone antibiotics, nonsteroidal anti-inflammatory drugs, selective serotonin reuptake inhibitors, omeprazole, lipid-lowering agents, amiodarone, and fluorouracil.

Interactions with a wide variety of food and herbal medications have also been described, including vitamin K intake, green tea, danshen, fish oil, soy milk, ginseng, and quilinggao.

While there are wide interindividual differences, Asians in general tend to achieve therapeutic benefit at lower warfarin doses than Caucasians. When assessing "what are the ethnic effects," studies have compared Caucasian with Asian subjects matched by CYP 2C9 allele variants and still found unexplained differences. Likewise, interethnic differences were present with or without consideration of CYP 2C9

coding region variants, suggesting other genetic, dietary, or environmental influences [41, 42].

In this case a simple explanation for ethnic differences remains elusive. Even accounting for the genetic heterogeneity in the primary metabolic pathway, multiple mechanisms appear to be involved. Undefined genetic factors, general dietary patterns, and other environmental influences are all ethnic factors, both intrinsic and extrinsic.

### 10.15.5 ETHNIC PHARMACEUTICAL SAFETY AND EFFICACY ISSUES: PHARMACODYNAMICS

As we discussed earlier, safety and tolerability are as important, if not more important than efficacy. Once the PK profile is understood and necessary dose adjustments taken into consideration, we are prepared to assess efficacy, safety, and tolerability of the selected doses.

Pharmacodynamic effects are often about the same in Japan as elsewhere but need to be evaluated in the context of the dose and related PK characteristics. Consider risedronate sodium as an example, approved in Japan in 2002 for the treatment of osteoporosis with use of a bridging strategy. The dose was set at 2.5 mg/day in Japan, compared with 5 mg/day in the United States and the European Union because of a PK finding that a 2.5-mg dose in Japanese produced a similar AUC to a 5-mg dose in Caucasians. In addition, the efficacy of 2.5 mg/day in Japan regarding the effect on bone density was comparable to the benefit of 5 mg/day in the United States and the European Union. The main difference in the PK was increased absorption. Although these findings primarily involve the PK profile, note that confirmation of the same PD effect at the lower yet equivalent dose was necessary for dose labeling.

The same was observed with rosuvastatin calcium, for the treatment of hypercholesterolemia. Interestingly, this compound was discovered by Shionogi in Japan but developed by AstraZeneca in the United States and the European Union, with subsequent bridging studies leading to approval in Japan. Rosuvastatin received U.S. approval in August 2003 compared with approval in Japan in December 2004. Japanese and other Asian subjects showed increased absorption, with  $C_{\max}$  about twice as high as Caucasians at the same doses. Efficacy at the lower doses paralleled the PK profile. Thus labeling in the United States is from 5 to 40 mg/day with a recommended starting dose of 10 mg/day, compared with labeling in Japan of 2.5 to 20 mg/day with a recommended starting dose of 5 mg/day.

In other cases, for reasons not always well understood, the PD effects and adverse event profile may differ in Japan from the United States and the European Union. Eletriptan hydrobromide, for the treatment of migraine headaches, is another medication approved based on a bridging strategy. The maximal doses were set at 40 mg/day in Japan, compared with 80 mg/day in the European Union, after the bridging study demonstrated higher rates of adverse events such as nausea and somnolence in the Japanese at 80 mg/day, although the  $C_{\max}$  and AUC of single doses were lower in the Japanese than the foreign population. Likewise, sildenafil citrate is recommended at 50 mg, ranging from 25 to 100 mg in the United States and the European Union, but limited to 25–50 mg in Japan as the 100-mg dose caused increased



adverse events (e.g., abnormal vision) in Japanese subjects but added no further effectiveness.

After allowance for ADME profiles, one possible mechanism for different PD responses is ethnic differences in the target receptor. Ethnic heterogeneity has been described for some important receptors. Selected examples include:

- Three functional  $\beta$ -adrenergic receptor polymorphisms have been described with clinical relevance to several disease, including asthma, hypertension, congestive heart failure, cystic fibrosis, and obesity, as well as response to  $\beta$ -agonist therapy such as salbutamol or  $\beta$ -blocker therapy such as propranolol. Distribution of these alleles varies in Caucasian, black, and Asian populations, providing at least a partial explanation for their different clinical responses to therapy [43–46].
- The effects of serotonin-related gene polymorphisms on central nervous system (CNS) serotonergic function vary as a function of both ethnicity and gender [47] and may affect response to therapy.
- The mu opioid receptor gene (OPRM1) has several known variations, with different distribution patterns in Caucasian, Asian, and black populations [48]. Ethnic variations in receptor function contribute to differences in the response to morphine [49], have been associated with methamphetamine psychosis [50], and may affect patterns of opiate addiction [51].

Receptor heterogeneity is an area of active investigation, and ethnic patterns are likely to be defined with relevance for numerous clinical domains [52].

### 10.15.6 BRIDGING STUDY DESIGN

Genetically determined metabolizing enzymes, transport proteins, and receptors are significant intrinsic ethnic factors. Diet is one of the major extrinsic ethnic factors. The data cited above show some reasons to believe drugs may behave differently in different populations or new regions. With the regulatory and physiological context as above, let us turn our attention to bridging study design.

For simplicity of discussion, this section will use a scenario where the bridging goes from the European Union or the United States to Japan, that is, where a compound is approved in the European Union or the United States, or a significant body of data exist in one of these jurisdictions and the compound is well on its way to regulatory approval, and Japan is the new jurisdiction of interest. In the first 5 years of E5 guideline implementation, this has been the most common scenario. Presumably, the same general principles would work when bridging from Japan to the United States or the European Union or bridging from an ICH jurisdiction to a non-ICH jurisdiction.

The comments that follow in this section and the next are based on experience between 1998 and 2005. Requirements for data acceptability are inherently a “moving target,” as we accrue experience with bridging programs in the context of emerging scientific data, medical practice changes, and political realities. In Japan, the Organization for Pharmaceutical Safety and Research (OPSR) offers consultative guidance to sponsors planning a bridging program, and sponsors are generally well advised to seek guidance prospectively.



### 10.15.6.1 Candidate Compound

As expected, the ADME characteristics as well as the intended clinical application of the candidate compound will determine the bridging strategy. The sponsor must assess the potential for intrinsic and ethnic effects based on the criteria discussed above.

### 10.15.6.2 Basic Study Design

The successful bridging study programs to date have often utilized a phase I style clinical pharmacology study to establish the dose selection, most commonly in normal healthy individuals, followed by patient-based longer term Phase II or Phase III style studies demonstrating safety and efficacy within the new jurisdiction. In the case of applying for registration in Japan, the safety and efficacy studies for successful NDAs so far have been performed in Japan. Utilizing ethnic Japanese populations residing out of Japan to demonstrate safety and efficacy has been problematical. As per the examples discussed above, we do not yet know all the variables and mechanisms affecting a compound's PK profile or PD effects. These considerations make it understandable to perform certain safety and efficacy studies within the new target region.

Study design needs to comply with good clinical practices and related ICH guidelines, with ethical principles consistent with the most current internationally recognized standard, currently the Declaration of Helsinki.

Single-dose PK/PD studies ordinarily include a range of at least two or three doses; the usual dose administered in the jurisdiction of reference and one or more lower or higher doses to clarify dose selection in the new jurisdiction. For compounds where single-dose kinetics do not reliably predict steady-state kinetics, a multiple-dose study (again, with a range of doses administered) will likely be advised in the new ethnic group. Pharmacokinetics may need to be confirmed in a patient population if the pivotal PK study was performed in healthy individuals.

The phase II/III efficacy studies may or may not employ a placebo, versus a range of two or three active doses or an already approved active comparator. Dose selection for these studies is based on the bridging ADME data. If the doses selected are lower for the Japanese than commonly used in the United States or the European Union, pharmacodynamic efficacy endpoints need to be confirmed as well as the safety and tolerability at the lower doses selected.

***Variables Affecting ADME Study Data*** As major decisions regarding the phase II/III program will depend on the pharmacokinetic profile, the reference PK study needs to be carefully designed.

A host of variables may affect ADME study data, some of which can be foreseen and anticipated, others are more subtle. Such variables are especially relevant when considering the decision to use historical data for comparison versus concurrent controls, and when considering whether to perform a study at one center or two. At some level, these variables may affect study conduct, data capture, or the regulatory agency's perception of data quality.

Examples include:

- Staff experience, training, and expertise
- Characteristics of the local population from which subjects are recruited
- Procedures for obtaining and documenting informed consent
- Overall GCP compliance
- Temperature, humidity, mold, lighting, ventilation, and other internal environmental variables within the study unit
- Staffing and work flow regarding study procedure performance
- Work habits and involvement of the principal investigator
- Interpretation and application of the protocol
- Source documentation data capture
- Methods and style for inquiry and documentation of adverse events
- Specimen processing and handling
- Calibration and maintenance of scales, centrifuges, and related devices
- Diet and beverages provided while confined, including food preparation techniques
- Interpretation and application of study restrictions
- Medical monitor approval of protocol deviations
- Quality and completeness of the regulatory file
- Monitoring and audit team findings, reports, and interaction with study staff
- Local lab methodology differences and reported ranges of normal
- Unexpected external events (e.g., local electrical power surges or failure, and storms, earthquakes, or other natural disasters)

Bridging PK studies have been successfully performed utilizing a one-protocol, two-center approach [53]. Some studies require placement at multiple centers for timely completion. Nevertheless, for the reasons listed above and many others, combining or comparing data across centers may involve unexpected findings or subtle differences. If unexpected findings occur—are they the result of ethnic differences or study conduct variables? Unexplained findings may require repeating the study under more rigorously controlled conditions.

***Caucasian ADME Data: Historical or Direct Comparison?*** The fundamental assumptions of bridging strategies involve use of historical data gathered in a foreign region, with assessment for applicability to the new region.

Partially for this reason, the use of historical Caucasian data seems like an efficient and cost-saving approach when planning the ADME study. However, if the new data show unexpected differences, it may be difficult to tell if this is due to “real” ethnic differences or the results of differences in study design, conduct, or other nonethnic variables. Considering the pivotal nature of the PK profile for the rest of the bridging program, it is generally preferable to plan the PK study with concurrent enrollment of subjects from both ethnic groups.

#### 10.15.6.3 Subject Population

***Definition of “Japanese”*** Intuitively, defining “Japanese” seems like a simple issue, but it is not. Every ethnic group contains a certain amount of heterogeneity.

Japan is no exception, that is, sumo wrestlers and geisha may both have unquestioned Japanese genealogy. Ancient migration patterns produced different subpopulations in regions spreading from the Kyushu District (Okinawa and other islands in the south) to Hokkaido (northern islands), and immigration within the past century or so has introduced many Korean, Chinese, and others to Japan. How many generations does it take for descendents of such individuals to become “Japanese?” Conversely, if an individual is born and raised in Japan, with a pedigree of pure Japanese parents, grandparents, and great-grandparents, does he or she cease to be “Japanese” as soon as he leaves the confines of Japan’s airspace?

Even more debatable is how many generations can pass after leaving Japan before an individual is no longer considered “Japanese.” Generational terms for Japanese are as follows, each assuming that the pedigree is pure Japanese ethnicity:

- Issei (first generation, subject was born in Japan)
- Nissei (second generation, subject born elsewhere but parents born in Japan)
- Sansei (third generation, subject and one or both parents born elsewhere but grandparents born in Japan)

My Japanese colleagues who perform phase I studies in Japan report that if a subject in Japan chooses to participate in a clinical pharmacology study conducted in one of their units in Japan, there is relatively little requirement to document Japanese ethnicity. Yet for many studies, if that same subject wishes to participate in a clinical study in Honolulu, we must document Japanese birth and citizenship by passport or visa, and fill a three-generation genealogy tree documenting that both parents and all four grandparents are Japanese and were born in Japan.

The stricter standard for “foreign data” is understandable, in that “Japanese” citizens born, raised, and living in Japan represent the population of interest to the MHLW—regardless of exact ethnic pedigree or regional origin. Subjects out of Japan may or may not represent their population, and it is important to know that the subjects enrolled as “Japanese” meet acceptable criteria. Consequently, virtually all bridging studies performed outside of Japan involving Japanese subjects require at least a “four grandparents” rule, that all four grandparents are known to be of Japanese ethnicity. This allows Issei, Nissei, and Sansei participants, but not those of mixed Japanese/non-Japanese descent.

Depending on the goals of the study, some protocols limit Japanese enrollment to Issei (with documentation of a three-generation genealogy chart), sometimes with a requirement that the subject has been outside of Japan for a limited time, for example, no longer than 5 or 10 years. Additionally, many protocols also inquire regarding dietary patterns and lifestyle, discussed in further detail below.

**Definition of “Caucasian”** Likewise, defining “Caucasian” may seem to be simple, but it is not. When we have asked sponsors for their definitions, responses vary considerably, including one to the effect “anyone with blonde hair, light skin, and blue or green eyes!” Fortunately, most sponsors include more geographically based definitions. Of course, many non-Caucasians are born or raised in Europe. However, without quibbling about ancient migration patterns or attempts to nar-

rowly define genetic origins, we deal with a reasonable sense of what it means to be of “European” origin genetically, that is, the people characteristically and traditionally known as English, German, French, Italian, Russian, and so forth. Specific protocols may or may not include those of the Middle East or North African descent, partially depending on the population in which the majority of clinical pharmacology data were originally generated: If performed in Europe the definition of “Caucasian” tends to reflect the participants of the original studies; if performed in the United States the definition of “Caucasian” is generally more broadly defined.

The Guidance for Industry Regarding Collection of Race and Ethnicity Data in Clinical Trials published by the U.S. FDA, January 2003, includes guidance for clinical trials conducted outside of the United States and recognizes racial designations of:

- American Indian or Alaska Native
- Asian
- Black, of African heritage
- Native Hawaiian or Other Pacific Islander
- White

The guidance also includes ethnic categories of “Hispanic or Latino” or “Not Hispanic or Latino.” Per this guidance document, the racial category “white” can reflect origins in Europe, the Middle East, or North Africa, and the category “Asian” can reflect origins from areas ranging from India to Japan.

**Healthy Normal Versus Patient Population** Most bridging studies follow the model established in the original jurisdiction and are most commonly performed in normal healthy volunteers. The exception usually involves medications with significant potential toxicities, with ethical constraints limiting exposure in healthy individuals who will not be receiving any direct medical benefit. Precedents exist for performing PK studies in patients as well as healthy volunteers, for measuring the PK profile in patients as part of a study to demonstrate efficacy, and for confirming the PK in patient studies as well as performing a PK study in healthy volunteers.

**Male and Female Participants** Unless the subject compound was intended exclusively for a female population (e.g., oral contraceptives), pharmacokinetic studies were traditionally performed in healthy male volunteers in all jurisdictions. Current regulatory guidance favors inclusion of female subjects, as long as there are appropriate measures to minimize the risk of in utero fetal exposure to an investigational drug. Bridging studies follow this same trend.

“Adequate” contraceptive measures vary with the protocol. Oral or injected hormonal contraception is generally allowed except when concerns exist regarding direct or indirect drug interactions. Requiring surgical sterilization or postmenopausal status often introduces an age difference between the male and female participants. Intrauterine devices are not commonly used. Abstinence or various forms of “double-barrier” contraception are increasingly allowed. Regardless of the method of conception, enrollment of women of child-bearing potential requires negative tests for pregnancy at the initial screening visit and again shortly prior to dose administration.

**Age** What is the difference between an 18-year old and a 20-year old? Not much, from a clinical pharmacology perspective. But this is an important distinction from a regulatory standpoint. In most of the United States, an 18-year old is legally an adult and has the autonomy and legal authority to sign his or her own informed consent. In Japan, this occurs at age 20. Consequently, many bridging studies define the lower age limit as 20 rather than 18.

The upper age limit is increasingly flexible. Through the 1990s many studies used age 45 as the upper limit for normal healthy studies. As the average population age increases—a trend occurring in Japan as well as the United States—there is growing interest in the inclusion of more mature participants, particularly for compounds whose clinical applications will largely apply to an elderly population.

Bridging studies will typically follow the trends established in the original jurisdiction, and more commonly include individuals up to age 50 or 55 in their “normal healthy” population, and sometimes to age 65. The trend established in the original jurisdiction is also followed for compounds that will be almost exclusively used by an elderly population, and bridging studies have been performed with a lower age limit of 50.

At the other end of the spectrum, pediatric trials have their own unique issues, including the requirement for the parent/guardian to sign the informed consent and for the participant to sign an age appropriate document indicating his or her assent. Bridging studies may also include a pediatric population. As of 2003 two compounds for pediatrics use were approved in Japan utilizing bridging studies: Palivizumab (for the treatment of respiratory syncytial virus in pediatrics) and oseltamivir phosphate for pediatric influenza infection. The pediatrics approval of oseltamivir followed the approval for use in adults at higher doses.

**Weight** Historically, the Japanese population has been smaller and lighter than the American population, and weight differences account for some of the differences in typical doses administered. Ideally, bridging studies will reflect the populations in both jurisdictions, and yet still provide comparable data. In a study enrolling concurrent Japanese and Caucasian participants, we ordinarily see a reasonable range for body mass index (BMI) applicable to both populations. Perceptions of appropriate body weight vary: In Japan a BMI over 25 is considered obese, whereas in the United States obesity is a BMI over 30. To better represent the Japanese population, the lower limit is often around 18, especially if the study enrolls female participants. To better represent the Caucasian population, the upper limit is often 27 or 29. Within these ranges, weight-adjusted pharmacokinetics can be determined.

#### 10.15.6.4 Diet and Food Effects

**Japanese Versus Western Diet** Paradoxically, as globalization progresses, diets worldwide are both increasingly diverse and increasingly similar. On a short tour of a major city in Japan, one observes Japanese consuming coffee, hamburgers and fries from American-based “fast food” outlets, and Americans have a growing taste for sushi, miso soup, and saimin. Both populations enjoy pizza and pasta.

Despite this globalization, we have discussed the significant differences between the traditional Japanese diet and the traditional American diet, and many generali-

ties still apply when we look at the overall population. To illustrate, I believe most international travelers would agree that menu choices in a typical local restaurant in Japan look quite different from menu choices in a typical local restaurant in America!

Do these differences affect a compound's pharmacokinetics? Data discussed previously regarding diet effects on metabolizing enzymes and transport proteins indicates it is prudent to assume "yes" until proven otherwise. Such effects may be related to specific diet contents and effects on pharmacologically related enzymes, beyond the usual fed versus fasted effects.

For this reason, establishing an appropriate diet for bridging study requires considerable attention. If differences exist, are they dietary or genetic? We have seen some studies structured where the Caucasians consume a typical "Western" diet, to match the conditions of the original studies, and the Japanese participants consume a more typical "Japanese" diet to match the conditions expected in Japan. Other studies require that all subjects consume the same diet, which may be Western, Japanese, or selected from common denominator food items shared by both cultures. Priorities vary with the compound, any known food effects, and the mechanics of absorption.

If a specific diet is not defined in the protocol, it may be prudent to provide a sample menu along with other study data when submitted for approval.

**Fed/Fasted Studies** Taking a medication with a meal may have no effect or may increase or decrease absorption. For compliance and optimal labeling, this is an important question for orally administered medications.

At this time, the OPSR has not defined a Japanese equivalent of the FDA's standard "high-fat" breakfast to evaluate food effects. Although other specific diet items may be of scientific interest, when it comes to regulatory submission, a high-fat meal conforming to the FDA's guidance is the place to start when assessing the effects of taking a medication fed versus fasted.

#### 10.15.6.5 Restrictions

The approach to study restrictions varies with the compound, the intended clinical application, and the conditions under which data were generated in the original jurisdiction. Often, sponsors will retain certain study restrictions from the original PK studies when they are planning bridging studies so the data will be more easily compared, even though subsequent experience has shown that the original design was overly conservative.

Common restrictions include the following examples.

**Concomitant Medications** Unless a drug interaction is the goal of the study, concomitant medications are usually restricted for anywhere from 1 to 2 weeks, or up to 5 half lives, whichever is longer. Medications known to inhibit or induce the metabolic enzyme pathways of the study compound are generally prohibited for a longer period of time. Both over-the-counter and prescription medications are commonly cited, and usually herbal products and nutritional supplements are limited or prohibited. Certain exceptions often apply, such as acetaminophen or continued use of oral contraceptives.

**Smoking and Use of Tobacco Products** Nicotine may not have a direct effect on the metabolism of the study drug but is well known to affect blood pressure and heart rate, and the occurrence of nicotine withdrawals during a clinical trial skews the adverse events profile. However, smoking up to 10 cigarettes a day is still quite common in a young Japanese population, and the exclusion of smokers may cause recruitment delays.

A compromise measure is to limit cigarette consumption to a total of 10 cigarettes a day and to prohibit smoking for 2 hours before vital signs or electrocardiograms are obtained.

**Caffeine/Xanthine Products** Intake of caffeinated food and beverages is nearly ubiquitous in both Japanese and Caucasian participants. Again, there may or may not be direct effects on the ADME profile, but there may be effects on urine output and vital signs, and caffeine withdrawal headaches and related symptoms affect the adverse events. Additionally, different forms and quantities consumed by different subjects adds another variable that could complicate data analysis.

For these reasons, use of xanthine-containing products is often excluded during a confinement period, and typically for 48–72 hours prior to the first dose and sometimes through the entire course of the study, even if there are outpatient visits following the confinement. However, use is so common and xanthines are contained in so many different products that it is difficult to get through a large study, especially with cross-over periods, multiple confinements, or multiple follow-up exams, without one or more subjects violating these conditions. So as to not have an unacceptable number of protocol deviations, or lose subjects prior to study completion, it is prudent to limit these restrictions or include language in the protocol such that a medical monitor may allow participation if the quantity consumed is deemed too small to have a clinically or pharmacologically significant effect.

**Exercise** At least in Hawaii, most young healthy study participants are quite active physically. While most medical authorities support an individual program of appropriate regular aerobic exercise, vigorous exercise while enrolled in a clinical pharmacology study introduces variables in cardiac output, renal and hepatic blood flow, hydration, and other circulation changes that can affect ADME data [54]. These are especially problematic if there are marked subject-to-subject differences in the intensity or duration of the exercise, especially if the subjects baseline level of fitness are dissimilar. Vigorous exercise can also cause elevation of creatinine phosphokinase (CPK), aspartate aminotransferase (AST), and other enzymes, which may raise unnecessary questions about the compounds potential to cause myositis, rhabdomyolysis, or cardiac or hepatic toxicity.

For these and related reasons, most bridging studies collecting ADME data limit or prohibit vigorous exercise during critical portions of the study conduct, and often for 48–72 hours before the first dose. When dealing with a young healthy population, it is also advisable to obtain a chemistry panel including CPK and AST at admission or prior to dose in addition to screening, to clarify exercise versus pharmacological effects if elevations are observed postdose exposure.



## 10.15.7 BRIDGING STUDY CONDUCT

### 10.15.7.1 Recruitment

If the study unit is located in an area with a large Caucasian population, recruitment of Caucasian subjects is about the same as recruiting for any other phase I study. Or sometimes easier, if the study compound is already approved by the FDA or there is a large body of safety data in Caucasians with dose exposure similar to the study of interest and the risk of significant adverse events is small.

Recruiting Caucasian subjects in Japan is limited by the available local Caucasian population, and performing bridging studies in Japan reflect the timelines and cost structure for other clinical pharmacology studies.

Recruiting Japanese participants outside of Japan is primarily limited by the local population of eligible Japanese subjects. Yet a sizeable population alone does not necessarily assure that they are willing to be study participants. Study participation is still viewed with considerable suspicion by many Japanese. When recruiting Issei participants, cultural biases exist against exposure to medications with unknown risks, as well as known adverse events even if the risks are small. Overcoming this bias requires establishing trust, credibility, and considerable discussion, a process usually facilitated with the assistance of Issei Japanese staff members and prior study participants.

As with any “special population” study, the narrower the population is defined the more difficult it is to recruit. Limiting the Japanese participants to Issei (versus four grandparents, allowing Sansei participants), restricting BMI to less than 25, prohibiting smokers, and restricting the upper age to 45 narrows the pool of eligible candidates, especially if the Issei subjects have a time restriction on how long they have been out of Japan. Inclusion/exclusion criteria allowing both male and female participants, but requiring the females to be Issei, out of Japan less than 5 years, under age 45, and of non-child-bearing potential means that for practical purposes, the study will only enroll male subjects.

“Matching” Japanese and Caucasian subjects by one or more criteria also complicates the recruitment process. Matching subjects by sex, weight, or age may increase the confidence in pharmacokinetic analysis but must be weighed against the greater time and expense required for recruitment.

### 10.15.7.2 Informed Consent

ICH and GCP standards require that the consent be truly “informed,” with the subject’s signature only after adequate discussion whereby the study staff is confident the subject truly understands the nature of the study, including its attendant risks, goals, benefits, procedures, and restrictions. The major concern with bridging study is language: If recruiting Sansei subjects who grew up with English as their primary language (for studies performed at American centers), obtaining informed consent is much the same as for English-speaking Caucasian subjects.

However, if the protocol requires Issei subjects, especially with a fairly short time restriction out of Japan, one cannot necessarily assume that a discussion in English and use of an English language consent is adequate. While Issei candidates may be conversationally fluent in English, the medical and legal terms in many consent

documents are not necessarily familiar to them, yet they may not want to admit they are uncertain about the meaning. At times, I have found this to be the case even when dealing with university students who attend classes conducted in English and read textbooks written in English. As mentioned earlier, informed consents are not a familiar item in Japanese culture, and signing a document listing possible risks in detail is uncomfortable for many Japanese.

For these reasons, it is usually advantageous to provide an informed consent document in Japanese, and utilize staff fluent in Japanese to facilitate the discussion. This means obtaining approval and maintaining two versions of the consent: one in English and one in Japanese. The institutional review board (IRB) must approve the Japanese translation and often require that the translation be performed by their own staff. Japanese sponsors often request that this translation be submitted to them for their review and approval before it is utilized. Maintaining dual-language informed consents adds time and cost, and needs to be considered in advance.

### **10.15.7.3 GCP Practices**

Good clinical practices are well specified elsewhere, especially ICH Section E6, and will not be discussed here other than to note that bridging studies must be compliant with these guidelines. Were it not for these consensus guidelines, bridging studies as we know them could not even be performed. Study planning and conduct compliance with these guidelines is what allows data from bridging studies to be acceptable when submitted to the new jurisdiction. Their importance cannot be overstated.

### **10.15.7.4 Adverse Events**

Definition and documentation requirements for adverse event collection and reporting are specified in the good clinical practices. Bridging studies follow these guidelines the same as other studies. There are two particular challenges: language and culture.

As explained when discussing the informed consent process, Japanese participants may not be familiar with English medical terminology even if fluent in ordinary conversation. And, as with all subjects, they may only understand and explain their symptoms using informal, colloquial terms with which they are familiar. If Issei subjects are enrolled, it is advisable to utilize staff fluent in Japanese to inquire regarding adverse events and translate their results.

Related to this, participant's use of concomitant medications from Japan often requires a search for drug names or components that can be coded for Western regulatory agencies.

Japanese culture generally downplays spontaneous adverse event reporting, as many feel that there is some level of shame associated with experiencing an adverse event. This makes active inquiry even more important. At a minimum, the "expected" adverse events would have been listed in the informed consent, and subjects should be made to feel comfortable describing these should they occur. A private environment with supportive staff fluent in Japanese should inquire at regular intervals regarding adverse events in general, make sure they are adequately characterized and "translated" into codeable terms, and continue follow-up until resolved.

### 10.15.7.5 Documentation

Documentation again follows GCP guidelines, generally utilizing source documents and separate case report forms (CRFs). The standard approach to “document everything” is especially important for bridging studies. Free-form progress notes and explanatory notes to file are important items even with excellent source documents, particularly for any protocol deviations or events out of the norm.

American study centers will generally utilize English language source documents. It is advisable to capture data using metric units for weights, measures, and temperature as much as possible. CRFs may be written or electronic. They will be in English if the study center uses English-speaking staff for data entry, but good CRFs will have already defined how the data will be coded in Japanese for their own NDA submission.

### 10.15.7.6 Monitoring and Audits

As the pharmaceutical industry becomes increasingly globalized, monitoring and audits are often shared between staff in Japan, the United States, and the European Union. These processes also follow GCPs. Adequate monitoring and detailed audits, with observation of study conduct and data verification, is a feature distinguishing ICH-compliant studies from clinical trials performed prior to ICH implementation.

Because bridging studies are pivotal by nature, most are audited, and may well be audited by more than one team. Bridging studies are also subject to audits from the regulatory agencies in both jurisdictions.

## 10.15.8 BRIDGING STUDY EXPERIENCE

Since the E5 guidelines were adopted in 1998, industry has responded by implementing bridging strategies and submitting NDAs utilizing bridging data. Starting with 3.2% in 1999, use of bridging studies increased steadily. In 2003, 25% of NDAs approved in Japan were based on a bridging strategy. Dr. Kato reported at the ICH Sixth International Conference on Harmonization (Osaka, Japan, November 2003) that at least 67 submissions had involved the application of the E5 guideline. Of course, this number does not include the bridging studies performed from 1998 to 2003 for development programs for which the NDA had not been submitted.

The experience of these early bridging studies remains instructive. We will cite examples of success and illustrate challenges encountered.

### 10.15.8.1 Examples of Success

The first bridging study for which I was the principal investigator was performed in July 1998, shortly after the E5 guidelines were adopted. We studied the single- and multiple-dose pharmacokinetic profiles of oseltamivir phosphate (Tamiflu), an antiviral product to treat acute influenza infection, in 24 healthy young males—12 Japanese and 12 Caucasians. The Japanese were all Issei. The data showed that

pharmacokinetics were similar in both populations and no safety issues were observed. These data allowed the sponsor to implement a clinical trial program in Japan to demonstrate safety and efficacy during the 1998–1999 and 1999–2000 influenza seasons, and this product was approved in Japan in 2000 with the bridging study allowing use of safety and efficacy data generated in the United States and the European Union coupled with data from the studies in Japan. Early approval and introduction into clinical use has potentially saved many lives in Japan, which would have been otherwise lost to influenza. Early approval would not have been possible without the use of bridging data.

A recent review selected 26 total NDAs approved in Japan from 1999 to 2003 based on a bridging strategy. This list only included new drugs. Target clinical domains vary considerably and include cancer, viral infections, diabetes, osteoporosis, rheumatoid arthritis, migraine headache, age-related macular degeneration, glaucoma, Parkinson's disease, renal transplant rejection, allergic rhinitis, erectile dysfunction, and *Helicobacter pylori* eradication. Six of these compounds start at lower doses or limit the highest dose. Others have differences in labeling. Examples have been cited previously in this chapter.

The bridging strategy for these compounds typically included an ADME study coupled with one or more phase II/III safety and efficacy studies performed in Japan, with their design based on foreign data and the results of the ADME study. The ADME study enrolled healthy subjects in 12 of these 26 cases, healthy subjects plus patients in 4 cases, and 5 cases used an independent ADME study plus measuring the PK in their efficacy trials.

The bridging strategy “worked” in terms of the goal to save time and money. Measuring the clinical development period from the time when a phase II/III efficacy and safety study began in Japan to the time of NDA submission, the median time for NDAs using a bridging strategy was 32 months versus 56 months for compounds without a bridging strategy, a savings of 24 months [55].

### 10.15.8.2 Challenges

Discussion at the ICH Sixth International Conference on Harmonisation also included some of the challenges and difficulties encountered with some bridging programs. All concerned recognized there was some misunderstanding and confusion regarding the E5 guidelines. Attempts to clarify included a Questions and Answers document approved in November 2003, based on actual experience.

In short, issues included differences of opinion between industry and the MHLW on what constituted intrinsic and extrinsic differences, assessment of a compound's “ethnic sensitivity,” how to interpret GCP compliance, definitions of disease and medical practice in defining patient populations, criteria for evaluating risk–benefit ratio if disease incidence varied between regions, labeling for multiple indications once bridging established efficacy for the primary indication, criteria for determining if foreign data met regulatory requirements for acceptability, and mutually acceptable surrogate endpoints when assessing efficacy. Still, industry felt that the E5 guidance overall saved them time and money in their development program.

### 10.15.9 SUMMARY AND FUTURE DIRECTIONS

Again we ask the question: “Do bridging studies work?” The examples provided illustrate that the answer to this is “yes,” that bridging studies have allowed use of existing data with an abbreviated and streamlined drug development program in the new jurisdiction, saving both time and money in NDA approval.

Do bridging studies always work? No, not all have produced the expected outcome for the resources invested.

Let us come back to the question posed at the beginning of this chapter: Are “ethnic differences” real?

My answer to this is “yes,” at least for certain compounds. Some compounds have similar properties across all ethnic groups studied, and individual intraethnic variation is greater than interethnic variation. Yet for other compounds, we see characteristic pharmacokinetic or pharmacodynamic differences in one ethnic group that are quite distinct from our observations in other ethnic groups. In some cases, we have defined one or more mechanisms that account for these findings. In other cases, there are multiple mechanisms, or we do not have a clear explanation for all of our findings. We may not be able to distinguish the particular contribution of genetics versus diet, environment, lifestyle, medical practice, or the interaction of these factors within a region, all cumulatively considered “ethnic differences.” Yet we know that certain medications need to be used differently in one region than another.

As we continue our investigations, it is prudent to respect the unique culture and traditions of each jurisdiction. While we are answering scientific questions, it is prudent to design our studies with awareness of the historical, social, political, and medical context affecting the respective regulatory agencies.

### REFERENCES

1. ICH Efficacy Topic ES (1998), U.S. *Federal Register*, June.
2. Kimura, R. (1995), History of medical ethics: contemporary Japan, *Encyclopedia of Bioethics*, rev. ed., edited by W. T. Reich, pp. 1496–1505.
3. FDA Backgrounder (2002), Milestones in U.S. Food and Drug Law History, updated Aug 5; [www.FDA.gov/opacom/backgrounders/miles.html](http://www.FDA.gov/opacom/backgrounders/miles.html).
4. Kim, R. B., Yamazaki, H., Chiba, K., et al. (1996), In vivo and in vitro characterization of CYP2E1 activity in Japanese and Caucasians, *J. Pharmacol. Exp. Ther.*, 279(1), 4–11.
5. Shu, Y., Cheng, Z. N., Liu, Z. Q., et al. (2001), Interindividual variations in levels and activities on cytochrome P-450 in liver microsomes of Chinese subjects, *Acta Pharmacol. Sin.*, 22(3), 283–288.
6. Mizutani, T. (2003), PM frequencies of major CYPs in Asians and Caucasians, *Drug Metab. Rev.*, 35(2–3), 99–106.
7. Solus, J. F., Arietta, B. J., Harris, J. R., et al. (2004), Genetic variation in eleven phase I drug metabolism genes in an ethnically diverse population, *Pharmacogenomics*, 5(7), 895–931.
8. Inoue, K., Yamazaki, H., and Shimada, T. (2000), Characterization of liver microsomal 7-ethoxycoumarin O-deethylation and chlorzaxozone 6-hydroxylation activities in

- Japanese and Caucasian subjects genotyped for CYP2E1 gene, *Arch. Toxicol.*, 74(7), 372–378.
- 8a. Inoue, K., Asao, T., and Shimada, T. (2000), Ethnic-related differences in the frequency distribution of genetic polymorphisms in the CYP1A1 and CYP1B1 genes in Japanese and Caucasian populations, *Xenobiotica*, 30(3), 285–295.
  9. Ham, X. M., Chen, S. P., Wu, Q. N., et al. (2000), G-2964A and C734A genetic polymorphisms of CYP1A2 in Chinese population, *Acta Pharmacol. Sin.*, 21(11), 1031–1034.
  10. Aitchison, K. J., Gonzalez, F. J., Quattrochi, L. C., et al. (2000), Identification of novel polymorphisms in the 5' flanking region of CYP1A2, characterization of interethnic variability, and investigation of their functional significance, *Pharmacogenetics*, 10(8), 695–704.
  11. Bartoli, A., Xiaodong, S., Gatti, G., et al. (1996), The influence of ethnic factors and gender on CYP1A2-mediated drug disposition: A comparative study in Caucasian and Chinese subjects using phenacetin as a marker substrate, *Ther. Drug Monit.*, 18(5), 586–591.
  12. Inoue, K., Yamazaki, H., Imiya, K., et al. (1997), Relationship between CYP2C9 and 2C19 genotypes and tolbutamide methyl hydroxylation and S-mephenytoin 4'-hydroxylation activities in livers of Japanese and Caucasian populations, *Pharmacogenetics*, 7(2), 103–113.
  13. Caro, A. A., and Cederbaum, A. I. (2005), Inhibition of CYP2E1 catalytic activity in vitro by S-adenosyl-L-methionine, *Biochem. Pharmacol.*, 69(7), 1081–1093.
  14. Kessova, I., and Cederbaum, A. I. (2003), CYP2E1: Biochemistry, toxicology, regulation and function in ethanol-induced liver injury, *Curr. Mol. Med.*, 3(6), 509–518.
  15. Le Marchand, L., Donlon, T., Seifried, A., et al. (2002), Red meat intake, CYP2E1 genetic polymorphisms, and colorectal cancer risk, *Cancer Epidemiol. Biomarkers Prev.*, 11(10Pt1), 1019–1024.
  16. Zhou, S., Yung Chan, S., Gher Goh, B., et al. (2005), Mechanism-based inhibition of cytochrome P450 3A4 by therapeutic drugs, *Clin. Pharmacokinet.*, 44(3), 279–304.
  17. He, P., Court, M. H., Greenblatt, D. J., et al. (2005), Genotype-phenotype associations of cytochrome P450 3A4 and 3A5 polymorphism with midazolam clearance in vivo, *Clin. Pharmacol. Ther.*, 77(5), 373–387.
  18. Eap, C. B., Buclin, T., Hustert, E., et al. (2004), Pharmacokinetics of midazolam in CYP3A4- and CYP3A5-genotyped subjects, *Eur. J. Clin. Pharmacol.*, 60(4), 231–236.
  19. Lamba, J. K., Lin, Y. S., Schuetz, E. G., et al. (2002), Genetic contribution to variable human CYP3A-mediated metabolism, *Adv. Drug Deliv. Rev.*, 54(10), 1271–1294.
  20. Harris, R. Z., Jang, G. R., and Tsunoda, S. (2003), Dietary effects on drug metabolism and transport, *Clin. Pharmacokinetic*, 42(13), 1071–1088.
  21. Yoshida, A. (1992), Molecular genetics of human aldehyde dehydrogenase, *Pharmacogenetics*, 2(4), 139–147.
  22. Pande, J. N., Pande, A., and Singh, S. P. (2003), Acetylator status, drug metabolism and disease, *Natl. Med. J. India*, 16(1), 24–26.
  23. Sekine, A., Saito, S., Iida, A., et al. (2001), Identification of single-nucleotide polymorphisms (SNPs) of human N-acetyltransferase genes NAT1, NAT2, AANAT, ARD1 and L1CAM in the Japanese population, *J. Human Genet.*, 46(6), 314–319.
  24. Krynetski, E. Y., and Evans, W. E. (2000), Genetic polymorphism of thiopurine S-methyltransferase: Molecular mechanisms and clinical importance, *Pharmacology*, 61(3), 136–146.



25. Mackenzie, P. I., Miners, J. O., and McKinnon, R. A. (2000), Polymorphisms in UDP glucyronosyltransferase genes: Functional consequences and clinical relevance, *Clin. Chem. Lab. Med.*, 38(9), 889–892.
26. Bailey, D. G., Malcom, J., Arnold, O., et al. (1998), Grapefruit juice-drug interactions, *Br. J. Clin. Pharmacol.*, 46(2), 101–110.
27. Zhou, S., Lim, L. Y., and Chowbay, B. (2004), Herbal modulation of P-glycoprotein, *Drug Metab. Rev.*, 36(1), 57–104.
28. Izzo, A. A. (2004), Drug interactions with St. John's Wort (*Hypericum perforatum*): A review of the clinical evidence, *Int. J. Clin. Pharmacol. Ther.*, 42(3), 139–148.
29. Zou, L., Harkey, M. R., and Henderson, G. L. (2002), Effects of herbal components on cDNA-expressed cytochrome P450 enzyme catalytic activity, *Life Sci.*, 71(13), 1579–1589.
30. Foster, B. C., Vandenhoek, S., Hanna, J., et al. (2003), In vitro inhibition of human cytochrome P450-mediated metabolism of marker substrates by natural products, *Phytomedicine*, 10(4), 334–342.
31. Unger, M., and Frank, A. (2004), Simultaneous determination of the inhibitory potency of herbal extracts on the activity of six major cytochrome P450 enzymes using liquid chromatography/mass spectrometry and automated online extraction, *Rapid Commun. Mass Spectrom.*, 18(19), 2273–2281.
32. Stupans, L., Tan, H. W., Kirlich, A., et al. (2002), Inhibition of CYP3A-mediated oxidation in human hepatic microsomes by the dietary derived complex phenol, gallic acid, *J. Pharm. Pharmacol.*, 54(2), 269–275.
33. Hu, Z., Yang, X., Ho, P. C., et al. (2005), Herb-drug interactions: A literature review, *Drugs*, 65(9), 1239–1282.
34. Dresser, G. K., Bailey, D. G., Leake, B. F., et al. (2002), Fruit juices inhibit organic anion transporting polypeptide-mediated drug uptake to decrease the oral availability of fexofenadine, *Clin. Pharmacol. Ther.*, 71(1), 11–20.
35. Wang, X., Wolkoff, A. W., and Morris, M. E. (2005), Flavonoids as a novel class of human organic anion transporting polypeptide OATP1B1 (OATP-C) modulators, *Drug Metab. Dispos.*, 33(11), 1666–1672.
36. Nabekura, T., Kamiyama, S., and Kitagawa, S. (2005), Effects of dietary chemopreventive phytochemicals on P-glycoprotein function, *Biochem. Biophys. Res. Commun.*, 327(3), 866–870.
37. Nguyen, H., Zhang, S., and Morris, M. E. (2003), Effect of flavonoids on MRP1-mediated transport in Panc-1 cells, *J. Pharm. Sci.*, 92(2), 250–257.
38. Zhang, S., and Morris, M. E. (2003), Effects of the flavonoids biochanin A, morin, phloretin, and silymarin on P-glycoprotein-mediated transport, *J. Pharmacol. Exp. Ther.*, 304(3), 1258–1267.
39. Yoshida, N., Takagi, A., Katagawa, H., et al. (2005), Inhibition of P-glycoprotein-mediated transport by extracts and monoterpenoids contained in *Zanthoxyli fructus*, *Toxicol. Appl. Pharmacol.*, 209(2), 167–173.
40. Wilkinson, G. R. (1997), The effects of diet, aging and disease-states on presystemic elimination and oral drug bioavailability in humans, *Adv. Drug Deliv. Rev.*, 27(2–3), 129–159.
41. Zhao, F., Loke, C., Rankin, S. C., et al. (2004), Novel CYP 2C9 genetic variants in Asian subjects and their influence on maintenance warfarin dose, *Clin. Pharmacol. Ther.*, 76(3), 210–219.
42. Takahashi, H., Wilkinson, G. R., Caraco, Y., et al. (2003), Population differences in S-warfarin metabolism between CYP2C9 genotype-matched Caucasian and Japanese patients, *Clin. Pharmacol. Ther.*, 73(3), 253–263.



43. Maxwell, T. J., Ameyaw, M. M., Pritchard, S., et al. (2005), Beta-2 adrenergic receptor genotypes and haplotypes in different ethnic groups, *Int. J. Mol. Med.*, 16(4), 573–580.
44. Xie, H. G., Kim, R. B., Wood, A. J., et al. (2001), Molecular basis of ethnic differences in drug disposition and response, *Annu. Rev. Pharmacol. Toxicol.*, 41, 815–850.
45. Xie, H. G., Stein, C. M., Kim, R. B., et al. (1999), Frequency of functionally important beta-2 adrenoceptor polymorphisms varies markedly among Africa-American, Caucasian and Chinese individuals, *Pharmacogenetics*, 9(4), 511–516.
46. Zhou, H. H., Koshakji, R. P., Silberstein, D. J., et al. (1989), Altered sensitivity to and clearance of propranolol in men of Chinese descent as compared with American whites, *N. Engl. J. Med.*, 320(9), 565–570.
47. Williams, R. B., Marchuk, D. A., Gadde, K. M., et al. (2003), Serotonin-related gene polymorphisms and central nervous system serotonin function, *Neuropsychopharmacology*, 28(3), 533–541.
48. Crowley, J. J., Oslin, D. W., Patkar, A. A., et al. (2003), A genetic association study of the mu receptor and severe opioid dependence, *Psychiatr. Genet.*, 13(3), 169–173.
49. Klepstad, P., Dale, O., Skorpen, F., et al. (2005), Genetic variability and clinical efficacy of morphine, *Acta Anaesthesiol. Scand.*, 49(7), 902–908.
50. Ide, S., Kobayashi, H., Tanaka, K., et al. (2004), Gene polymorphisms of the mu opioid receptor in methamphetamine abusers, *Ann. N. Y. Acad. Sci.*, 1025, 316–324.
51. Tan, E. C., Tan, C. H., Karupathivan, U., et al. (2003), Mu opioid receptor gene polymorphisms and heroin dependence in Asian populations, *Neuroreport*, 14(4), 569–572.
52. Evans, D. A., McLeod, H. L., Pritchard, S., et al. (2001), Interethnic variability in human drug responses, *Drug Metab. Dispos.*, 29(4 Pt 2), 606–610.
53. van Gerven, J. M., Uchida, E., Uchida, N., et al. (1998), Pharmacodynamics and pharmacokinetics of a single oral dose of nitrazepam in healthy volunteers: An interethnic comparative study between Japanese and European volunteers, *J. Clin. Pharmacol.*, 38(12), 1129–1136.
54. Lenz, T. L., Lenz, N. J., and Faulkner, M. A. (2004), Potential interactions between exercise and drug therapy, *Sports Med.*, 34(5), 293–306.
55. Uyama, Y., Shibata, T., Nagai, N., et al. (2005), Successful bridging strategy based on ICH E5 guideline for drugs approved in Japan, *Clin. Pharmacol. Ther.*, 78(2), 102–113.

# 10.16

---

## *Brief History of Clinical Trials on Viral Vaccines*

**Megan J. Brooks,<sup>1</sup> Joseph J. Sasadeusz,<sup>1</sup> and Gregory A. Tannock<sup>2</sup>**

<sup>1</sup>*Victorian Infectious Diseases Service, Centre for Clinical Research Excellence in Infectious Diseases, The Royal Melbourne Hospital, Parkville, Victoria, Australia*

<sup>2</sup>*Department of Biotechnology and Environmental Biology, RMIT University, Bundoora, Victoria, Australia*

### **Contents**

|                                                                                                                                 |     |
|---------------------------------------------------------------------------------------------------------------------------------|-----|
| 10.16.1 Introduction                                                                                                            | 769 |
| 10.16.2 Recent Trends in Human Vaccine Manufacture                                                                              | 770 |
| 10.16.3 Events over 60 Years that Have Contributed to Changes to Regulatory Environment for Development of Human Viral Vaccines | 771 |
| 10.16.4 Expanded Programs for Use of Vaccines in Developing Countries                                                           | 774 |
| 10.16.5 Clinical Studies Necessary for Development of New Vaccines                                                              | 775 |
| 10.16.6 The Future                                                                                                              | 776 |
| References                                                                                                                      | 777 |

### **10.16.1 INTRODUCTION**

Although our recognition of viruses as the etiologic agents of infectious disease only dates from the early years of the twentieth century, it is interesting to note that vaccines against smallpox and rabies were introduced much earlier. In the case of smallpox the procedure referred to as *variolation*, in which skin lesions from infected individuals were administered to those without lesions, dates back to the late Middle Ages [1]. Despite this, the observation by Jenner in 1798 that milkmaids continuously exposed to cows infected with cowpox were largely resistant to endemic

smallpox is rightly regarded as the origin of modern vaccination practice. Rabies vaccines, consisting of formolized suspensions of infected animal brain material and introduced in the late nineteenth century, were highly reactogenic but were used as recently as the 1960s when they were replaced by vaccines prepared from viruses grown in cell culture [2].

Most viral vaccines in use today were developed as a consequence of our ability to cultivate viruses in embryonated eggs (from the 1930s) or cell cultures (the 1950s and 1960s) and so fulfill the second of Koch's postulates for establishment of the etiology of an infectious disease. Until their discontinuation for general use in the early 1980s, vaccines against smallpox were prepared from vaccinia virus grown on the skin of calves, a procedure that would now be completely unacceptable. In vitro cultivation of bacteria was first achieved in the late nineteenth century, and many highly effective bacterial vaccines used today (notably those for the prevention of tetanus and diphtheria) were available by the 1920s. For other significant human viruses, such as hepatitis B, C, and the human papilloma viruses, the second of Koch's postulates has never been fulfilled; until the late 1980s the antigens that comprise the active ingredients of hepatitis B vaccines were actually prepared from chronic hepatitis B carriers. Fortunately, over the past 15 years hepatitis B vaccines have been prepared by the application of molecular techniques and are far more acceptable from the standpoint of safety and at the same time much less expensive. Molecular technology has allowed the realization of preventive vaccines against human papilloma viruses [3]. However, despite much effort and an urgent need, vaccines against hepatitis C (HCV) or the human immunodeficiency virus (HIV) seem much further away. For other reasons, urgently needed vaccines against herpes simplex and cytomegaloviruses, both of which can be grown in cell culture, are still unavailable, and the same could have been said until recently of rotaviruses.

### 10.16.2 RECENT TRENDS IN HUMAN VACCINE MANUFACTURE

Greater concerns about safety over the past 30 years have increased the costs of viral vaccine development to the point where all but a handful of very large multinational companies have the resources to fully comply with all the steps necessary for registration in the two major regulatory jurisdictions, the Food and Drug Administration (FDA) and the European Union. These costs have been largely driven by the willingness of U.S. courts to grant extraordinary damages against manufacturers for claims, some inspired by negligence but others due to historical gaps in our knowledge at the time of the original product registration. Examples of the latter will be given below. Until the last 20 years human viral and bacterial vaccines were prepared in Western countries by about 30–40 medium to large sized companies and a comparable number of state-run institutions or statutory corporations. The *state serum institutes* of several European countries were founded early in the twentieth century and are good examples of the latter. Their existence followed advances in microbial cultivation and the identification of blood groups around the turn of the twentieth century and the consequent development of vaccines and serum therapies, largely for use within the same country. Similar institutions had a

prominent role in the countries of the former Soviet Bloc, with their strong emphasis on immunization and public health planning. In several countries of the developing world similar institutes undertook vaccine manufacture under the aegis of colonial powers, as exemplified by the various Pasteur Institutes of some Francophone countries of Africa and Asia. The importance of these institutes in the provision of biologics was greatly enhanced by two world wars. In Australia, for example, the Commonwealth Serum Laboratories (now CSL Ltd) was instituted by the government of the day in 1917 for the express purpose of supplying smallpox vaccine for use by the Australian army in World War I in response to a likely shortfall in supply from manufacturers in the United Kingdom. CSL Ltd has greatly expanded its range of vaccines and other biologics since 1917 but is no longer a government-owned institution. As with comparable organizations, nearly all of its output, until comparatively recently, was reserved for use within Australia. The *raison d'être* for their survival over almost a century has been based primarily on perceptions of national interest!

Over the past 20–30 years most small manufacturers of human vaccines have either disappeared and/or have been subsumed by or developed strategic alliances with four giant multinational companies (*Bigpharma*). Only Bigpharma has the resources to undertake the complete development of human vaccines from early research to final clinical application, although, for many vaccines, early research relating to proof of concept is frequently undertaken by smaller start-up companies—often in association with research institutes. Because of the international nature of their operations, vaccine manufacture and accreditation of existing vaccines by Bigpharma may be carried out in more than one country. The cornerstone of modern manufacture has been the emergence of rigidly enforced international codes of good manufacturing and laboratory practice (GMP and GLP). Most countries with only limited capacity to manufacture human viral vaccines may, nevertheless, impose regulatory conditions for efficacy, safety, and standardization that are relevant to local needs. In practice, because of the sheer cost and complexity of manufacture, requirements for registration by the federal drug authority in the United States and the European Union [often with guidelines set by the World Health Organization (WHO)] remain points of reference for most national regulatory authorities.

### 10.16.3 EVENTS OVER 60 YEARS THAT HAVE CONTRIBUTED TO CHANGES TO REGULATORY ENVIRONMENT FOR DEVELOPMENT OF HUMAN VIRAL VACCINES

1. *Contamination of Yellow Fever Vaccine by Hepatitis B by Serum Preservatives (1940s)* The development of live vaccines against yellow fever and, in particular, the introduction of the egg-grown attenuated 17D strain by Theiler's group at the Rockefeller Institute just prior to World War II was a major achievement in human medicine. By April 1942 some 7 million doses had been administered to the U.S. Armed Forces alone, using vaccines prepared by the Rockefeller Institute. Yellow fever viruses are highly labile, and thermostability was conferred by the addition of human serum to the vaccine, at a time when nothing was known of the etiology of

the hepatitis B virus. By December 1942 over 50,000 cases of hepatitis and 84 deaths were recognized from infection arising from contaminated lots, representing some 2.5 million administered doses. The involvement of human serum was quickly recognized, despite the absence of definitive tests (which only became available in the 1970s), and the problem was overcome by elimination of the serum from the vaccine [4].

2. *Cutter Incident: Release of Inactivated Poliomyelitis Vaccines (IPV) Containing Residual Amounts of Live Poliovirus (1955)* Soon after the commencement of programs for the mass immunization of children in the United States, a batch of IPV, prepared by Cutter Laboratories in California, was released and subsequently shown to be responsible for poliomyelitis in 79 vaccinated children, 105 family members, and 20 community members [5]. Despite the overwhelming success of the immunization program, the *Cutter incident* and the investigations that followed provided the catalyst for the introduction of powerful national agencies for the registration of vaccines.

3. *Contamination of Early Batches of IPV by Simian Virus 40 (SV<sub>40</sub>)* Between 1955 and 1963 successive batches of IPV prepared in primary monkey kidney cultures were shown, retrospectively, to have been contaminated with SV<sub>40</sub>, a commensal of certain species of monkeys and referred to as the vacuolating agent from the cytopathic effect produced after inoculation to certain cell lines. In the United States alone, it was estimated that 10–30 million individuals were inoculated with contaminated vaccines. Because SV<sub>40</sub> virus is a member of the family Papovaviridae, members of which are oncogenic for several animal species, concerns were expressed about long-term risks to vaccine recipients. Some 30 years later and, after comprehensive epidemiological studies in the United States and Europe, the balance of evidence indicates no causal relationship between the receipt of contaminated vaccine and cancer [6].

4. *Enhancement of Pathogenicity in Infants Receiving Experimental Formalin-Inactivated Respiratory Syncytial Virus (RSV) Vaccine* Respiratory syncytial virus is the most significant respiratory pathogen of the neonatal child for which a vaccine is urgently required. In a traditional approach to control, a formalin-inactivated vaccine was administered parenterally to infants in 1968 at the Children's Hospital in Washington, D.C. Results of the study indicated that not only was the vaccine unprotective but recipients who were subsequently subjected to natural challenge experienced enhanced immunopathology in the form of Arthus-type hypersensitivity reactions [7]. These results have had profound consequences for the development of effective vaccines against RSV, and most efforts since have been directed toward the use of intranasally administered live attenuated vaccines. Experimental vaccines for use in neonatal children present unusual difficulties both in immunologic terms and from an ethical standpoint. Despite considerable progress, vaccines against RSV are still unavailable almost 40 years after the Washington study.

5. *Molecular Instability of Oral Poliomyelitis Vaccines (OPVs): Recognition of Vaccine-Associated Paralysis (1970s)* Oral poliomyelitis Vaccines have been crucial for the elimination of poliomyelitis from all but a few regions of a very small number of countries, under programs sponsored by the WHO in recent years [8].

These vaccines were developed in the early 1960s at a time virtually nothing was known as to the genetic stability of their live virus components. Molecular techniques developed over the past 20 years now indicate instability for some of these viruses following human passage, which is associated with reversion to neurovirulence [9]. If molecular evidence for instability had been available in the 1960s, OPVs would almost certainly not have been licensed and there would have been little prospect of eliminating poliomyelitis from the world some 40 years later. IPV's can only be afforded in developed countries where they have largely replaced OPVs in order to overcome concerns associated with instability.

6. *Association between Guillain-Barré Syndrome (GBS) and Receipt of Swine Influenza Vaccine (1976–1977)* In 1976 following the death of a soldier at Fort Dix, New Jersey, after infection with an H1N1 (swine) virus with apparent antigenic similarities to the agent responsible for the 1918–1919 pandemic, a decision was made to offer the entire U.S. population specially prepared monovalent vaccines. The program was curtailed after an association was recognized that resulted in the attribution of 8–10 excess cases of GBS per million during the first 6 weeks after administration of the A/swine vaccine [10]. For reasons unknown, the incidence of GBS in subsequent vaccines has been much lower, and egg-grown inactivated vaccines remain the principal public health measure for the prevention of influenza.

7. *Possible Association between Intussusception and Administration of Oral Rotavirus Vaccines (1999)* A quadrivalent reassortant simian rotavirus vaccine was licensed in 1998 for manufacture and distribution by Wyeth Laboratories and approved by the FDA after extensive testing in several countries. The vaccine was prepared for administration to infants in 3 doses at 2, 4, and 6 months. Within 12 months, subsequent evaluation of U.S. clinical data suggested a small increase in the frequency of intussusception in infants who received the vaccine, especially after the first dose [11]. Despite an urgent public health need for effective preventive measures against rotaviruses and developmental research programs extending over many years that were largely publicly funded, the vaccine was withdrawn. Fortunately, a reassortant vaccine of bovine origin and prepared by Merck was approved by the FDA in 2006, which, it is hoped, will have an impact on neonatal rotavirus infections, especially in the developing world [12].

8. *Association between Intranasal Administration of Inactivated Influenza Vaccine and Bell's Palsy (2000)* An influenza vaccine was prepared by a Swiss company (Berna Biotech) in which vaccine surface antigens were incorporated into carrier liposomes. The heat-labile *Escherichia coli* enterotoxin, a powerful mucosal adjuvant, was incorporated into the formulation. After its introduction in Switzerland, the risk of Bell's palsy was increased by a factor of 19 among those who received the vaccine intranasally. An increase did not occur in recipients who received the vaccine by parenteral injection [13]. These adverse events were not detected in prelicensure trials and the vaccine has since been withdrawn.

The list is incomplete but these events have profoundly affected the regulatory environment and have been accompanied over the same time by a vast increase in our understanding of viral replication and pathogenesis, with flow-on consequences

for product liability. Consequent advances over the period have also occurred in vaccine safety testing and in the development of models for vaccine evaluation in test populations that allow the power of a study to be estimated well in advance of its commencement. We now have a vastly increased array of standardized tests, both biological and molecular for estimating (1) the active ingredient(s) of a vaccine, (2) the presence of known adventitious contaminants, and (3) the likely toxicological consequences of its use in humans. However, the need still exists for extensive and costly clinical testing in a climate where ethical constraints on the use of volunteers for clinical trials have been greatly increased in the years immediately before and after the Declaration of Helsinki in 1964 [14]. Many trials that were conducted on earlier vaccines would be simply disallowed these days by institutional review boards, which place strong emphasis on need to provide comprehensive information on the nature of a trial to each volunteer at the point of entry. They also require to be adequately briefed concerning inducements, financial and otherwise, that are provided to volunteers. Overall, the manufacture of viral vaccines and other biologics has become commercially unattractive for all but a handful of very large manufacturers. As a consequence, the role of governments in sponsoring and meeting the costs of developing new vaccines has increased greatly.

Many highly successful vaccines for the prevention of measles, poliomyelitis, rubella, mumps, and yellow fever were developed in a much less stringent regulatory environment, and it is arguable whether many would have been registered today. The situation with OPVs has already been discussed but, to take another example, measles vaccines were first developed in the early 1960s at a time when very little was known about subacute sclerosing panencephalomyelitis (SSPE), a rare and unpredictable immunologic complication of measles associated with chronic neural infection. Effective vaccines against measles consist of live virus preparations and, if SSPE had been widely recognized at the time, the onus would have been placed on vaccine developers to prove that mass administration was not associated with SSPE. Given that there are still no satisfactory *in vitro* markers for SSPE, this would have been an impossible task and measles vaccines would probably never have become available.

Over the same period and in a very different regulatory environment, the countries of the former Soviet Bloc have made enormous and often unrecognized contributions to vaccinology. Nowhere was this more evident than in the adoption of OPV for the prevention of poliomyelitis in Western countries in the early 1960s, following the conduct of very large field trials under the direction of Smorodintsev and his colleagues in the former Leningrad [15]. The recent reemergence of childhood infectious diseases, such as diphtheria, in former Soviet countries is ample testimony to the success of former state-sponsored programs, many of which have not been maintained over the past 10–15 years. [16].

#### **10.16.4 EXPANDED PROGRAMS FOR USE OF VACCINES IN DEVELOPING COUNTRIES**

The global program sponsored by the WHO for the elimination of smallpox between the mid-1960s and the early 1980s is widely regarded as perhaps the greatest medical achievement of the twentieth century. Its success has removed the need for smallpox



vaccination—often associated with unacceptable adverse reactions—and delivered significant economic benefits, mainly to the developed world, by eliminating the need to maintain vaccination programs.

Because of the success of that program, several initiatives have commenced over the past 20 years for the elimination of pediatric viral diseases, using vaccines widely used in Western countries. These programs have been supported by philanthropic organizations, such as the Gates and Rockefeller Foundations, usually in close collaboration with the WHO. Mention has already been made of programs for the elimination of poliomyelitis by the strategic use of OPV. Another program is concerned with immunization against measles, a cause of high mortality in developing countries especially in association with malnutrition. Logistical problems in this and similar programs arise from difficulties in vaccine delivery due to poor infrastructure and problems in maintenance of the *cold chain*. Protective responses to the vaccine may be diminished by the presence of maternal antibody that is associated with high levels of measles endemicity. Here, the window of opportunity to vaccinate a child after the decline of maternal antibody but before natural exposure to the virus is often relatively narrow. It should be remembered that the use of combination measles–mumps–rubella vaccines in infants was only possible in developed countries by the elimination of endemic measles by earlier vaccination programs! Reservoirs of measles, poliomyelitis, and other preventable pediatric viruses are much more widespread, and these viruses are responsible for much higher rates of sub-clinical disease than smallpox viruses. Vaccination programs against these viruses will, therefore, need to be maintained until well after the elimination of reported clinical disease.

### 10.16.5 CLINICAL STUDIES NECESSARY FOR DEVELOPMENT OF NEW VACCINES

Figure 1 provides an indication of the cost and time scale of any program for the development and registration of a hypothetical new vaccine for the prevention of a specific viral disease. The entire process is broadly divided into four phases of

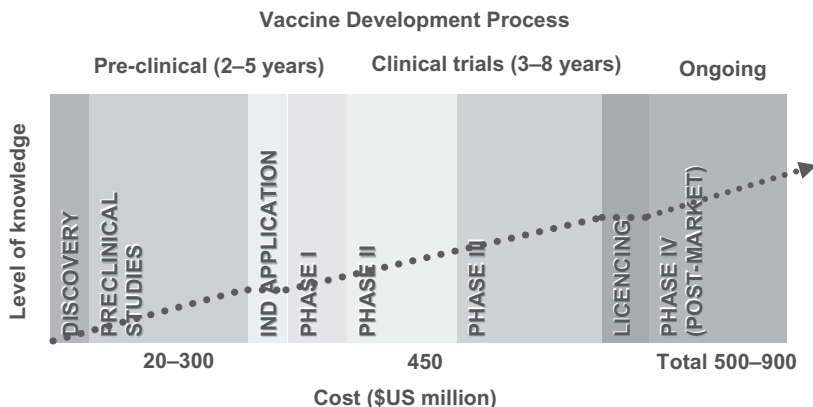


**FIGURE 1** Vaccine development process.

variable length and complexity, which relate to the virus itself, its pathogenicity, and technical impediments that may stand in the way of evaluating its effectiveness as a preventive measure. For the registration of many vaccines, data from phases II and III or II and IV trials are considered as single categories by regulators. These studies involve the participation of increasing numbers of volunteers in clinical trials and are the most expensive part of the registration process. They are generally beyond the resources of most small-to medium-sized manufacturers and research institutes and can only be undertaken by BigPharma to achieve near-simultaneous registration in many countries.

Before clinical studies can commence, a dossier of preclinical study results, and the proposed protocols for human studies are submitted to the regulatory agencies. For the FDA, an investigational new drug (IND) application is made. Clinical trials are classified as phase I, II, III, or IV, depending on the stage of clinical development of a vaccine. Phase I studies are *first-in-human* studies in which the drug is administered to small numbers of healthy volunteers to determine safety and tolerability, as well as pharmacological activity. Results can also provide information as to appropriate dose ranges for future studies. If there is a high possibility of toxic side effects, then the drug may be only administered to the patients for whom it is primarily intended. Therapeutic vaccines against HIV and hepatitis C could be examples of such an approach. Phase I studies are usually conducted in purpose-built facilities that allow close monitoring of patients.

The principal purpose of a phase II study is to determine vaccine efficacy in small groups of the targeted patients, although safety data is also collected. A range of doses is generally used to determine the maximum tolerated dose and the therapeutic range. Phase III studies are conducted in large groups of patients if phase II studies indicate that the drug is safe and has potential clinical benefit. Efficacy data continues to be collected in phase III studies, together with data on the incidence and nature of adverse effects. Phase IIIb studies may be undertaken while the application for licensing is being assessed to provide further safety and/or efficacy data (periapproval studies).

Once approval for marketing is granted, phase IV studies are conducted to determine efficacy in the typical clinical setting (using inclusion/exclusion criteria less stringent than those in the protocols for the phase III studies). Postmarketing surveillance is conducted to track adverse effects and studies comparing the new drug to existing therapies, or its role in combination therapy, may be conducted as part of phase IV studies.

### 10.16.6 THE FUTURE

The advent of recombinant deoxyribonucleic acid (DNA), together with the emergence of HIV and, more recently, HCV viruses as major human pathogens have given new impetus over the past 20 years to technologies associated with viral vaccine development. Ironically, the very success of commonly used vaccines for the prevention of poliomyelitis, rubella, mumps, and yellow fever was a major reason for a diminution of interest in vaccine technology from the mid-1960s to the early 1980s. Both HIV and HCV have been targets for an unprecedented increase in research funding on vaccines in recent years. However, for several technical reasons

not discussed here, vaccines are still unavailable against either agent, either for preventive or therapeutic use.

In recent years great changes have occurred in most countries following the acceptance by governments and other public bodies of the need to maintain vaccination programs as a basic human right. These changes have been accompanied by commensurate improvements to vaccine production, which should have resulted in improved efficiencies and lower costs and a reduction in the time and cost required for the registration of new vaccines. Unfortunately, this has not been the case. Regulatory agencies are increasingly able to fast-track approvals for new products for use in situations of immediate need, such as for the development of avian influenza vaccines, that make use of newer molecular and older technologies. However, the majority of new vaccines are only approved many years after an etiology has been established for a previously unrecognized disease or a novel approach has been put forward as an alternative to a new vaccine. Contemporary examples of these delays include (1) live attenuated vaccines against influenza, which were first proposed in the 1970s and achieved registration for restricted use by the FDA in 2002 and (2) rotavirus vaccines, referred to above, in 2006—some 33 years after an etiology was established! Costs associated with bringing a new vaccine to the community are reported to be \$US500–900 million, although BigPharma often avoid disclosing specific details, citing commercial sensitivities. National regulatory agencies are highly risk averse and are responsible for ensuring the quality, efficacy, safety, and timely availability for vaccines for the community, through a uniform system of regulatory controls. They are involved in all aspects of vaccine development from initial discovery to postapproval surveillance for adverse effects in close consultation with vaccine developers. The reasons for their increased involvement are historic and have been outlined above. However, it is fair to state that many developers of highly successful pediatric viral vaccines would have experienced great difficulties in achieving registration today. It is also highly problematic whether the manufacture of such vaccines in the current regulatory environment would be profitable, even allowing for the economies of scale enjoyed by BigPharma.

While the prevention of viral diseases by vaccination has been a singular triumph for medical science, many challenges still exist. Quite apart from HIV and HCV, vaccines are still unavailable for the prevention of most respiratory virus infections, and improvements to existing influenza vaccines are universally regarded as a matter of urgency. Vaccines are also unavailable for the prevention of all human (but not veterinary) herpesvirus infections, again despite an urgent need. Much work needs to be done to expand the number of adjuvants to improve the performance of existing and future vaccines (at present only alum is acceptable), despite the formidable problems that have been encountered over the past 20 years in this area. It would be a great pity if the constraints referred to above gave rise to a diminution in the forces of innovation that have been responsible for so many improvements to public health over the past century.

## REFERENCES

1. Blake, J. B. (1959), *Public Health in the Town of Boston, 1630–1832*, Harvard University Press, Cambridge, MA.

2. Dreesen, D. W. (1997), A global review of rabies vaccines for human use, *Vaccine*, 15(Suppl), S2–6.
3. U.S. Food and Drug Administration (2006), FDA licenses new vaccine for prevention of cervical cancer and other diseases in females caused by human papillomaviruses, FDA 2006-06-08.
4. Seeff, L. B., Beebe, G. W., Hoofnagle, J. H., et al. (1987), A serologic follow-up of the 1942 epidemic of post-vaccination hepatitis in the United States Army, *N. Engl. J. Med.*, 316(16), 965–970.
5. Nathanson, N., and Longmuir, A. D. (1963), The Cutter incident. Poliomyelitis following formaldehyde-inactivated poliovirus vaccination in the United States during the spring of 1955. II. Relationship of poliomyelitis to Cutter vaccine, *Am. J. Hyg.*, 78, 29–60.
6. Strickler, H. D., Rosenberg, P. S., Devesa, S. S., et al. (1998), Contamination of poliovirus vaccines with simian virus 40 (1955–1963) and subsequent cancer rates, *JAMA*, 279(4), 292–295.
7. Kapikian, A. Z., Mitchell, R. H., Chanock, R. M., et al. (1969), An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with inactivated RS virus vaccine, *Am. J. Epidemiol.*, 89, 405–421.
8. World Health Organization (2003), *Global Polio Eradication Initiative: Strategic Plan 2004–2008*, Geneva.
9. Evans, D. M., Dunn, G., Minor, P. D., et al. (1985), Increased neurovirulence associated with a single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome, *Nature*, 314, 548–550.
10. Houff, S. A., Miller, G. L., Lovell, C. R., et al. (1977), The Guillain-Barre syndrome and swine influenza vaccination, *Trans. Am. Neurol. Assoc.*, 102, 120–123.
11. Centers for Disease Control and Prevention (CDC) (1999), Intussusception among recipients of rotavirus vaccine—United States, 1998–1999, *MMWR Morb. Mortal Wkly. Rep.*, 48(27), 577–581.
12. Glass, R. I., and Parashar, U. D. (2006), The promise of new rotavirus vaccines, *N. Engl. J. Med.*, 354, 75–77.
13. Mutsch, M., Zhou, W., Rhodes, P., et al. (2004), Use of the inactivated intranasal influenza vaccine and the risk of Bell's palsy in Switzerland, *N. Engl. J. Med.*, 350(9), 896–903.
14. World Medical Association (2000), *Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects*, Geneva.
15. Smorodintsev, A. A., Davidenkova, E. F., Drobyshevska, Y. A., et al. (1959), Results of a study of the reactogenic and immunogenic properties of live anti-poliomyelitis vaccine, *Bull. World Health Organ.*, 20, 1053–1074.
16. Vitek, C. R., and Wharton, M. (1998), Diphtheria in the former Soviet Union: Reemergence of a pandemic disease, *Emerg. Infectious Dis.*, 4(4), 539–550.

## *Methods of Randomization*

**Gladys McPherson and Marion Campbell**

*Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland*

### **Contents**

|        |                                                 |     |
|--------|-------------------------------------------------|-----|
| 11.1   | Patients Registration                           | 780 |
| 11.1.1 | Patient Identification and Recruitment          | 781 |
| 11.1.2 | Confirmation of Eligibility                     | 781 |
| 11.1.3 | Agreement to Randomize                          | 781 |
| 11.1.4 | Informed Patient Consent                        | 782 |
| 11.1.5 | Formal Entry to Trial                           | 782 |
| 11.1.6 | Treatment Assignment                            | 782 |
| 11.1.7 | Completion of Relevant Documentation            | 784 |
| 11.1.8 | Commencement of Treatment                       | 784 |
| 11.2   | Generation the Random Sequence                  | 784 |
| 11.2.1 | Simple Randomization                            | 785 |
| 11.2.2 | Restricted Randomization                        | 789 |
| 11.3   | Covariate-Adaptive Randomization                | 793 |
| 11.3.1 | Stratification                                  | 793 |
| 11.3.2 | Random Permuted Blocks within Strata            | 794 |
| 11.3.3 | Minimization                                    | 796 |
| 11.3.4 | Minimization Compared with Simple Randomization | 799 |
| 11.3.5 | Other Constrained Methods                       | 799 |
| 11.4   | Response-Adaptive Methods                       | 800 |
| 11.5   | Special-Case Scenarios                          | 800 |
| 11.5.1 | Unequal Randomization                           | 800 |
| 11.5.2 | Cluster Randomization                           | 801 |

|              |     |
|--------------|-----|
| 11.6 Summary | 802 |
| Appendix I   | 802 |
| Appendix II  | 803 |
| Appendix III | 804 |
| Appendix VI  | 805 |
| References   | 805 |

The randomized controlled trial (RCT) is one of the simplest, most powerful, and revolutionary tools of research. In essence, the RCT is a comparison of two or more interventions in which people are allocated at random to a treatment group. Randomization is a method of eliminating bias in the way that treatments are allocated to patients, but it does not guarantee that the characteristics of the groups will be completely balanced; in that if imbalance occurs it will have occurred by chance rather than by the introduction of some systematic bias. Some factors are known in advance to be important risk factors associated with the outcome of the patient. Imbalance in these prognostic variables could have a marked effect on the results of a trial and on their credibility. However, if the randomization is performed fairly, then any difference between groups is due to chance, but the groups might differ in a way that might affect their response to treatment (this is a question of clinical importance rather than statistical significance). While it is possible to modify the statistical analysis to take account of any differences between groups at baseline, it is better to control the problem at the design stage.

There are four classes of randomization procedure: complete (simple) randomization, restricted randomization, covariate-adaptive randomization, and response-adaptive randomization. The simplest randomization procedure is complete randomization with equal allocation and does not depend on prognostic factors or previous treatment assignments. Restricted randomization is used when it is desired to have equal numbers of patients assigned to each treatment group. Covariate-adaptive randomization is used to achieve similar numbers of patients in each treatment group and also to ensure that patient groups are similar with respect to prognostic factors such as age or gender. In response-adaptive randomization the treatment assignments depend upon previous patient responses to treatment. These different approaches to generating the random sequence will be discussed in detail in Sections 11.2–11.4.

## 11.1 PATIENT REGISTRATION

In a randomized controlled trial patients are identified for entry to the trial and, if they are eligible and consent to trial entry, they are then randomized to one of the study treatments. For a patient to be included into a clinical trial the following steps are necessary:

- Patient identification and recruitment
- Confirmation of eligibility
- Agreement to randomize
- Informed patient consent

- Formal entry to trial
- Treatment assignment
- Completion of relevant documentation
- Commencement of treatment

### **11.1.1 Patient Identification and Recruitment**

How potentially eligible patients are to be identified and how and where they are to be approached should be outlined in the study protocol. The study design team should have made these decisions so as to ensure that the study population will be as representative as possible of the particular disease area under investigation to ensure the generalizability of the results. All appropriate patients should be logged even if they are ineligible or missed so that a full record of potential recruits is obtained. This can then be used as a check that systematic selection bias of patients into the trial had not occurred. The recruitment process should be outlined fully and could consist of a script for the recruitment officer to follow.

### **11.1.2 Confirmation of Eligibility**

Eligibility criteria should be agreed upon before recruitment begins and should be recorded in the study protocol. The recruitment officer should check that a potential participant satisfies the eligibility criteria before asking them to consent to the trial. If the patient is recruited and then subsequently found to be ineligible, this should be logged and the decision whether to include them in an intention-to-treat analysis, or exclude them from analysis according to some consistent algorithm, made at a later date. Postrandomization exclusions unrelated to noncompliance, withdrawal, or losses to follow up occur when patients are inappropriately randomized into a clinical trial [1]. These patients can be removed from both study arms without risking bias. The number of patients randomized but not included in the primary analysis should be clearly stated along with the reasons for exclusion.

### **11.1.3 Agreement to Randomize**

It is vitally important that the patient understand the concept of randomization and agree to accept the outcome of the treatment assignment. The recruitment officer will have to explain this very carefully and be satisfied that the patient has a good understanding of the process. Patients may have a preference for one treatment over another, and it is sometimes desirable to record this as it can help to better understand subsequent noncompliance. However, the recruitment officer must be assured that this preference is not so strong that the patient will then not comply with the allocated treatment. In this situation patients are uncertain about the treatment choices and therefore should not agree to be randomized. This concept also applies to any clinician who has a strong treatment preference or who is unwilling to randomize the patients. There are study designs that allow treatment allocation according to either patient or collaborator preference, which may be adopted in order to increase recruitment especially in surgical trials.



### 11.1.4 Informed Patient Consent

There is a strong ethical and legal requirement to obtain informed consent to a randomized control trial before the patient is randomized. Therefore, it is important that the recruitment officer ensures that the patient fully understand the purpose of the study as well as agreeing to the randomization process. A well-designed patient information leaflet will describe these concepts. It should also outline what patient involvement is expected for the duration of the study (medication to take or surgery to be undertaken, clinics to attend, questionnaires to be filled in, etc.) and what steps should be taken if, at any time, the patient wishes to withdraw from the study. Patients are usually ineligible if they cannot give informed consent. For example, they cannot understand English (if there are no translations available) or they have mental health problems. Such exclusions would be stipulated in the protocol.

### 11.1.5 Formal Entry to Trial

If all the previous steps have been taken and a patient is identified as being eligible for entry to the trial, then it is necessary at this stage to record some basic demographic details about that patient. The patient's name and address can now be logged for entry into the study database. If any important prognostic factors have been identified and are to be used in the randomization process, the values should be recorded on a designated pro-forma prior to randomization. The patient might be allocated a study number at this stage (or this might be allocated at the time of randomization) if randomization is done by an external randomization service.

### 11.1.6 Treatment Assignment

Adequate concealment of treatment allocation is widely recognized to be of paramount importance to the scientific validity of any clinical trial [2]. As such, randomization should not be performed by the recruitment staff. Randomization may be performed by staff in the central study office or by an independent trials center offering this service. It is common to use either a randomization list or computer program to do this, and the process involves either sending paper forms to the randomization center for manual data entry and treatment allocation, by telephoning an automated system, or using a Web-based randomization system. Treatment allocation by sealed opaque envelopes is not recommended as this method has been shown not to be tamper-proof (e.g., envelopes have been observed to have been opened out of sequence; and there has been evidence that envelopes have been opened in advance to seek prior knowledge of treatment assignment).

Some common study scenarios are as follows:

(a) *Multicenter Trial with Central Study Office* If the trial design requires immediate treatment assignment (or randomization outside normal working hours), then it is preferable to use an automated telephone randomization service often referred to as an IVR (interactive voice response) system. A benefit of this system is that a full audit of the randomization process is automatically created. The date and time

of randomization is logged, along with at least one patient identifier so that treatment allocations can be proven to be fair and unbiased. IVR systems also help to maintain blinding since there is no staff involvement and no need for code break envelopes. It is necessary, however, to ensure that backup procedures are documented in case of very occasional system failure. It is now routine that these systems can be configured and maintained relatively cheaply, with minimal denial of service achieved for affordable cost.

This system may be backed up by IWR (interactive web response) systems, which may help overcome language difficulties (e.g., understanding of written English can often be easier than listening to a voice over the telephone). The person performing the randomization also has more time to read and digest instructions, and helpful tips can also be included on the page. It is also possible to construct multilingual Web or telephone systems. The use of an IWR system offers further benefits as it is possible to extend the system to include real-time reporting (giving up-to-date information about recruitment, randomization, and patient status). Some recent trials have also used this technology for the collection of electronic patient-reported outcome data [3].

If randomization is required only within normal working hours, then treatment assignments can be read off the randomization list and given to the recruitment officer over the telephone. Ideally, a small, customized computer program would be used to store the list and output the next treatment assignment as required to minimize breaking allocation concealment. If sufficient funds are available, then an automated system is preferred as this does not rely on someone being available to answer the telephone.

If patients are screened first and then only considered for randomization later, for example, after responding appropriately to a screening questionnaire, then an automated system may not be necessary and a computerized randomization system may be written into the trial management program. If the code is kept secure, then this would offer all the benefits of an automated telephone response system without the expense.

The randomization scheme itself could consist of an allocation list constructed from permuted blocks or might incorporate a program using minimization if important prognostic variables had been identified.

(b) *Double-blind Trial of Drug Therapy* For this type of trial design the pharmacist preparing the drugs needs to have access to the treatment allocation. While the randomization lists continue to be created by the study office or the trials unit performing the randomization, the treatment allocation should now also include a corresponding drug treatment identifier that will be used to label the drug. These lists are then sent to the pharmacy where the drugs are prepared and packaged according to the list. Treatment unblinding can be performed by the pharmacy and/or the central randomization service.

When a patient is randomized into the trial, the recruitment officer telephones the automated service. The patient is allocated a treatment allocation number, which is known to be in stock at that site. If the randomization algorithm uses random permuted blocks, then the allocated treatment will simply be the next available one on the list for that site. If a minimization algorithm is being used, then the allocation will be the next available pack of the appropriate drug type. The randomization center must be aware at all times of the current stock levels at each site and an early

warning system can be written into the IVR system to alert the pharmacy when stocks are low.

(c) *Single-Center Trial* For a single-center trial there are several possible solutions. The randomization can be performed by an independent person who is not involved with the trial; a small local computer program that incorporates a full audit trail can be written to perform the randomization; or a central randomizing service may be used if funds permit.

### 11.1.7 Completion of Relevant Documentation

At this stage in the study the following documentation should be available:

- A log sheet of all patients identified as appropriate for the study
- Ineligibility forms for those not eligible
- Eligibility forms for those considered eligible for entry to the study (with the study number)
- A consent form appropriately signed and dated (with the study number)
- A baseline data form containing basic demographic data (with the study number), previous medical history, relevant information about their clinical condition, and any baseline questionnaires (such as SF36, EQ-5D)
- (*Optional*) A randomization form containing the values for any prognostic factors, the treatment allocation code, and the study number

Any forms should be piloted before the study commences and committed to final printing only after this has proved successful. Consideration must be given as to where these forms are to be stored, and for what length of time, as any patient identifiable material should be kept in a secure environment.

### 11.1.8 Commencement of Treatment

Treatment should commence as soon as possible after randomization; so in a multitherapy trial each randomization should take place only when the patient is ready to receive that treatment. In a drug therapy trial each dispatch of study treatment requires rigorous control, and batch numbers should be checked and recorded. It is not sufficient to record only the drug pack allocated at randomization; the actual drug pack received should also be recorded.

## 11.2 GENERATING THE RANDOM SEQUENCE

In Sections 11.2.1–11.5 a description of each randomization method including advantages and disadvantages is given. This outlines the resources required, ease of use of method, allocation concealment, and predictability of the method.

In most clinical trials the complete list of participants is not available at the start of the trial, and patients will be recruited and randomized one at a time as they are identified. For both single-center and multicenter trials it is preferable that the randomization is performed by an independent organization or a central study

office to ensure that the randomization is carried out according to the required specification.

### 11.2.1 Simple Randomization

With simple randomization each patient has a known chance, usually equal, of being given each treatment, and the treatment to be given cannot be predicted in advance. The simplest method is tossing a coin, but one can also use random tables, or a random-number generator in a calculator or computer. Table 1 gives an example of a table of random digits that were generated by computer (see Appendix I for the code). To randomize a patient, (a) choose an arbitrary starting point in the list and (b) work along the rows assigning treatments by using the numbers on the list.

**Example** For *two* treatments assume the digits 0–4 correspond to treatment A and digits 5–9 correspond to treatment B. Therefore the top row of Table 1 would produce the following treatment assignments:

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 |
| B | B | B | A | A | A | B | A | A | A | B | B | A | B | A | A | B | A | A | B | A | A | B |

For *three* treatments assume the digits 1–3 correspond to treatment A, digits 4–6 correspond to treatment B, and digits 7–9 correspond to treatment C. Ignore 0 (ignoring a digit has no effect on the inherent “randomness” of the sequence). The top row of Table 1 would then produce the following treatment assignments:

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 |
| B | B | B | A | A | A | C | A | A | B | B | B | B | C | A | — | C | A | B | C | A | B | C |

For *four* treatments assume the digits 1–2 correspond to treatment A, digits 3–4 correspond to treatment B, digits 5–6 correspond to treatment C, and digits 7–8 correspond to treatment D. Ignore 0 and 9. The top row of Table 1 would then produce the following treatment assignments:

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 |
| C | C | C | B | A | B | — | A | A | B | C | C | B | D | A | — | D | B | B | D | B | B | D |

The list should be produced before the start of the trial and should be large enough to complete the trial.

#### Advantages of This Method

- Each patient assignment is completely unpredictable; and, if the numbers are large enough (say  $\geq 500$ ), then probability theory allows us to be confident that the numbers in each treatment group will be similar.
- The method is very straightforward and easy to implement.

#### Disadvantages of This Method

- Table 2 illustrates the difference in treatment numbers that may occur in a two-treatment trial with probability at least 0.05 and at least 0.01. The probability of  $\geq 10\%$  overall imbalance between treatment groups is also shown.

TABLE 1 Table of Random Numbers

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 3 | 1 | 3 | 9 | 1 | 1 | 4 | 5 | 5 | 4 | 8 | 1 | 0 | 7 | 3 | 4 | 8 | 3 | 4 | 7 | 9 | 3 |
| 5 | 3 | 0 | 6 | 0 | 0 | 5 | 2 | 5 | 2 | 3 | 8 | 3 | 5 | 4 | 8 | 7 | 1 | 6 | 0 | 9 | 8 | 0 | 3 | 6 |
| 0 | 8 | 0 | 4 | 3 | 9 | 9 | 6 | 0 | 1 | 2 | 0 | 6 | 8 | 8 | 8 | 9 | 3 | 5 | 6 | 2 | 4 | 4 | 7 | 8 |
| 8 | 7 | 2 | 5 | 3 | 0 | 7 | 7 | 1 | 6 | 7 | 0 | 9 | 6 | 9 | 4 | 4 | 9 | 0 | 3 | 8 | 2 | 8 | 3 | 6 |
| 1 | 6 | 8 | 7 | 1 | 7 | 2 | 9 | 1 | 2 | 1 | 3 | 7 | 9 | 1 | 1 | 3 | 8 | 7 | 2 | 3 | 9 | 0 | 4 | 5 |
| 3 | 9 | 9 | 9 | 8 | 9 | 6 | 8 | 9 | 2 | 7 | 7 | 0 | 8 | 9 | 7 | 2 | 8 | 4 | 3 | 0 | 8 | 2 | 0 | 8 |
| 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 |
| 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 |
| 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 |
| 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 |
| 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 |
| 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 |
| 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 |
| 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 |
| 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 |
| 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 |
| 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 |
| 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 |
| 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 |
| 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 |
| 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 |
| 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 |
| 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 |
| 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 |
| 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 |
| 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 |
| 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 |
| 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 |
| 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 |
| 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 |
| 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 |
| 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 |
| 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 |
| 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 |
| 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 |
| 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 |
| 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 |
| 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 |
| 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 |
| 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 |
| 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 |
| 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 |
| 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 |
| 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 |
| 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 |
| 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 |
| 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 |
| 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 |
| 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 |
| 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 |
| 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 |

TABLE 1 *Continued*

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 |
| 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 |
| 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 |
| 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 |
| 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 |
| 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 |
| 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 |
| 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 |
| 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 |
| 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 |
| 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 |
| 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 |
| 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 | 9 | 0 | 2 | 0 | 6 | 7 |
| 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 | 0 | 6 | 7 | 9 | 9 | 5 |
| 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 | 0 | 1 | 1 | 3 | 6 | 9 |
| 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 | 8 | 2 | 7 | 7 | 0 | 8 |
| 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 | 5 | 1 | 5 | 1 | 5 | 9 |
| 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 | 2 | 5 | 1 | 3 | 6 | 8 |
| 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 | 3 | 2 | 0 | 4 | 2 | 0 |
| 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 | 3 | 8 | 2 | 6 | 8 | 7 |
| 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 | 5 | 2 | 1 | 2 | 8 | 0 |
| 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 | 0 | 5 | 9 | 9 | 4 | 1 |
| 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 | 9 | 4 | 3 | 8 | 9 | 5 |
| 9 | 0 | 2 | 0 | 6 | 7 | 7 | 7 | 8 | 3 | 4 | 5 | 1 | 4 | 3 | 7 | 7 | 8 | 6 | 1 | 4 | 3 | 9 | 6 | 6 |
| 0 | 6 | 7 | 9 | 9 | 5 | 8 | 3 | 3 | 8 | 0 | 2 | 8 | 1 | 7 | 2 | 5 | 0 | 6 | 8 | 7 | 0 | 7 | 1 | 8 |
| 0 | 1 | 1 | 3 | 6 | 9 | 0 | 1 | 2 | 7 | 6 | 2 | 2 | 8 | 9 | 3 | 5 | 3 | 9 | 8 | 8 | 8 | 9 | 6 | 7 |
| 8 | 2 | 7 | 7 | 0 | 8 | 8 | 7 | 2 | 7 | 3 | 3 | 9 | 7 | 1 | 9 | 7 | 0 | 2 | 9 | 9 | 6 | 6 | 1 | 3 |
| 5 | 1 | 5 | 1 | 5 | 9 | 1 | 9 | 8 | 7 | 1 | 0 | 4 | 1 | 3 | 4 | 4 | 6 | 9 | 9 | 4 | 4 | 6 | 0 | 6 |
| 2 | 5 | 1 | 3 | 6 | 8 | 6 | 2 | 1 | 1 | 1 | 2 | 9 | 8 | 9 | 5 | 0 | 7 | 0 | 1 | 1 | 0 | 5 | 0 | 6 |
| 3 | 2 | 0 | 4 | 2 | 0 | 3 | 5 | 2 | 2 | 9 | 7 | 2 | 6 | 9 | 1 | 8 | 1 | 8 | 9 | 4 | 2 | 1 | 0 | 4 |
| 3 | 8 | 2 | 6 | 8 | 7 | 9 | 0 | 3 | 7 | 7 | 9 | 1 | 0 | 5 | 8 | 2 | 6 | 0 | 8 | 6 | 2 | 4 | 4 | 4 |
| 5 | 2 | 1 | 2 | 8 | 0 | 0 | 3 | 4 | 5 | 3 | 8 | 1 | 9 | 6 | 3 | 3 | 7 | 3 | 4 | 6 | 8 | 2 | 5 | 2 |
| 0 | 5 | 9 | 9 | 4 | 1 | 4 | 2 | 2 | 7 | 2 | 5 | 3 | 7 | 6 | 1 | 5 | 0 | 8 | 7 | 0 | 3 | 6 | 0 | 7 |
| 1 | 6 | 5 | 1 | 1 | 7 | 1 | 2 | 4 | 2 | 8 | 9 | 0 | 9 | 0 | 5 | 6 | 7 | 3 | 6 | 6 | 9 | 9 | 0 | 8 |
| 3 | 6 | 5 | 1 | 8 | 8 | 2 | 8 | 9 | 1 | 1 | 7 | 0 | 5 | 5 | 0 | 2 | 4 | 0 | 3 | 9 | 4 | 8 | 3 | 8 |
| 0 | 9 | 2 | 9 | 3 | 0 | 2 | 3 | 3 | 5 | 8 | 1 | 2 | 3 | 4 | 9 | 8 | 4 | 4 | 0 | 2 | 5 | 7 | 5 | 1 |
| 0 | 0 | 0 | 1 | 8 | 9 | 0 | 4 | 9 | 9 | 2 | 0 | 0 | 9 | 4 | 9 | 5 | 5 | 1 | 9 | 3 | 1 | 9 | 2 | 4 |
| 1 | 1 | 8 | 8 | 3 | 5 | 7 | 3 | 7 | 3 | 7 | 1 | 3 | 1 | 0 | 9 | 3 | 2 | 6 | 3 | 5 | 6 | 6 | 8 | 1 |
| 1 | 6 | 6 | 8 | 2 | 9 | 4 | 7 | 3 | 5 | 8 | 0 | 8 | 4 | 3 | 3 | 3 | 4 | 1 | 0 | 1 | 7 | 2 | 9 | 2 |
| 3 | 4 | 2 | 7 | 2 | 8 | 5 | 4 | 2 | 6 | 4 | 2 | 5 | 7 | 4 | 4 | 1 | 9 | 4 | 8 | 1 | 3 | 0 | 3 | 0 |
| 1 | 6 | 4 | 3 | 2 | 6 | 5 | 0 | 4 | 8 | 0 | 9 | 1 | 2 | 5 | 9 | 9 | 1 | 3 | 2 | 7 | 0 | 4 | 8 | 2 |
| 0 | 8 | 5 | 6 | 6 | 6 | 7 | 4 | 3 | 4 | 0 | 2 | 2 | 5 | 6 | 7 | 5 | 0 | 3 | 1 | 8 | 5 | 5 | 9 | 5 |
| 6 | 8 | 0 | 4 | 7 | 4 | 2 | 7 | 1 | 1 | 7 | 3 | 6 | 4 | 4 | 9 | 4 | 7 | 5 | 9 | 8 | 3 | 7 | 2 | 0 |
| 9 | 2 | 5 | 8 | 2 | 9 | 1 | 6 | 5 | 1 | 1 | 7 | 1 | 2 | 4 | 2 | 8 | 9 | 0 | 9 | 0 | 5 | 6 | 7 | 3 |
| 6 | 6 | 9 | 9 | 0 | 8 | 3 | 6 | 5 | 1 | 8 | 8 | 2 | 8 | 9 | 1 | 1 | 7 | 0 | 5 | 5 | 0 | 2 | 4 | 0 |
| 3 | 9 | 4 | 8 | 3 | 8 | 0 | 9 | 2 | 9 | 3 | 0 | 2 | 3 | 3 | 5 | 8 | 1 | 2 | 3 | 4 | 9 | 8 | 4 | 4 |
| 0 | 2 | 5 | 7 | 5 | 1 | 0 | 0 | 0 | 1 | 8 | 9 | 0 | 4 | 9 | 9 | 2 | 0 | 0 | 9 | 4 | 9 | 5 | 5 | 1 |
| 9 | 3 | 1 | 9 | 2 | 4 | 1 | 1 | 8 | 8 | 3 | 5 | 7 | 3 | 7 | 3 | 7 | 1 | 3 | 1 | 0 | 9 | 3 | 2 | 6 |
| 3 | 5 | 6 | 6 | 8 | 1 | 1 | 6 | 6 | 8 | 2 | 9 | 4 | 7 | 3 | 5 | 8 | 0 | 8 | 4 | 3 | 3 | 3 | 4 | 1 |

**TABLE 2 Possible Imbalance between Treatment Groups in Simple Randomization with Two Treatments**

| Total Number of Patients | Difference in Numbers   |                         | Probability of $\geq 10\%$<br>Treatment Imbalance |
|--------------------------|-------------------------|-------------------------|---------------------------------------------------|
|                          | Probability $\geq 0.05$ | Probability $\geq 0.01$ |                                                   |
| 10                       | 2:8                     | 1:9                     | 0.744                                             |
| 20                       | 6:14                    | 5:15                    | 0.819                                             |
| 50                       | 18:32                   | 16:34                   | 0.477                                             |
| 100                      | 40:60                   | 37:63                   | 0.371                                             |
| 200                      | 86:114                  | 82:118                  | 0.188                                             |
| 500                      | 228:272                 | 221:279                 | 0.036                                             |
| 1000                     | 470:530                 | 459:541                 | 0                                                 |

**TABLE 3 Possible Imbalance between Treatment Groups in Simple Randomization with Three Treatments**

| Total Number of Patients | Lowest Observation in a Group |                         | Probability of $\geq 10\%$<br>Treatment Imbalance |
|--------------------------|-------------------------------|-------------------------|---------------------------------------------------|
|                          | Probability $\geq 0.05$       | Probability $\geq 0.01$ |                                                   |
| 10                       | 1 in any group                | 0 in any group          | 1                                                 |
| 20                       | 2 in any group                | 2 in any group          | 1                                                 |
| 50                       | 9 in any group                | 8 in any group          | 0.732                                             |
| 100                      | 24 in any group               | 21 in any group         | 0.474                                             |
| 200                      | 54 in any group               | 49 in any group         | 0.196                                             |
| 500                      | 144 in any group              | 137 in any group        | 0.022                                             |
| 1000                     | 300 in any group              | 290 in any group        | 0.001                                             |

**TABLE 4 Possible Imbalance between Treatment Groups in Simple Randomization with Four Treatments**

| Total Number of Patients | Lowest Observation in a Group |                         | Probability of $\geq 10\%$<br>Treatment Imbalance |
|--------------------------|-------------------------------|-------------------------|---------------------------------------------------|
|                          | Probability $\geq 0.05$       | Probability $\geq 0.01$ |                                                   |
| 10                       | 0 in any group                | 0 in any group          | 1                                                 |
| 20                       | 1 in any group                | 0 in any group          | 0.944                                             |
| 50                       | 6 in any group                | 4 in any group          | 0.806                                             |
| 100                      | 15 in any group               | 13 in any group         | 0.530                                             |
| 200                      | 37 in any group               | 33 in any group         | 0.193                                             |
| 500                      | 104 in any group              | 98 in any group         | 0.004                                             |
| 1000                     | 220 in any group              | 213 in any group        | 0.001                                             |

- For three or four treatments, Tables 3 and 4 show the probability of no patients being allocated to one of the treatment groups, and the probability of  $\geq 10\%$  overall imbalance between treatment groups is also shown.
- There is also no guarantee that the groups are similar with respect to important prognostic factors, and for small to medium sized trials they are unlikely to be balanced. Therefore, in practice, some sort of restricted randomization is almost always used.



### 11.2.2 Restricted Randomization

Restricted randomization is used to balance treatment assignments assigned to each treatment group. Three possible approaches are described here.

**Random Permuted Blocks** A list of allocations is prepared before the start of the trial to be assigned in sequence order. This method is used to keep numbers in each treatment group closely balanced at all times, so there is a greater chance of balance whenever recruitment ends (which may be early). Blocks are made up of all possible permutations of treatment assignments. Therefore block sizes that are multiples of the number of treatments are easier to manage.

For two treatments, blocks of an even size would be used. It is unusual to use a block size as small as 2 as this makes guessing easier, but for a small trial with few patients it may be justified. The random numbers in Table 1 can be used to produce the randomization list.

*Example: Two Treatments, Block Size 2* Let 0–4 correspond to block AB and 5–9 correspond to block BA. The top row of Table 1 would then produce the following treatment assignments:

|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 5  | 5  | 6  | 3  | 1  | 3  | 9  | 1  | 1  | 4  | 5  | 5  | 4  | 8  | 1  |
| BA | BA | BA | AB | AB | AB | BA | AB | AB | AB | BA | BA | AB | BA | AB |

*Example: Three Treatments, Block Size 3* Let 1 correspond to block ABC, 2 to block ACB, 3 to block BAC, 4 to block BCA, 5 to block CAB, and 6 to block CBA. Ignore 0 and 7–9. The top row of Table 1 would then produce the following treatment assignments:

|     |     |     |     |     |     |   |     |     |     |     |
|-----|-----|-----|-----|-----|-----|---|-----|-----|-----|-----|
| 5   | 5   | 6   | 3   | 1   | 3   | 9 | 1   | 1   | 4   | 5   |
| CAB | CAB | CBA | BAC | ABC | BAC | — | ABC | ABC | BCA | CAB |

*Example: Two Treatments, Block Size 4* Let 1 correspond to block AABB, 2 to block ABAB, 3 to block ABBA, 4 to block BBAA, 5 to block BABA, and 6 to block BAAB. Ignore 0 and 7–9. The top row of Table 1 would then produce the following treatment assignments:

|      |      |      |      |      |      |   |      |      |      |
|------|------|------|------|------|------|---|------|------|------|
| 5    | 5    | 6    | 3    | 1    | 3    | 9 | 1    | 1    | 4    |
| BABA | BABA | BAAB | ABBA | AABB | ABBA | — | AABB | AABB | BBAA |

It is also common practice to make the second block the mirror image of the first block, for example, Block 1 (A B A B) becomes Block 2 (B A B A). Producing a list of randomly ordered numbers is easily done within Microsoft Access. To randomly order the numbers 0–19 first create a table, say called TblRandom20, with one field in it (called ID) and add the data values 0–19 into this table. Then create a query with the following SQL (Structured Query Language) syntax:

```
SELECT TblRandom20.ID
FROM TblRandom20
ORDER BY Rnd([ID]);
```

Each time this query is opened in the datasheet view it will produce the numbers 0–19 in a random order. For 20 digits the number of possible permutations (and therefore the number of differently ordered blocks) are:

$$20! = 20 \times 19 \times 18 \times \cdots \times 3 \times 2 \times 1 = 2432902008176640000$$

This can be used by a computer program (see Appendix II) to produce a list of random permutations as illustrated in Table 5, which can then be used to produce the randomization list.

**Example: Two Treatments, Block Size 20** Let 0–9 correspond to treatment A and 10–19 to treatment B.

The top row of Table 5 would then produce block 1:

|   |    |    |   |   |    |   |   |    |    |    |    |   |   |    |   |    |   |   |    |
|---|----|----|---|---|----|---|---|----|----|----|----|---|---|----|---|----|---|---|----|
| 9 | 18 | 17 | 5 | 3 | 10 | 2 | 4 | 11 | 13 | 12 | 14 | 6 | 7 | 15 | 0 | 19 | 8 | 1 | 16 |
| A | B  | B  | A | A | B  | A | A | B  | B  | B  | B  | A | A | B  | A | B  | A | A | B  |

And the second row would produce block 2:

|   |   |    |    |   |    |   |    |    |   |   |    |   |    |   |    |   |    |   |    |
|---|---|----|----|---|----|---|----|----|---|---|----|---|----|---|----|---|----|---|----|
| 4 | 1 | 15 | 10 | 5 | 14 | 3 | 18 | 17 | 9 | 0 | 19 | 2 | 12 | 6 | 11 | 8 | 13 | 7 | 16 |
| A | A | B  | B  | A | B  | A | B  | B  | A | A | B  | A | B  | A | B  | A | B  | A | B  |

When using a fixed block size, it is possible that the block size may be guessed (resulting in allocation concealment being subverted). If the block size has been correctly guessed, then the last assignment in each block could be predicted with certainty, and potentially half the treatments in a block might be guessed. Varying the block length can be used to prevent this occurring. Implementing a variable-block length design is most easily done by a computer program. The program given in Appendix III was used to produce a list of treatment assignments for a study design with two treatments and random block sizes of 4 and 6 for an intended study population of 200 patients. The output from this program was stored in a table that, when read sequentially, gave the following treatment allocation schedule (assigning treatment A to 0 and treatment B to 1):

B A B A B A A A B B A A A B B B A B B A B A B A B A B B A A B  
 A B A B A A A B B A B A B B A A B A A B B A A B B B A B A B A  
 A B A A A B B B A A B A B B A B B A B B A B A A A B A B A A A  
 B B B A A A B B B A B B B A A B B A A A B A B B A A B A B A B  
 B A A B B A B B A A B A B B A A B A B A A A B B B B A A A A B  
 A B B B A A B A B B A B B A B A A B A A B A B B A B B A A A B  
 B A B A A B B B A A B A A B A B B A A B A B B B A A A B A B B  
 A A B A B A B B A B A B A A A B B A A A B B A

### Advantages of This Method

- The list can be prepared in advance and will ensure roughly equal numbers in each treatment group even if recruitment stops early.

TABLE 5 Random Permutations of 20 Numbers (0–19)

|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 9  | 18 | 17 | 5  | 3  | 10 | 2  | 4  | 11 | 13 | 12 | 14 | 6  | 7  | 15 | 0  | 19 | 8  | 1  | 16 |
| 4  | 1  | 15 | 10 | 5  | 14 | 3  | 18 | 17 | 9  | 0  | 19 | 2  | 12 | 6  | 11 | 8  | 13 | 7  | 16 |
| 19 | 0  | 5  | 11 | 12 | 4  | 7  | 1  | 13 | 16 | 3  | 10 | 17 | 9  | 18 | 14 | 15 | 6  | 8  | 2  |
| 11 | 10 | 7  | 12 | 18 | 1  | 15 | 8  | 17 | 2  | 3  | 0  | 19 | 4  | 6  | 14 | 9  | 13 | 5  | 16 |
| 7  | 5  | 3  | 16 | 2  | 6  | 12 | 4  | 14 | 1  | 17 | 19 | 0  | 11 | 18 | 8  | 9  | 13 | 15 | 10 |
| 3  | 2  | 12 | 8  | 17 | 16 | 19 | 0  | 6  | 5  | 11 | 9  | 15 | 14 | 13 | 1  | 10 | 4  | 18 | 7  |
| 6  | 17 | 19 | 0  | 3  | 15 | 7  | 13 | 8  | 5  | 4  | 18 | 2  | 16 | 9  | 11 | 1  | 14 | 12 | 10 |
| 5  | 9  | 11 | 10 | 4  | 14 | 15 | 2  | 8  | 12 | 18 | 3  | 17 | 0  | 19 | 13 | 6  | 7  | 16 | 1  |
| 12 | 15 | 19 | 0  | 3  | 2  | 9  | 13 | 16 | 1  | 5  | 17 | 18 | 4  | 6  | 11 | 10 | 8  | 7  | 14 |
| 10 | 12 | 4  | 11 | 1  | 3  | 8  | 5  | 16 | 9  | 2  | 14 | 15 | 13 | 17 | 6  | 7  | 18 | 0  | 19 |
| 8  | 13 | 10 | 18 | 19 | 0  | 5  | 12 | 9  | 17 | 3  | 2  | 14 | 11 | 15 | 4  | 1  | 7  | 16 | 6  |
| 8  | 6  | 19 | 0  | 13 | 16 | 7  | 1  | 9  | 3  | 2  | 12 | 18 | 14 | 10 | 11 | 4  | 5  | 15 | 17 |
| 3  | 7  | 12 | 10 | 4  | 6  | 9  | 17 | 15 | 2  | 1  | 5  | 14 | 8  | 13 | 18 | 11 | 16 | 0  | 19 |
| 17 | 12 | 4  | 2  | 16 | 15 | 5  | 10 | 6  | 13 | 8  | 19 | 0  | 9  | 3  | 18 | 7  | 1  | 14 | 11 |
| 18 | 17 | 10 | 7  | 12 | 11 | 4  | 13 | 6  | 16 | 0  | 19 | 1  | 2  | 15 | 3  | 9  | 8  | 5  | 14 |
| 11 | 4  | 3  | 9  | 18 | 8  | 7  | 14 | 10 | 5  | 16 | 2  | 13 | 6  | 12 | 15 | 1  | 0  | 19 | 17 |
| 5  | 13 | 9  | 4  | 18 | 7  | 1  | 12 | 14 | 2  | 16 | 3  | 17 | 10 | 11 | 6  | 15 | 8  | 0  | 19 |
| 4  | 7  | 9  | 6  | 18 | 13 | 11 | 10 | 16 | 3  | 17 | 2  | 5  | 8  | 14 | 1  | 12 | 15 | 0  | 19 |
| 2  | 9  | 19 | 0  | 5  | 17 | 6  | 15 | 14 | 10 | 4  | 1  | 12 | 3  | 16 | 8  | 7  | 13 | 11 | 18 |
| 9  | 3  | 13 | 11 | 14 | 5  | 8  | 10 | 18 | 7  | 16 | 0  | 19 | 12 | 1  | 17 | 15 | 2  | 4  | 6  |
| 14 | 10 | 15 | 2  | 13 | 4  | 11 | 17 | 1  | 16 | 18 | 6  | 0  | 19 | 7  | 3  | 12 | 8  | 9  | 5  |
| 2  | 14 | 18 | 11 | 5  | 13 | 6  | 16 | 17 | 10 | 8  | 7  | 3  | 15 | 9  | 0  | 19 | 1  | 12 | 4  |
| 11 | 16 | 7  | 3  | 9  | 1  | 6  | 10 | 15 | 18 | 12 | 14 | 13 | 0  | 19 | 4  | 17 | 8  | 2  | 5  |
| 8  | 19 | 0  | 16 | 2  | 3  | 6  | 1  | 11 | 4  | 5  | 9  | 7  | 12 | 10 | 18 | 13 | 15 | 17 | 14 |
| 13 | 10 | 5  | 8  | 16 | 9  | 7  | 14 | 4  | 15 | 3  | 0  | 19 | 6  | 17 | 2  | 12 | 18 | 1  | 11 |
| 18 | 7  | 9  | 14 | 8  | 5  | 19 | 0  | 15 | 3  | 13 | 11 | 10 | 2  | 17 | 1  | 16 | 4  | 12 | 6  |
| 16 | 19 | 0  | 14 | 15 | 9  | 6  | 8  | 2  | 11 | 12 | 4  | 1  | 7  | 5  | 17 | 3  | 18 | 13 | 10 |
| 16 | 6  | 15 | 13 | 9  | 14 | 0  | 19 | 2  | 1  | 18 | 4  | 8  | 7  | 17 | 5  | 3  | 11 | 12 | 10 |
| 12 | 4  | 16 | 7  | 10 | 5  | 8  | 1  | 11 | 13 | 3  | 15 | 14 | 9  | 6  | 2  | 18 | 17 | 0  | 19 |
| 12 | 6  | 5  | 14 | 10 | 2  | 18 | 4  | 17 | 1  | 3  | 15 | 0  | 19 | 16 | 13 | 7  | 11 | 8  | 9  |
| 19 | 0  | 9  | 13 | 5  | 12 | 16 | 18 | 8  | 7  | 4  | 10 | 11 | 6  | 15 | 17 | 2  | 1  | 3  | 14 |
| 19 | 0  | 6  | 9  | 4  | 11 | 13 | 1  | 3  | 10 | 7  | 14 | 5  | 16 | 18 | 17 | 2  | 12 | 15 | 8  |
| 5  | 2  | 12 | 4  | 3  | 10 | 1  | 11 | 17 | 15 | 14 | 6  | 16 | 0  | 19 | 8  | 13 | 18 | 9  | 7  |
| 17 | 2  | 3  | 6  | 14 | 12 | 11 | 8  | 19 | 0  | 18 | 16 | 15 | 13 | 4  | 10 | 7  | 9  | 1  | 5  |
| 18 | 2  | 11 | 9  | 3  | 5  | 10 | 12 | 16 | 4  | 19 | 0  | 7  | 8  | 1  | 13 | 15 | 6  | 17 | 14 |
| 5  | 10 | 12 | 19 | 0  | 18 | 15 | 17 | 2  | 14 | 3  | 1  | 9  | 8  | 13 | 7  | 11 | 4  | 16 | 6  |
| 13 | 6  | 18 | 14 | 3  | 19 | 0  | 8  | 9  | 10 | 7  | 16 | 17 | 4  | 2  | 15 | 11 | 5  | 12 | 1  |
| 18 | 9  | 17 | 6  | 2  | 8  | 3  | 16 | 10 | 5  | 11 | 12 | 4  | 7  | 15 | 1  | 0  | 19 | 13 | 14 |
| 16 | 18 | 13 | 4  | 10 | 12 | 14 | 3  | 9  | 0  | 19 | 17 | 2  | 11 | 15 | 5  | 7  | 6  | 8  | 1  |
| 7  | 13 | 9  | 8  | 19 | 0  | 6  | 12 | 2  | 10 | 11 | 1  | 15 | 16 | 5  | 17 | 3  | 18 | 14 | 4  |
| 3  | 5  | 14 | 4  | 11 | 9  | 12 | 15 | 13 | 8  | 18 | 2  | 1  | 16 | 7  | 17 | 6  | 10 | 0  | 19 |
| 14 | 16 | 15 | 18 | 12 | 17 | 5  | 11 | 7  | 10 | 2  | 6  | 1  | 13 | 0  | 19 | 8  | 9  | 3  | 4  |
| 11 | 6  | 10 | 14 | 5  | 13 | 9  | 15 | 17 | 1  | 12 | 2  | 7  | 18 | 3  | 19 | 0  | 16 | 8  | 4  |
| 2  | 3  | 16 | 7  | 12 | 5  | 13 | 4  | 14 | 6  | 19 | 0  | 18 | 15 | 17 | 1  | 9  | 10 | 11 | 8  |
| 8  | 12 | 2  | 9  | 15 | 13 | 14 | 5  | 16 | 18 | 3  | 17 | 4  | 19 | 0  | 6  | 7  | 1  | 10 | 11 |
| 12 | 6  | 5  | 13 | 3  | 19 | 0  | 2  | 7  | 9  | 11 | 18 | 10 | 4  | 15 | 8  | 17 | 1  | 14 | 16 |
| 9  | 12 | 8  | 5  | 14 | 19 | 0  | 7  | 6  | 11 | 4  | 10 | 13 | 18 | 1  | 15 | 2  | 17 | 16 | 3  |
| 8  | 14 | 4  | 16 | 13 | 17 | 7  | 19 | 0  | 12 | 5  | 11 | 10 | 15 | 18 | 3  | 9  | 2  | 6  | 1  |
| 14 | 12 | 13 | 19 | 0  | 18 | 10 | 5  | 15 | 8  | 6  | 4  | 17 | 11 | 16 | 9  | 7  | 2  | 3  | 1  |

### Disadvantages of This Method

- The smaller the block size, the greater the risk of predicting the next treatment assignment.
- Also, this method does not attempt to balance prognostic factors.

**The Biased Coin Method** Although blocking has been widely used, it may not be necessary to enforce such strict equality. To prevent major inequalities in treatment numbers, the biased coin design [4] can be used. This fairly simple method limits selection bias and maintains groups of roughly equal size. When a patient is to be randomized the probability of assignment to the treatment group with the least patients so far is assigned a probability  $p > \frac{1}{2}$ . If the two treatment groups are equal, then the probability  $p = \frac{1}{2}$ . Efron favored  $p = \frac{2}{3}$  and maintained that although  $p = \frac{3}{4}$  will maintain stricter balance, there is a greater chance of predicting the next allocation when there is an inequality. For a trial with  $N = 100$ , an inequality of 45:55 is not excessive but is extremely unlikely to occur with  $p = \frac{2}{3}$  [5].

*Example: Two Treatments,  $p = \frac{2}{3}$*  Let 1–6 correspond to assignment to the treatment with least patients, 7–9 to assignment to the treatment with most patients, and ignore 0.

If treatment numbers are equal, use simple randomization: Let 0–4 correspond to treatment A and 5–9 to treatment B. The top row of Table 1 would then produce the following treatment assignments:

|    |   |    |   |    |   |    |   |    |   |    |   |    |   |   |   |   |   |   |    |   |
|----|---|----|---|----|---|----|---|----|---|----|---|----|---|---|---|---|---|---|----|---|
| 5  | 5 | 6  | 3 | 1  | 3 | 9  | 1 | 1  | 4 | 5  | 5 | 4  | 8 | 1 | 0 | 7 | 3 | 4 | 8  | 3 |
| B* | A | B* | A | A* | B | B* | A | A* | B | B* | A | A* | A | B | — | A | B | B | B* | A |

where \* indicates simple randomization has been used.

To keep the allocation process as “random” as possible and to minimize the risk of prediction, it is possible to use a combination of methods. For example, use simple randomization unless the treatment imbalance exceeds some specified limit and then introduce a biased coin method to address the imbalance.

**Urn Method** Using Efron’s design, the bias of the coin,  $p$ , is a constant whatever the degree of imbalance. Wei [6] developed an adaptive biased coin method where the probability of assignment depends upon the magnitude of imbalance and number of patients treated.

For the urn design, assume an urn contains an equal number of balls of two types, A and B. When a patient is randomized, a ball is drawn and replaced. If the ball is type A, then the patient receives treatment A, and type B balls are added to the urn. If the ball is type B, then the patient receives treatment B, and type A balls are added to the urn. So the probability of assignment to the treatment with fewer numbers is increased.

The urn design is designated as UD ( $\alpha, \beta$ ) with  $\alpha$  being the number of balls of each type in the urn at the start and  $\beta$  being the number of balls added to the urn after each allocation. The simplest model has one ball of each type in the urn initially ( $\alpha = 1$ ), and one ball of the opposite type is added to the urn following each treatment allocation ( $\beta = 1$ ).

The program listed in Appendix IV was used to produce the following randomization list:

B A A A B B A A B B B A B B B A A B B B B A B A B  
A B A B A A B A A A A A B B A A B A A B B A A B B

### Advantages of These Methods

- The biased coin and urn designs control the likelihood of treatment imbalance without imposing strict balance. Therefore, they are less predictable than blocked designs and less susceptible to selection bias.

### Disadvantages of These Methods

- These methods are slightly more complex, requiring that totals of previous assignments are kept in order to calculate the probabilities to be used for the current assignment. This is best done with the aid of a computer program.
- Does not attempt to balance prognostic factors.

## 11.3 COVARIATE-ADAPTIVE RANDOMIZATION

In any clinical trial it is often desirable not only to achieve similar numbers of patients in each treatment group but also to ensure that patient groups are similar with respect to important prognostic factors such as age or gender. When using simple randomization or blocks, there is a chance that the treatment groups will differ with respect to important prognostic variables. Stratifying the randomization by the prognostic factors will reduce this imbalance. This will ensure that roughly equal numbers of patients will be allocated within each stratum to each of the treatment options either by simple randomization or by using permuted blocks. However, stratified randomization becomes unworkable as the number of prognostic factors increases because the number of strata required can quickly exceed the number of patients in the trial (see next section).

### 11.3.1 Stratification

Each combination of prognostic factors represents one stratum within the allocation list, and lists of random allocations are drawn up for each stratum. Multiple strata can become unmanageable and may lead to empty cells (the number of cells are the product of all the levels of the prognostic factors), therefore this method may not be suitable for small trials with several factors. The maximum desirable number of strata is unknown but according to Therneau [7] the number of strata in a two-treatment trial should be

$$\text{No. of strata} < \frac{N}{\text{block size}} \quad \text{and maybe} \quad \text{No. of strata} < \frac{N}{\text{block size} \times 2}$$

Imbalance between treatment groups can occur if there are strata with incomplete randomized blocks, and there is a possibility of overall treatment imbalance even

though each stratum has more or less equal numbers in each treatment. Retrospective balancing by means of an analysis involving covariates can take care of possible imbalances, but balancing provides a more efficient comparison of treatments and trials in which the prognostic factors balanced are far more convincing than sophisticated analysis alone.

International Conference on Harmonisation (ICH) Guideline E9 [8] recommends stratifying by center and recognizes that balancing by prognostic variables may sometimes be valuable, having greater potential benefits in small trials. If stratifying by center is not possible, the Committee for Proprietary Medicinal Products (CPMP) [9] recommend stratifying by country or region. However, if there are a large number of strata compared to the number of patients, there is no guarantee of balance at site level and no guarantee of treatment balance.

### 11.3.2 Random Permuted Blocks within Strata

The most commonly used method of stratification is to categorize patients into several types (or strata) and then to allocate patients within strata using permuted blocks. The block size within each stratum will normally need to be somewhat smaller than for unstratified random permuted blocks, since the number of patients within a given stratum may be quite small. The smaller the size of the trial and the larger the number of strata, the smaller the block size should be. However, blocks of size 2 should be avoided if there is any chance that an investigator could predict the next assignment.

**Example** In a recent study to investigate whether vaginal progesterone gel is effective in preventing preterm delivery in twin pregnancy (before 34 weeks of pregnancy), patients were randomized either to daily vaginal progesterone gel 90 mg or to placebo. The chorionicity of the pregnancy (monochorionic or dichorionic) was considered to be of major prognostic importance and was therefore included as a stratification factor along with the center. There were 6 recruiting centers that produced 12 ( $6 \times 2$ ) strata consisting of all combinations of center and chorionicity.

Before the trial started, a randomization list would be produced for each of the patient strata using the method described in Section 11.2.2. In this example a block size of 4 could be used, or a mix of block sizes 2 and 4, or 4 and 6, to prevent predictability of assignment. Table 6 shows the randomization lists for 240 patients using the output from the random permuted blocks of sizes 4 and 6 given in Section 11.2.2.

As patients enter the trial, they are identified as belonging to one stratum. Each patient is then allocated the next available treatment assignment from the corresponding randomization list. For example, if the first patient was from Center 1 and had monochorionicity, she would be allocated to treatment B. If the second patient was from Center 5 and had dichorionicity, she would be allocated to treatment A.

Some studies will have just one major prognostic factor for stratified randomization while others may have several. However, using too many strata increases the number of lists required and so increases the administrative burden. Another problem with having too many strata is the possibility of empty cells or cells with very few patients. Table 7 illustrates the characteristics of the first 136 patients randomized by showing their distribution across the 12 strata.

**TABLE 6** Example of Random Permuted Blocks within Strata for a Trial in Twin Pregnancies<sup>a</sup>

| Center       | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
|--------------|---|---|---|---|---|---|---|---|---|---|---|---|
| Chorionicity | M | M | M | M | M | M | D | D | D | D | D | D |
|              | B | B | A | B | B | B | A | B | A | B | A | A |
|              | A | A | B | A | B | B | B | A | B | B | B | B |
|              | B | B | A | A | A | A | A | B | A | A | B | A |
|              | A | A | B | B | B | B | B | A | B | A | A | B |
|              | B | B | B | A | A | B | B | A | B | A | A | B |
|              | A | A | A | A | A | A | A | A | A | B | B | A |
|              | A | B | A | A | B | A | A | B | B | B | A | B |
|              | A | B | B | B | B | A | B | B | B | A | B | A |
|              | B | A | A | B | A | B | B | B | A | B | B | B |
|              | B | A | A | B | B | B | A | B | B | A | B | A |
|              | A | B | B | A | A | A | B | A | A | A | A | A |
|              | A | A | B | A | A | A | B | A | A | B | A | A |
|              | A | B | A | B | A | A | A | A | B | B | A | B |
|              | B | A | A | A | B | B | A | A | A | B | B | B |
|              | B | B | B | B | B | A | B | B | A | A | A | A |
|              | B | A | B | B | B | B | A | A | B | A | B | A |
|              | A | A | B | A | A | B | B | B | A | B | B | A |
|              | B | A | A | B | A | A | B | B | B | A | A | B |
|              | B | B | B | B | A | A | A | B | B | A | A | B |
|              | A | B | A | A | B | B | A | A | A | B | B | A |

<sup>a</sup>M = Monochorionic, D = Dichorionic; A = Intervention; B = Control.**TABLE 7** Distribution of Patients across Strata in a Twin Pregnancy Trial

| Center No.    |    |   |    |    |    |    |
|---------------|----|---|----|----|----|----|
|               | 1  | 2 | 3  | 4  | 5  | 6  |
| Chorionicity  |    |   |    |    |    |    |
| Monochorionic | 4  | 1 | 5  | 5  | 2  | 2  |
| Dichorionic   | 33 | 8 | 17 | 13 | 32 | 14 |

Such an uneven distribution of patients across strata is typical and may possibly lead to imbalance. For a small trial with even more strata it is obvious that the chance of empty cells is even greater. In multicenter trials “center” is very often included as a stratifying variable as it is in this example. Indeed, this is recommended by the ICH Guidelines [8] and the Food and Drug Administration (FDA). However, for a trial involving a large number of centers this could lead to a very large number of strata. Therefore, in these cases it is worth considering using center as a minimization factor. This also reduces predictability of the next treatment allocation as centers are unaware of the characteristics of patients recruited by other centers. Overstratification can be particularly evident in small trials, but at the same time the possibility of serious imbalance is greater in small trials without stratification. An alternative approach would be to use minimization.

#### Advantages of this method

- Randomization lists can be made up before the trial starts.
- Treatment allocation is unpredictable as long as the block size is kept secret.



- Allows for interaction between variables (treatment groups are balanced across *combinations* of prognostic factors).

### Disadvantages of This Method

- The number of strata can quickly become large compared to the number of patients especially if center is one of the stratifying variables.
- When stratifying by center, it is desirable to keep the block size to a minimum as there is a risk of severe imbalance resulting from the cumulative effect of imbalance within each stratum. Recruitment officers do not need to know that blocked assignment is being used, and they should never be made aware of the block size.

### 11.3.3 Minimization

Minimization was first described by Taves in 1974 [10] and Pocock and Simon in 1975 [11] and aims to ensure treatment arms are balanced with respect to predefined patient factors as well as for the number of patients in each group. Minimization can be used when it is really important to achieve close similarity between treatment groups for several variables as it ensures balance between groups for several prognostic factors, even in small samples. Simulations show that minimization provides better balanced treatment groups when compared with restricted or unrestricted randomization, and it can incorporate more prognostic factors than stratified randomization methods such as permuted blocks within strata. Some more computationally complex methods may result in an even better performance. The loss in power resulting from treatment groups of unequal size is not likely to compromise a trial, but a serious imbalance in treatment groups with regard to a factor of prognostic importance can have severe consequences. Adjustment should always be made for minimization factors in the analysis (e.g., analysis of covariance).

**Example** For a hypothetical trial with two treatments, intervention and control, assume there are three prognostic factors: sex (male or female), age (<40, 40–60 or >60), and previous surgery (yes or no). A minimization program was prepared and 10 hypothetical patients randomized. Table 8 demonstrates how minimization works by considering the treatment assignment of the next patient. If the 11th patient considered is male, aged 45, and has had previous surgery, the decision to allocate to the treatment or control group can be made by comparing the totals in these categories for both treatments and allocating the patient to the treatment group that gives most balance overall with a high degree of probability. Using Taves' method, the patient would always (i.e., with 100% probability) be assigned to the treatment group that gives best balance, and the allocation is said to be deterministic as no random factor has actually been used.

Pocock and Simon [11] defined a more general method where treatment assignment depends on three functions: the amount of variation among assignments for any given factor level, a measure of the total imbalance in treatment numbers and assignment probabilities to the  $k$  arms of the trial  $\{p_i\}$  (where  $p_1$  is the probability of assignment to the arm that would lead to the least overall imbalance). Total imbalance is usually calculated by taking the sum of the individual imbalances, but

**TABLE 8 Example of How Minimization Works Using Hypothetical Trial Data**

| Prognostic Factor | Intervention | Control |
|-------------------|--------------|---------|
| Sex               |              |         |
| Male              | 4            | 3       |
| Female            | 1            | 2       |
| Age group         |              |         |
| <40               | 2            | 2       |
| 40–60             | 1            | 1       |
| >60               | 2            | 2       |
| Previous surgery  |              |         |
| Yes               | 2            | 4       |
| No                | 3            | 1       |

a weighted sum can be used if some factors are considered more important than others. If  $p_1 = 1$ , the assignment is deterministic (as in Taves) and goes automatically to the treatment with least overall imbalance, but the  $\{p_i\}$  can also be chosen to decrease the predictability of assignment. Table 8 provides a demonstration of how the range method works in practice.

If the 11th patient has factors: Male, Age 45, Yes, Has had previous surgery:

#### **Taves' Minimization**

Total in intervention group  $4 + 1 + 2 = 7$

Total in control group  $3 + 1 + 4 = 8$

The patient is allocated to the group with the lowest marginal totals.

Therefore the 11th patient is allocated to the intervention group because  $7 < 8$ .

#### **Pocock and Simon's Range Method** (Using Unweighted Sum and $p_1 = 1$ )

If allocated to intervention group, total imbalance is  $|(4 + 1) - 3| + |(1 + 1) - 1| + |(2 + 1) - 4| = 4$

If allocated to control group, total imbalance is  $|4 - (3 + 1)| + |1 - (1 + 1)| + |2 - (4 + 1)| = 4$

The patient is allocated to the group that would lead to less overall imbalance. However, in this case the groups are equal. Therefore, simple randomization would be used to allocate a treatment to the 11th patient.

Although minimization does not ensure balance between combinations of factors, it does consider frequently occurring combinations of subcategories indirectly since they tend to work together, giving them a greater chance of being equally distributed than if they were independent.

This method cannot be implemented simply by preparing a randomization list since it is necessary to keep a record of all patient factors and allocations to date, as in Table 8. The easiest way to keep track is to have a custom-designed program (there are some generic minimization programs available). Alternatively, many clinical trials units and commercial companies sell minimization services, and the use of these can help to eliminate allocation or investigator bias.

### Advantages of the Method

- Balances treatment assignment across several prognostic factors simultaneously and across treatment groups overall.
- There is increased credibility by presenting data where prognostic variables are closely balanced within each treatment group [12].
- Planning to use minimization is a good discipline for making trialists think about prognostic factors before a study starts and for helping ensure adherence to the protocol as the trial progresses [13].

### Disadvantages of the Method

- For these dynamic allocation methods assignment lists cannot be prepared in advance.
- Although balance in the whole trial will be good, balance in a subgroup may not be. This is because minimization regards all the factors as independent (no interaction).
- If  $p = 1$ , then the assignment is deterministic, but if no investigator can predict the assignment, this may not be relevant. Decreasing the probability value can lead to treatment imbalance, so there is a need to select that  $p$  that strikes a balance between minimizing imbalance and avoiding predictability of assignment. There is a counter argument that states that even with  $p = 1$  it would still be possible for any individual to be allocated to any treatment; this is determined by the order in which they are recruited to the study, in effect by chance. The random element in the assignment arises because the factors of the next patient come from a random distribution.
- It may be more difficult to administer than other methods especially when the number of prognostic factors increases.
- A computer program may be necessary to perform the randomization and, this has potential implications for cost, programming error, system failure, and training. If a well-tested generic program is not available, it is important to use simulations to check that the method is working correctly before starting the trial.

ICH E9 recognizes that dynamic allocation procedures such as minimization can achieve balance across a number of factors simultaneously but recommend avoiding fully deterministic assignment. The ICH guidelines state that deterministic allocation procedures should be avoided, and an appropriate element of randomization should be incorporated, and that details of the randomization that facilitate predictability should not be contained in the protocol. Minimization may be used with a biased coin in favor of the winner with a suggested value of 0.8. Further extensions to the method include using combinations of factors, or some factors may be weighted more heavily if they are considered to be more important. However, minimization does consider frequently occurring combinations of subcategories indirectly since they tend to work together giving them a greater chance of being equally distributed than if they were independent [10]. The CPMP states that not only does balancing protect against chance imbalance but it also increases the statistical power for small trials and increases the power and credibility of subgroup analyses.

### 11.3.4 Minimization Compared with Simple Randomization

For varying trial sizes, the imbalance that might occur following simple randomization was simulated using the characteristics of the first 3000 participants recruited to a UK multicenter trial evaluating two interventions for the secondary prevention of bone fractures as the population. Key prognostic factors for this trial were sex, age group, presenting fracture, and time since fracture. The trial size was varied from small (100 patients) to large (2000 patients). Imbalance was deemed to occur if the percent of participants in any of the prognostic categories varied by more than 10% between trial groups. The rate of imbalance decreased nonlinearly with trial size. Small trials ( $N \leq 100$ ) regularly displayed imbalance (over 80% of occasions), but in trials of 1000 participants or more, imbalance rarely occurred. Minimization ensured that balance was obtained on all occasions. There is, therefore, evidence of a trade-off between trial size and the need for minimization. Minimization is particularly useful for small to medium sized trials, but simple randomization may suffice when the trial is large. However, even large trials are small at the time of interim analysis and when producing reports for committees such as the Data and Safety Monitoring Committee (DSMC), so there is an argument that we should always use minimization.

### 11.3.5 Other Constrained Methods

Extensions to the minimization method include Atkinson's method, which uses optimum design theory [14]. Here the probability of allocation to the under-represented treatment responds to increasing imbalance rather than just being an arbitrary value. Klotz's method [15] addresses the question of a trade-off between uncertainty and imbalance. His method is similar to that of Pocock and Simon but is more computationally intensive and describes a function for calculating optimal treatment randomization probabilities. Titterton proposes a similar method [16] that involves minimizing a quadratic criterion subject to a balance constraint. The method aims to compromise between unpredictability of assignment and good stratification balance and claims to be simpler than Klotz's method. Signorini and co-workers outlined a dynamic balancing method [17] described as an easily implemented method that balances treatment allocations both within strata and across the trial as a whole. The method keeps a tally of total treatment allocations across all strata. Simulations have been used to demonstrate that the major imbalances that are possible with other schemes do not occur using this method, and the potential for selection bias is much reduced.

However, within the United Kingdom, these methods are very seldom used [18], and the results of a recent survey of trialists conducted by our group was to keep the method of randomization as simple as possible. When selecting a method, one must only balance on scientifically justifiable factors. If a stratified method can cope, then this is the method of choice; if there are too many factors, then use a dynamic method. Predictability may be an important factor as there is a risk of selection bias if the recruitment officer can guess the next allocation. Dynamic techniques help to make the allocation less predictable.

## 11.4 RESPONSE-ADAPTIVE METHODS

As a trial progresses, it may become evident that one treatment is performing better than the other. It may be possible to make use of this information by varying the allocation ratio to ensure more patients are allocated to the apparently better treatment. In the “play-the-winner” rule of Wei and Durham [19] the probability of assignment varies according to the success of the treatments. Initially, they conceptualize an urn containing equal numbers of balls representing the two treatments, A and B. A ball is drawn and replaced and the next patient is allocated to that treatment. If a treatment produces a successful outcome, then a ball of that type is added to the urn. If a treatment produces a nonfavorable outcome, then a ball of the other type is added to the urn. Therefore the probability of assignment is skewed to favor the treatment performing “better”.

This adaptive approach to treatment assignment results in a substantial reduction of patients on an inferior treatment compared to a conventional randomized trial. However, this design is not often used as the analysis of such designs is problematic. It is recommended that assignment procedures be simple, objective, and foolproof, and this is difficult to achieve with an adaptive design.

## 11.5 SPECIAL-CASE SCENARIOS

### 11.5.1 Unequal Randomization

In the majority of trials patients are randomized equally between the experimental and control groups. Sometimes it is preferable to favor the experimental treatment where anecdotal evidence suggests there may be a great health benefit, and patients may be reluctant to have only a 50% chance of receiving the new treatment. Such unequal allocation schemes may help to improve recruitment. It should be cautioned, however, that if the experimental treatment is harmful, unequal randomization in this situation may subject more patients to a harmful treatment.

If research costs between the treatments differ, it may be more cost effective to randomize more patients to the cheaper treatment. However, economic reasons are not the only criterion for selecting unequal randomization ratios, and in some cases it may even be preferable to favor the more expensive treatment. For example, when evaluating the learning curve for a new surgical intervention, more patients may be randomized to the experimental (or expensive) treatment [20]. For many trials there is a substantial amount of previous experience relating to the control treatment, and there is more to be learnt about the new treatment [5].

When experimental treatments differ in their costs, then for a given statistical power, unequal randomization will produce the least-cost trial. Unequal randomization is unnecessary if there is no cost difference between the treatments. However, it may be possible to alter the randomization ratio as the trial proceeds if large cost differences become apparent. If unbalanced allocation is used, it is generally justified on the grounds of ethics or cost. Randomization ratios of 2:1 or 3:2 do not greatly alter the statistical properties of a trial, but loss of power appears to be substantial for a ratio of 3:1 or more extreme [5]. The chance of obtaining a statistically significant difference between the two treatments is not significantly

reduced, as long as the chosen ratio is less extreme than 70:30 [21]. Formulas exist to adjust standard sample size estimates to accommodate for unequal randomization [22].

### 11.5.2 Cluster Randomization

A cluster-randomized trial is one in which clusters of individuals, rather than individuals themselves, are randomized to different treatment groups. Cluster-randomized trials have become particularly widespread in the evaluation of non-therapeutic interventions such as lifestyle modification or educational programs. The units of randomization range from relatively small clusters such as families to entire communities such as neighbourhoods, hospitals, or medical practices. In a trial to investigate if training in a certain consultation procedure is effective, then some physicians would receive training and others would not. The patients being seen by a physician would form a cluster. If the individual patients were randomized, then they may have to be seen by another physician as it would prove difficult for the physician to switch consultation practice between patients and contamination could occur.

The clusters are often decided in advance and randomized before the intervention is applied. However, the subjects themselves are sometimes not identified at the time of randomization, and so the randomization will not balance the prognostic factors as in individual randomization. The purpose of randomization in cluster trials is to try and balance the confounding factors associated with the cluster [23]. Cluster unit characteristics will balance as the number of clusters increases, and so the patient characteristics will balance as a result. The methods described in Sections 11.2–11.4 can all be used to allocate clusters.

The design and conduct of cluster-randomized trials requires special considerations [24]. They are often more complex to design, require more participants, and require more complex analyses than individually randomized trials. Cluster randomization is less efficient in a statistical sense than randomizing individuals to treatment groups [25] because the responses of individuals within a cluster tend to be more similar than responses of individuals in different clusters. The theory of experimental design assumes that the experimental unit that is randomized is also the unit of analysis, and this is not necessarily the case in a cluster randomization trial. Analyses that do not take the clustering into account may report significance when none exists. Increasing the number of clusters enhances power more efficiently than increasing the number of subjects within a cluster.

#### **Advantages of This Method**

- Cluster randomization is often used for practical or ethical reasons when patient randomization is not possible.
- Clustering helps to avoid contamination.

#### **Disadvantages of This Method**

- Loss of statistical efficiency
- Need to recruit more participants

## 11.6 SUMMARY

In conclusion, the most important factors when choosing a randomization method are the size of the trial, practicality of the method, maintaining the blind, and reducing predictability. Smaller trials may require some form of restricted allocation to maintain balance, whereas simple randomization may suffice for large trials. Keeping the randomization method as simple as possible will reduce time, cost, and programming errors.

If the number of important prognostic factors and layers within these is sufficiently small, then the preferred method of randomization, recommended by ICH, is permuted blocks of varying random length within strata. This method eliminates the problem of predictability and at the same time balances across combinations of factors, whereas minimization only balances over single factors. If the number of prognostic factors is large, then minimization can be used to provide treatment balance as well as balance over these factors. However, only those factors known to affect outcome should be considered.

Web-based systems may become more widely used in the future, but there are some problems to be overcome, including access to the Internet, speed and reliability of access, and the lack of willingness in trial centers to adopt new technologies.

## APPENDIX I

Visual Basic for Applications (VBA) code to produce a table of 2500 random numbers.

```
Private Function GenerateList()
    Dim Num As Integer
    Dim MyValue
    Dim Numbers(25) As Integer
    Dim intI As Integer

    Open "random.txt" For Output As #1 ' Open file.
    For Num = 1 To 100

        For intI = 0 To 24
            Randomize ' Initialize random-number
            generator.
            MyValue = Int((10 * Rnd)) ' Generate random
            value between 0 and 9.
            Numbers(intI) = MyValue
        Next

        Print #1, Numbers(0); " "; Numbers(1); " ";
        Numbers(2); " "; Numbers(3); " "; Numbers(4); " ";
        Numbers(5); " "; Numbers(6); " "; Numbers(7); " ";
```



```

Numbers(8); " "; Numbers(9); " "; Numbers(10);
" "; Numbers(11); " "; Numbers(12); " ";
Numbers(13); " "; Numbers(14); " "; Numbers(15);
" "; Numbers(16); " "; Numbers(17); " ";
Numbers(18); " "; Numbers(19); " "; Numbers(20);
" "; Numbers(21); " "; Numbers(22); " ";
Numbers(23); " "; Numbers(24);
Print #1,

Next Num
Close #1

End Function

```

## APPENDIX II

Visual Basic for Applications (VBA) code to produce a table of 100 permutations of 20 numbers (0–19).

```

Private Function Permutations()
    Dim Num As Integer
    Dim intI As Integer
    Dim db As Database
    Dim rst As Recordset
    Dim rstAdd As Recordset

    Set db = CurrentDb
    Set rstAdd = db.OpenRecordset("TblPermute20")

    For Num = 1 To 100
        rstAdd.AddNew
        Set rst = db.OpenRecordset("QrySort20")
        For intI = 0 To 19
            rstAdd(intI) = rst!ID
            rst.MoveNext
        Next intI
        rstAdd.Update
    Next Num

    rst.Close
    rstAdd.Close

    MsgBox "Done!", vbOKOnly

End Function

```

**APPENDIX III**

Visual Basic for Applications (VBA) code to produce a table of approximately 200 treatment allocations in random blocks of 2 and 4.

```

Private Function Blocks()
    Dim db As Database
    Dim arrBlock() As Integer
    Dim intBlock As Integer
    Dim intBlockSize As Integer
    Dim intCount As Integer
    Dim intResult As Integer
    Dim strInsert As String

    Set db = CurrentDb

    For intBlock = 1 To 45
        Randomize

        If (Rnd > 0.5) Then
            intBlockSize = 6
        Else
            intBlockSize = 4
        End If

        ReDim arrBlock(intBlockSize - 1)

        For intCount = LBound(arrBlock) To UBound(arrBlock)
            arrBlock(intCount) = 0
        Next

        intCount = 1
        While intCount <= intBlockSize / 2
            intResult = Int((intBlockSize - 1 + 1) * Rnd + 1)
            If arrBlock(intResult - 1) <> 1 Then
                arrBlock(intResult - 1) = 1
                intCount = intCount + 1
            End If
        Wend

        For intCount = LBound(arrBlock) To UBound(arrBlock)
            strInsert = "INSERT INTO TblAllocate(Treatment)
            VALUES(" & arrBlock(intCount) & ")"
            db.Execute strInsert
        Next
    Next
End Function

```

## APPENDIX IV

Visual Basic for Applications (VBA) code to produce a table of 50 treatment allocations using the Urn method.

```
Private Function Urn()
    Dim db As Database
    Dim strInsert, Treat As String
    Dim Num, x As Integer
    Dim randnum, p As Double

    Set db = CurrentDb

    Num = 0
    For x = 1 To 50
        Randomize

        If x = 1 Then
            p = 0.5
        Else
            p = Num / (x - 1)
        End If
        If (Rnd < p) Then
            Treat = "A"
        Else
            Treat = "B"
        End If
        If Treat = "B" Then
            Num = Num + 1
        End If

        strInsert = "INSERT INTO TblUrn(Treatment) VALUES('" &
            Treat & "')"
        db.Execute strInsert
    Next x

End Function
```

## REFERENCES

1. Fergusson, D., Aaron, S. D., Guyatt, G., et al. (2002), Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis, *BMJ*, 325, 652–654.
2. Altman, D. G., and Schulz, K. F. (2001), Statistics notes: Concealing treatment allocation in randomised trials, *BMJ*, 323, 446–447.
3. Byrom, B. (2004), Electronic diary solutions: Enhanced collection of patient reported outcome data, *EBR*, Autumn, 90–94.

4. Efron, B. (1971), Forcing a sequential experiment to be balanced, *Biometrika*, 58, 403–417.
5. Pocock, S. J. (1979), Allocation of patients to treatment in clinical trials, *Biometrics*, 35, 183–197.
6. Wei, L. J. (1978), An application of an urn model to the design of sequential controlled clinical trials, *J. Am. Statist. Assoc.*, 73, 559–563.
7. Therneau, T. M. (1993), How many stratification factors are “too many” to use in a randomization plan? *Controlled Clin. Trials*, 14, 98–108.
8. ICH Steering Committee (1998), ICH Harmonised Tripartite Guideline E9; Statistical Principles for Clinical Trials.
9. Committee for Proprietary Medicinal Products (CPMP) (2003), The European Agency for the Evaluation of Medicinal Products (EMA). Points to consider on adjustment for baseline covariates. CPMP/EWP/2863/99 at [www.emea.europa.eu](http://www.emea.europa.eu).
10. Taves, D. R. (1974), Minimization: A new method of assigning patients to treatment and control groups, *Clin. Pharmacol. Ther.* 15, 443–453.
11. Pocock, S. J., and Simon, R. (1975), Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, *Biometrics*, 31, 103–115.
12. Brown, B. W. (1980), Statistical controversies in the design of clinical trials, *Controlled Clin. Trials*, 1, 13–27.
13. Day, S. (1999), Commentary: Treatment allocation by the method of minimisation, *BMJ*, 319, 947–948.
14. Atkinson, A. C. (1982), Optimum biased coin designs for sequential clinical trials with prognostic factors, *Biometrics*, 69, 61–67.
15. Klotz, J. H. (1978), Maximum entropy constrained balance randomization for clinical trials, *Biometrics*, 34, 283–287.
16. Titterton, D. M. (1983), On constrained balance randomization for clinical trials, *Biometrics*, 39, 1083–1086.
17. Signorini, D. F., Leung, O., Simes, R. J., et al. (1993), Dynamic balanced randomization for clinical trials, *Stat. Med.*, 12, 2343–2350.
18. Scott, N. W., McPherson, G. C., Ramsay, C. R., et al. (2002), The method of minimization for allocation to clinical trials: A review. *Controlled Clin. Trials*, 23, 662–674.
19. Wei, L. J., and Durham, S. (1978), The randomized play-the-winner rule in medical trials, *J. Am. Statist. Assoc.*, 73, 840–843.
20. Torgerson, D., and Campbell, M. (1997), Unequal randomisation can improve the economic efficiency of clinical trials, *J. Health Services Res. Policy*, 2, 81–85.
21. Peto, R., Pike, M. C., Armitage, P., et al. (1976), Design and analysis of randomized clinical trials requiring prolonged observation of each patient, *Br. J. Cancer*, 34, 585–612.
22. Campbell, M. J., Julious, S. A., and Altman, D. G. (1995), Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons, *BMJ*, 311, 1145–1148.
23. Machin, D., and Campbell, M. J. (2005), *Design of Studies for Medical Research*, Wiley, Chichester, p. 274.
24. Campbell, M. K., Grimshaw, J. M., and Elbourne, D. R. (2004), Intraclass correlation coefficients in cluster randomized trials: Empirical insights into how they should be reported, *BMC Med. Res. Method.* 4.
25. Donner, A., and Klar, N. (2000), *Design and Analysis of Cluster Randomization Trials in Health Research*, Arnold, London.

# 12

---

## *Randomized Controlled Trials*

**Giuseppe Garcea<sup>1</sup> and David P. Berry<sup>2</sup>**

*<sup>1</sup>Cancer Studies and Molecular Medicine, The Leicester Royal Infirmary,  
United Kingdom*

*<sup>2</sup>Department of Hepatobiliary and Pancreatic Surgery, The Leicester General Hospital,  
United Kingdom*

### **Contents**

|                                                               |     |
|---------------------------------------------------------------|-----|
| 12.1 Introduction                                             | 807 |
| 12.2 Uncontrolled Trials                                      | 809 |
| 12.3 Problems with Uncontrolled Trials                        | 810 |
| 12.4 Historical Controls                                      | 811 |
| 12.5 Problems with Historical Controls                        | 812 |
| 12.6 Concurrent Nonrandomized Controlled Trials               | 814 |
| 12.7 Problems with Concurrent Nonrandomized Controlled Trials | 816 |
| 12.8 Is Randomization Feasible?                               | 817 |
| 12.9 Conclusion                                               | 819 |
| References                                                    | 820 |

### **12.1 INTRODUCTION**

A randomized controlled trial is a study in which subjects are allocated at random to receive one of two possible interventions. The outcomes measured are events or quantifiable changes that are present or absent following the proposed interventions. These outcomes are known as endpoints. In general, one of the interventions chosen is a standard of comparison: the control. The control intervention can be a drug or procedure with no known clinical effect: a placebo. Alternatively, the

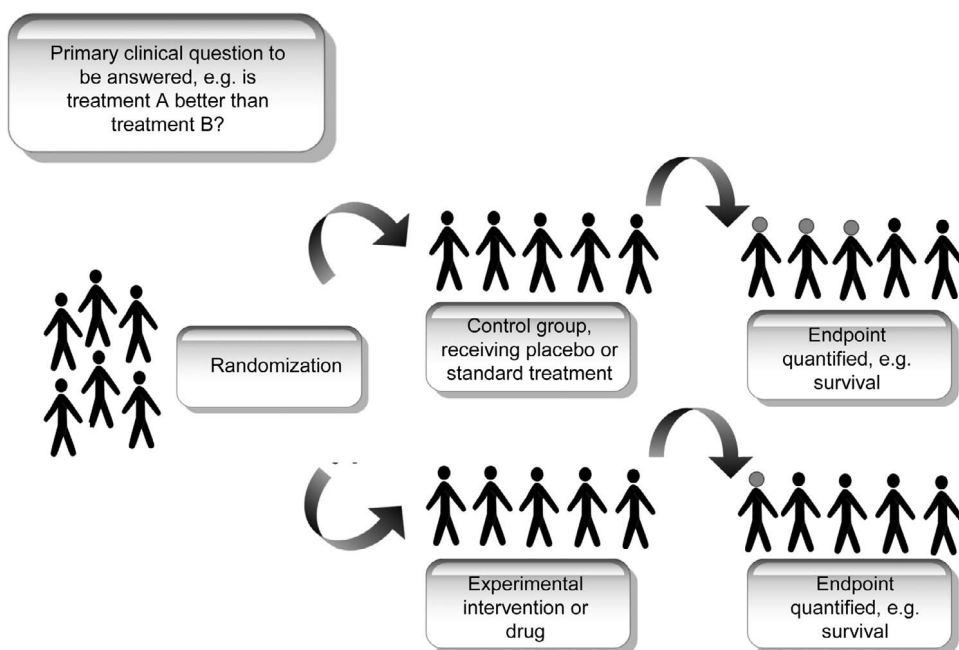


**FIGURE 1** Structure of randomized controlled trials.

control can be a therapeutic measure that is currently the accepted treatment or intervention for a disease process. The other group is the experimental intervention or trial drug (Fig. 1).

Randomization refers to the random allocation of the subjects to one of the intervention groups. Therefore, the allocation of patients is not determined by the clinical investigators or the subjects within the study. The random allocation of subjects to the treatment arms of the study ensures that the demographics of the subjects are likely to remain similar. Randomization also removes investigator bias in the allocation of patients to the various treatment arms. Hence, randomization reduces the risk of imbalance between the subjects receiving the trialled intervention or drug and those receiving the standard intervention or placebo drug. This balancing of prognostic factors between the two groups increases the likelihood that any observed difference is attributable solely to the proposed treatment.

A number of different endpoints can be employed to quantify the response to a trialled intervention. Endpoints can vary from laboratory-measured responses, such as changes in cell-signaling molecules or histological changes, known as biomarkers, through to clinical responses. Ideally, randomized controlled trials should have one primary objective, asking one specific question and one primary outcome or endpoint. Both the primary objective and endpoints should be well defined and easily measured. A quantifiable objective endpoint is superior to a subjective one. Many randomized controlled trials will also have secondary objectives and endpoints. In general, conclusions drawn from these secondary objectives should be regarded as tentative since the study may lack the power to address them.

A randomized controlled trial (RCT) should answer a clinical need (Fig. 2). Any intervention studied should be feasible in clinical practice and remain a viable treat-

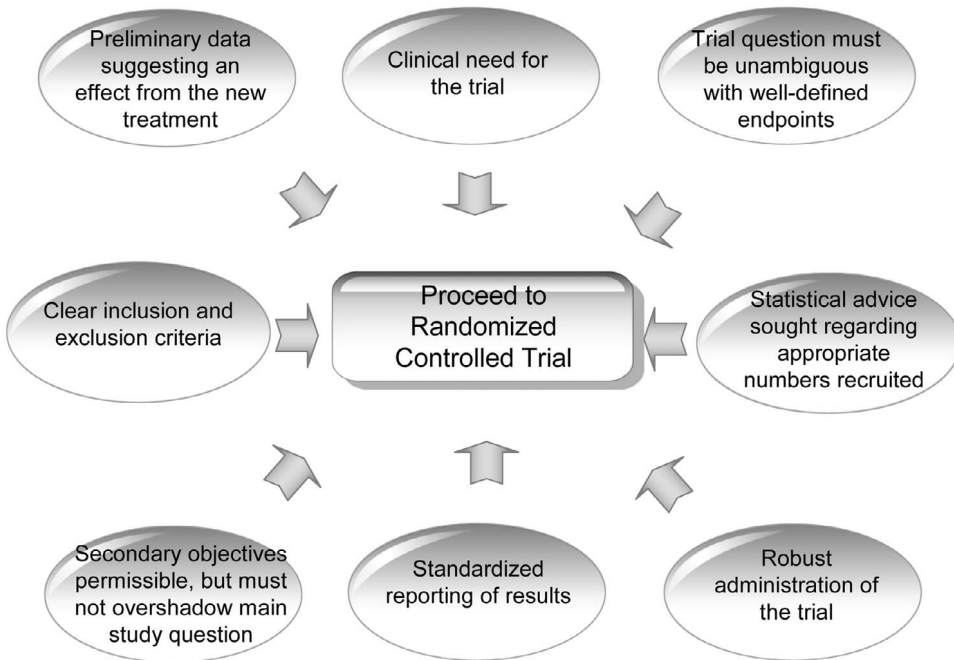


**FIGURE 2** Issues in designing a randomized controlled trial.

ment following the results of the study. The disease process in question should be of a serious nature or impact upon a significant proportion of the population. There must be preliminary evidence (from phase I and phase II trials) that the trial intervention has a promise of efficacy. However, there also needs to be a genuine uncertainty regarding the efficacy of the new proposed treatment over the standard treatment or placebo treatment, known as equipoise. The design of the RCT should be robust and proper administration of the study is essential. Standardized reporting of the results of any RCT, combined with valid statistical analyses allows data from multiple RCTs to be compared for a more precise estimate of the effects of a treatment.

In practice other study methodologies have been used to answer the same clinical questions tackled by RCTs. This chapter will examine these methods and discuss their potential flaws.

## 12.2 UNCONTROLLED TRIALS

In some circumstances it may not be possible to conduct randomized controlled trials. This may be due to ethical problems (e.g., examining the effect of cessation of smoking and the risk of lung cancer), financial considerations, or, as occurs frequently in surgical practice, problems in recruiting adequate numbers of patients. In these scenarios, uncontrolled trials may give important clinical information. Uncontrolled studies are essentially observational trials where the effect of an intervention is assessed on a population of patients (a cohort) over a period of time. In uncontrolled trials, there is no control group and hence all endpoints are essen-



tially qualitative (in respect to the patient's initial symptom or disease status). In drug therapy, uncontrolled trials may be a useful bridging step from translating in vitro and in vivo evidence of drug efficacy into a clinical setting. Hence, the results from uncontrolled clinical trials may be used to rationalize and plan subsequent RCTs of the same intervention. In particular, uncontrolled clinical trials can provide important safety and tolerability data in the setting of drug trials.

Surgery is one field where the RCT has suffered a notable decline [1–4]. In one report, less than 40% of the published hypotheses involving surgical operations for gastrointestinal surgery could have been answered successfully by an RCT [5]. The main problems encountered included accrual of patients due to a low disease incidence and patient preference. For example, the mastectomy versus lumpectomy RCT for breast cancer suffered from significant poor accrual into the latter treatment group [6]. In such circumstances a noncontrolled cohort study may be the only means of evaluating the efficacy of a surgical intervention.

Uncontrolled trials are also useful for chronic conditions when compared to placebo controlled RCTs. In this group of patients long-term follow-up during RCTs is plagued by poor patient adherence to treatment regimens, especially in the placebo control group. There are also ethical considerations in administering a nontherapeutic drug for long periods to patients with chronic illness, which is severely restricting their quality of life. In such scenarios patients taking their preferred (and therapeutic) treatments are more likely to comply with the trial, and hence longer-term follow-up can be achieved with minimal loss of patients recruited to the trial. Uncontrolled observational studies can also recruit larger numbers of patients with fewer inclusion and exclusion criteria than RCTs. One criticism of RCTs is that they often recruit relatively small numbers of patients (in comparison to the disease burden) who are then treated and monitored under conditions that do not reflect everyday practice [7, 8].

### 12.3 PROBLEMS WITH UNCONTROLLED TRIALS

While we have discussed the clinical pressures that necessitate the undertaking of noncontrolled trials, there are major disadvantages in this type of study and assessing the reliability of the evidence reported. The lack of a control group makes it impossible to compare any new treatment with no treatment or with the “gold standard” of treatment currently adopted with any particular disorder. An example would include a hypothetical series of laparoscopic hernia repairs reported by surgeon A, which reveals a low recurrence rate, a high rate of same-day discharge, and a high level of patient satisfaction following the procedure. With no control group, we cannot compare this treatment with the current standard method of treating hernias (i.e., an open repair). As a result, although the results may indicate that laparoscopic surgery is a viable technique in dealing with hernias, we cannot assess if it is more efficacious than the currently adopted treatment regimen. In the field of testing new drugs, this may have important implications since many new drugs available on the market would incur significantly greater costs if their use was to be adopted universally, when compared to cheaper and better established medication.

Furthermore, with the lack of a control, any observed effect of drug or intervention is subject to systematic error or bias within the study. This bias may arise from

subject selection, outcome measurement, or other confounding factors. For example, in the case of a new drug treatment for insomnia, the patients selected may be well-motivated individuals who, in addition to taking the new trial drugs, have also instituted other measurements to ease stress and hence improve their ability to sleep at night. The investigator is clearly aware that all individuals assessed following the drug treatment have received the trial drug and this may affect their data collection. In studies where qualitative assessments are to be used (such as the quality of sleep overnight), this observer bias is likely to be of particular importance. Finally, the placebo effect of the new drug has not been accounted for. The placebo effect is a well-recognized response of patients to new treatments or drugs. The fact that these patients with insomnia have been selected to trial a new and potentially efficacious medication may in itself help to ease their symptoms, even if the trial drug had no pharmaceutical action at all.

Uncontrolled trials can be a useful adjunct in determining drug tolerance and acceptability in phase I and phase II clinical trials. However, results from these trials, however encouraging, should be interpreted with caution pending control group studies.

## 12.4 HISTORICAL CONTROLS

It is possible to compare outcome from new treatments in trials with data from previous studies examining alternative treatments or no treatment. Hence, the control group is named a historical control (Fig. 3). A major advantage in using

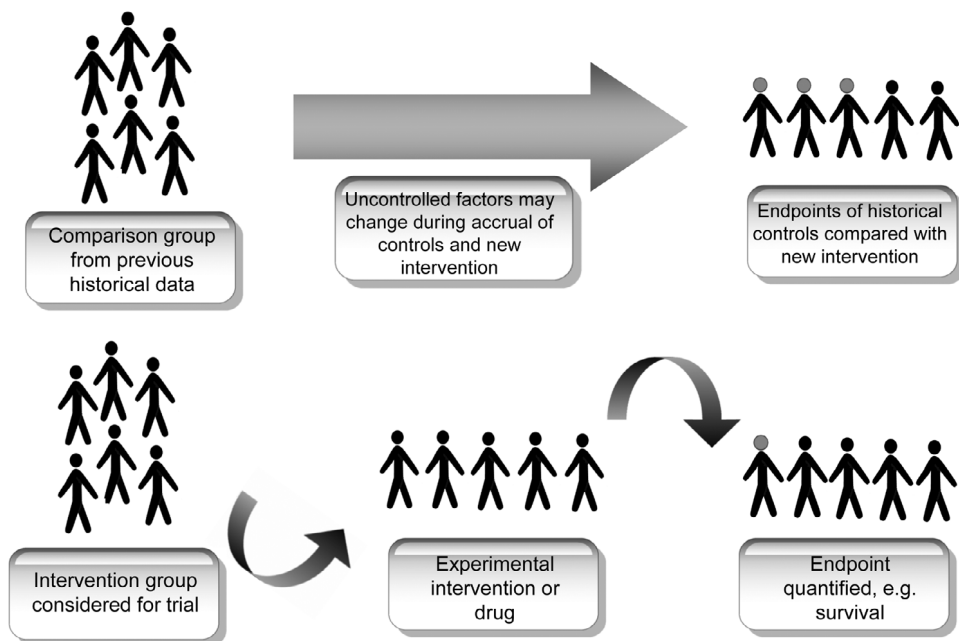


**FIGURE 3** Structure of historical control trials.

historical controls is that the studies are easier to perform, with fewer patients and shorter study period duration than studies that use concurrent controls. The treatment of diabetes mellitus with insulin, chemotherapy for childhood leukemia, and vitamin B<sub>12</sub> therapy for pernicious anemia all represent major medical advances that have been instituted as a result of data obtained from historical controls [9–12].

Historical controls may also be instituted for conditions that are sufficiently rare to result in a low rate of accrual of eligible patients for concurrent control studies. By comparing data from historical controls, it may be possible to evaluate new treatments in a much shorter time period than would otherwise be possible. In areas where financial considerations may be a limiting factor, historical control groups are an attractive proposition since they avoid the need to recruit control groups and randomization for treatment allocation. In doing so, historical controls reduce the need of support personnel required and cost.

The final argument for historical controls is an ethical one. In general, RCTs compare a newer and presumed better drug with standard therapy. If no effective established therapy exists for a particular condition, it may be argued that ethically all patients should receive the new treatment. For example, if comparing insulin treatment for the treatment of diabetes when compared to no treatment at all, those patients allocated to the control arm of an RCT would be denied a life-saving treatment.

## **12.5 PROBLEMS WITH HISTORICAL CONTROLS**

The major drawback to using historical controls is that many factors may change in the time period from accrual of data from historical controls to the implementation of the new drug or intervention in the intervention subjects. These differences may reflect the continuous improvement of medical care, changes in the demographics of patients referred, changes in referral patterns (such as earlier referral due to increased awareness of a particular condition), and improvements in epidemiological factors (better diet, increased exercise). For the intervention group, the increased monitoring and attention will also have a strong placebo effect, which will not be adequately controlled for among the historical controls.

In the setting of cancer drug trials, for example, the use of historical controls relies on the assumption that the survival rate among the controls would be the same as that among the control group in a concurrent RCT. The factors discussed in the preceding paragraph suggest that this may not be true. These confounding factors can be minimized if the data from the historical controls is from patients who were all treated at the same institution. This would ensure that referral patterns, medical care, nursing care, and other support would be as similar as possible to the intervention group [10]. It is probable, however, that some differences will inevitably exist between the control and intervention group.

Another potential flaw in the use of historical controls is that for most trialled treatments there are restrictions on which patients are selected for the intervention group. For example, in the case of chemotherapy agents the patients with worst prognosis may be excluded from the trial since they may be least likely to benefit

from the therapy. These exclusion criteria cannot be applied to the historical controls. Since exclusion and inclusion criteria cannot be adequately controlled for in the historical group, this invalidates any comparison between them.

Due to the overall improvements in medical and nursing care over the time interval separating intervention groups from their historical controls, there is a tendency for historical control trials to overemphasize the magnitude of any particular intervention. For example, the use of anticoagulant therapy following acute myocardial infarction using data from historical controls revealed a 54% reduction in mortality [13]. Subsequent concurrent RCTs showed a 21% reduction in mortality. This discrepancy occurred in spite of broad database consisting of 18 studies and 9000 patients in all. While it may not be possible to eliminate this “historical error,” it is possible to give an approximation as to the degree of error for a historically controlled study. This may be achieved by portioning the period of control entry into different intervals. The results obtained between those groups can be then compared to the overall outcome. If a significant reduction in the benefit from a new intervention is observed across the different groups, then the degree of error due to changes in clinical practice is significant (Fig. 4).

In a study by Diehl and Perry the survival times and disease-free survival times of historical controls from chemotherapy trials were compared with those from control groups derived from matched RCTs [9]. The RCTs were matched for location of disease, stage, and follow-up. This study showed that 42% of studies varied by more than 10 percentage points, 21% varied by more than 20 percentage points, and 5%

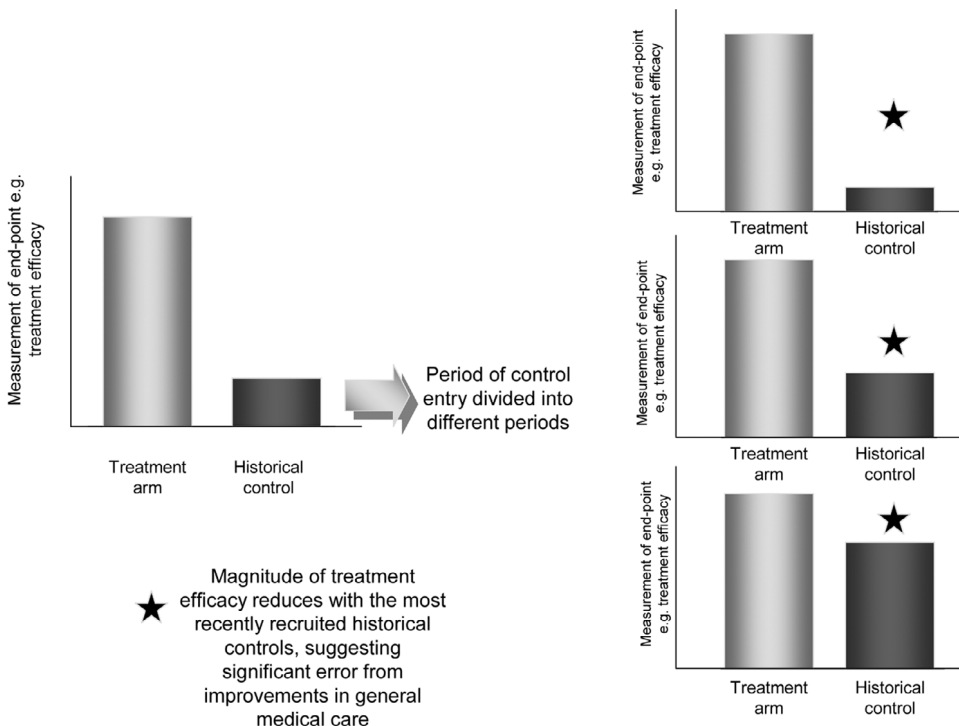


**FIGURE 4** Schematic representation from error occurring in historically controlled trials.

varied by more than 30 percentage points. It also revealed that even for institutions that had both a historical control and randomized control group, there were still significant differences between survival and disease-free survival. In summary, although historical controls offer certain advantages over RCTs in terms of simplicity, speed, and cost, they cannot be used to reliably control treatment studies.

## 12.6 CONCURRENT NONRANDOMIZED CONTROLLED TRIALS

In certain scenarios a nonrandomized design may be applicable. Nonrandomized trials still include a control and intervention group. However, allocation of patients to control or treatment arms is determined by the investigator or the patient themselves (Fig. 5). There are numerous valid reasons for the use of nonrandomized controlled trials in clinical practice. Surgery is an area where the conduction of RCTs is especially problematic. These relate to the difficulty in patients allowing themselves to be randomized. Patients who may be eligible for randomization may withhold consent to be randomized. This may be due to the fact that the new treatment under trial may be less cumbersome, less traumatic, or less disfiguring than the standard therapeutic procedure. Patients undergoing surgery will need to coordinate the timing of their procedure and recovery time around their employment and hence may be unwilling to be randomized into a treatment arm that would cause more disruption for them.

A similar circumstance can be envisaged in a hypothetical chemotherapy trial comparing an intravenous chemotherapy regime, with one using orally administered medication. The former treatment arm will necessitate the positioning of

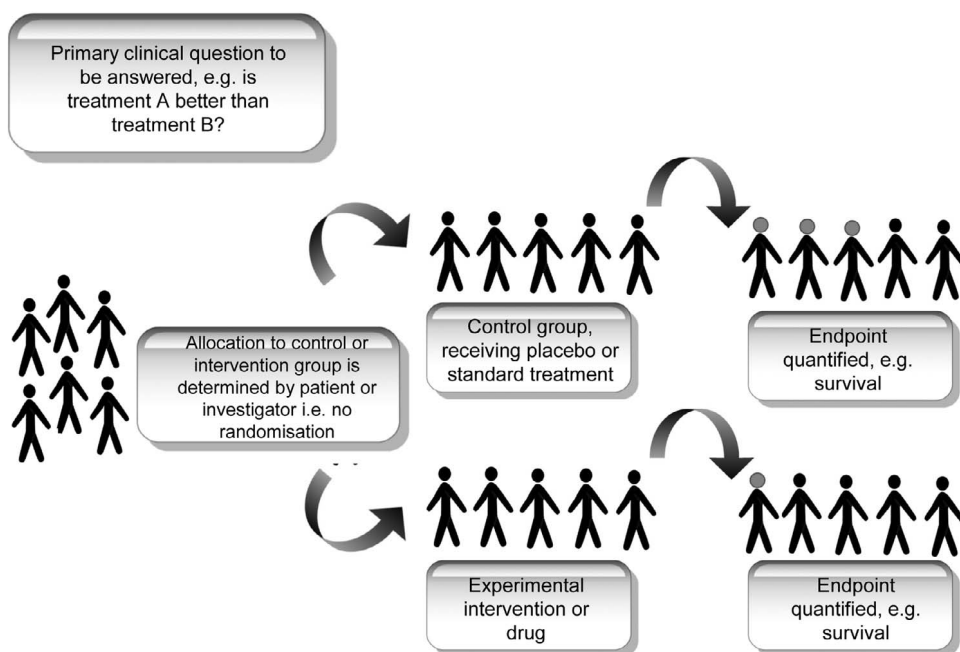


**FIGURE 5** Structure of nonrandomized concurrent control trials.

intravenous cannulae and a visit to a hospital daycare unit, while the oral preparation may be taken at home. Many patients would prefer the less inconvenient option, particularly if the oral preparation is the newer and perceived (if only by the patients) more effective treatment for their cancer.

For certain diseases, the population of patients may be too small to accrue sufficient numbers for an effective RCT. In such situations, every patient is valuable, and hence the investigators may allow individuals to choose their treatment in order to accumulate sufficient numbers of patients. Finally, there may be ethical considerations that make randomization unfeasible. For example, in the case of the hypothetical orally administered chemotherapy agent discussed above, patients with extensive carcinomatosis and a poor prognosis, randomization into the standard treatment arm will have a significant impact on their remaining quality of life. It may be a better option to allow patients to choose their treatment, which, in turn, may result in better compliance among those patients who have chosen to be in the treatment arm examining standard therapy.

The perceived disadvantages in losing randomization has been challenged by Abel and Koch [14]. The argument for randomization is that it leads to comparable and balanced groups and so allows comparison of the new treatment arm of the study with the standard treatment or placebo arm. Having ensured that baseline characteristics have been eliminated by randomization, any differences in effect are attributable to the interventions applied. However, this assumption is inaccurate as it does not take into account the many variables that may affect outcome following allocation to treatment or control groups. Figure 6 summarizes the causes of differ-

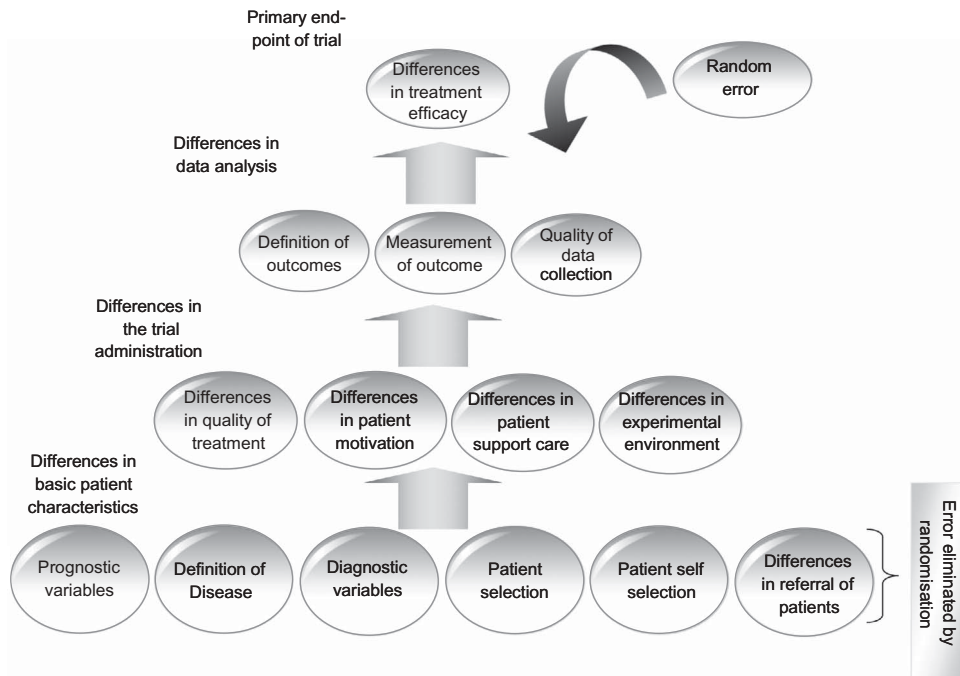


**FIGURE 6** Pyramid of possible reasons for differences observed between treatment in clinical trials and the effect of randomization on eliminating errors. (Adapted from Abel and Koch [14].)

ences between control and treatment groups and which ones are eliminated by randomization. It is clear that there are still many more confounding factors that may alter the findings of a study erroneously.

Blinding is a concept in which patients and physicians are unaware of which treatment arm the patient has been assigned. Blinding prevents patients from overstating the benefits of a new medication in order to build rapport or please their attending doctor. It also prevents physicians from introducing observer bias when recording and interpreting the results of a tested intervention. Another stated disadvantage to nonrandomized controlled trials is that blinding is not possible. While randomization is essential for blinding of patients, it is not a prerequisite for blinding of treatment evaluation.

Nonrandomized controlled studies are a response to a clinical need to evaluate a new therapy, where the applicability of a conventional RCT is in doubt. A comparison of 30 randomized and nonrandomized controlled studies comparing treatment with no treatment found no influence of randomization on the mean observed treatment effect [15]. Other studies, however, have shown that the magnitude of effect observed may vary with randomization [16–19] for both medical and surgical clinical trials. In summary, while nonrandomized controlled trials can probably reliably predict a treatment effect, the magnitude of the effect may vary from that observed in RCTs.

## **12.7 PROBLEMS WITH CONCURRENT NONRANDOMIZED CONTROLLED TRIALS**

While the loss of the traditional advantages of randomization do not necessarily excessively penalize a nonrandomized study, there are other benefits from randomization that are now discussed. Figure 6 demonstrates that there are many other points of error that are not eliminated by randomization; however, randomization of the population study group does at least remove one tier of possible error in that prognostic factors, both known and unknown are balanced between intervention and comparison groups.

In the conduct of clinical trials, nonconsent bias is removed. For example, in our hypothetical orally administered chemotherapy agent, it may be that patients who are not willing to travel to the hospital for the intravenous standard treatment are those too frail or immobile to make such repeated journeys. In such circumstances, the true benefit of the new oral agent may not be fully appreciated due to the confounding factors of age and general fragility affecting long-term survival of the intervention arm. In addition, it would be relatively easy for an investigator to sway a patient's decision making by recommending one intervention over another. In this manner, patients with the smallest disease burden or better prognosis may be chosen for the trial arm, which would of course affect the validity of the study.

Randomized controlled trials require a significant support framework to administer treatment to patients and for appropriate supervision and data collection. In this manner, randomization compels the investigator to plan the conduct of the trial and introduce written protocols. While this is not a direct advantage of randomization, it is an advantage offered by the RCT trial framework. A second indirect benefit from introducing an RCT is that in many cases multicenter trials are required



due to a number of reasons, such as ensuring adequate recruitment of patients, and pooling of financial resources. By using a multicenter framework, many of the differences between the deliveries of patient care between centers is diluted. In this manner, the population group of the RCT is more representative of the population of patients with the disease out in the community.

The implementation of an RCT may have other beneficial effects within the trial center. RCTs may be a learning experience for clinicians in the planning, data collection, and analysis of trials. The RCT teaches a methodological approach to implementing trial therapies. The skills learnt from this can be used in the critical evaluation of other published evidence by clinicians and can directly cross over to the daily ward management of patients. The written protocols for RCTs will stipulate the manner in which the trial intervention is administered and how it is monitored. These trial tests need to run along the normal administration of care. As a result, this requires medical and nursing staff reevaluate the patient pathway through their unit. This may frequently be an opportunity to further improve the delivery of care, even for nontrial patients.

In conclusion, the pitfalls of not randomizing controlled studies are illustrated by a trial of vitamin C in patients with advanced colorectal cancer [20]. One hundred patients with terminal colorectal cancer were given high-dose vitamin C. The survival of these patients was then compared with 1000 control patients who received no vitamin C. The study found that the mean survival time in the vitamin C patients was 4.2 times greater than those in the control group. Despite these results, two further RCTs using the same inclusion and exclusion criteria as the previous study found no survival benefit from vitamin C therapy, with no differences in symptoms, performance status, appetite, or weight loss [21].

## 12.8 IS RANDOMIZATION FEASIBLE?

The final point to consider in this analysis is how applicable is randomization in providing the best medical care to patients. Earlier in this chapter the technical considerations on why randomization is not always feasible in practice were discussed. These may include poor patient recruitment due to dislike of one of the intervention options, for example, the mastectomy versus lumpectomy RCT [6], where patients preferred the less disfiguring operation over a mastectomy. Trials evaluating treatments for chronic illness may also suffer from low recruitment rates and high dropout rates, due to failure to comply with less effective medication or medication with greater side effects. It is possible to minimize these issues using the randomized discontinuation design [22]. The purpose behind these study designs is to select a subset of enrolled patients who are relatively homogenous with respect to their prognostic factors and randomize only this subgroup to treatment or intervention. These patients may be selected by their ability to tolerate medication without excessive side effects or their ability to comply with treatment regimens.

In chemotherapy trials, it is common for patients to deviate from their randomized treatment during follow-up. This is often due to worsening of their disease, which requires rescue medication, either in addition to their trial medication or in place of it. In such cases, the additional medication plays a major confounding factor since the rescue medication will doubtless contribute to outcome differences

between the intervention and control arms. In scientific and statistical terms, the best approach to this problem is to avoid making any changes to medication once the patient has been enrolled and randomized, however, this would be unethical. Excluding patients from the trial outcome analysis would also mask a proportion of patients who may not be responding to the trial therapy and hence may give a false impression of the true efficacy of the drug under evaluation. In addition, any synergistic effect of trial drug plus rescue medication combination would be lost.

Finally, there are ethical issues in the process of randomization of patients in clinical trials. In the Declaration of Helsinki it states that a subject may be enrolled in an RCT if there is genuine uncertainty (equipoise) about which treatment (standard or trial treatment) would benefit the patient [23]. New drug treatments undergo extensive preclinical, phase II and phase III testing before entering RCTs. This laborious evaluation process occurs as a result of huge expenditure on the part of pharmaceutical companies, who would only pursue such clinical testing if there was clear and dramatic evidence of the trial drug offering significant benefits over standard treatment. Doubtless, the principal investigator will be familiar with the evidence. In light of these observations, how much uncertainty is there that the new trial drug is no better than the control drug that patients will receive? The counter-argument to this observation concerns whether it is ethical to leave a patient's treatment to chance, or should he or she receive the best treatment available according to the knowledge of the day?

Another important ethical issue when considering RCTs is what to do if the trialled treatment is significantly worse or better than the standard treatment for a particular condition, during the trial period. There is provision for calling RCTs to a halt in either circumstance, but at what point would it be ethical to intervene? The azidothymidine (AZT) study was a double-blind randomized controlled study evaluating the efficacy of AZT in patients with AIDS (acquired immune deficiency syndrome) [24]. The study was terminated early before full recruitment of patients had occurred, when 19 patients receiving placebo had died and only one patient receiving AZT. The study confirmed the efficacy of AZT in the management of AIDS, but 19 patients who had been randomized to the study were unable to benefit from the new treatment.

Ultimately, the physician has a duty to provide the optimum care for his or her patient, but what if this means breaking the study protocol? The first RCT in Britain in the 1940s was assessing streptomycin in the treatment of tuberculosis. During the conduct of the trial, one of the senior research council members contracted tuberculosis and received supplies of streptomycin outside of the trial. This example highlights the fact that clinical judgment can be at conflict with the rigid protocols of RCTs.

The ethical issues touched upon do not, however, preclude the use of RCTs or nullify their usefulness in modern medicine. Sensitivity to these issues is important to the principle investigator in these trials and can be used to optimize study design. With regard to the subject of equipoise, there are numerous examples of RCTs where the trialled drug treatment has been found to be ineffective. In such circumstances, had it not been for the unbiased data collection afforded by RCTs, the ineffectual experimental treatment would be that chosen by many clinicians, having been seduced by the wealth of preclinical data supporting the efficacy of the drug. There is a subtle distinction between theoretical equipoise, as defined by the

Helsinki declaration, and clinical equipoise. Clinical equipoise is best defined as a lack of consensus among the medical community over which treatment is the most effective. Hence, although a new drug treatment is theoretically more efficacious (supported by preclinical and phase I/II data), it is still not actually proven to be so.

Most RCTs compare new treatments with standard treatments and hence are powered to pick a 5–10% improvement in efficacy. For this reason, examples such as the AZT treatment in AIDS, where a new effective treatment was compared with the standard (i.e., no treatment or placebo), are rare. In such circumstances, a well-designed and monitored RCT can be stopped and the efficacious new drug offered to all trial participants and the disease population. The AZT trial is a good example of an RCT that was well monitored by a central oversight group that intervened suitably when it became clear that the AZT treatment should be available to all participants. Although an RCT may not be appropriate or applicable to all clinical scenarios, where possible it can provide unbiased assessment of new drugs or treatments.

The technical details of randomization are discussed elsewhere in this book. In this chapter the phenomenon of “code breaking” (deciphering the allocation sequence) will be briefly touched upon. Investigators rarely document code breaking, but in anonymous questionnaires, there have been investigators who admit to sabotaging randomization codes [25]. Awareness that investigators may actively attempt to unravel the allocation sequence is important in the design of clinical trials, and hence robust steps must be taken to ensure that this not occur. RCTs may frequently alienate the clinicians involved by frustrating their clinical inclinations, and allocation deciphering is an attempt to provide the perceived best care for their patient. It is unlikely that investigators are attempting to deliberately sabotage the results of the trial. As we have seen, randomization, when possible, offers key advantages and credibility to the results of a clinical trial. It is essential then that the word *randomization* is not used in vain, and concealment of allocation codes is essential in the planning of any RCT.

## 12.9 CONCLUSION

Randomized controlled trials are the best statistical method of evaluating one treatment over another. In general, a well-conducted and monitored RCT provides an unbiased, balanced, and reliable method of determining the efficacy of new drugs or devices. Figure 7 is a schematic illustration of the level of evidence provided by different trial methodology. RCTs also carry with them important logistic and ethical issues, which must be thoroughly addressed before commencing the trial. A review of phase III RCTs evaluating chemotherapy agents observed that the methods used were satisfactory in 67.5% of cases, but 30% of authors were not respecting fundamental ethical principles [26]. In certain circumstances, RCTs may not be possible, in which case recourse to alternative trial designs may be necessary, but this must be with realization that the level of evidence provided will be diminished.

Green and Raley reported that a well-designed clinical trial should incorporate a proper randomization procedure allocating patients to treatment groups in an unbiased manner with a sample size large enough to provide adequate statistical

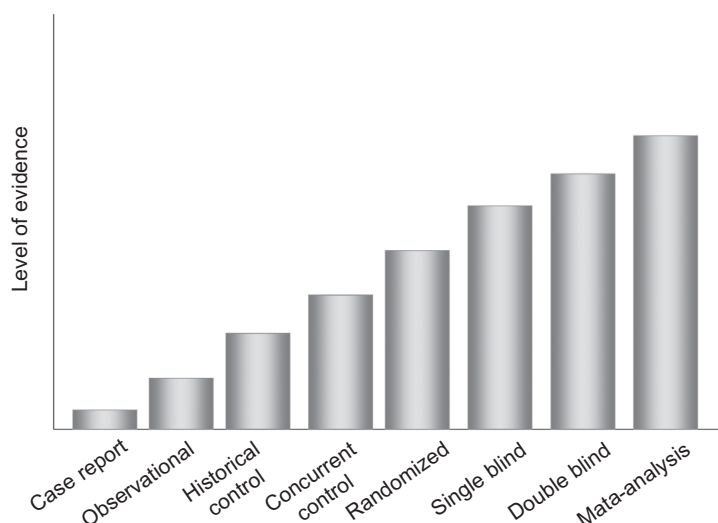


**FIGURE 7** Schematic representation of the tiers of evidence provided by different trial methodology. (Reproduced with permission from Lyman and Kudere [28].)

power [27]. In addition simple endpoints are required and an intention-to-treat analysis that includes all individuals randomized to the treatment groups regardless of whether or not they completed treatment. Results should be presented clearly with a confidence interval that quantifies uncertainty. A final point to consider is the quality of trial results reporting. Well-reported and designed controlled trials can have their results combined to produce a more precise estimate of the effects of treatment.

## REFERENCES

1. Russell, P. S. (1989), *Theoret. Surg.*, 4, 169–170.
2. Haines, S. J. (1979), *J. Neurosurg.*, 4, 163–170.
3. Oettinger, W., and Berger, H. G. (1989), *Theoret. Surg.*, 4, 170.
4. Pollack, A. V. (1989), *Theoret. Surg.*, 4, 163–170.
5. Solomon, M. J., and McLeod, R. S. (1995), *Surgery*, 118, 459–467.
6. Taylor, K. M., Mangolese, R. G., and Soskolne, C. L. (1984), *N. Engl. J. Med.*, 310, 1363–1367.
7. Mant, D. (1999), *Lancet*, 353, 743–746.
8. Charlson, M. E. (1984), *BMJ*, 289, 1281–1284.
9. Diehl, L. F., and Perry, D. J. (1986), *J. Clin. Oncol.*, 4, 114–1120.
10. Gehan, E. A., and Feicrich, E. J. (1974), *N. Engl. J. Med.*, 290, 198–203.
11. Cranberg, L. (1979), *BMJ*, 2, 1265–1266.
12. Gehan, E. A. (1978), *Biomedicine*, 28, 13–19.
13. Chalmers, T. C. (1977), *N. Engl. J. Med.*, 297, 285.
14. Abel, U., and Koch, A. (1999), *J. Clin. Epidemiol.*, 52, 487–497.
15. Ottenbacher, K. (1991), *Controlled Clin. Trial*, 13, 50–61.

16. Pak, C. Y., Adams-Huet, B., Sakhaee, K., et al. (1996), *J. Bone Mineral Res.*, 11, 160–168.
17. Miller, J. N., Colditz, G. A., and Mosteller, F. (1989), *Stat. Med.*, 8, 455–454.
18. Colditz, G. A., Miller, J. N., and Mosteller, F. (1989), *Stat. Med.*, 4, 441–454.
19. Chalmers, T. C., Celano, P., Sacks, H. S., et al. (1993), *N. Engl. J. Med.*, 309, 1358–1361.
20. Cameron, E., and Pauling, L. (1976), *Proc. Natl. Acad. Sci. U.S.A.*, 73, 3685–3689.
21. Creagan, E. T., Moertel, C. G., O’Fallon, J. R., et al. (1979), *N. Engl. J. Med.*, 27, 687–690.
22. Kopec, J. A., Abrahamowicz, M., and Esdaile, J. M. (1993), *J. Clin. Epidemiol.*, 46, 959–971.
23. Edwards, S. J. L., Lilford, R. J., Braunholtz, D. A., et al. (1998), *Health Technol. Assessment*, 2, 1–32.
24. Fischl, M. A., Richman, D. D., and Grieco, M. H. (1987), *N. Engl. J. Med.*, 387, 185–191.
25. Schulz, K. F. (1995), *JAMA*, 274, 1456–1458.
26. Tuech, J. J., Pessaux, P., Moutel, G., et al. (2005), *J. Med. Ethics*, 31, 251–255.
27. Green, S. B., and Raley, P. L. (2000), *Sci. Ed.*, 23, 157.
28. Lyman, G. H., and Kudere, N. M. (1997), *Cancer Control*, 4, 413–418.



# 13

---

## *Cross-Over Designs*

**Raphaël Porcher and Sylvie Chevret**

*Département de Biostatistique et Informatique Médicale, Hôpital Saint-Louis, France*

### **Contents**

|        |                                                       |     |
|--------|-------------------------------------------------------|-----|
| 13.1   | Introduction                                          | 824 |
| 13.2   | Main Issues                                           | 825 |
| 13.2.1 | Advantages                                            | 825 |
| 13.2.2 | Disadvantages                                         | 825 |
| 13.2.3 | Should Cross-Over Trials Be Used and When?            | 826 |
| 13.3   | Analysis of AB/BA Design                              | 826 |
| 13.3.1 | Basic Analysis                                        | 827 |
| 13.3.2 | Adjustment for Period Effect                          | 829 |
| 13.3.3 | Model-Based Analysis                                  | 830 |
| 13.3.4 | Two-Stage Analysis                                    | 831 |
| 13.3.5 | Carry-Over                                            | 833 |
| 13.3.6 | Use of Baselines                                      | 833 |
| 13.3.7 | Bayesian Analysis                                     | 834 |
| 13.4   | Other Endpoints                                       | 838 |
| 13.4.1 | Nonparametric Analyses                                | 838 |
| 13.4.2 | Binary or Categorical Data                            | 839 |
| 13.4.3 | Censored Endpoints                                    | 841 |
| 13.5   | Power and Sample Size Considerations                  | 841 |
| 13.5.1 | Sample Size                                           | 842 |
| 13.5.2 | Efficiency of Cross-Over versus Parallel-Group Design | 843 |
| 13.6   | More Complicated Designs                              | 844 |
| 13.6.1 | Higher Order Designs for Two Treatments               | 844 |
| 13.6.2 | Designs for Three or More Treatments                  | 844 |
| 13.7   | Summary                                               | 846 |
|        | References                                            | 847 |



### 13.1 INTRODUCTION

Cross-over designs refer to clinical trials where patients sequentially receive several of the compared treatments, instead of receiving only one of them as in the classical parallel-arm designs. Each subject hence acts as his (her) own control. The main purpose of such a design is thus to perform within-subject comparison of treatment effects instead of between-subject comparisons, which increases efficiency. As a consequence, a cross-over trial usually requires less patients than a parallel-arm design.

The simplest cross-over design is the two-treatment two-period cross-over design, also referred to as the AB/BA design, since patients receive either treatment A followed by treatment B or treatment B followed by treatment A. A schematic representation of the AB/BA design is given Figure 1. As depicted in Figure 1, patients included in an AB/BA cross-over trial are randomized between each *group* (or sequence)  $A \rightarrow B$  or  $B \rightarrow A$ . After a potential *run-in* period where subjects are given neither A nor B, the *first period* (period 1 on the figure) begins with the administration of one of the two treatments, according to the randomization group. During the run-in period, baseline measurements can be performed. Between the two treatment periods, a *wash-out* period is generally used to remove any residual effects of the treatment received in the first period. As for the run-in period, neither A nor B is administered during the wash-out period. Then, patients switch to the treatment they did not receive yet.

Provided the disease condition of subjects is similar at the beginning of each treatment period, the cross-over design is intuitively appealing to compare treatments. Indeed, each patient will serve as his (her) own control. However, the effect of the first treatment received may persist during the second treatment period, leading to the so-called *carry-over* effect of the treatment. Several considerations on both the use of two-stage designs to cope with carry-over and the low power of the test for carry-over have generated a large literature on cross-over designs [1–6] and somewhat contradictory guidelines from regulatory agencies [7, 8]. More recently, three textbooks dedicated to cross-over trials have been published, where details and discussion regarding cross-over trials may be found [9–11].

In this chapter, we will review the main issues concerning the use and analysis of cross-over designs. We will focus on the AB/BA design and more particularly on the analysis of such trials. More complex design will be mentioned later. Another type of cross-over design can be frequently encountered, for example, in malignant diseases. It is basically a parallel-arm design where patients who fail to respond to the

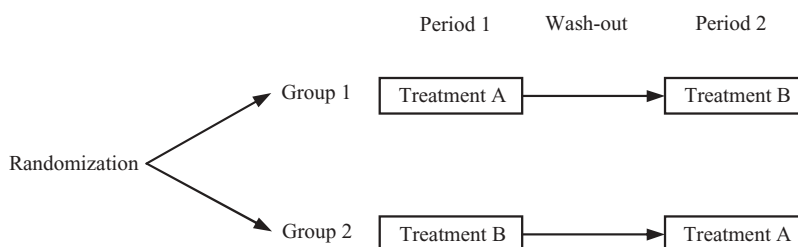


**FIGURE 1** Representation of an AB/BA cross-over trial.

first treatment they were assigned to are switched to the other one. These treatment cross-overs are planned for ethical reasons in order to give each patient a chance to receive the “superior” treatment, and the main analysis relies on comparing the initial treatments only. This type of cross-over design will thus not be considered in this chapter.

## 13.2 MAIN ISSUES

### 13.2.1 Advantages

The cross-over design is intuitively well suited to compare treatments as all, or at least some, are administered to each subject. Nevertheless, the main advantage of cross-over trials is that inference on treatment effect is based on within-subject information, each subject being used as his (her) own control. Between-subject variability is eliminated, which leads to increased efficiency. In general, fewer patients, and sometimes much fewer, will be required to obtain the same precision in estimation than classical parallel designs. Cross-over trials thus may save considerable resources.

Another advantage of cross-over trials is their acceptability. As all subjects know they will receive the superior treatment at some point, their willingness to participate may increase, as well as their compliance.

### 13.2.2 Disadvantages

The principal disadvantages of cross-over trials stem from carry-over. Carry-over is defined as the persistence of the effect of the treatment received in a period into the subsequent periods. In the presence of carry-over, the outcome measured after the second treatment will be also affected by the first treatment, which is likely to bias the estimate of the treatment effect. For example, a drug agent may physiologically persist during the next period. The first treatment could also change permanently the condition of the patient, for example, by curing him (her). If the first case can be corrected by using an appropriate wash-out period between treatment periods, the second will preclude using a cross-over design. This is the case of acute infections, for instance. Carry-over has focused most concerns toward cross-over trials, and some approaches have dealt with removing carry-over, or at least accounting for it, as will be seen below. The best approach is, however, to design the study in order to avoid carry-over. Another problem related to carry-over is treatment-by-period interaction, that is, when the treatment effect is not the same in all treatment periods. As will be seen in the following, treatment-by-period interaction cannot be separated from carry-over in the AB/BA design, but both can be estimated separately in more complex designs.

Dropouts can also be more problematic in cross-over trials than in parallel-group trials, as patients who do not participate to the second study period do not provide any information on the treatment effect. This is the case for patients who die during the study.

Cross-over trials can also be used in particular situations of stable diseases, where the patient condition will not be permanently modified by the treatment.

Other less important disadvantages of cross-over trials are that all patients will receive the inferior treatment (at some point). Moreover, the duration of participation to cross-over trials is longer, due to the several period design particularly for high-order cross-over designs. This can cause some inconvenience to patients and decrease their willingness to participate.

At last, it should be underlined that cross-over trials are more complex to analyze than parallel-group designs.

### 13.2.3 Should Cross-Over Trials Be Used and When?

As already seen, cross-over trials are obviously not adapted to all situations. As each patient receives each treatment sequentially, a cross-over trial cannot be reasonably considered in diseases where patients have a nonnegligible probability of dying or considerably deteriorating during the trial period. Cross-over trials can also not be used if one of the treatments is supposed to have a permanent, for instance, curative, effect. Moreover, carry-over leads to biased estimates of treatment effect. Therefore, the Biometric and Epidemiology Methodology Advisory Committee (BEMAC) of the Food and Drug Administration (FDA) recommended during the 1970s that cross-over trials should be avoided as much as possible. One should prefer completely randomized designs where the estimate of treatment effect is unbiased, without any modeling assumption [7].

Since then, considerable research has been published on the design and analysis of cross-over trials. In contrast to this recommendation, several authors have published in defense of cross-over trials, at least in certain cases [5]. Actually, cross-over trials may be very useful but require more careful planning than parallel-group trials, as it is important to avoid carry-over [8, 11]. They should be used only in the appropriate disease settings and when dropouts are also expected to be small. For instance, it is well suited for chronic diseases such as hypertension or chronic renal failure. When planning a cross-over trial, everything should be done to avoid carry-over. This will not arise from a particular modeling strategy but from design only. It is therefore of primary importance to plan sufficient wash-out periods [11]. Additionally, random and balanced allocation to each group is advisable, with adjustment for period effect in the analysis. At last, baseline measurements may be used to increase precision. A satisfactory use of cross-over trials can also be found in demonstrating the bioequivalence of two formulations of the same drug. This is particularly the case in trials involving healthy volunteers, where carry-over is likely to be properly eliminated using an adequate wash-out period.

## 13.3 ANALYSIS OF AB/BA DESIGN

Even for the apparently simple AB/BA design, analysis of cross-over designs has generated a wide literature in the last three decades. The main reason has to do with carry-over. This section introduces the most common methods used for analysis of cross-over designs and sketches the implications of the use of these methods. More details may be found in dedicated monographs [9, 11] or in the issue of *Statistical Methods in Medical Research* dedicated to cross-over trials (1994, volume 4, issue 3).

To simplify the notations and the comparison between the different methods reviewed here, we propose to use the general model formulation for the cross-over design with a continuous outcome. Such a model has been very commonly used in the cross-over literature [1, 12–14]. Let us then note  $Y_{ijk}$  as the outcome measure of the  $j$ th subject within the  $i$ th treatment group (or sequence) at the  $k$ th period. For the AB/BA design,  $i = 1, 2; j = 1, \dots, n_i; k = 1, 2$ . The model for  $Y_{ijk}$  we consider is

$$Y_{ijk} = \mu + \pi_k + \tau_l + \eta_k \lambda_i + s_{ij} + \varepsilon_{ijk} \quad (1)$$

where  $l = 1$  if  $i = k$  and  $2$  if  $i \neq k$ , and  $\eta_k = 0$  if  $k = 1$  and  $1$  if  $k = 2$ .

In this model,  $\mu$  is the overall mean,  $\pi_k$  is the period effect,  $\tau_l$  the treatment effect and  $\lambda_i$  the first-order carry-over effect. The subject effect is represented by  $s_{ij}$  and may be regarded as a fixed or a random effect. We will adopt the latter formulation and consider that the  $s_{ij}$  are normally and independently distributed with mean 0 and variance  $\sigma_s^2$ . The within-patient errors  $\varepsilon_{ijk}$  are also assumed to be normally distributed, with mean 0 and variance  $\sigma^2$ , and both independent from each other and from the  $s_{ij}$ . We will note  $\rho$ , the correlation between observations on the same patient, which is  $\rho = \sigma_s^2 / (\sigma_s^2 + \sigma^2)$ . In this model, the whole variance of observations is  $\sigma_T^2 = \sigma_s^2 + \sigma^2$ , which can also be rewritten as  $\sigma^2 / (1 - \rho)$ . Table 1 presents the expected cell means for model (1). This model explicitly assumes that the only effects are due to treatments or periods and that they do not vary from a patient to another. It also assumes that the total variability can be split into a between-subject and a within-subject variability, that are both equal for all patients. It is thus clearly not the only possible model for the data. Freeman presents the same model with baseline measurements corresponding to a third period noted  $k = 0$  [14].

Let us note the treatment difference, or treatment effect,  $\tau = \tau_2 - \tau_1$ , the carry-over difference  $\lambda = \lambda_2 - \lambda_1$ , thereafter referred as carry-over for simplicity, and the period effect between periods 2 and 1,  $\pi = \pi_2 - \pi_1$ . Let  $y_{ijk}$  denote the observation of the random variable  $Y_{ijk}$  thereafter.

### 13.3.1 Basic Analysis

The basic analysis of the AB/BA cross-over design overlooks any period or carry-over effect. As each patient in a cross-over trial acts as its own control, the natural idea is to use within-patient differences. The basic analysis consists in computing for each patient the difference in outcome between both treatments, which is called the cross-over difference. Suppose that we use treatment A as a control for B, the cross-over difference will be  $Y_{1j2} - Y_{1j1}$  for a patient of group 1 and  $Y_{2j1} - Y_{2j2}$  for a patient

TABLE 1 Expected Cell Means for the Two-Period Cross-Over Model

| Randomization Group |               | Period                 |                                    |
|---------------------|---------------|------------------------|------------------------------------|
|                     |               | First Treatment (1)    | Second Treatment (2)               |
| 1                   | Treatment     | A                      | B                                  |
|                     | Expected mean | $\mu + \pi_1 + \tau_1$ | $\mu + \pi_2 + \tau_2 + \lambda_1$ |
| 2                   | Treatment     | B                      | A                                  |
|                     | Expected mean | $\mu + \pi_1 + \tau_2$ | $\mu + \pi_2 + \tau_1 + \lambda_2$ |

of group 2, as treatment B is given in the second period in group 1 and in the first period in group 2. Once these cross-over differences are calculated, the analysis is carried-out ignoring the distinction between both groups. This is equivalent to comparing the outcomes obtained with treatment A to the outcomes obtained with treatment B using an appropriate paired test statistic, irrespective of the randomization group. The parametric approach relies on the matched paired  $t$  test, while the nonparametric approach is based on Wilcoxon's signed rank test. This is illustrated on the example displayed in Table 2. These data have been used by Grieve [15] and come from a study carried out by CIBA-GEIGY to assess the effectiveness of transdermal nitroglycerine patches in the prophylaxis of angina in general practice [16]. A total of 63 patients were randomized between either 3-week treatment with placebo (PL) followed by transdermal nitroglycerine (TN) (group 1, PL  $\rightarrow$  TN) or vice-versa (group 2, TN  $\rightarrow$  PL). The main endpoint was the weekly anginal attack rate during the third week of treatment.

**TABLE 2 Weekly Anginal Attack Rates in Transdermal Nitroglycerine Trial**

| Group 1 |          |          |                       | Group 2 |          |          |                       |
|---------|----------|----------|-----------------------|---------|----------|----------|-----------------------|
| Patient | Period 1 | Period 2 | Cross-Over Difference | Patient | Period 1 | Period 2 | Cross-Over Difference |
| 19      | 3        | 10       | 7                     | 20      | 12       | 16       | -4                    |
| 22      | 8        | 6        | -2                    | 21      | 4        | 11       | -7                    |
| 26      | 6        | 4        | -2                    | 23      | 6        | 5        | 1                     |
| 35      | 1        | 0        | -1                    | 36      | 7        | 14       | -7                    |
| 38      | 12       | 6        | -6                    | 37      | 13       | 25       | -12                   |
| 39      | 4        | 2        | -2                    | 40      | 9        | 11       | -2                    |
| 42      | 6        | 3        | -3                    | 41      | 1        | 1        | 0                     |
| 59      | 11       | 3        | -8                    | 43      | 4        | 0        | 4                     |
| 64      | 3        | 4        | 1                     | 56      | 4        | 10       | -6                    |
| 73      | 11       | 3        | -8                    | 57      | 2        | 5        | -3                    |
| 76      | 8        | 8        | 0                     | 60      | 0        | 8        | -8                    |
| 78      | 8        | 9        | 1                     | 61      | 17       | 13       | 4                     |
| 80      | 18       | 4        | -14                   | 65      | 1        | 1        | 0                     |
| 81      | 12       | 5        | -7                    | 67      | 6        | 8        | -2                    |
| 84      | 12       | 2        | -10                   | 75      | 8        | 8        | 0                     |
| 85      | 3        | 1        | -2                    | 77      | 7        | 4        | 3                     |
| 115     | 1        | 3        | 2                     | 79      | 3        | 19       | -16                   |
| 122     | 12       | 4        | -8                    | 82      | 4        | 19       | -15                   |
| 124     | 8        | 6        | -2                    | 83      | 3        | 12       | -9                    |
| 126     | 7        | 12       | 5                     | 86      | 2        | 4        | -2                    |
| 128     | 1        | 1        | 0                     | 87      | 2        | 1        | 1                     |
| 140     | 2        | 0        | -2                    | 121     | 4        | 7        | -3                    |
| 142     | 3        | 0        | -3                    | 123     | 3        | 1        | 2                     |
| 146     | 21       | 10       | -11                   | 125     | 3        | 3        | 0                     |
| 147     | 17       | 7        | -10                   | 127     | 1        | 0        | 1                     |
| 150     | 12       | 5        | -7                    | 130     | 41       | 36       | 5                     |
| 201     | 4        | 5        | 1                     | 145     | 10       | 24       | -14                   |
| 209     | 0        | 1        | 1                     | 148     | 9        | 18       | -9                    |
| 211     | 7        | 0        | -7                    | 149     | 4        | 13       | -9                    |
| 233     | 11       | 0        | -11                   | 210     | 8        | 1        | 7                     |
| 236     | 18       | 7        | -11                   | 234     | 5        | 7        | -2                    |
|         |          |          |                       | 235     | 0        | 9        | -9                    |

In the example, the cross-over differences are displayed for each patient. Their mean would produce the basic estimator of the treatment effect. Numerically, we obtain  $\hat{\tau}^b = -3.651$ , with standard error 0.70. Using a  $t$  test, the observed  $t$  statistics is  $-5.188$  with 62 degrees of freedom, leading to a  $P$  value of  $2.5 \times 10^{-6}$  (for a two-sided test). The 95% confidence interval (CI) of the treatment effect is  $-5.058$  to  $-2.244$ .

This is a simple way to analyze cross-over trials. It is, however, only valid under particular conditions. Actually, from model (1), the cross-over differences will have expectation  $\tau + \lambda_1 + \pi$  in group 1 and  $\tau - \lambda_2 - \pi$  in group 2. With  $n_1$  patients in group 1 and  $n_2$  patients in group 2, the basic estimator has expectation  $\tau + \pi(n_1 - n_2)/(n_1 + n_2) + (n_1\lambda_1 - n_2\lambda_2)/(n_1 + n_2)$ . It is unbiased when there is no period effect, that is, no global time trend and no carry-over. However, for an unbalanced design, a period effect is sufficient to produce a biased estimator. Moreover, and even more seriously, a period effect would also result in an inflated variance of the estimator, as deterministic variations would have been attributed to random variation. Senn lists three other factors affecting the value of this basic analysis, namely a treatment by period interaction, patient by treatment interaction, and patient by period interaction [11]. Of these factors, carry-over and period by treatment interactions have been regarded as the major problems with cross-over trials.

### 13.3.2 Adjustment for Period Effect

Hills and Armitage [4] used a stratified analysis to adjust for a period effect. They still use within-patient differences but take the unweighted average of the estimator of the treatment effect obtained in each group. This defines a new estimator of the treatment effect,  $\hat{\tau}^c = \{(\bar{Y}_{1,2} - \bar{Y}_{1,1}) + (\bar{Y}_{2,1} - \bar{Y}_{2,2})\}/2$ , where the subscript “.” indicates averaging with respect to the corresponding index. This method thus combines some of the cell means of Table 1. Under the assumptions of model (1), this estimator has variance  $\sigma^2(1/n_1 + 1/n_2)/2$ , where  $\sigma^2$  is the within-subject variance. For a balanced trial where  $n_1 = n_2 = n$ , this variance reduces to  $\sigma^2/n$ . In such a case, the estimate is equal to the basic estimate, but with a different estimate of the variance. Actually, the within-subject variance  $\sigma^2$  is half the variance of a cross-over difference, which may be estimated by pooling the two within-groups sum of squares and dividing the sum by the appropriate  $n_1 + n_2 - 2$  degrees of freedom. This allows to calculate a  $t$  statistic, which also has  $n_1 + n_2 - 2$  degrees of freedom. This can be illustrated on the preceding example, with useful cell means presented Table 3.

For the two randomization groups, we obtain means equal to  $-3.839$  and  $-3.469$ , respectively, and estimated variances  $S_1 = 26.740$  and  $S_2 = 36.451$ . The estimate of the treatment effect is thus simply  $\hat{\tau}^c = -3.654$ , and the pooled variance estimate

**TABLE 3 Cell Means for Transdermal Nitroglycerine Trial Data**

| Randomization Group |           | Period 1 | Period 2 | Difference TN – PL |
|---------------------|-----------|----------|----------|--------------------|
| 1                   | Treatment | PL       | TN       |                    |
|                     | Cell mean | 8.065    | 4.226    | –3.839             |
| 2                   | Treatment | TN       | PL       |                    |
|                     | Cell mean | 6.344    | 9.813    | –3.469             |

$$2\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 31.675$$

Hence, the within-subject variance is  $\hat{\sigma}^2 = 15.837$ , the estimated standard error of  $\hat{\tau}^c$  is 0.709, and the resulting  $t$  statistic is  $-5.152$  with 61 degrees of freedom. The corresponding  $P$  value is  $2.9 \times 10^{-6}$ . These results are very close to those obtained with the basic estimator.

This estimator is the one recommended by Senn [11]. It also corresponds to the standard weighted least-squares estimate of the treatment effect in model (1) [17]. Freeman referred to it as the CROS estimator [14], and the term was subsequently used in several reviews [11, 18]. We will thus use this terminology thereafter.

### 13.3.3 Model-Based Analysis

The alternative is to fit directly model (1) to the data. Note that, up to now, we did not consider baselines. Whether to allow separate variation for groups and patients within groups or not does not modify the estimate of treatment effect but allows to estimate a carry-over effect. The detailed analysis of variance (ANOVA) table for model (1) with fixed group, period, and treatment effects and random patient effect nested in group is not straightforward and is not detailed here. It can be found in the textbook by Jones and Kenward [9].

Numerical values for the ANOVA table together with the fixed effects estimate are presented in Table 4 for the weekly angina attack data. Using this model, the treatment effect estimator is identical to the CROS estimate presented above. The model also allows to test for carry-over or period effects, although the latter is generally of minor interest in a cross-over trial. We will detail the test of carry-over effect later. Results suggest a strong evidence for a reduction of angina attacks, about three or four per week. More precisely, the 95% confidence interval of the treatment effect is estimated at  $-5.072$  to  $-2.235$ . Moreover, there seems to be little evidence of a carry-over (group) effect nor a period effect. For purpose of comparison, note that model (1) uses a different parameterization than that used in Grieve [15], the effects estimated being minus twice the ones he chose to parameterize the model. Also note that the group effect is half the carry-over effect  $\lambda$ . The results of Table 4 perfectly match those presented by Grieve [15].

Another parameterization leading to exactly the same model would be to replace the group effect by a treatment-by-period interaction. The model used is obviously unable to attribute the difference of treatment effect in the second period to a

**TABLE 4** Simplified ANOVA Table and Parameter Estimates for Transdermal Nitroglycerin Trial<sup>a</sup>

| Source                | df    | F Value | P Value              | Parameter Estimate | SE    |
|-----------------------|-------|---------|----------------------|--------------------|-------|
| Group (2 vs. 1)       | 1, 61 | 1.693   | 0.198                | 1.933              | 1.486 |
| Period (2 vs. 1)      | 1, 61 | 0.068   | 0.795                | -0.185             | 0.709 |
| Treatment (TN vs. PL) | 1, 61 | 26.545  | $2.9 \times 10^{-6}$ | -3.654             | 0.709 |

<sup>a</sup>df = degrees of freedom; SE = standard ends.



**TABLE 5** Distribution Parameters for Contrasts in Model (1)

| Contrast        | Mean                             | Variance                                          |
|-----------------|----------------------------------|---------------------------------------------------|
| $\hat{\tau}^c$  | $\tau - \frac{\hat{\lambda}}{2}$ | $\frac{m\sigma^2}{2}$                             |
| $\hat{\lambda}$ | $\lambda$                        | $2m\sigma^2 \left( \frac{1+\rho}{1-\rho} \right)$ |
| $\hat{\tau}^p$  | $\tau$                           | $m\sigma^2 \left( \frac{1}{1-\rho} \right)$       |

carry-over [as parameterized in model (1) formulation] or to an interaction between treatment and period. Both can thus not be separated using the observed data.

Let us consider two additional contrasts:

$$\hat{\lambda} = (\bar{Y}_{2.1} + \bar{Y}_{2.2}) - (\bar{Y}_{1.1} + \bar{Y}_{1.2}) \quad \text{and} \quad \hat{\tau}^p = \bar{Y}_{2.1} - \bar{Y}_{1.1}$$

where  $\hat{\lambda}$  is the difference between the two groups, and has expectation  $\lambda$ , of which it is the weighted least-squares estimate. The contrast  $\hat{\tau}^p$  is simply the between-patient contrast for treatment effect using the first period data only. They are linked to the CROS estimator by  $\hat{\tau}^p = \hat{\lambda}/2 + \hat{\tau}^c$ . These three contrasts have a Gaussian distribution, with respective mean and variance given in Table 5, where  $m$  denotes  $(n_1 + n_2)/(n_1 - n_2)$ .

It appears in Table 5 that the CROS estimator is biased in the presence of carry-over, except if the carry-over is the same after each treatment ( $\lambda_1 = \lambda_2$ ), which does not seem much more plausible than the absence of any carry-over in practice. An unbiased estimator of  $\tau$  would be obtained using  $\hat{\tau}^p$  instead, but this would imply discarding all data recorded during the second period. This analysis has been referred to as PAR (for parallel, as no cross-over data are used). Consequently, the variance of such an estimator would be at least twice that of the CROS estimator. The contrast  $\hat{\lambda}$  provides an estimate and test for the carry-over effect. Given the variance of  $\hat{\lambda}$  as compared to that of the test of treatment effect (based either on CROS or PAR), it is clear that the test for carry-over will not be very powerful. This is, however, the basis for the general two-stage strategy proposed by Grizzle [12].

### 13.3.4 Two-Stage Analysis

Initially proposed by Grizzle [12], a two-stage approach has been recommended for many years and was commonly used for analysis. This approach begins with a formal test for carry-over, usually at a 10% level, given the limited power of this test. If no evidence of a carry-over effect is brought by this test, the treatment effect is estimated from within-patient differences using the CROS estimator. However, if the test for carry-over is considered as significant, a between-patient estimator is used instead. This estimator is the one referred as PAR above and only used the first period of the trial. Since it cannot be known in advance which test will be used,

some authors have advised the trial should always be adequately powered for the PAR analysis [1]. At first sight, this design was considered ingenious and regarded as the method of choice. Freeman conducted a thorough analysis of the two-stage design as a whole and showed that it had highly inflated type I error rates, mainly because the test for carry-over and the PAR test are not independent [14]. While some work was more recently published in defense of this procedure [5], main monographs on cross-over trials do not recommend it anymore [9–11].

For the weekly anginal attack rates of the transdermal nitroglycerin data, no significant evidence of carry-over was found at the 0.10 level. We would thus simply use the CROS analysis and conclude at a significant reduction of the weekly anginal attack rate with nitroglycerin. The estimated rate reduction would be 3.654 per week (95%CI 2.235–5.072). For comparison, if a more liberal level was used for the test of carry-over, say 0.30 for instance, the PAR analysis would be preferred. The estimated anginal attack reduction would be 1.721 per week (95%CI –5.056 to +1.615), and we would have concluded to a nonsignificant anginal attack rate reduction ( $P = 0.306$ ). Although the CROS analysis yields here a treatment effect estimate very close to the basic estimate, this two-stage procedure leads to a clearly different conclusion.

Freedman showed that if CROS behaved badly for carry-over effects departing from zero, substantial carry-over making it almost certain that spurious treatment differences would be reported [14]. The two-stage procedure had possibly highly inflated type I error rates, even in the case of absence of carry-over. For example, if  $\rho = 0.6$ , and using a 0.1 level for the test of carry-over effect and a 0.05 level for the second stage test, the actual type I error rates of the two-stage procedure would range from 0.085 in case of no carry-over effect up to 0.5 for increased carry-over effect (though it has been argued that high carry-over difference values may be quite unrealistic in case of no treatment effect). Nevertheless, the two-stage procedure always has increased significance level. On the contrary, in case of no cross-over, the two-stage procedure shows a better power than PAR, but purchased at the expense of increased type I error rates. These properties are explained by the joint distribution of the three contrasts  $(\hat{\tau}^c, \hat{\lambda}, \hat{\tau}^p)$ . If  $\hat{\lambda}$  and  $\hat{\tau}^c$  are independent, the correlation between  $\hat{\lambda}$  and  $\hat{\tau}^p$  is  $\sqrt{(1+\rho)/2}$ , which can be remarkably high, for example, about 0.866 if the within-patient variability  $\sigma^2$  is equal to the between-patient variability  $\sigma_s^2$ . The tests based on PAR and the test for carry-over are thus not independent. Actually, Senn showed that the conditional error of PAR given a significant test for carry-over at the 0.1 level lies between 0.25 and 0.5, depending on the balance between within-subject and between-subject variability [19]. On the contrary, the type I error rate of the CROS analysis is 0.05 when the test for carry-over is nonsignificant. The global type I error of the whole procedure thus lies between  $0.9 \times 0.05 + 0.1 \times 0.25 = 0.07$  and  $0.9 \times 0.05 + 0.1 \times 0.5 = 0.095$ .

Another particular property of the two-stage design pointed out by Freeman is that its performance does not increase when the sample size increases [14]. This makes worthless the proposal of some authors that the trial should be adequately powered for the carry-over test (see, e.g., Brown [1]). Actually, as results of the first stage test are more likely to yield reliable conclusions, results of the second stage will be all the more sensitive to small carry-over values. The overall performance thus worsens as sample size increases, with inflated type I error rates. For example, with  $\rho = 0.6$  and a 0.1 level for the test of carry-over effect and a 0.05 level for the

second stage test, an actual type I error rate of 0.5 occurs at  $\lambda = 1.25\sigma$  when  $n (= n_1 = n_2)$  is about 20,  $\lambda = \sigma$  when  $n$  is about 30 and  $\lambda = 0.55\sigma$  when  $n = 100$ . At last, it should be noted that if  $\hat{\tau}^c$  is a biased estimator of  $\tau$ , with bias depending on carry-over  $\lambda$ ,  $\hat{\tau}^c$  even unbiased, is correlated to  $\hat{\lambda}$ , the estimate of  $\lambda$ .

Based on these considerations, both Senn [11, 19] and Jones and Kenward [9] recommend not using the two-stage procedure. They advise the best should be done to remove carry-over, using, for instance, adequate wash-out periods. Then, a sensible analysis can be performed by CROS. Freeman would recommend the use of a Bayesian approach [15, 20, 21].

### 13.3.5 Carry-Over

As seen in Section 13.3.3, carry-over and treatment-by-period interaction cannot be estimated separately using the observed data. Thus, carry-over or treatment-by-period interaction is rather a modeling point of view [14]. If carry-over cannot be ruled out by design, then it may be preferable to perform a parallel-group trial [11, 14]. On the contrary, if a treatment-by-period interaction is suspected, any form of trial is questionable, as estimating an absolute difference between treatments is meaningless.

Of note, both may also exist in parallel-group trials. For example, in a trial comparing an experimental treatment to a standard treatment of a chronic disease, patients allocated to the experimental group may have received the standard treatment in the past, and carry-over could occur in the same way as in a cross-over trial if a (run-in) wash-out period is not planned. Conversely, one may also imagine that some patients already treated by the standard therapy and allocated to the control group may have developed some form of resistance to this treatment, which would amplify the difference between treatment arms.

Different approaches concerning carry-over may be found in the literature on cross-over trials. For instance, Senn recommends to design the trial at best to avoid carry-over and then to not consider further this issue [6, 11, 22]. Willan and Pater would also likely use a CROS analysis, but in cases where between-subject variability exceeds within-subject variability and the carry-over effect is small [13]. This would be consistent with the point of view of Poloniecki and Pearce, who argue that, since the test for carry-over has a low power, it will allow to detect only large values of carry-over [23].

### 13.3.6 Use of Baselines

Carry-over has been the focus of much concern in preceding sections. Baseline measurements may be used to add information and eliminate parameters such as carry-over from the contrasts of interest. Baselines may be intended as genuine baseline measurements of the outcome (i.e., before treatment) or more generally as other covariates. Additionally, baselines may be measured at several time points, as during a run-in period, before the first treatment period, at the end of the wash-out period, or even at the end of a second wash-out period after the last treatment period [9–11]. For example, the formulation for model (1) used in Freeman only considers baseline measurements during a run-in period [14]. Such a situation can be treated in practice in the same way as for genuine covariates. In both cases, if

the response is related to the covariate (or baseline measurement), the between-subject residual variance may be sensibly reduced by adjusting for the covariate in the model. This is the same as for parallel group trials. Considering two or more baselines requires a modification of the way data are to be analyzed [9]. Several approaches have been proposed to incorporate baseline measurements into the analysis. They differ with regard to the assumptions that may be made or not [9, 11, 24]:

- Is the period effect applicable to a baseline measurement equal to the one applying to the following outcome measurement?
- Does carry-over also apply to baseline measurements and, if so, is it the same as carry-over for the subsequent treatment?
- Do the observations on the same subject have a uniform covariance structure?

To summarize, it has been shown that incorporating baselines can reduce the variance of the treatment effect estimate, as expected, and Kenward and Jones showed how to consider these baselines without assuming a particular covariance structure for the measurements [24]. They also showed that if information from baseline could increase the power of the test for carry-over, it remained unlikely that it would be increased up to the power of the test for direct treatment effect. The methods are not detailed here but can be found in appropriate references [9, 24].

### 13.3.7 Bayesian Analysis

The reader unfamiliar with Bayesian analysis should perhaps read first the introduction to Bayesian methods in health science by Spiegelhalter et al. [25]. A first Bayesian analysis of the  $2 \times 2$  cross-over trial as used for bioequivalence assessment was proposed by Selwyn et al. [26] Grieve and co-workers have proposed a Bayesian approach to the analysis of the AB/BA cross-over design, which has been extended to incorporate baseline measurements and to other cross-over designs [15, 20, 21, 27]. According to Freeman, this is the only satisfactory analysis of the two-period two-treatment cross-over design [14]. We here present a brief summary of this approach.

The assumed model is model (1). Grieve considers a joint uninformative prior distribution for the model parameters [20]. Let us note  $p(\tau, \lambda, \sigma^2, \rho)$  the prior distribution of model parameters of interest (for simplicity, the period effect is not explicitated here). The prior distribution has the form [28]

$$p(\tau, \lambda, \sigma^2, \rho) \propto \frac{1}{\sigma^2(1+\rho)}$$

Noting  $Y$  the observed data, the following posterior distributions are obtained for the location parameters  $\tau, \lambda$  given the observations and the second-order parameters  $\sigma^2, \rho$  are obtained from standard properties of the normal distribution as:

$$\begin{aligned}
p(\tau|\lambda, \sigma^2, \rho, Y) &= N\left(\hat{\tau}^c + \frac{\lambda}{2}; \frac{m\sigma^2}{2}\right) \\
p(\tau|\sigma^2, \rho, Y) &= N\left(\hat{\tau}^p; \frac{m\sigma^2}{1-\rho}\right) \\
p(\lambda|\sigma^2, \rho, Y) &= N\left(\hat{\lambda}; 2m\sigma^2 \frac{1+\rho}{1-\rho}\right) \\
p(\tau, \lambda|\sigma^2, \rho, Y) &= N\left(\begin{matrix} \hat{\tau}^p \\ \hat{\lambda} \end{matrix}; \frac{m\sigma^2}{1-\rho}\right) \begin{bmatrix} 1 & 1+\rho \\ 1+\rho & 2(1+\rho) \end{bmatrix}
\end{aligned}$$

Unconditional distributions are computed by reparameterizing the model with  $\sigma_1^2 = \sigma^2 + 2\sigma_s^2 = \sigma^2(1-\rho)(1+\rho)$  and  $\sigma^2$  instead of  $\sigma^2$  and  $\rho$ . They depend on the sum of squares associated with residual error (SSE) and with subject within group (SSS) of the ANOVA table. The following posterior distributions are obtained:

$$p(\lambda|Y) = t\left(\hat{\lambda}; \frac{2mSSS}{n_1 + n_2 - 2}; n_1 + n_2 - 2\right) \quad (2)$$

$$p(\tau|\lambda, Y) = t\left(\hat{\tau}^c + \frac{\lambda}{2}; \frac{mSSE}{2(n_1 + n_2 - 2)}; n_1 + n_2 - 2\right) \quad (3)$$

where  $t(\mu, \theta, \nu)$  denotes the noncentral and scaled  $t$  distribution with  $\nu$  degrees of freedom, location parameter  $\mu$  and scale parameter  $\theta^{1/2}$ . The integral necessary to obtain the posterior density of  $\tau$  unconditional on  $\lambda$  is intractable, but the following approximation was derived by Grieve:

$$p(\tau|Y) \approx t\left(\hat{\tau}^p; \frac{m\phi}{2\nu}; \nu\right) \quad (4)$$

where

$$\begin{aligned}
\nu &= \frac{(SSE + SSS)^2 (n_1 + n_2 - 6)}{SSE^2 + SSS^2} + 4 \\
\phi &= \frac{(\phi - 2)(SSE + SSS)}{n_1 + n_2 - 4}
\end{aligned}$$

Grieve shows how to additionally take the constraint  $\rho > 0$  into account in this analysis [15, 20]. However, if negative correlations between different measurements on a same patient were expected, a cross-over design would unlikely be conducted. Additionally, the implied modifications only make little difference, so that these refinements are not further considered here.

The results of a Bayesian analysis of the transermal nitroglycerin trial are displayed in Figure 2. As shown in Grieve, the posterior distribution  $P(\tau < 0|\lambda = 0, Y)$  and  $P(\tau < 0|Y)$  differ markedly [15]. For example,  $P(\tau < 0|Y) = 0.85$ , whereas  $P(\tau < 0|\lambda = 0, Y) > 0.9999$ . Note that despite the modified parameterization, these

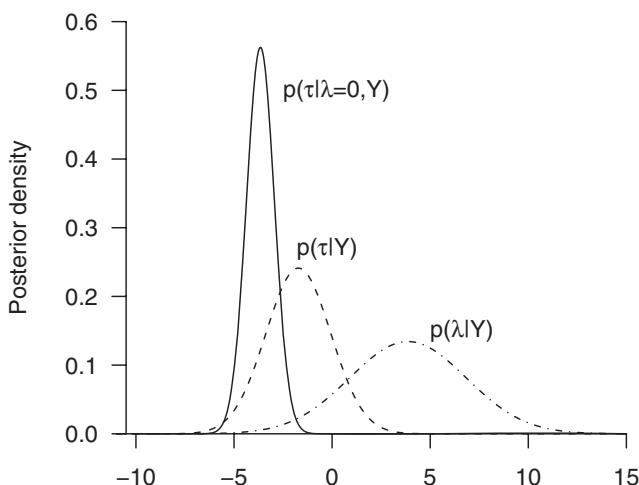


**FIGURE 2** Posterior marginal distribution of  $\lambda$  and  $\tau$  for transdermal nitroglycerin trial data.

values are identical to Grieve's analyses. Recall that the  $P$  value associated to a (frequentist) test of carry-over is 0.198, thus not supporting a model with a carry-over effect in the nitroglycerin trial data. On the contrary, the posterior density of  $\lambda$  seems rather in favour of a carry-over, with  $P(\lambda > 0|Y) = 0.90$ .

Results of the example well illustrate that a model without carry-over, noted  $M_0$ , and a model with carry-over,  $M_1$ , can lead to different conclusions. Grieve proposed a Bayesian approach of differentiating between the models  $M_0$  and  $M_1$ , using a prior specification allowing to incorporate information on prior belief on carry-over [20]. The proposed method models the set up as a mixture of  $M_0$  and  $M_1$ , and then uses a Bayes factor. The posterior probability of each of  $M_0$  and  $M_1$  can be written as

$$P(M_i|Y) = P(Y|M_i) \frac{P(M_i)}{P(Y)} \quad i = 0, 1$$

Noting the Bayes factor

$$B_{01} = \frac{P(M_0|Y)P(M_1)}{P(M_0)P(M_1|Y)} = \frac{P(Y|M_0)}{P(Y|M_1)}$$

and the prior odds on  $M_0$ ,  $\kappa = P(M_0)/P(M_1)$ , the posterior probability of the two models are

$$P(M_0|Y) = \frac{\kappa B_{01}}{1 + \kappa B_{01}} \quad \text{and} \quad P(M_1|Y) = \frac{1}{1 + \kappa B_{01}}$$

Inference on the treatment effect can then use the mixture posterior distribution

$$p(\tau|Y) = \frac{\kappa B_{01}}{1 + \kappa B_{01}} p(\tau|M_0, Y) + \frac{1}{1 + \kappa B_{01}} p(\tau|M_1, Y)$$

Where  $p(\tau|M_0, Y)$  and  $p(\tau|M_1, Y)$  are given by Equations (3) and (4), respectively. The Bayes factor  $B_{01}$  can be calculated following Spiegelhalter and Smith as [29]:

$$B_{01} = \sqrt{\frac{3}{2m}} \left( 1 + \frac{F_\lambda}{n_1 + n_2 - 2} \right)^{(n_1 + n_2)/2}$$

where  $F_\lambda$  is the  $F$ -test statistic for assessing significance of carry-over effect  $\lambda$ . The value of  $\kappa$  one chooses represents his (her) prior belief in carry-over. A value of  $\kappa = 1$  indicates indifference to the choice of a model.

A graph of posterior treatment effect as a function of prior belief in carry-over can be plotted. It represents either the posterior mean and the 95% highest posterior density (HPD) interval or either  $P(\tau < 0|Y)$  or  $P(\tau > 0|Y)$  according to model parameterization against  $P(M_1) = (1 + \kappa)^{-1}$ . Both are given in Figure 3.

Figure 3 displays some of the possible summaries of the posterior distribution of  $\tau$  according to the prior belief in carry-over for the example data. These are the posterior mean with 95% HPD interval and the probability of a benefit of transdermal nitroglycerin in the prevention of anginal attacks,  $P(\tau < 0|Y)$ . Details on the interpretation of such graphs are well described in the works of Grieve [15, 20, 27]. Briefly, the results show here that the conclusions strongly depend on the prior belief in carry-over, but it has been pointed out that it should be due to the relatively small value of  $B_{01}$  (2.052, in the present case) [20].

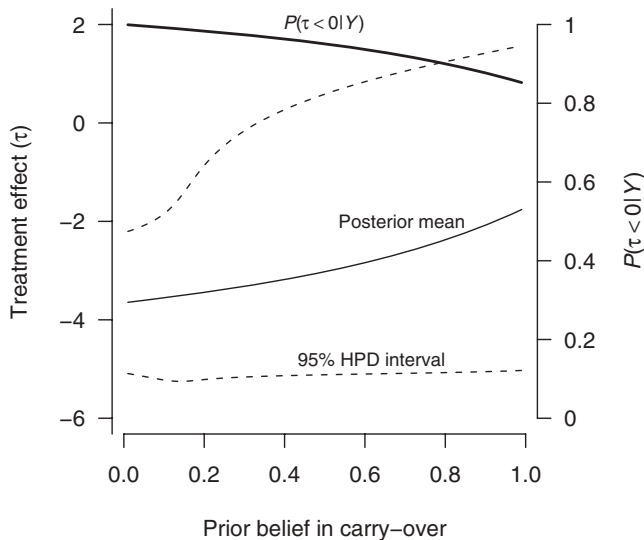


**FIGURE 3** Relationship between posterior inference on treatment effect  $\tau$  and prior belief in a carry-over effect (expressed as probability of a model with carry-over) for the weekly anginal attack rates.



The main advantage of Bayesian analysis of a cross-over trial is that it allows to stand in between the options of CROS and PAR analyses of a classical two-stage design. Several extensions of this technique have been considered but not reviewed here. Notably, it has been developed to consider baseline measurements [27] or nonuniform covariance matrices [30]. The use of informative priors has also been outlined [15].

## 13.4 OTHER ENDPOINTS

Analysis of AB/BA cross-over trials has generated a wide literature for normally distributed outcomes, as summarized in Section 13.3. We review here some methods for other situations.

### 13.4.1 Nonparametric Analyses

Model (1) and  $t$  tests rely on the assumption that the outcome measurements  $Y_{ijk}$  are normally distributed random variables. In some cases, however, such an assumption is clearly unverified. A first approach is to perform some variance stabilizing transformation, for example, taking the logarithm of data, in order to obtain a distribution closer to a normal distribution. Another choice is to prefer a nonparametric analysis that does not rely on a particular distribution for data. Nonparametric analyses are also well suited for ordinal variables with more than 4 or 5 classes. An excellent review of nonparametric analyses of cross-over trials is given by Tudor and Koch [31].

If a nonparametric analysis is found better adapted, the basic analysis described in Section 13.3.1 can be very simply carried out using a nonparametric test based on ranks such as Wilcoxon signed rank-sum test or a sign test, instead of the paired  $t$  test. This corresponds to test whether cross-over differences obtained for each patients are centered around zero.

Several nonparametric methods accounting for a period effect are also available [11]. Koch, in an important study on nonparametric analysis of cross-over trials proposed to use a Wilcoxon rank-sum test (or equivalently a Mann–Whitney U test) to compare period differences across groups [32]. Contrary to the basic analysis, the period differences are constructed for each patient, that is,  $Y_{ij2} - Y_{ij1}$ , whatever  $i$  is. A nonparametric test for equality of the distribution of period differences across both groups is then performed. A significant difference results in the conclusion that the average period difference is not the same in both group, which would imply a treatment effect. An estimator of this treatment effect can be obtained using Hodges–Lehman estimator, and 95% confidence limits may also be computed [33]. Senn has underlined the parallel between such a procedure and the CROS estimator of Hills–Armitage [11].

A more general procedure was also proposed by Koch that resembles the two-stage design discussed in previous sections [9, 31, 32]. In case of a significant test for treatment effect, as described above, a similar nonparametric test for carry-over is performed. Contrary to Grizzle’s two-stage procedure, this test is not carried out first but has a supportive role in a second stage, in order to address the concerns

regarding the classical two-stage procedure [11, 14]. A test of period effect given no carry-over can also be performed.

Nonparametric methods have been extended to account for multiple centers or strata, or baseline measurements. For further reading, see Tudor and Koch [31] or Jones and Kenward [9].

### 13.4.2 Binary or Categorical Data

For binary data, analysis has long relied on more or less appropriate methods for contingency tables, with the exception of the Mainland–Gart test. More recent works share the common feature to use an explicit model [34–37] while former methods often relied on implicit models [38]. We briefly present here some of the possible approaches to the analysis of binary AB/BA cross-over data, as well as their extension to count data.

**Mainland–Gart Test** The Mainland–Gart test examines the association between the period and group for patient preference [39, 40]. Subjects are classified according to their group into three categories: those who responded to treatment in period 1 and not in period 2, whose response is noted (1,0); those who responded to treatment in period 2 but not in period 1, that is, response (0,1); and those who had identical response after each period, noted (1,1) or (0,0) and who exhibited no preference. An example is given Table 6.

Under the null hypothesis, the preference (i.e., response in period 1 or in period 2) should only reflect the period effect. Ignoring patients who exhibited no preference such as in McNemar’s test, that is, ignoring the counts  $b$  and  $e$  in Table 6, a difference between treatment should lead to an association between group and the category of response (1,0) or (0,1). Such an association can be tested using any test for unpaired proportions, such as Fisher’s exact test or a chi-square test with or without continuity correction.

One of the disadvantages of such a procedure is that the number of subjects actually contributing to the test procedure may be small, as observations with the same response to both treatments are simply discarded. On the contrary, the Mainland–Gart test does not depend on the randomization of subjects into groups [9]. A test relying on randomization, but exploiting all information, was subsequently proposed by Prescott [41]. It will not be detailed here.

**Model-Based Methods** As shown for continuous outcomes, cross-over trials rely on within-subject information for the comparison of treatments. As a consequence, they result in correlated observations. Methods for modeling correlated binary data, although theory exists for many years, have long suffered from unavailability in statistical software packages. Recent advances in this domain have led to a rapid

**TABLE 6 Contingency Table for Mainland–Gart Test**

| Randomization Group | Overall Response |                |       |
|---------------------|------------------|----------------|-------|
|                     | (1,0)            | (1,1) or (0,0) | (0,1) |
| 1                   | $a$              | $b$            | $c$   |
| 2                   | $d$              | $e$            | $f$   |

diffusion of new methods for analysis of clustered binary data in the last years [34].

Contrary to the case of a linear model such as (1) for normally distributed data, the effect of a covariate on binary correlated data may have two different interpretations. The marginal effect corresponds to the average difference between two populations defined by differing covariate values, whereas the subject-specific or conditional effect corresponds to the expected difference for a same subject under different covariate values. These effect interpretations correspond to different possible models, namely the marginal or population-averaged model and the marginal or subject-specific model [34, 36, 42, 43]. Actually, for a linear model as (1), the outcome and the linear combination of covariates are on the same scale, so that parameters in both modeling approaches have the same interpretation, the difference lying in the error structure. For binary data, where a generalized model with a nonlinear link function relating the linear combination of covariates to the probability of response has to be used, both models cannot have the same interpretation anymore [36, 38].

Let us briefly detail these models and summarize their condition of use for AB/BA cross-over data. Paralleling model (1), a marginal logistic model can be expressed as

$$\text{logit } P(Y_{ijk} = 1) = \log \left[ \frac{P(Y_{ijk} = 1)}{P(Y_{ijk} = 0)} \right] = \mu + \pi_j + \delta_k \tau_l \quad (5)$$

where  $\pi_j$ ,  $\delta_k$ , and  $\tau_l$  have the same definition as for Equation (1). Other choices for the link function, such as a probit, for instance, are possible. Subject effect does not appear explicitly in model (5). Rather, dependence among observations from the same subject is accounted for via a robust covariance matrix. Model parameters are only related to the whole population, and estimates will be similar to classical logistic regression. The difference will lie in the standard errors of the parameter estimates. Several approaches have been proposed for model inference, one of the most popular being the generalized estimating equations (GEE) approach of Liang and Zeger [44]. These population-averaged models have been favored for practical computational aspects and availability in standard statistical software packages [38]. However, they are best adapted to assess the effects of cluster-level, that is, subject level in the case of cross-over trial data, covariates [36].

A conditional model can be written in a form even closer to model (1), as

$$\text{logit } P(Y_{ijk} = 1) = \mu + s_{ij} + \pi_j + \delta_k \tau_l \quad (6)$$

with  $s_{ij}$  denoting the effect associated with subject  $ij$ . As for model (1), observations are assumed independent conditionally of the subject effect, and the subject effect can be considered as either fixed or random. Two approaches to conditional modeling can thus be considered in practice. The conditional likelihood approach eliminates the subject effect [45] and corresponds to the Mainland–Gart test for the AB/BA design [38]. The other approach assumes a particular distribution for the subject effect, usually a normal distribution, which leads to a generalized linear mixed model (GLMM). This is the closest approach to model (1). In a conditional model, the covariate effects are measured conditionally on the subject effect  $s_{ij}$ . Thus,

regression parameters corresponding to covariates that do not change within subject may suffer from interpretation problems when using a conditional likelihood approach, as such effects are actually not observable [42, 43]. Conversely, GLMMs are based on both within-subject and between-subject comparisons, at the cost of distributional assumptions for the subject effects [36, 42]. For cross-over trials, however, where treatment effect estimation relies on inference about individual differences, conditional models are particularly well suited [36]. It must be noted that GLMMs have been long unavailable for use in routine practice, as major statistical packages did not offer this type of model. This was due in particular to the intractable likelihood function leading to very computer-intensive algorithms involving numerical integration. More recently, there has been a large development of methods, making the use of these models easier, though it still remains a topic of current research.

**Categorical Data** For categorical data, the same types of model as described above for binary data exist. More detailed can be found in appropriate references [9, 38, 46].

### 13.4.3 Censored Endpoints

Few works in the literature have concerned the analysis of censored endpoints in cross-over trials [47–49]. At first, cross-over trials and censored data may seem not to fit with each other because one usually thinks of survival. However, a cross-over may be adapted if the treatments are aimed for recurrent events, such as infection or headache. Another domain could be measures of some form of endurance [49].

For a single duration response, France et al. [48] have proposed to use a Cox proportional hazards model stratified by patient [50]. Such a model cannot incorporate a group effect and thus explicitly does not address carry-over. A further analysis of the same example by Bristol proposed to use a group effect without stratification [51]. Both methods led, however, to very different results. Another approach used a parametric log-linear model for a Weibull process [49, 52]. The results showed that a stratified model should be preferred to an unstratified model with period effect.

A markedly different approach was adopted by Feingold and Gillespie [47]. It used an additive model instead of the proportional hazards model of France et al. [48], which was additionally based on the usual methods for complete (uncensored) data. Two tests were considered, namely the Wilcoxon test, which only applies to two-period designs, and a more general but less powerful score transformation procedure. It was shown that these methods were more powerful than that of France et al. [48].

Multiple duration responses were only considered by Lindsey, through the log-linear model evoked above [49].

## 13.5 POWER AND SAMPLE SIZE CONSIDERATIONS

It is sometimes stated that cross-over trials require half the sample size of a parallel-arm trial to demonstrate a given (direct) treatment effect because each patient

receives both treatments. Recent work based on a review of published cross-over studies has shown that a parallel design is expected to need between 4 and 10 times more subjects than a cross-over design in order to achieve the same power, leading to costs between 2 and 5 times higher [53]. In this section, we give sample size equations for a two-period two-treatment cross-over trial with a continuous outcome. We then provide some discussion on the comparative efficiency of cross-over and parallel trials. We place ourselves in the situation where there is no carry-over and analysis is performed using ANOVA in model (1) or, equivalently, CROS.

### 13.5.1 Sample Size

Sample size is usually computed to ensure the trial will have sufficient power against a given alternative hypothesis. Under this alternative, both treatments A and B have different direct effects and the treatment effect  $\tau = \tau_2 - \tau_1$  is supposed equal to  $\Delta$ . This value is often referred as the minimal clinically relevant difference or the expected treatment effect. Sample size for cross-over trials are computed using usual methodology for the test of the null hypothesis  $\tau = 0$ . Let  $\alpha$  and  $\beta$  denote the type I and II error rates, that is, the probability of rejecting the null hypothesis when it is actually true and the probability of not rejecting the null hypothesis when the alternative is actually true, respectively. These error rates are usually fixed by convention at 0.05 for  $\alpha$  and 0.10 or 0.20 for  $\beta$  [54], though some different values can be sometimes found, such as lower  $\beta$  or  $\alpha = 0.025$  for one-sided tests. Moreover, the type II error rate is more frequently expressed using its counterpart  $1 - \beta$ , called power, which is the probability of rejecting the null hypothesis when there is actually a treatment effect. The null hypothesis can be rejected when  $\hat{\tau}^c > 0$  or  $\hat{\tau}^c < 0$  by a sufficient amount. Here, we consider the case when both chances of rejecting the null hypothesis are taken into account, which is referred to as a two-tailed or two-sided test.

As noted above, the null hypothesis of no treatment effect can be tested using  $\hat{\tau}^c$ . Recalling Table 1 for a balanced cross-over trial, that is, when  $n = n_1 = n_2$ ,  $\hat{\tau}^c$  has variance  $\sigma^2/n$ , and expectation 0 under the null hypothesis and  $\Delta$  under the alternative. Assuming the within-subject variance  $\sigma^2$  is known,  $\hat{\tau}^c$  has a normal distribution and usual sample size for Gaussian variables can be used [9, 11, 54]:

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\Delta^2} \quad (7)$$

where  $z_{1-\gamma}$  denotes the  $(1 - \gamma)$  quantile of the standard normal distribution.

If  $\sigma^2$  is unknown, it has to be estimated from the trial's data, and  $\hat{\tau}^c$  has a Student distribution. One could use appropriate modification of Equation (7) using the noncentral  $t$  distribution. As an approximation, a correction factor  $1/4 z_{1-\alpha/2}$  may be added to  $n$ , which improves the approximation of Equation (7) [54]:

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\Delta^2} + \frac{z_{1-\alpha/2}}{4} \quad (8)$$

To use Equation (7) or (8) in practice, one has to choose a value for  $\Delta$  but also to obtain an a priori estimate for  $\sigma^2$ . This is naturally also the case when planning

any type of trial, and such values are usually chosen using existing trials or studies in the domain considered. However, for cross-over trials, it might be a little more difficult to obtain a value for  $\sigma^2$ , as it represents the within-subject variance only. It may be worth noting that  $\sigma^2$  is half the variance of cross-over (or within-subject) differences, which may be more easily obtained using prior published information, for instance. The within-subject variance is also  $(1-\rho)\sigma_T^2$ , where  $\sigma_T^2$  is the total variance of an observation.

As an example, consider the transdermal nitroglycerin trial used for illustration throughout this chapter. Suppose it was considered at the planning stage that a mean diminution of two attacks per week would be clinically relevant and that the within-subject variance was not known, but that in prior early phase studies where patients received transdermal nitroglycerin the within-subject before–after differences had variance of about 30. A value of  $\sigma^2 = 15$  would then seem reasonable. With  $\Delta = -2$  and, for example,  $\alpha = 0.05$  and  $\beta = 0.20$ , formula (8) would lead to

$$n = \frac{(1.96 + 0.84)^2 \times 15}{4} + \frac{1.96}{4} = 29.89$$

By rounding up to 30, it would then be decided to recruit 30 patients in each group.

Sample size for cross-over trials in other situations, such as an equivalence or a noninferiority trial, or for bioequivalence, have been considered in detail in the review of Julious for normally distributed data [54]. A method for binary endpoints is given in Ezzet and Whitehead [46].

### 13.5.2 Efficiency of Cross-Over Versus Parallel-Group Design

A complete review of efficiency considerations for cross-over and parallel designs is given in Chapter 9 of Senn [11]. We sketch here some of these issues.

Consider a cross-over trial where  $n$  patients are randomly assigned to each of the sequences AB and BA and a parallel-group trial where  $p$  patients are randomly assigned to treatment by A and  $p$  patients are randomly assigned to treatment by B. As considered higher, the estimate of treatment effect in the cross-over trial is  $\hat{\tau}^c$ , which has variance  $\sigma^2/n$ . Conversely, the estimate of treatment effect in the parallel design is analogous of  $\hat{\tau}^p$ , as all data will be recorded in a single (first) period. From Table 5, replacing  $n_1$  and  $n_2$  by  $p$ , the variance of the estimator of treatment effect in this design is  $2\sigma^2/[p(1-\rho)] = 2\sigma_T^2/p$ . To obtain equal variance estimates, it is then necessary to recruit  $p = 2n/(1-\rho)$  patients per arm in the parallel-group trial. In the variance component model (1),  $\rho \geq 0$ , so that  $p$  will always be larger than  $n$ , even in the case where between-subject variance is negligible as compared to within-patient variance. This is because the parallel-group design is also subject to within-subject variation, which is often overlooked, as remarked by Senn [11]. Once again, if the variance component model does not hold and negative values of  $\rho$  are expected, it is anyway most improbable that a cross-over trial would be considered at all.

For choosing, whenever possible, between a parallel and a cross-over design, one may also take the duration of the trial into account, and not only the sample size. This should account for the two treatment periods of the cross-over trial but also for the

possible run-in and wash-out periods. Senn shows that when there is no run-in nor wash-out, the cross-over trial duration is always shorter than that of a parallel trial [11]. With no run-in but a wash-out duration equal to treatment period duration, the cross-over trial is shorter as soon as  $1/(1 - \rho) > 3/2$ , which is the case when between-subject variation is at least half the within-subject variation, a reasonable situation in practice. For more general cases, the reader may directly refer to Senn [11].

## 13.6 MORE COMPLICATED DESIGNS

There has been some work done toward finding the optimal design for a particular situation [9]. In particular, designs for two treatments but with more than two periods, have been studied, as a way of estimating carry-over and treatment-by-period interactions. Cross-over trials for more than two treatments, and usually at least three periods, have also been considered.

### 13.6.1 Higher Order Designs for Two Treatments

The main limitation of the AB/BA design without baseline measurements is that carry-over effects, treatment-by-period interaction, and group effects cannot be separated. Higher order designs, under specific and possibly disputable assumptions, may allow estimation of carry-over or treatment-by-period interactions.

Several assumptions or models for carry-over have been considered to examine design optimality. Some are, for example, the *simple carry-over* model, where the carry-over lasts exactly for one period and only depends on the last treatment, or the *steady-state carry-over* model, where the carry-over also lasts exactly for one period but treatment does not induce carry-over into itself.

Assuming either model, there have been numerous works on “optimality” for a wide range of cross-over designs with more groups and two periods (e.g., AB/AA/BA/BB) or with more periods (e.g., ABB/BAA) or both, such as ABA/BAB/AAB/BBA or AABB/BBAA/ABBA/BAAB [55–58]. The choice between these designs is not easy. Generally, increasing the number of periods increases the precision of the estimation of carry-over effects, but this gain is not always more efficient in terms of costs than including more patients. If the major cost is recruitment, then it has been advised to keep patients as many periods as is practical and ethical [9].

However, a greater number of periods may also lead to consider issues such as second-order carry-over or treatment-by-carry-over interactions. Overall, if a particular design should be recommended, the AABB/BBAA/ABBA/BAAB or the ABB/BAA designs would have the favor of Jones and Kenward [9].

These approaches, and particularly the assumptions concerning carry-over, have been quite sharply criticized [2, 11]. Some authors thus do not recommend multiperiod cross-over trials to be conducted except if prior knowledge on carry-over supports the underlying assumptions.

### 13.6.2 Designs for Three or More Treatments

Cross-over designs have been extended to compare more than two treatments. In practice, such designs have been preferred in situations where intervals between



treatments were short (e.g., some days). They have been quite popular in early phase I trials, with treatments also intended as several doses of the same drug.

Planning of such trials often relies on choosing sequences that form a Latin square. An example of such a design for four treatments (A, B, C, and D) may be the one displayed in Table 7, where patients would be randomly assigned in equal number to each of the four groups.

This Latin square ensures that every subject receives all four treatments, and each treatment occurs only once in each group. Moreover, the example given Table 7 also ensures that each ordered pair of treatments (e.g., A followed by B) occurs once and only once. The literature refers to these Latin squares as balanced. The efficiency of Latin square designs is the highest possible if the Latin square is balanced. As an example, the Latin square given in Table 8 is not balanced because treatment A is followed by treatment B more than once.

These designs are also most efficient if the simple carry-over model holds. Alternative models for other carry-over situations are considered in Jones and Donev [59].

Additionally, to achieve balance, it may be desirable that each treatment be followed by each other exactly once. This is not always possible in a simple Latin square, and sets of orthogonal Latin squares have been used for this purpose. Often, very few patients are included in each group. An example of a complete set of orthogonal Latin squares is presented in Table 9.

Such designs show the advantage of being balanced over all preceding treatments, so they are balanced for all first-order, second-order, and so on carry-over effects. Conversely, they may require much more subjects. For instance, using the design given in Table 9 implies including multiples of 12 subjects. Moreover, having an increased number of groups, they have a higher probability of a sequence being incorrectly administered to a patient. Williams showed that balance could be achieved by using a particular Latin square for an even number of treatments and

**TABLE 7** Balanced Latin Square Design for Four Treatments

| Group | Period |   |   |   |
|-------|--------|---|---|---|
|       | 1      | 2 | 3 | 4 |
| 1     | A      | D | B | C |
| 2     | B      | A | C | D |
| 3     | C      | B | D | A |
| 4     | D      | C | A | B |

**TABLE 8** Another Latin Square Design for Four Treatments

| Group | Period |   |   |   |
|-------|--------|---|---|---|
|       | 1      | 2 | 3 | 4 |
| 1     | A      | B | C | D |
| 2     | B      | C | D | A |
| 3     | C      | D | A | B |
| 4     | D      | A | B | C |

**TABLE 9** Orthogonal Latin Squares Design for Four Treatments

| Group | Period |   |   |   |
|-------|--------|---|---|---|
|       | 1      | 2 | 3 | 4 |
| 1     | A      | B | C | D |
| 2     | B      | A | D | C |
| 3     | C      | D | A | B |
| 4     | D      | C | B | A |
| 5     | A      | D | B | C |
| 6     | B      | C | A | B |
| 7     | C      | B | D | A |
| 8     | D      | A | C | B |
| 9     | A      | C | D | B |
| 10    | B      | D | C | A |
| 11    | C      | A | B | D |
| 12    | D      | B | A | C |

two particular squares for an odd number of treatments [60, 61]. Such designs require less subjects than using complete sets of orthogonal squares.

All these designs have the same number of treatments and periods. Sometimes, more treatments are investigated than can be reasonably given to a single patient. Some designs thus may have less periods than treatments. Conversely, to better estimate carry-over effects, more periods than treatments could be necessary, as in the two-treatment case. Details on both types of designs can be found, for example, in Jones and Kenward [9]. Briefly, in the situation where the number of treatments is greater than the number of periods, the simplest method is to delete periods from a full set of orthogonal Latin squares [62, 63]. Another popular method is the balanced incomplete block design, where each treatment is replicated the same number of times, and each pair of treatments occurs in the same block the same number of times [63]. These designs are, however, not balanced for carry-over, as each treatment does not follow every other with the same frequency. More precisions can be found in cross-over textbooks [9, 11].

### 13.7 SUMMARY

Cross-over trials are trials where subjects subsequently receive all the treatments being compared (or at least some of them). Between the treatment periods, wash-out periods are used to avoid a residual or carry-over effect of a treatment into the next period. The estimate of treatment effect being based on within-subject rather than between-subject information, cross-over trials are usually more efficient, but they can only be used in situations where treatments will not durably modify patient conditions.

Although the use of complex models for cross-over trials was long limited by availability in common statistical software packages, all analyses presented here can be easily performed using standard software such as SAS (SAS Institute Inc., Cary, NC) or R (The R Foundation for Statistical Computing, Vienna, Austria).

## REFERENCES

1. Brown, B. W. (1980), The crossover experiment for clinical trials, *Biometrics*, 36, 69–79.
2. Fleiss, J. L. (1989), A critique of recent research on the two-treatment crossover design, *Controlled Clin. Trials*, 10, 237–243.
3. Grieve, A. P. (1994), Crossover studies, *Clin. Pharmacol. Ther.*, 56, 112–114.
4. Hills, M., and Armitage, P. (1979), The two-period cross-over clinical trial, *Br. J. Clin. Pharmacol.*, 8, 7–20.
5. Jones, B., and Lewis, J. A. (1995), The case for cross-over trials in phase III, *Statist. Med.*, 14, 1025–1038.
6. Senn, S. J., and Hildebrand, H. (1991), Crossover trials, degrees of freedom, the carryover problem and its dual, *Statist. Med.*, 10, 1361–1374.
7. Food and Drug Administration. (1977), A report on the two-period crossover design and its applicability in trials of clinical effectiveness. Minutes of the Biometric and Epidemiology Methodology Advisory Committee (BEMAC) meeting.
8. ICH International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. (1998). *Note for the Guidance on Statistical Principles in Clinical Trials (ICH-E9)*. Geneva, Switzerland. <http://www.ich.org/LOB/media/MEDIA485.pdf>.
9. Jones, B. J., and Kenward, M. G. (2003), *Design and Analysis of Cross-Over Trials*, 2nd ed., Chapman & Hall, London.
10. Ratkowsky, D. A., Evans, M. A., and Alldredge, J. R. (1993), *Cross-over Experiments, Design, Analysis and Application*, Marcel Dekker, New York.
11. Senn, S. (2002), *Cross-over Trials in Clinical Research*, 2nd ed., Wiley, Chichester.
12. Grizzle, J. E. (1965), The two-period change-over design and its use in clinical trials, *Biometrics*, 21, 467–480.
13. Willan, A. R., and Pater, J. L. (1986), Carryover and the two-period crossover clinical trial, *Biometrics*, 42, 593–599.
14. Freeman, P. R. (1989), The performance of the two-stage analysis of two-treatment, two-period crossover trials, *Statist. Med.*, 8, 1421–1432.
15. Grieve, A. P. (1994), Bayesian analyses of two-treatment crossover studies, *Stat. Methods Med. Res.*, 3, 407–429.
16. Wheatley, D. (1987), Transdermal nitroglycerin in angina pectoris, *Stress Med.*, 3, 199–203.
17. Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis*, 2nd ed., Wiley, New York.
18. Senn, S. (1994), The AB/BA crossover: past, present and future? *Stat. Methods Med. Res.*, 3, 303–324.
19. Senn, S. (1998), Crossover designs, in Armitage, P., and Colton, T., Eds. *The Encyclopedia of Biostatistics*, 2 vols, Wiley, New York, pp. 1033–1049.
20. Grieve, A. P. (1985), A Bayesian analysis of the two-period crossover design for clinical trials, *Biometrics*, 41, 979–990.
21. Racine, A., Grieve, A. P., Flühler, H., et al. (1986), Bayesian methods in practice: Experiences in the pharmaceutical industry (with discussion), *Appl. Statist.*, 35, 93–150.
22. Senn, S. J. (1991), Problems with the two stage analysis of crossover trials, *Br. J. Clin. Pharmacol.*, 32, 133.
23. Poloniecki, J. D., and Pearce, A. C. (1983), Interaction in the two-way crossover trial (Letter to the Editor), *Biometrics*, 39, 798.

24. Kenward, M. G., and Jones, B. (1987), The analysis of data from  $2 \times 2$  cross-over trials with baseline measurements, *Statist. Med.*, 6, 911–926.
25. Spiegelhalter, D. J., Myles, J. P., Jones, D. R., et al. (1999), Methods in health service research. An introduction to Bayesian methods in health technology assessment, *BMJ*, 319, 508–512.
26. Selwyn, M. R., Dempster, A. R., and Hall, N. R. (1981), A Bayesian approach to bioequivalence for the  $2 \times 2$  changeover design, *Biometrics*, 37, 11–21.
27. Grieve, A. P. (1994), Extending a Bayesian analysis of the two-period crossover to allow for baseline measurements, *Statist. Med.*, 13, 905–929.
28. Geisser, S. (1964), Estimation in the uniform covariance case, *J. Roy. Stat. Soc., B*, 26, 477–483.
29. Spiegelhalter, D. J., and Smith, A. F. M. (1982), Bayes factors for linear and log-linear models with vague prior information, *J. Roy. Stat. Soc., B*, 44, 377–387.
30. Grieve, A. P. (1992), Implementation of Bayesian methods in the pharmaceutical industry, PhD Thesis, University of Nottingham.
31. Tudor, G., and Koch, G. G. (1994), Review of nonparametric methods for the analysis of crossover studies, *Stat. Methods Med. Res.*, 3, 345–381.
32. Koch, G. G. (1972), The use of non-parametric methods in the statistical analysis of the two-period change-over design, *Biometrics*, 28, 577–584.
33. Hauschke, D., Steinijans, V. W., and Diletti, E. (1990), A distribution-free procedure for the statistical analysis of bioequivalence studies, *Int. J. Clin. Pharmacol. Therapy Toxicol.*, 28, 72–78.
34. Carlin, J. B., Wolfe, R., Brown, C. H., et al. (2001), A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes, *Biostatistics*, 2, 397–416.
35. Kenward, M. G., and Jones, B. (1992), Alternative approaches to the analysis of binary and repeated measurements, *J. Biopharma. Stat.*, 2, 137–170.
36. Neuhaus, J. M. (1992), Statistical methods for longitudinal and clustered designs with binary responses, *Stat. Methods Med. Res.*, 1, 249–73.
37. Zeger, S. L., and Liang, K.-Y. (1992), An overview of methods for the analysis of longitudinal data, *Stat. Med.*, 11, 1825–1839.
38. Kenward, M. G., and Jones, B. (1994), The analysis of binary and categorical data from crossover trials, *Statist. Methods Med. Res.*, 3, 325–344.
39. Gart, J. J. (1969), An exact test for comparing matched proportions in crossover designs, *Biometrika*, 56, 75–80.
40. Mainland, D. (1963), *Elementary Medical Statistics*, 2nd ed., Saunders, Philadelphia.
41. Prescott, R. J. (1981), The comparison of success rates in cross-over trials in the presence of order effect, *Appl. Stat.*, 30, 9–15.
42. Hu, F. B., Goldberg, J., Hedeker, D., et al. (1998), Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes, *Am. J. Epidemiol.*, 147, 694–703.
43. Neuhaus, J. M., Kalbfleisch, J. D., and Hauck, W. W. (1991), A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data, *Int. Stat. Rev.*, 59, 25–35.
44. Liang, K.-Y., and Zeger, S. L. (1986), Longitudinal analysis using generalized linear models, *Biometrika*, 73, 13–22.
45. Breslow, N. E., and Day, N. E. (1980), *Statistical Methods in Cancer Research I: The Analysis of Case Control Studies*, International Agency for Research on Cancer, Lyon.

46. Ezzet, F., and Whitehead, J. (1991), A random effects model for ordinal responses from a crossover trial, *Stat. Med.*, 10, 901–906; discussion 906–907.
47. Feingold, M., and Gillespie, B. W. (1996), Cross-over trials with censored data, *Stat. Med.*, 15, 953–967.
48. France, L. A., Lewis, J. A., and Kay, R. (1991), The analysis of failure time data in cross-over studies, *Stat. Med.*, 10, 1099–1113.
49. Lindsey, J. K., Jones, B., and Lewis, J. A. (1996), Analysis of cross-over trials for duration data, *Stat. Med.*, 15, 527–535.
50. Cox, D. R. (1972), Regression models and life tables (with discussion), *J. Roy. Stat. Soc., B*, 34, 187–220.
51. Bristol, D. R. (1992), The analysis of failure time data in crossover studies (Letter to the Editor), *Stat. Med.*, 11, 975–977.
52. Lindsey, J. K. (1994), Fitting parametric counting process by using log linear models, *Appl. Stat.*, 44, 201–212.
53. Garcia, R., Benet, M., Arnau, C., et al. (2004), Efficiency of cross-over designs: An empirical estimation, *Stat. Med.*, 23, 3773–3780.
54. Julious, S. A. (2004), Sample sizes for clinical trials with normal data, *Stat. Med.*, 23, 1921–1986.
55. Ebbutt, A. F. (1984), Three-period crossover designs for two treatments, *Biometrics*, 40, 219–224.
56. Kershner, R. P., and Federer, W. I. (1981), Two-treatment crossover designs for estimating a variety of effects, *J. Am. Stat. Assoc.*, 76, 612–619.
57. Laska, E. M., and Meisner, M. (1985), A variational approach to optimal two-treatment crossover designs: Applications to carryover effects models, *J. Am. Stat. Assoc.*, 80, 704–710.
58. Laska, E. M., Meisner, M., and Kushner, H. B. (1983), Optimal crossover designs in the presence of carryover effects, *Biometrics*, 39, 1087–1091.
59. Jones, B., and Donev, A. N. (1996), Modelling and design of cross-over trials, *Stat. Med.*, 15, 1435–1446.
60. Williams, E. J. (1949), Experimental designs balanced for the estimation of residual effects of treatments, *Austral. J. Scient. Res.*, 2, 149–168.
61. Williams, E. J. (1950), Experimental designs balanced for pairs of residual effects, *Austral. J. Scient. Res.*, 3, 351–363.
62. Patterson, H. D. (1951), Change-over trials, *J. Roy. Stat. Soc., B*, 13, 256–271.
63. Patterson, H. D. (1952), The construction of balanced designs for experiments involving sequences of treatments, *Biometrika*, 39.



# 14.1

---

## *Biomarkers*

**Michael R. Bleavins,<sup>1</sup> Claudio Carini,<sup>2</sup> Malle Jurima-Romet,<sup>3</sup>  
and Ramin Rahbari<sup>4</sup>**

<sup>1</sup>*Michigan Technology and Research Institute, Ann Arbor, Michigan*

<sup>2</sup>*Fresenius Biotech of North America, Waltham, Massachusetts*

<sup>3</sup>*MDS Pharma Services, Montreal, Quebec*

<sup>4</sup>*Innovative Scientific Management, New York, New York*

### **Contents**

|                                          |     |
|------------------------------------------|-----|
| 14.1.1 Biomarker Definition and Criteria | 851 |
| 14.1.2 Biomarkers and Decision Making    | 853 |
| 14.1.3 Life Cycle of Biomarker           | 855 |
| 14.1.4 Drug Safety                       | 859 |
| 14.1.5 Efficacy                          | 861 |
| 14.1.6 Patient Stratification            | 864 |
| References                               | 867 |

### **14.1.1 BIOMARKER DEFINITION AND CRITERIA**

Biomarkers are valuable tools for drug development that provide greater accuracy or more complete information regarding drug performance and/or disease progression. Developments in genomics, genetics, proteomics, and imaging have highlighted the importance of biomarkers as useful clinical indicators. The Biomarkers Definitions Working Group [1], sponsored by the National Institutes of Health, has defined a biomarker as:

A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.



In most cases, a biomarker should satisfy two criteria: (1) be associated with the biological mechanisms involved within a disease or treatment of a disease and (2) have the potential to correlate statistically with clinical outcomes. This broad definition means that a vast array of tests falls within this scope, spanning methods as varied as measurements of serum troponins and cardiac computed tomography. It is also important to clarify that most biomarkers are *not* surrogate endpoints. Under regulatory agency definitions, surrogate endpoints are findings or measurements that may be substituted for disease endpoints in clinical trials to evaluate the safety or effectiveness of a medical therapy for treating disease. These well-defined and accepted biomarkers represent a small subcategory of the biomarker world.

While there are different categories of biomarkers, three categories are most relevant to this discussion: target, mechanism, and outcome. Although a biomarker used in any stage of drug development may be from the target, mechanism, or outcome categories, in general, target biomarkers are more often part of discovery applications and outcome biomarkers are more likely to be seen in clinical situations. Qualification and acceptance of biomarkers, in applications paralleling the test's ultimate use and in cost-effective manners that minimize the possibility of patient harm, also are key considerations [2]. The primary goals of incorporating biomarkers into a drug development program, either clinically or preclinically, are to make better informed decisions earlier, more accurately, and to reduce the overall cost of drug development. The basic strategy is to define the critical question to be answered and then employ an appropriate biomarker in a relevant biological system that will answer that question as early as possible. When the purpose of the biomarker is to confirm hitting the putative therapeutic target, it then is possible to assess the relevance of the presumed pathophysiological mechanism and if altering this mechanism will affect clinical status.

Strategies for incorporating new biomarkers into clinical practice form the basis for translational medicine. One of the most tangible problems that scientists have been facing in recent years is finding biomarkers that are predictive, or "translate well," from animal or *in silico* models to humans. For example, inhibiting an enzyme in animal models may have a dramatic effect in the animal, whereas inhibiting the same enzyme in humans may have very limited or no clinical impact. The reasons for lack of correlation are many and varied: (1) the therapy can impact the biomarker chosen, but the parameter is irrelevant to the disease, (2) the biomarker may be indicative of the disease but not reflect clinically important effects of the treatment, and (3) a drug may work in humans through several different pathways not reflected in animal models. It is important to remember that a drug or other intervention should treat the disease, not a biomarker. Proper biomarker progress requires close dialog between biomarker scientists and practicing physicians, and this ideally can be accomplished under the banner of translational medicine.

The cost and duration of clinical trials are two of the primary drivers of biomarker research. Improving the understanding of a new drug's effects also is fundamental in using biomarkers within the pharmaceutical industry and for current regulatory focus on improving the success rates in drug development. Development of new therapeutic agents is estimated to cost between \$800 million [3] and \$1.7 billion [4] and requires 7–12 years. Reducing the high attrition rate of drugs by selecting better molecules, and being able to make decisions on compound viability earlier, have tremendous potential for reducing the high costs of bringing new

therapeutic agents to patients. The poor success rate for many compounds makes this an especially attractive goal. Attrition rates in phase II clinical research programs are as high as 55% [5] to 90% [6], with lack of efficacy as the primary cause. Thus, biomarkers that accurately predict human efficacy at the preclinical or phase I clinical trial stages are particularly valuable in reducing costs and increasing confidence in the compound. Similarly, biomarkers of chronic disease progression/resolution would substantially enhance clinical trial design and characterization of research efforts in diseases such as rheumatoid arthritis, type II diabetes, Alzheimer's disease, and Parkinson's disease.

Although many think of *omics* (proteomics, genomics, metabonomics, etc.) when the term *biomarkers* is used, many biomarkers are less exotic or "high-tech" and yet have contributed significantly to our understanding of disease processes. In 1733, Stephen Hales first measured blood pressure [7], and subsequent to this initial measurement, many investigators found blood pressure to be elevated in some disorders and reduced in others. Over many years, blood pressure was proven to have such clear clinical utility and predictivity to warrant acceptance as a surrogate for many forms of cardiovascular disease. Upon first description, however, blood pressure characterization was a new tool with unknown real potential. Although now deeply ingrained in medical practice, blood pressure initially presented a situation akin to the current status of C-reactive protein (CRP). The utility and significance of CRP [8, 9] are yet to be defined for inflammation and other diseases and only will be understood after more clinical experience is gained.

Although CRP is becoming a standard marker for inflammation, in the absence of known sources of inflammation, CRP is associated with the pathogenic inflammatory process in several diseases such as rheumatoid arthritis. Clearly, CRP and other newly discovered biomarkers need to stand the test of time in large clinical trials and long-term population studies. In addition, a biomarker like CRP must be reliably and conveniently measured. Clinical laboratories must be able to reproduce the measure, and national standards must be established so that consistent interpretation of measures from different laboratories can be achieved [10]. As the significance of inflammation has been identified in a variety of diseases, other inflammatory biomarkers are also the subject of current research interest. Lind [11] has described fibrinogen, leukocyte count, interleukin-6, tumor necrosis factor- $\alpha$  (TNF $\alpha$ ), immunoglobulins, cell-adhesion molecules, complement, serum amyloid A, phospholipase-A2, and neopterin as other biomarkers of inflammation that might be useful in predicting coronary diseases. Like blood pressure, if these biomarkers prove important in mediating disease processes, they also will be targets for drugs that block their action, and thereby reduce the risk for disease.

### 14.1.2 BIOMARKERS AND DECISION MAKING

Biomarkers constitute a rational approach that, at its most optimal, reflects both the biology of the disease and the effect of the drug candidate. Proper incorporation of biomarkers in drug development strategy enables the concept of "fail fast, fail early," allowing early identification of the extremely high proportion of compounds that fail during drug development. In addition to minimizing human exposure to drugs unlikely to be effective or with safety concerns, substantial cost savings can

be achieved by shifting resources to those molecules most likely to become effective new medicines. A properly selected and characterized biomarker also facilitates the choice of the proper critical path toward approval and can differentiate the new product from approved drugs in a competitive marketplace. The challenge, therefore, is to identify relevant biomarkers early enough to implement them for “go, no-go” decisions at critical stages of the development process.

Traditional clinical trial endpoints, such as morbidity and mortality, often require extended time frames and may be difficult to evaluate. Imaging-based biomarkers are providing objective endpoints that may be confidently evaluated in a reasonable time frame. Imaging techniques tend to be expensive but can be cost effective when used in well-defined situations where subjective assessment has been the only approach available. Examples of therapeutic areas where imaging is reshaping clinical trial design include Alzheimer’s disease, pain, and osteoarthritis. In affected joints, as well as in tissues expressing specific receptors, magnetic resonance imaging (MRI), computed tomography (CT), and position emission tomography (PET) imaging are delivering new information to clinicians and researchers. These biomarker technologies are rapidly developing, requiring close collaboration between basic researchers and physicians treating patients.

Biomarker development forms one of the cornerstones of a new working paradigm in the pharmaceutical industry by increasing the importance of linking diagnostic technologies with the use of drugs. The position of the Food and Drug Administration (FDA) is rather clear on this. Biomarkers are crucial to generate safe and efficacious drugs and are essential for deciding what patients should receive which treatment. It appears likely that development of pharmaceuticals also will drive the development of biomarkers as diagnostic probes for clinical decisions and stratification of patients. Preclinical biomarkers are essential to identifying compounds with the greatest potential for therapeutic intervention. Work at this stage of the process has tight controls of environmental conditions, diet, and genetics of the laboratory animals and provides access to tissues that cannot be collected during normal clinical trials. These studies provide opportunities for high-content data gathering using functional genomics, proteomics, metabonomics, simulations, computational methods/informatics, analytical technologies, and interventions. Results from the preclinical studies then can be distilled and used in the clinical phases of the most promising compounds.

In early-stage clinical trials, the use of suitable biomarkers should be incorporated to assist in demonstrating proof of concept (POC) and to identify appropriate dosage regimens for safe and efficacious clinical trials. Assessing which subpopulations are most likely to benefit from a new treatment can help in the planning of subsequent efficacy clinical trials. Although this approach may appear to be straightforward, integrating biomarkers can involve several practical problems and pitfalls. Clinical observations in a given patient may serve multiple purposes, with implications for the way in which the observed characteristics are documented, validated, interpreted, and used.

Aside from their use in diagnosis and prognosis, biomarkers currently have applications in drug discovery and development, aiding in formulating dosing and in guiding pharmacokinetic/pharmacodynamic (PK/PD) indices, and assisting in determining efficacy. These PK/PD indices are important factors in determining efficacious doses, particularly for evaluation in early-phase clinical trials. A deter-

mination of efficacy using biomarkers serves an important and cost-effective role in lieu of traditional large-scale clinical trials with standard clinical endpoints. Another important area that is well served by biomarkers is drug safety and the characterization of toxicity. Safety biomarkers can ensure exclusion of high-risk patients before treatment is administered, and monitor the need for early withdrawal of treatment before adverse side effects are manifested. These biomarkers also can be critical for assessing the relevance of preclinical findings and preventing or monitoring for their occurrence in humans.

### 14.1.3 LIFE CYCLE OF BIOMARKER

Biomarkers do not come to preclinical or clinical practice fully formed, fully validated, applicable across all patient populations, and with the guidance as to when to use or not use the particular marker. The tests are introduced poorly characterized with limited directions for understanding their limitations, potential, and applicability, much like an infant enters the life cycle. With care and further research, over time the biomarker can become a powerful tool to be utilized in drug development or patient care. In addition to the biological component of the biomarker requiring nurturing, the testing platform utilized to measure this biomarker usually does not exist as a point of care device that can be administered at the bedside. Rather, it is usually and necessarily by definition in the early stages, an untried, unreliable, and briefly tested platform that will be appropriately labeled “exploratory.”

To illustrate the life cycle of a biomarker, it can be viewed from late maturity, the stage at which most individuals outside of the research laboratory will first encounter the test, working backwards to inception. At its most mature stage, an accepted biomarker is commonly referred to as “validated,” meaning the biology is solid and the testing platform is reliable. But for the purpose of making what sort of decision? The term *validated* also conjures quite varied images in the minds of different people, with the same assay’s robustness assumed to be greater as one ascends each step on the management ladder. Currently, the concept of “fit-for-purpose” is gaining acceptance in the pharmaceutical industry to provide more precise terminology and recognizing that the requirements for a biomarker when screening compounds is fundamentally different than when deciding to escalate doses in a clinical trial. Earlier in this chapter the different classifications of biomarkers were described (target, mechanism, outcome), in addition to the decisions that can be made with the information the biomarker provides and the types of testing platforms that can be utilized. Since the nature of questions asked in different situations varies greatly, the real answer for when a given biomarker is appropriate lies in whether it can be expected to reliably provide the information necessary to confidently make a decision. The risks and benefits to be weighed when conducting *in vitro* testing of a chemical library using high-throughput screening are very different than determining whether to recommend angioplasty or choosing between orally active compounds acting through Factor Xa inhibition versus coumarin therapy.

Determining a biomarker to be fit-for-purpose in the indication for which it is to be used is a more precise designator of the confidence in the combination of the

biology, platform, patient population, and the proposed action. This approach also takes into account that the magnitude of a decision can be substantially different even when the underlying questions are similar. For example, the acceptable false-positive/false-negative rates for a biomarker may be higher when one is prioritizing a large number of early-stage molecules than when determining whether to cancel research on a potential therapeutic target for which only one or two candidate compounds have been identified.

The degree of testing also can be affected by knowledge of similar drugs in the same class or current clinical practice. A biomarker supporting a new statin compound with a safety profile comparable to the marketed products is unlikely to require the same degree of characterization as a biomarker for drugs with a completely new indication (i.e., neovascularization). The potential of a biomarker to provide actionable information is directly proportional to its maturity, fitness for the intended purpose, and the gravity of the decision to be made using this information.

The development of a new biomarker minimally takes 6–9 months and can span many years (Fig. 1). Testing and building confidence in the biology can strengthen or wane, and the testing platform can improve by becoming smaller, more portable, and more widely available. In many ways, a biomarker goes through phases similar to the development of a child. The life cycle of a biomarker can be viewed as progressing from infant to toddler, then to child, preteen, adolescent, young adult, independent individual, and finally a mature person. The progression of blood pressure as a biomarker through these stages is illustrated below.

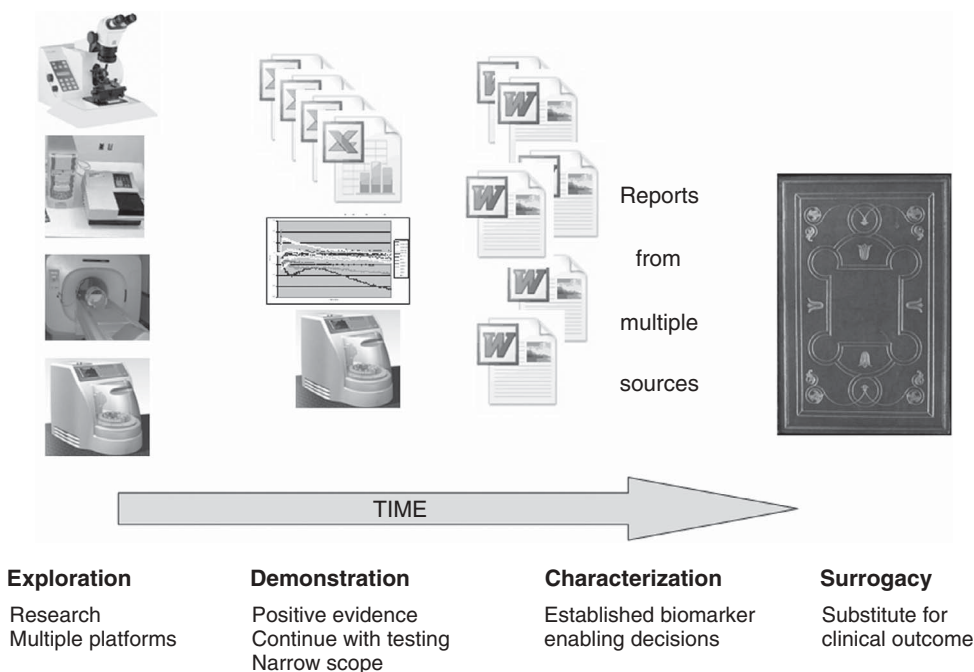


**FIGURE 1** Life cycle of a biomarker: exploration, demonstration, characterization, and surrogacy.

In 1913, Janeway published a well-documented study where blood pressure was measured via sphygmomanometer in a large group of patients [12]. This study was made possible because of the blood pressure cuff developed by Scipione Riva-Rocci [13] and a large relevant patient population (~8000) examined over nearly a 10-year period. Janeway noted that 11% of the patients had systolic blood pressures over 160 mmHg, and this group of patients survived for only 4–5 years after being identified with the condition he named “hypertensive cardiovascular disease” [12]. The underlying physiology of hypertension was poorly understood at this time, and therapy options largely were limited to rest and reduced salt intake. Extensive research efforts in multiple laboratories led researchers in numerous directions, but it was not until 1939 that the renin–angiotensin–aldosterone system (RAAS) became the coherent focus of primary research [14]. Much more bench research and clinical findings, in addition to a consensus opinion of the community, led to the National High Blood Pressure Education Program Task Force I (NHLBI/NIH); *Report to the Hypertension Information and Education Advisory Committee* in 1973, and the report of the Joint National Committee (JNC I) on *Detection, Evaluation, and Treatment of High Blood Pressure* [15, 16]. These reports address reference ranges for the biomarker of blood pressure, compare recommended therapies and pharmaceutical interventions, and in the most recent release discuss patient stratification. This developing application and acceptance of blood pressure as a biomarker beyond its original use demonstrates its increasing maturity to the current status as one of the very few accepted surrogate biomarkers by both clinical practitioners and regulatory agencies. The life cycle of blood pressure is nearly 100 years and continues to grow as knowledge in the field accumulates and scientific and medical confidence in the biomarker increases.

Further illustrating that biomarkers are not a new concept, and their progression through various stages of development/acceptance, can be seen in routine medical practice. The likely tests performed at an annual physical examination include several biomarkers that have changed over the years, as has the clinical confidence in using them to make decisions. Among the tests administered will be body weight, diastolic/systolic blood pressure, heart rate, and serum lipid profile to provide an initial diagnosis of cardiovascular health. If any of these tests are skewed or abnormal, additional biomarker measurements would be initiated. Electrocardiogram (EKG), stress tests, troponins, CRP, and a host of other tests can characterize the presence or absence of heart disease, as well as inform the physician on appropriate therapy. Tests/biomarkers such as glucose, glycated hemoglobin, and alanine aminotransferase (ALT) provide information on other possible clinically important alterations and diseases that may indirectly impact cardiovascular health. Among these biomarkers, some can be considered “adolescent” (CRP) as “young adult” (prostate-specific antigen), “independent individual” (troponin, stress test, body mass index, glycated hemoglobin), and “mature person” (blood pressure, EKG, lipid profile, glucose, ALT).

The two initial reports on hypertension mentioned above can be considered the forerunners of biomarker initiatives currently underway as consortia between the pharmaceutical industry and the FDA, International Life Sciences Institute/Health and Environmental Sciences Institute (ILSI/HESI), and National Institutes of Health (NIH). The driver for these present-day programs clearly is the economics of drug development. The process of drug development has grown significantly in



cost at the same time as the number of drugs submitted and approved per year has steadily decreased. The decline has resulted from a wide range of causes from the complexity of new disease targets, multifactorial diseases, increased regulatory requirements, and higher expectations of drug safety. As the acute causes of mortality and morbidity have been addressed by the medical community, the population now is experiencing the ill effects of more chronic diseases, and novel therapeutic targets and methods have become necessary. Many of these new therapeutic areas are still in need of predictive animal models for early-stage testing, as well as more comprehensive tools for defining the condition. Due to the greater uncertainty in these areas, the compound attrition rates have been higher. Inefficiencies in the drug development process and marketing expenses also have contributed to the increases in drug development costs. The marketplace also has a substantial impact on whether a new drug, even if approved, will actually be distributed to patients. A compound that cannot be sufficiently differentiated from similar approved molecules in terms of safety, efficacy, or cost, is unlikely to be included in the formularies of large insurers or become an approved course of treatment in countries with publicly funded health care systems. Biomarkers can be used to establish differentiation in class and this use will likely increase.

In the early-attrition paradigm designed to identify nonviable compounds at the first possible stage of development, many new biomarkers are being utilized by the pharmaceutical industry to achieve the cost-containment goals. These biomarkers are only sufficiently characterized to be considered a discovery tool and are generally limited to uses that prioritize similar early-stage molecules. Nevertheless, the tests are valuable for proving viability or proof of nonviability of a therapeutic target or class of compounds. These biomarkers most often are mechanism of action or safety markers. In contrast, disease-related outcome biomarkers can be difficult to validate in the absence of sizable efforts like the one for hypertension. Without additional research involving multiple clinical situations, possibly necessitating consortia to share resources, very few biomarkers currently being developed will mature past the point in their life cycle where they are used for internal decision making during pharmaceutical testing.

Biomarkers are seen as among the most promising tools for reducing the cost of developing drugs and ensuring that the appropriate patient populations are treated. Proper use of these tests can improve the quality and quantity of mechanistic information obtained at each step of the drug development process, allow better understanding of other results and the factors relevant to their interpretation, aid in the allocation of scarce resources, and provide clearer indications of a molecule's activity or potential safety liabilities. Clearly there is a wide spectrum of technologies encompassing biomarkers, with individuals from many specialties involved in the field with varying degrees of success. There also are numerous considerations to keep in mind when selecting a biomarker, and which must be balanced against each other, to ensure the test developed will meet the intended goal and actually be used. Among the critical biomarker parameters to be defined are:

- Nature of the biomarker's use (activity, efficacy, safety, differentiation)
- Throughput, capacity, and turn-around time requirements
- Reproducibility, sources of physiologic variation, and sensitivity
- Sample requirements and practicality of collection from the patients or pre-clinical species being evaluated



- Stability of parameter and compatibility with the study design
- Safety considerations in collecting and/or processing sample
- Existence of similar assay or approach that can be modified to the situation
- Time frame in which an appropriate method needs to be developed
- Defining how the biomarker adds value

Biomarkers of pharmacological or therapeutic activity and safety can be critical indicators of an early compound's activity and safety. This is especially important as drug development costs rise and the number of new compounds registered decreases. Appropriate selection and application of biomarkers can allow companies to terminate nonviable programs earlier, and reallocate resources to compounds with the highest probability of helping patients. Biomarkers also can be important for small companies in highlighting a new chemical entity's attractiveness and potential when looking for partnerships or licensing opportunities with big pharmaceutical companies.

#### 14.1.4 DRUG SAFETY

Biomarkers in drug safety can be key determinants in expediting decisions and may foretell the survival of a molecule, or even an entire program. Serious effects seen in preclinical testing, particularly at low multiples of presumed efficacious exposure, necessitate either a biomarker that allows safe progression or discontinuation of the compound. Many drug safety biomarkers are clinical pathology parameters, with long histories and years of clinical use. Alanine aminotransferase (ALT), troponins, urea nitrogen, bile acids, creatinine, electrolytes, and a host of other tests allow monitoring of tissue injury and have good correlations between humans and pre-clinical species. Safety departments within the pharmaceutical industry also are actively developing new biomarkers using the older platforms, as well as the technologies of proteomics, genomics, and metabonomics.

Discovering, developing, validating, and utilizing safety biomarkers present a variety of challenges in advancing pharmaceutical compounds. Safety has implications at all stages of the drug development process and carries significant relevance for clinical trial design and regulatory approval. Since the purpose of preclinical toxicology testing is to assist in defining relevant safety concerns for the clinical trials, biomarkers developed and evaluated in these experiments can be adapted for studies conducted in healthy volunteers and patients. Acceptable tests for the adverse effects seen in the preclinical testing that can be used in healthy human volunteers or patients are essential for safely advancing a compound into clinical trials. This connection between drug safety and clinical groups is one of the most important aspects of translational medicine.

Early identification of safety concerns, or indications that they will not be an issue, allows more effective prioritization and resource allocation. It also is often necessary to work backwards from a phenotype (liver injury, vasculitis, renal damage, etc.) caused by an unknown mechanism to a specific parameter, or group of markers, that caused the injury. The “wealth of opportunities and potential mechanisms” can be the most difficult and valuable aspect of the problem. In

contrast to many activity/efficacy biomarkers, biomarkers of specific tissue damage often can be applicable for monitoring toxicity of a range of molecules acting via a variety of mechanisms, but on the same tissue or subset of tissues.

Often even more challenging to manage are effects seen in preclinical species for which the clinical significance is unknown. Vasculitis has been an issue for a variety of therapeutic targets [antivirals, phosphodiesterase inhibitors, endothelin antagonists, norepinephrine reuptake inhibitors]. In these instances effects seen in animals may or may not be predictive of human risk, with the further complication that no good measures of the injury exist for monitoring in clinical trials. A biomarker panel capable of clearly monitoring vascular injury and early changes would allow compounds to be safely tested in human volunteers and patients, as well as having significant value in improving clinical practice related to vasculitis.

Identification, development, and validation of a safety biomarker follow approaches very similar to that required for efficacy assays. These include:

- Based on relevant parameters, with a foundation in a known mechanism.
- Coordinated engagement across departments (discovery, clinical, pharmacokinetics and drug metabolism, safety/toxicology) to optimize acceptance.
- Confidence in the test's performance that it can be used to make decisions.
- Use technology to efficiently demonstrate role in the toxicity, while remembering that the platforms are the tools and not the goal.
- Keep in mind that the question to be answered/characterized may be tissue rather than specific enzyme based—do not limit options.
- Comparative biology may be important in determining if the toxicity is relevant for human volunteers/patients or unique to the preclinical species.
- Good science is essential, as is the ability to show why the assay should be accepted as premonitory or able to detect changes before serious and irreversible injury has occurred.
- Ability to translate from preclinical to clinical application, as well as reverse engineering clinical tests, is critical for acceptance when entering new areas.

In many instances it is possible to utilize an efficacy biomarker to predict undesirable effects, particularly in areas such as anticoagulants. Thrombin inhibitors, low-molecular-weight heparins, Factor Xa inhibitors, and vitamin K antagonists are valuable therapeutic agents. Since most of these molecules at high doses exhibit only minimal toxicity other than those mediated through the same pathways as their pharmacologic methods of action (bruising, internal bleeding, etc.), the same assays that predict their activity can be used to assess toxicity. Tests such as ecarin clotting time, activated partial thromboplastin time, prothrombin time, and Factor Xa activity all represent useful safety and efficacy biomarkers. Since these assays show considerable translatability across species, and are already within the repertoire of most hospital laboratories, their acceptance by clinicians and regulatory agencies is enhanced. An added benefit for the development of drugs for which toxicity is generally limited to exaggerated pharmacology is that frequently early indications of efficacy and maximum tolerated doses can more accurately be estimated in phase I clinical trials than for most other compounds.

Additionally, it is highly desirable to have a biomarker that precedes overt tissue injury. Many of the oldest and most accepted biomarkers of toxicity represent leakage of cellular components and therefore only are seen at significant levels after at least minimal damage to an organ system has occurred.

In many instances, significant corporate overhead is automatically assumed whenever the situation requires a safety biomarker. Concerns are frequently voiced regarding regulatory agency acceptance, interpretation of new tests, and clinical practice. While at Pfizer, two of the authors of this chapter (MRB and RR) were among the scientists who advanced and advocated the “fit-for-purpose” approach (Bleavins and Rahbari, personal communication). Careful consideration of the decisions to be made for the test results, nature of the toxicity, and population being evaluated are all critical. However, open discussion across toxicology and clinical departments, as well as early engagement of the relevant regulatory agencies has been productive. It has been our experience that if the science is solid, and the rationale for use of the biomarker as part of the compound development strategy is well communicated, the physicians involved with the clinical trials and the regulatory agency reviewers will feel comfortable with the monitoring, and compound development is expedited.

#### 14.1.5 EFFICACY

One of the most important areas in drug development for the application of biomarkers is defining and characterizing efficacy in clinical trials. As previously described, few validated biomarkers of disease efficacy and even fewer surrogate endpoints currently exist. Nevertheless, biomarkers are increasingly being incorporated into clinical trials in various therapeutic areas, often at early stages (e.g., phase I, phase IIa), and sometimes relying on advanced technologies such as genomic, proteomic, or imaging analyses. At this stage of drug development, there is a strong desire to have very quick assurances of hitting the target and a justification for continued development of the compound. In the later phases of development, especially after the initial proof of concept has been established, the more established measures of efficacy (“gold standard”) may be employed.

Initially seen as an extension of the mantra of “fail fast, fail early” that was widely adopted by large pharmaceutical companies in the 1990s in preclinical development programs, the value of biomarker use in clinical development is increasingly being recognized for more than just identifying potential drug failures at an earlier stage. Integration of a drug’s molecular response profile in animal disease models with clinical response enables better drug development decisions and efficiency. The current drug development paradigm is being reshaped by patient stratification using biomarker profiles.

An example of a clinical efficacy biomarker is an individual’s BCR-ABL tyrosine kinase genotype. Characterization of the chromosomal rearrangement creating this abnormal protein, and the multiple subsequent mutations that can occur, are used to identify patients with chronic myeloid leukemia who are likely to respond to the drug, imatinib mesylate (Gleevec), which inhibits this kinase [17]. Additionally, imatinib also is an inhibitor of the C-KIT tyrosine kinase. Based on convincing clinical biomarker data, the U.S. FDA approved the use of imatinib to treat BCR-

ABL-positive chronic myeloid leukemia and gastrointestinal stromal tumors associated with activating mutations in C-KIT.

An additional well-characterized example is the anticancer drug trastuzumab (Herceptin), which achieved global sales in excess of \$1.3 billion despite being efficacious in only a subset of patients [18]. Herceptin is an antibody that targets the HER2 receptor protein. Amplification of the human epidermal growth factor receptor (*HER*)2 (ErbB2) gene in approximately 25–30% of patients with breast cancer results in elevated expression of HER2. The clinical response rate to treatment with Herceptin is high at 35% in patients with overexpression of HER2-while the drug is ineffective in patients who do not have increased protein levels. Therefore, clinical genotyping of breast cancer patients to identify HER2-positive cancers and determine a patient's suitability for Herceptin therapy is now commonly practiced. This approach maximizes the drug's use in patients who will benefit from the drug, while minimizing the exposure and cost to other patients of potential toxicity when those individuals are unlikely to respond. Although not a perfect biomarker, HER2 expression is a valuable tool for the oncologist, and additional research is being conducted to refine the predictivity of this measure, either by alternative tests or better characterization of key variables.

Efficacy biomarkers are, of course, not limited to genotypes and can be based on gene expression patterns, individual proteins or proteomic patterns, metabolomics, imaging, histology, physicians' clinical observations, or patients' self-reported observations (e.g., pain intensity). Thus, an efficacy biomarker is not defined by technology but by its reliable and predictive correlation to a differential response in patient subpopulations. It should be noted that rarely is the genomic or other molecular biomarker profile used alone to determine the drug therapy and dosage regimen. Many other factors such as age, body weight, concurrent diseases, and concomitant medications need to be considered as well, and the ultimate decision has to be made by physicians who are cognizant of all of these factors as they apply to the individual patient and his/her biomarker profile. The biomarker offers a valuable tool for enhancing and supplementing existing options, and "individualizing" the patient's treatment.

Interindividual differences in drug absorption, distribution, metabolism, and excretion (ADME) are known to be associated with differential efficacy or toxicity responses to many compounds. Genetic polymorphisms in drug metabolizing enzymes (DMEs) are a major cause of this substantial variability among individuals. This is especially important for drugs with narrow therapeutic windows that undergo extensive biotransformation to pharmacologically inactive metabolite(s) by one or more polymorphic drug metabolizing enzymes. Poor metabolizers (PMs) are typically at increased risk of adverse drug reactions, while ultra-rapid metabolizers may not be able to achieve therapeutic exposures when the parent compound is the active form. Prodrugs requiring metabolic activation by a polymorphic enzyme to an active metabolite also may have reduced efficacy or higher incidences of treatment failure in PMs. Three clinically important cytochrome P450s (CYP2D6, CYP2C9, CYP2C19) collectively are involved in the metabolism of approximately 40% of marketed drugs [19], and have a large number of functionally significant genetic polymorphisms. One example of an efficacy failure in PMs of CYP2D6 pertains to codeine, a central analgesic and cough suppressant that is metabolized to its pharmacologically active metabolite, morphine, by CYP2D6. The analgesic

effects of codeine are dramatically reduced in CYP2D6 PMs [20], whereas subjects who have multiple copies of the *CYP2D6* gene, so-called ultra-rapid metabolizers (UMs), have an increased analgesic response to normal doses of codeine, and in some cases, experience severe toxicity [21].

For the past several decades, pharmaceutical sponsors and clinical investigators have been conducting clinical pharmacology studies to evaluate the effects of genetic polymorphisms in DMEs using subpopulations of extensive and PMs (and UMs where appropriate). In some situations, this has led to postapproval labeling changes as new information in specific populations has become available. Several recent examples include the addition of deoxyribonucleic acid (DNA)-based DME biomarkers to the labeling of 6-mercaptopurine (Purinethol), azathioprine (Imuran), and irinotecan (Camptosar) [22].

More recently, regulatory agencies have recognized that pharmacogenomic testing has evolved to the point that regulatory policies and guidance are required. In January 2005, the FDA approved the first clinical diagnostic test, Roche's AmpliChip for the rapid genotyping of *CYP2D6* and *CYP2C19* variants. To encourage more extensive sharing of genomic data, increase understanding on how and when to use this type of information, and determine opportunities for genomic studies to expedite drug development, the FDA published its *Guidance for Industry: Pharmacogenomic Data Submissions* in March 2005 [23].

Whether intended to improve efficacy or to avoid toxicity, biomarkers for selecting appropriate patients to enroll in clinical trials are becoming increasingly prospective rather than retrospective. Historically, as mentioned above, stratification has occurred after market approval when variation in the drug's effects has become better understood and could be correlated to results of a biomarker assay or diagnostic test. For example, the clinical utility of the kinase inhibitor gefitinib (Iressa) was defined after the drug failed to show efficacy in an unselected population of patients with lung cancer. The drug was, however, associated with marked responses in a small subset of patients with non-small-cell lung carcinoma [24]. There are several other examples where a nonstratified approach to drug therapy failed because of poor efficacy or unacceptable toxicity, and subsequent patient stratification based on biomarkers restored drug efficacy in a subset of patients. Human immune deficiency virus (HIV) treatment regimens are based on matching therapies to the viral strain present in the patient in consideration of resistance-conferring mutations, and antibiotics are matched to specific resistant infections. The anticancer drug irinotecan (Camptosar) has been associated with severe toxicity (diarrhea and leukopenia) in patients with a genetic variant of the UDP-glucuronosyltransferase 1A1 enzyme (UGT1A1) [25]. An FDA-approved diagnostic test based on the UGT1A1 biomarker is now available to identify patients potentially at higher risk for developing toxicity [26].

From a drug development perspective, lessons learned from these retrospective applications of biomarkers are increasingly leading to more prospective incorporation of safety and efficacy biomarkers into clinical trials. By enriching patient populations in clinical trials with likely responders, and excluding patients at greater risk of developing toxicity, drug development risks are reduced at the same time as the sample size necessary to demonstrate efficacy is reduced. Most importantly for diseases with traditional long-term endpoints such as survival, efficacy biomarkers allow for shortened endpoint observation times. In chronic and slowly progressive

diseases (arthritis, Alzheimer's disease, diabetes, lysosomal storage diseases), biomarkers can be evaluated after several weeks/months of drug treatment versus years for traditional clinical endpoints. The escalating cost of drug development and poor survival of drug candidates in phase II are causing pharmaceutical companies to reevaluate programs for chronic diseases for which there are no predictive biomarkers of efficacy. This is an even more critical concern for biotechnology and emerging pharmaceutical companies, reliant on venture capital or other short-term financing, for which it is not financially feasible to conduct multiyear clinical trials with compounds of unknown activity. Demonstrating proof-of-concept in trials of shorter duration by employing efficacy biomarkers can make the difference between continuing to pursue a new molecular target and/or molecular entity and eliminating a research program due to poor clinical promise. Indeed, the use of biomarkers has become common practice in phase IIa first-in-patient and even in some phase I first-in-human trials. This approach has the potential to provide useful information to validate preclinical disease models, assist in dose and biomarker selection to take forward into larger late-stage clinical trials, as well as excite management and investors with early "hints" of efficacy. However, the complexity and cost of early clinical trials involving biomarkers increases, and one should be careful that "more information" does not quickly become "too much information." Particularly for molecular profiling or multivariant biomarker strategies, the amount of data generated even for relatively small patient sample sizes can be enormous, and may require sophisticated and time-consuming data analysis and interpretation. For novel biomarkers, data on the normal variation in baseline (predose) levels may be limited, and a preliminary "survey" study may be necessary to define normal ranges. Unless the biomarker incorporated into the study design is decision making for the compound's future, biomarkers are rarely useful when they result in equivocal findings.

It is apparent that the increasing use of efficacy biomarkers will continue to make important contributions to clinical drug development by lowering late-stage failure rates, improving decision making and the efficiency of clinical trials, improving patient care, and paving the way toward personalized medicine.

#### **14.1.6 PATIENT STRATIFICATION**

In traditional clinical trial designs, patients with a particular condition were enrolled in clinical trials assuming a near homogenous group (Fig. 2), with random selection to prevent bias. The goal of random sampling employed in traditional drug development was to ensure the general population was well represented. In reality, the groups were heterogeneous with respect to stage of disease, drug metabolizing capabilities, environmental conditions (e.g., diet, smoking status, lifestyle), concurrent or previous medication exposures, and often even underlying cause of the symptomology. By using biomarkers to better characterize the biological makeup of the participants in clinical trials, drug developers are streamlining the testing process. The most obvious and often quoted use of patient stratification is to select only patients who would be expected to respond to a particular investigational new drug based on the mechanism of action of the drug. The drug's statistically significant benefits, in this targeted subset of patients, could be shown in a smaller and shorter clinical trial than those needed for randomly selected patients. Under some



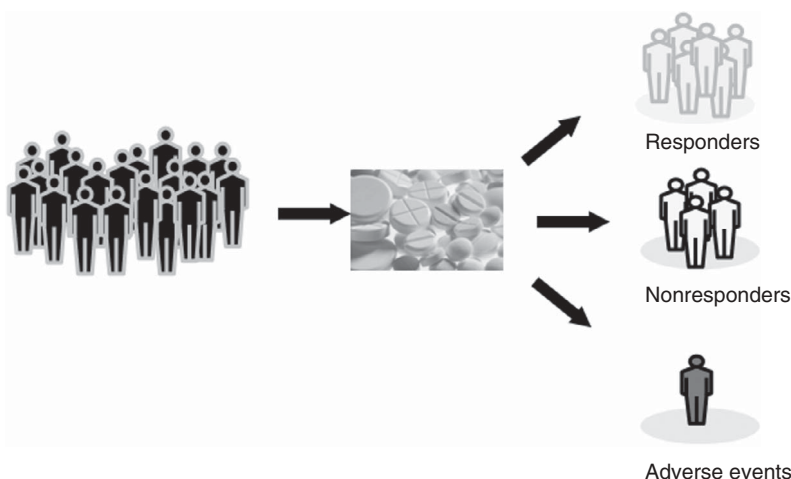


**FIGURE 2** Traditional treatment approaches often assume a relatively homogeneous patient population, with limited differentiation of individuals before administration of a new drug.

circumstances, stratification of responders could also be performed retrospectively on the data collected to demonstrate efficacy of a compound in a subpopulation. While generally not acceptable for registration, retrospective findings can be important for identifying potential new therapeutic applications and redirecting programs (e.g., Neurontin for neuropathic pain, Rogaine for hair growth, Viagra for erectile dysfunction).

Minimally, stratification can speed approval for drug candidates intended for a subset of patients while leaving the door open for further testing and market expansion in the more heterogeneous population of patients. Maximally, it can unmask a useful therapeutic agent that would be lost in the noise generated by the nonresponders in the study, as described above for trastuzumab and gefitinib.

Since the challenge of testing new drugs in chronic diseases is one of the most difficult aspects of tackling these devastating conditions, an example in this area seems appropriate. Alzheimer's disease (AD) is a major health care concern that will only become more pressing as the baby-boomer generation ages. However, clinical studies for AD are difficult due to the generally slow progression of the disease, subjective nature of diagnostic tests, and limited ability of the affected individuals to actively participate in their care. Selecting early-onset AD patients may provide advantages in studying this disease because their symptomology presents with faster progression. Clinical trials could be shorter, thereby providing a quicker determination on a new molecule's efficacy. Unfortunately, this approach also has some rather severe limitations. Initial clinical trials still need to be many months in duration to demonstrate a compound's activity, and the early-onset form of AD may or may not effectively predict effects in the much larger population developing AD later in life. Far more valuable to identifying new therapies for AD would be a biomarker that accurately predicts disease progression or resolution in weeks or months. A biomarker of this type would revolutionize clinical trial design and allow significantly more new chemical entities to be tested for activity. It also would provide desperately needed information on the relevance of animal models,



further increasing the speed and productivity of basic research into this disease. AD is at the forefront of biomarker development for neurodegenerative diseases, and several promising candidate biomarkers have emerged in recent years, including isoprotanes, tau, A $\beta$ , homocysteine, apolipoprotein E genotype, and several others [27]. Imaging technologies such as resting glucose PET and volumetric MRI are also being investigated as biomarkers for AD.

The focus of pharmacogenetics is to identify the molecular causes of differential therapeutic response across patient populations. It is generally appreciated that patients and their disease show significant variations in response to a given treatment based on their genetic background in addition to the better recognized environmental causes. Advances in understanding the mechanisms underlying diseases, as well as drug response, are increasingly creating opportunities to match the right patients with the right therapies that are more likely to be effective and safe. At the extreme of patient matching is personalized medicine (Fig. 3). Currently, there are few examples of patient matching to therapies for which some form of biomarkers are available to indicate that a given patient is likely to show a response to a specific therapy. In addition to the BCR-ABL example described above, for the chemokine receptor 5 (CCR5) antagonist anti-HIV drug Maraviroc, prospective stratification based on CCR5 and HIV-1 genotyping results was included in the new drug application (NDA) submission. These data also were key components in the successful registration of the compound in 2007.

Patient stratification strengthens traditional medicine by associating an individual patient with a specific therapy. Patient stratification is already practiced in several contexts, though often after first-line empirical treatment proves unsatisfactory because of poor efficacy or intolerable toxicity. Examples can be seen in antibiotics that are matched to specific resistant infections, or the often protracted process of selecting the appropriate medicine for patients with schizophrenia. Biomarkers enable the progression from empirical therapy to stratification therapy, linking patient subsets with therapies that have an increased likelihood of success. In breast cancer, patients with different levels of receptor tyrosine kinase HER2 expression characterize not only their cancer type but even their candidacy for receiving treatment. Similarly, HIV therapies that are tailored to the viral variant

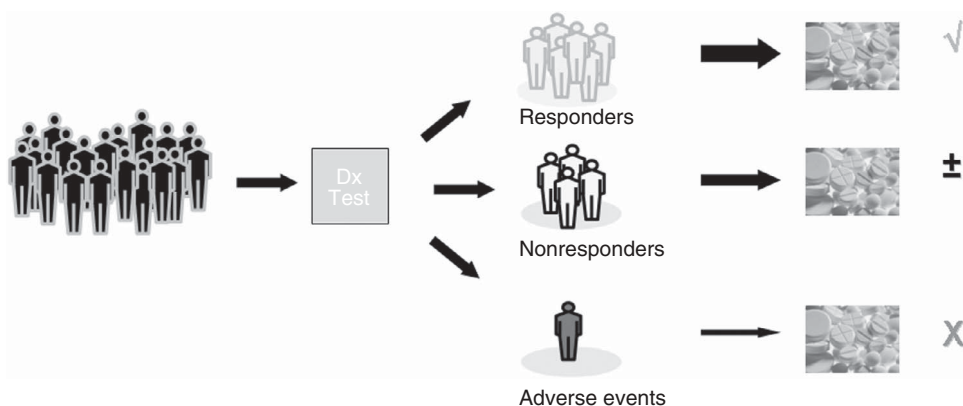


**FIGURE 3** Personalized medicine utilizes a diagnostic test/biomarker to optimize therapeutic decision making.

present and oncology treatment regimes prioritize agents based on histological and molecular tests. In the absence of a biomarker, patient stratification may not be possible. For instance, the lack of accepted biomarkers in depression that are reflective of disease status or progression often results in patients receiving several different therapies to identify one that leads to a satisfactory response. At present the treatment for depression remains limited to empirical, trial-and-error therapy.

Patient's stratification has considerable economic impact on the model of the pharmaceutical industry and society. By identifying the populations most likely to benefit from a new therapy, drug development costs can be reduced at the same time as treatment of people unlikely to respond is minimized. Exposure to drugs that do not address a patient's underlying disease is minimized, thereby also reducing the likelihood of adverse events in those for whom the compound does not provide help. Overall, the risk–benefit assessment is substantively refined when the potential for adverse drug–drug interactions is reduced, and the process enhanced when biomarker data enabling patient stratification are available. Patient compliance also will be improved as treatment becomes more closely linked with disease resolution in a higher proportion of the individuals treated. Ultimately, better targeting of specific disease and responsive patient through biomarkers can reshape clinical practice by providing clinicians with new opportunities for better patient care.

## REFERENCES

1. Biomarkers Definitions Working Group (2001), Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
2. Williams, S. A., Slavin, D. E., Wagner, J. A., et al. (2006), A cost-effective approach to the qualification and acceptance of biomarkers, *Nature Rev. Drug Dis.*, 5, 897–902.
3. DiMasi, J. A., Hansen, R. W., and Grabowski, H. D. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econ.*, 22, 151–158.
4. Bain & Co. (2003), Cost estimates quoted in *Chem. Eng. News*, 81(50), 8.
5. Hurko, O. (2006), Understanding the strategic importance of biomarkers for discovery and early development phases, *Drug Disc. World*, 16, 63–74.
6. Bleavins, M. R. (2006), Personal communication.
7. Hamilton, W. F., and Richards, D. W. (1982), The output of the heart, in Fishman, A. P., and Richards, D. W., Eds., *Circulation of the Blood: Men and Ideas*, Bethesda, MD: American Physiological Society, pp. 83–85.
8. Ridker, P. M., Cushman, M., Stampfer, M. J., et al. (1997), Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men, *N. Engl. J. Med.*, 336, 973–979.
9. Ridker, P. M., Buring, J. E., Cook, N. R., et al. (2003), C-reactive protein, the metabolic syndrome, and risk of incident cardiovascular events: An 8-year follow-up of 14,719 initially healthy American women, *Circulation*, 107, 391–397.
10. Pearson, T. A., Mensah, G. A., Alexander, R. W., et al. (2003), Markers of inflammation and cardiovascular disease: Application to clinical and public health practice. A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association, *Circulation*, 107, 499–511.
11. Lind, L. (2003), Circulating markers of inflammation and atherosclerosis, *Atherosclerosis*, 169, 203–214.

12. Janeway, T. C. (1913), A clinical study of hypertensive cardiovascular disease, *Arch. Intern. Med.*, 12, 755–798.
13. Riva-Rocci, S. (1896), Un nuovo sfigmomanometro, *Gazetta medica di torino* 47, 981–996; 1001–1017. [English translation: Faulconer, A. Jr., Keys, T. E., Eds., *Foundations of Anesthesiology*, Vol. 2, Charles C Thomas, Springfield, IL, 1965, pp. 1043–1075.]
14. Vertes, V., Tobias, L., and Galvin, S. (1991), Historical reflections on hypertension, *Prim. Care*, 18(3), 471–482.
15. National High Blood Pressure Education Program, Task Force I (1973), *Report to the Hypertension Information and Education Advisory Committee*, Recommendations for a national high blood pressure data base for effective antihypertensive therapy, U.S. Department of Health, Education and Welfare Publication No. (NIH) 74-593, Government Printing Office, Washington, DC.
16. JNC I (1977), Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure (JNC I), *JAMA*, 237, 255–261.
17. Druker, B. J., Talpaz, M., Resta, D. J., et al. (2001), Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia, *N. Engl. J. Med.*, 344, 1031–1037.
18. Burstein, H. J. (2005), The distinctive nature of HER2-positive breast cancer, *N. Engl. J. Med.* 353, 1652–1654.
19. Xie, H.-G., and Frueh, F. W. (2006), Pharmacogenomics steps toward personalized medicine, *Personalized Med.*, 2, 325–337.
20. Sindrup, S. H., and Brosen, K. (1995), The pharmacogenetics of codeine hypoalgesia, *Pharmacogenetics*, 6, 335–346.
21. Gasche, Y., Daali, Y., Fathi, M., et al. (2004), Codeine intoxication associated with ultra-rapid CYP2D6 metabolism, *N. Engl. J. Med.*, 351, 2827–2831.
22. Huang, S.-M., Goodsaid, F., Rahman, A., et al. (2006), Application of pharmacogenomics in clinical pharmacology, *Toxicol Mechanisms Methods*, 16, 89–99.
23. U.S. FDA (2005), *Guidance for Industry: Pharmacogenomic Data Submissions*; available at: <http://www.fda.gov/cder/genomics>.
24. Dowell, J. E., and Minna, J. D. (2005), Chasing mutations in the epidermal growth factor in lung cancer, *N. Engl. J. Med.*, 352, 830–832.
25. Andersson, T., Flockhart, D. A., Goldstein, D. B., et al. (2005), Drug-metabolizing enzymes: Evidence for clinical utility of pharmacogenomic test, *Clin. Pharmacol. Ther.*, 78, 559–581.
26. U.S. FDA (2005), FDA clears genetic test that advances personalized medicine: Test helps determine safety of drug therapy; available at: <http://www.fda.gov/bbs/topics/news/2005/new01220.html>.
27. Shaw, L. M., Korecka, M., Clark, C. M., et al. (2007), Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics, *Nature Rev. Drug Disc.* 6, 295–303.

# 14.2

---

## *Biomarkers in Clinical Drug Development: Parallel Analysis of Alzheimer Disease and Multiple Sclerosis*

**Christine Betard,<sup>1</sup> Filippo Martinelli Boneschi,<sup>2</sup> and Paulo Caramelli<sup>3</sup>**

<sup>1</sup>*Global Strategic Drug Development Unit, Quintiles, Levallois-Perret Cedex, France*

<sup>2</sup>*Neuro-Rehabilitation Unit, Department of Neurology,  
San Raffaele Scientific Milano, Italy*

<sup>3</sup>*Cognitive Neurology Unit, Department of Internal Medicine, Faculty of Medicine,  
Federal University of Minas Gerais, Belo Horizonte, Brazil*

### **Contents**

|          |                                                                                     |     |
|----------|-------------------------------------------------------------------------------------|-----|
| 14.2.1   | Introduction                                                                        | 870 |
| 14.2.1.1 | Working Definitions                                                                 | 872 |
| 14.2.2   | When to Use Biomarkers                                                              | 873 |
| 14.2.2.1 | Diagnosing AD and MS and Biomarkers to Identify Disease: Diagnostic Biomarkers      | 873 |
| 14.2.2.2 | Identifying Risk of Developing MS and AD: Presymptomatic Biomarkers                 | 874 |
| 14.2.3   | Clinical Development Application: Drug Efficacy and Safety                          | 882 |
| 14.2.3.1 | Using Biomarkers to Select Patient Subgroups                                        | 882 |
| 14.2.3.2 | Predicting Changes in Response to Therapy and Disease Course: Predictive Biomarkers | 883 |
| 14.2.4   | Conclusion                                                                          | 884 |
|          | Acknowledgments                                                                     | 885 |
|          | References                                                                          | 885 |

### 14.2.1 INTRODUCTION

Biomarkers are being used increasingly in clinical drug development by research-based pharmaceutical and biotechnology companies. Examples of their application to the development of new medicines to treat chronic disorders will be discussed in two central nervous system (CNS) diseases: multiple sclerosis (MS) and Alzheimer disease (AD). Combined, these two neurological disorders are creating great suffering to the affected patients and their families and represent a public health burden to societies. MS and AD are common and devastating neurological diseases for which pathophysiological processes leading to inflammation and neurodegeneration are thought to begin long before early clinical symptoms emerge and thus can be detected. For many years, AD and MS have been considered as two completely different diseases with no common features, MS being characterized by diffuse inflammation and demyelination, inflammation resulting in sclerotic scars (MS plaques) during healing [1], and AD clinically characterized by an insidious and progressive loss of memory and cognitive abilities, with *definite* diagnosis requiring pathological confirmation of amyloid plaques and neurofibrillary tangles associated with brain atrophy resulting from neuronal and synaptic loss [2–4].

Despite substantial progress, existing therapeutic interventions are only moderately effective in improving clinical symptoms, and research efforts are focusing on the development of disease-modifying therapies to impede disease progression. There is a widely shared perception that biomarkers are critical to identify the presymptomatic phase of the disease and to allow the development of these new therapies, since limitations in animal models and inadequate understanding of the pathophysiology still persist [U.S. Food and Drug Administration (FDA) Critical Path Initiative, European Medicines Agency (EMA) Innovative Medicine Initiative].

In recent years, transcriptomics and neuroproteomics analyses as well as genetic association studies have revealed that AD and MS pathophysiologies share common features of inflammation and neurodegeneration [1, 5–8] and also glucose metabolism impairment [9]. One of the main challenges in improving care of MS and AD patients resides in the poor performance of clinical diagnostic criteria that, in their current state, do not allow MS or AD diagnosis before the appearance of characteristic clinical symptoms but instead many years after the neurodegenerative and inflammatory processes have been started (Figure 1). Consequently, the diagnosis is made and the therapy is initiated when the disease process has reached irreversible steps, rendering the impact of potential disease-modifying therapies almost negligible in terms of clinical benefit. As basic molecular events in the initiation and progression of the diseases are still poorly understood, a rather long gap between disease onset and diagnosis remains. In addition, MS and AD show a high degree of individual variability in severity and clinical course, highlighting the existence of disease subtypes that current diagnostic criteria do not allow differentiation and which could necessitate different treatments.

A large panel of clinical biomarkers is beginning to be used in early drug development to determine whether the drug is reaching and affecting the appropriate molecular target in humans, allowing comparisons with preclinical data and providing measurable endpoints that predict desired or undesired clinical effects. The ultimate goal of this approach is to increase the success rate in the confirmatory stage of clinical development. Noninvasive neuroimaging techniques, particularly magnetic reso-

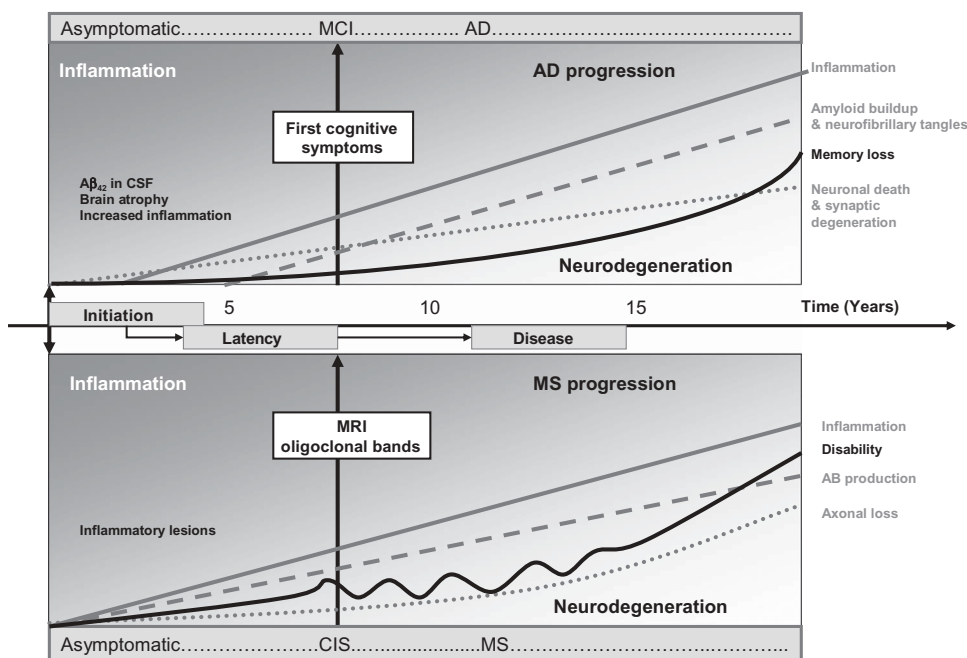


**FIGURE 1** Different stages of MS and AD related to time frame and window for therapeutic intervention. MCI: mild cognitive impairment; CIS: clinically isolated syndrome.

nance and positron emission tomography (PET), are promising methods applied to clinical trials to assess potential therapeutic mechanisms in MS and AD, in addition to molecular biomarkers emerging from genomics and proteomics research.

In this chapter, molecular and neuroimaging biomarkers are reviewed in parallel for MS and AD, focusing on inflammatory and neurodegenerative common features and highlighting unique biomarkers specific to each condition. Possible applications of biomarkers in clinical development are also discussed.

Very intensive research efforts in genetic research in these two multifactorial and genetically complex disorders have led to the identification of a number of genetic risk factors associated with disease, such as apolipoprotein E (ApoE) polymorphism in sporadic AD, amyloid precursor protein (APP), presenilin 1 (PSEN1) and presenilin 2 (PSEN2) mutations in familial forms of AD, and human leukocyte antigen (HLA) locus in MS. None of the biological indicators tentatively associated with the disorders has demonstrated enough sensitivity and specificity to allow their use as symptomatic or presymptomatic diagnostic tools.

This well-known limitation to diagnose MS and AD patients in a clinical course window between disease onset and clinical manifestation is undoubtedly linked to the absence of cure for these neurological disorders. Analysis of the clinical success rate from first-in-man to registration is showing that CNS compounds have one of the lowest success rates compared to other therapeutic areas, the lack of efficacy being the major cause of attrition in the clinic and accounting for approximately 30% of failures, with high failure rates in phases II and III trials [10] (Figure 2).

Different approaches have been proposed to reduce attrition in drug development, starting from the earliest stages of discovery. These include getting very strong

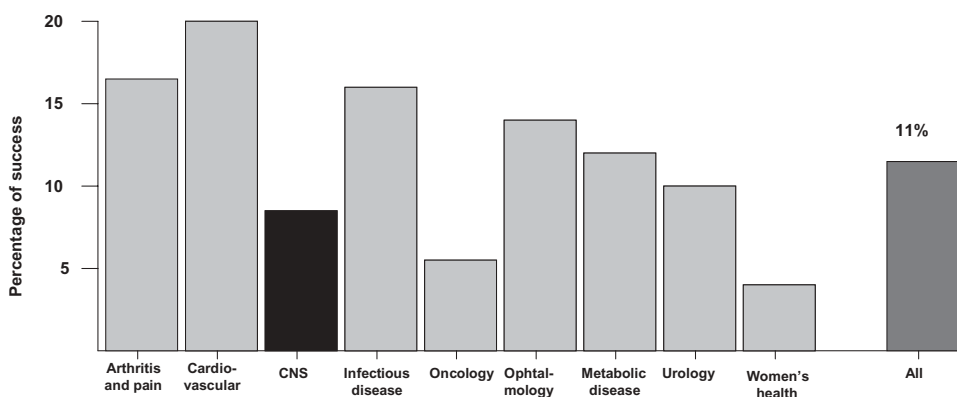


**FIGURE 2** Success rates from first-in-man to registration. The overall success rate is 11%. However, if the analysis is carried out by the therapeutic areas, big differences emerge. The data are from the ten biggest drug companies during 1991–2000. (The companies are AstraZeneca, Bristol-Myers Squibb, Eli Lilly, F. Hoffman-LaRoche, GlaxoWellcome Johnson & Johnson, Novartis, Pfizer, Pharmacia, Schering-Plough, and SmithKline Beecham; data were obtained by Datamonitor in the Pharmaceutical Benchmarking Study.) CNS: central nervous system. Reprinted by permission from Macmillan Publishers Ltd: *Nature Reviews Drug Discovery*, vol. 3, Kola, I., and Landis, J., *Can the pharmaceutical industry reduce attrition rate?* 711–715, copyright 2004.

evidence for proof of mechanism with modulation of target involved in relevant disease pathways, eliminating compounds with poor toxicity profiles, and using appropriate animal models for efficacy testing in preclinical studies. It is interesting to note that CNS pharmacological research, one of the areas with the highest attrition rate, is also the area where animal models are very poor predictors of human pathophysiology for AD as well as for MS.

In this context, clinical biomarkers have emerged as a critical tool to help define disease and to determine pharmacological response [11–13]. Biomarker research is intensively focusing on diagnostic biomarkers to track disease at its initiation or when only mild symptoms occur, as disease-modifying therapies could more effectively slow down the disease process if prescribed early. Intensive biomarker research is also conducted on prognostic and predictive biomarkers to evaluate disease severity, progression rate, and patients at risk who would most benefit from novel treatments to anticipate efficacy and/or safety of symptomatic and/or disease-modifying neurotherapies [14, 15].

#### 14.2.1.1 Working Definitions

The National Institutes of Health Biomarkers Definitions Working Group has agreed on the classification of biomarkers allowing a clear distinction between biomarkers and clinical endpoints [16], with surrogate endpoints being a subset of biomarkers:

*Biological Marker = Biomarker* A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention.

*Surrogate Endpoint* A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm



or lack of benefit or harm) based on epidemiological, therapeutic, pathophysiological, or other scientific evidence.

**Clinical Endpoint** A characteristic or variable that reflects how a patient feels, functions, or survives.

The optimal biomarker should identify a universal and fundamental feature of the pathophysiology of the disease and should be validated in patients who have developed the disease, such as in neuropathologically confirmed patients for AD or in morphologically characterized MS patients with demyelination, inflammation, gliosis, and axonal damage. The optimal biomarker should also be independent of the clinical symptoms, have high specificity and sensitivity, and possess specific technical characteristics, such as being noninvasive, allowing repeated measurements, and being easy to test and inexpensive. A combination of biomarkers may be necessary to achieve all/most of these components.

### 14.2.2 WHEN TO USE BIOMARKERS

#### 14.2.2.1 Diagnosing AD and MS and Biomarkers to Identify Disease: Diagnostic Biomarkers

**Diagnosing AD in Clinical Practice** Over the last two decades, a wealth of evidence has supported the amyloid hypothesis in AD, with a central role attributed to the accumulation and deposition of fibrillar  $\beta$ -amyloid driving neurodegeneration and cognitive decline and leading to dementia [17–19]. Advances in understanding the neurobiology of the normal aging brain as well as the mechanisms of AD have allowed increasing agreement between clinical (antimortem) and autopsy (postmortem) diagnoses in demented aged individuals.

The diagnosis of AD is currently based on clinical criteria [2–4]. According to the NINCDS-ADRDA criteria [4], the diagnosis of AD is classified as probable, possible, or definite. Definite AD depends on the observation of amyloid neuritic plaques and neurofibrillary tangles at neuropathological examination. The diagnoses of probable and possible AD rely on proper clinical information. Probable AD is defined by progressive memory decline associated with impairment of at least one additional cognitive function, such as language, visuospatial, or executive abilities, in the absence of other brain or systemic diseases that may cause cognitive decline. The diagnosis of possible AD admits the occurrence of an atypical clinical course, such as plateaus, or the presence of a concomitant clinical abnormality capable of producing cognitive impairment but not considered by the clinician to be sufficient to produce dementia.

Although being essentially based on exclusion, the diagnostic criteria of probable AD displays good sensitivity (81%), although with lower specificity (70%), in relation to the gold-standard pathological diagnosis, and appropriate clinical follow-up substantially increases this diagnostic accuracy [20]. Disease duration is typically 8–10 years but ranges from 2 to 25 years.

Some ancillary tests have been found to increase the diagnostic accuracy of AD. Hence, atrophy of the hippocampus and of the entorhinal cortex, indicated either by volumetric analysis [21–23] or by visual inspection [24, 25], on magnetic resonance imaging (MRI) and PET scan showing decreased metabolic rates in posterior temporoparietal, posterior cingulate, and prefrontal cortices [26, 27], increased levels

of total and hyperphosphorylated tau and reduced concentrations of amyloid beta ( $A\beta$ ) in the cerebrospinal fluid (CSF) [28, 29], and high  $A\beta_{42}$  plasma levels [30] have all repeatedly shown to improve the clinical diagnosis of AD. Moreover, the presence of one or two copies of the  $\epsilon 4$  allele of the *APOE* gene also increases diagnostic specificity [31, 32].

**Diagnosing MS in Clinical Practice** MS is the most frequent chronic inflammatory demyelinating disorder of the CNS, and its diagnosis is based on the demonstration of a temporal and spatial dissemination of inflammatory lesions in the CNS. This demonstration is obtained on a clinical basis and also using surrogate outcome measures like brain and spinal cord MRI and CSF examination. A recent revision of diagnostic criteria has been performed by a panel of experts [33], with recent modifications [34]. This is further complicated by the lack of existence of a definite histopathological diagnosis, like in AD, and by the relatively low number of autopsic cases so far analyzed. These criteria are universally accepted, but their validity and concordance among different examiners are yet to be established. On MRI, findings of multifocal lesions of various ages, especially those involving the periventricular white matter, the brainstem, the cerebellum, and the white matter of the spinal cord, support the diagnosis, and the appearance of gadolinium-enhancing lesions on MRI indicates sites of active inflammation and rupture of the blood–brain barrier. CSF analysis supports the diagnosis if it shows an increased intrathecal synthesis of immunoglobulins which are not present in the serum—so-called oligoclonal band (OB) positivity at the isoelectrofocusing. In the absence of gold standard for the diagnosis of MS, it is difficult to give specific accuracy features of MRI and CSF OB methods. There is a considerable clinical heterogeneity of the disease, for which there are essentially two different modes of presentation. The most common is the relapsing–remitting (RR), which occurs in about 85% of the patients, and it is characterized by the occurrence of episodes of neurological manifestations followed by complete or partial recovery. About 80% of RR patients develop a secondary progressive (SP) course, with or without superimposed relapses at a certain time after the onset: nearly 50% within 10 years and more than 80% within 25 years from the onset [35], while the remaining 20% are fully functional after 15 years of disease and are defined as benign cases. Ten to 15% of patients are affected by the primary progressive (PP) form, which is characterized by a progressive accumulation of irreversible disability since the disease onset. In MS, between 55 and 65% of patients experience cognitive decline as revealed by neuropsychological test battery, which is more frequent and severe in patients with the progressive phase [36–39] and in association with the *APOE4* genotype in severe cases of cognitive impairment [40]. The cognitive decline resembles the pattern seen in subcortical dementias, and it is clinically characterized by slow information processing, impaired memory retrieval, deficient problem solving, and personality and mood disturbances as a consequence of the disruption of the white matter connections between the frontal lobes and the subcortical structures.

#### 14.2.2.2 Identifying Risk of Developing MS and AD: Presymptomatic Biomarkers

Many studies have suggested that pathological processes occurring in MS and AD begin long before the onset of clinical symptoms. For instance, in relation to AD, neuropathological studies have demonstrated that a large proportion of cognitively

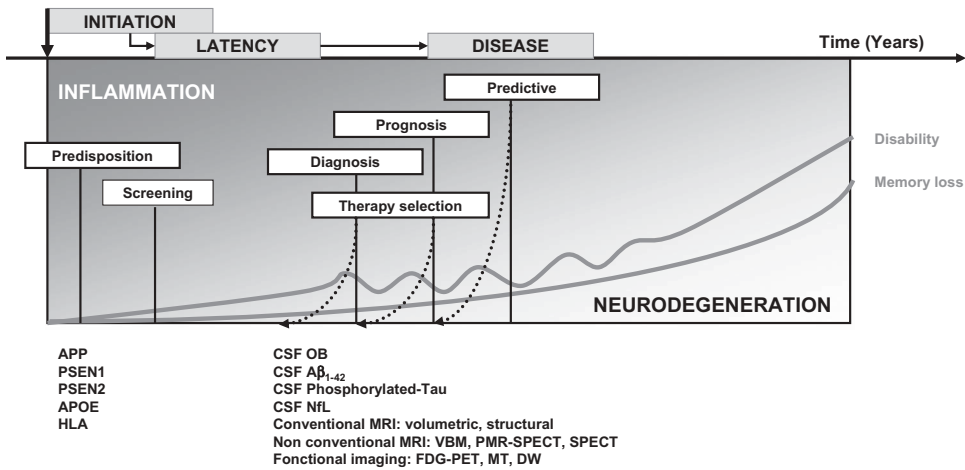


**FIGURE 3** Biomarkers in AD and MS. The goal of the next generation of biomarkers is to allow a shift in the detection of disease stages at earlier stages: from symptomatic AD or MS to presymptomatic MCI or CIS to asymptomatic, as indicated by the dotted line arrows for each biomarker subtype making disease-modifying therapies effective in preventing disease or slowing down its progression.

normal elderly individuals have amyloid plaques when autopsy is performed [41, 42]. These presymptomatic phases have been characterized as clinically isolated syndrome (CIS) for MS, which is defined as the occurrence of a first acute episode of neurological worsening due to a single white matter lesion and usually happens in about 85% of patients [43] and mild cognitive impairment (MCI) for AD [44]. The identification of these early disease stages are critically important for the development of disease-modifying therapies designed to slow the underlying progression of the disease, therapies likely to be the most valuable for patients with mild symptoms or asymptomatic patients. The development and validation of biological markers, neuroimaging tests, and other objective indicators of preclinical AD and preclinical MS are regarded as the most important research areas to aid in identifying individuals at high risk for developing symptomatic AD or MS [45–47] (see Figure 3).

### Neuroimaging Biomarkers

**Structural Imaging: Conventional and Nonconventional Biomarkers** Structural neuroimaging techniques, especially MRI, have been intensely investigated as diagnostic tools to identify patients in the very early stages of AD or in individuals with MCI, who are at high risk of conversion to AD. Great attention has been focused to medial temporal lobe structures, especially the hippocampus and the entorhinal cortex, since these areas are known to be the first to be affected in AD [48]. Some studies have shown that direct (volume of the structures) or indirect measures (such as volume of the temporal horn of the lateral ventricle) of medial temporal lobe involvement, particularly when longitudinal information is available, may be useful as indicators of increased risk of progression from MCI to dementia [49–51]. Even the assessment of medial temporal lobe atrophy by visual rating has been shown to be a significant independent predictor of conversion from MCI, especially when memory is impaired, to AD [52].

In a longitudinal study in which patients with MCI and cognitively normal controls were followed over two years, longitudinal hippocampal volume losses were

associated with increasing hyperphosphorylated tau and decreasing A $\beta$ <sub>42</sub> CSF levels, suggesting that the increasing atrophy over time of the hippocampal formation might be related to underlying pathological abnormalities that occur in AD [50].

In another study, in which amnesic MCI patients and normal controls were submitted to serial MRI, two quantitative MRI measures, namely whole-brain and ventricle atrophy rates, obtained 1–2 years before baseline were associated with an increased likelihood of conversion from MCI to AD. The combination of these two atrophy rates with cross-sectional hippocampal volume at baseline provided additional predictive information regarding conversion from MCI to AD. Nonetheless, some degree of overlap between converters and nonconverters was observed, indicating that these measures cannot be used as definite prognostic indicators for individual patients [49].

An additional MRI-based method evaluated in prospective studies of MCI is voxel-based morphometry (VBM). In two recent studies using this technique, MCI patients who converted to AD on follow-up (from 18 to 24 months) showed significant reductions in gray matter density in brain regions known to be affected early by AD pathology, such as the hippocampus and posterior cingulate, and also in other neocortical areas in the frontal, temporal, and parietal lobes [53, 54]. MRI spectroscopy seems to be another promising tool for early diagnosis of AD (in the MCI stage), although the studies published so far were all cross-sectional [55, 56].

As previously stated, CIS is considered to be the presymptomatic phase of MS. From a clinical point of view, according to a large database of patients affected by CIS, 52% presented with long-tract symptoms, 18% with an optic neuritis, 9% with a brainstem syndrome, and 21% with multifocal abnormalities [57]. By definition, the CIS phase evolves into MS if there is a demonstration of dissemination in time of inflammatory lesions. Differently from the past when the occurrence of a second clinical relapse resulted in a diagnosis of clinically definite MS [58], the recent McDonald criteria [33] with further revisions [34] consider that the presence of a gadolinium-enhanced lesion on an MRI done at least 3 months after the first episode or a new T2 lesion confirmed on two MRI scans done no less than 3 months apart is sufficient to demonstrate a dissemination in time of lesions and hence a diagnosis of MS. As regards to the sensitivity, specificity, and accuracy of the McDonald criteria in predicting the 3-year conversion to MS, Dalton et al. [59] found sensitivity, specificity, and accuracy of 83%, while Tintore et al. [60] reported respectively figures of 74, 86, and 80%. Therefore, conventional MRI is considered to be extremely important in making MS diagnosis and in helping to reduce the latency between the disease onset and the clinical diagnosis. However, conventional MRI techniques, namely the assessment of the number of T2 lesions, T1 lesions, and gadolinium-enhancing lesions, are of limited value for prognosis in the early phases of the disorder, while nonconventional MRI techniques are more useful.

We can distinguish among different techniques which can be used in the assessment and prognosis of CIS patients. It is now known that brains of MS patients show early axonal pathology that correlates with immune cell damage [61] and that even in a classical neuroinflammatory disorder like MS there is an unexpected collateral damage to neurons, which however do not represent the target of the immune attack, which is the myelin sheet. It is speculated that the extent and size of irreversible axonal loss in CIS patients are predictive of the long-term disability, and the measurement of progressive brain atrophy is a marker of irreversible tissue damage [62].

Dalton et al. [63] performed a longitudinal 3-year follow-up study on 55 CIS patients and demonstrated that patients who progressed into MS (53% of the total) developed more important gray matter atrophy than patients without MS diagnosis. Interestingly enough, the white matter volume did not change over time in both converters and nonconverters, and the number of T2 lesions was moderately correlated with atrophy measures, showing that they are less sensitive and less indicated to longitudinal changes. Another study [64] found a decrease in the normalized brain parenchymal volume and percentage brain volume change of CIS patients. Magnetization transfer (MT) and diffusion-weighted (DW) MRI has the potential to detect and quantify the extent of tissue damage present in the brain and cervical cord, which is not visible by using other techniques. A low MT ratio can be detected in normal-appearing white matter and brain tissue of CIS patients, but it is still unclear whether it can be used as a measure of disability prediction, as only one study found that the measure was predictive of subsequent disease progression [65], while others did not [66, 67].

Proton magnetic resonance spectroscopy can add information on the biochemical nature of changes in the CNS. It has been used in a recent study including 96 patients with CIS who have been tested within 6 months of their clinical episode [68], and high myoinositol and creatine levels have been found in the white matter of diseased patients. It is interesting to notice that a lower ratio of *N*-acetylaspartate and creatine and a lower brain parenchymal fraction have been found in patients with CIS without cognitive impairment versus patients affected by CIS and cognitive impairment, possibly suggesting a more severe axonal loss in the latter group.

Another interesting research area is represented by the MRI of the optic nerve as a model of damage and recovery in demyelination. The best model of demyelination is represented by CIS patients who experience a retrobulbar optic neuritis in which there is a visual loss which is determined by a demyelination of the optic nerve.

**Functional Imaging** As functional neuroimaging methods, particularly functional MRI (fMRI) and PET, became more widely available in recent years, they have been increasingly investigated as potential biomarkers for preclinical or very early AD and MS diagnosis.

In MCI, fMRI studies assessing the risk of conversion to AD are still scarce. However, a recent study in which two carriers of a *PSEN1* mutation were scanned (one presenting subtle memory problems on comprehensive neuropsychological testing and another with amnesic MCI), significant differences related to the pattern of brain activation during an episodic memory task were observed in both subjects in comparison to healthy controls [69]. It is noteworthy that the very mild symptomatic carrier was 20 years old, almost 30 years younger than the mean age of clinical manifestation of the disease in the family. This observation supports further investigation of fMRI as a possible biomarker in very early AD. Moreover, fMRI may also be viewed as a promising tool for monitoring the neurochemical and brain function changes induced by pharmacological agents [70].

PET scan, especially using fluoro-2-deoxy-*d*-glucose (FDG), has proven to be an efficient method for early diagnosis of AD, even prior to the MCI stage. Indeed, in a 3-year longitudinal study enrolling 48 cognitively healthy elderly, from whom 11 declined to MCI, De Leon et al. [71] found that reductions in the metabolic rate in the entorhinal cortex at baseline accurately predicted this conversion. Moreover,

among those subjects who declined, baseline metabolism anticipated the emergence of memory deficits as well as of hippocampal and temporal neocortex hypometabolism. Reduced metabolic rates in additional brain regions, such as the temporoparietal cortex, posterior cingulate, and prefrontal areas, have also shown to be good predictors of conversion from MCI to AD [72–74]. FDG PET was found to be superior to neuropsychological testing as an indicator of subsequent global cognitive decline in patients with MCI [53], although some degree of heterogeneity in the patterns of brain glucose metabolism reduction has been reported in these individuals, a feature that probably will have to be taken into account in future studies [75].

Perhaps the most promising neuroimaging method for very early AD diagnosis is the amyloid PET tracer named Pittsburgh Compound-B [(11)C-PIB], which has been shown to detect in vivo amyloid plaques in the brain of patients with mild AD [76]. Subsequent studies found an inverse correlation between in vivo amyloid load and the concentration of A $\beta$ <sub>42</sub> in the CSF in AD [77], a positive correlation between whole-brain atrophy rates and whole-brain and regional (11)C-PIB uptake [78] and a negative correlation between amyloid load in the parietal cortex and the performance in a verbal memory test [79]. More interesting and promising, though, is the finding of elevated (11)C-PIB uptake in some nondemented individuals, suggesting that it may turn out to be a sensitive method for detection of preclinical AD in MCI or even in totally asymptomatic individuals [80]. On the other hand, the observation of a relatively stable (11)C-PIB uptake in mild AD patients after a 2-year follow-up suggests that amyloid deposition in the brain tissue reaches a plateau by the early stages of the disease and thus the method may not be ideal for tracking amyloid reduction induced by future treatments [79].

As regards to MS, fMRI studies on CIS patients are extremely useful in assessing the cortical reorganization and plasticity secondary to lesions in the CNS. A 1-year follow-up study of patients with a CIS found that those who developed clinically definite MS had a different response on motor tasks at fMRI than those who did not convert [81]; more specifically, the nonconverters had more substantial activations in the contralateral primary somatomotor cortex, supplementary motor area, ipsilateral paracentral lobule, and cerebellar hemisphere, while MS patients who converted had a more widespread recruitment of additional areas not involved in motor response.

**Predisposition Biomarkers: Susceptibility Genes and Modifiers** Genetic epidemiology studies have demonstrated that genetic factors are clearly involved in the etiology of AD and MS [82–85]. Despite the fact that only about 25% of AD is familial (FAD), tremendous progress has been made in understanding the etiology of AD by dissecting the molecular genetics of the disease. Mutations in the A $\beta$  precursor protein ( $\beta$ APP) coding gene, *PSEN1* [86], and *PSEN2* are associated with early-onset familial AD (EOFAD). Mutations in these genes result in altered processing of  $\beta$ APP and the relative overproduction of all forms of the  $\beta$ -amyloid peptide or specific overproduction of isoforms ending at residue 42. The relative proportion of EOFAD patients carrying mutations are 10–15% for *APP*, 20–70% for *PSEN1*, and <1% for *PSEN2* [82–84, 87]. Other genes are causing EOFAD as kindreds with autosomal dominant FAD with no known mutations in *PSEN1*, *PSEN2*, or *APP* have been described [88]. About 1–6% of all AD cases are early onset (<60 years) and about 60% of early-onset AD is familial, with 13% appearing



to be inherited in an autosomal dominant manner [82, 89]. Therefore, the EOFAD group comprises less than 2% of all AD cases [90].

Despite the application of the same molecular tools to MS, identification of specific or definitive MS genes has failed. Association of MS with HLA haplotype within the major histocompatibility complex (MHC) on chromosome 6, extending over a long interval containing over 200 genes, has consistently been reported by several independent studies for more than three decades [91, 92]. However, the vast majority of MS patients do not share the same haplotype, the strength of association is declining from Northern to Southern Europe, and calculations suggest that the MHC could explain between 17 and 62% of the genetic etiology of MS [93]. Genetic data reject a single-locus model of MS susceptibility but suggest that MS depends on independent or epistatic effects of several genes, with a small contribution of each of them. As efforts in genetic research are pursued, it is likely that a limited number of genetic profiles determined by genetic variants in different genes can clearly identify MS patients and be used for presymptomatic diagnosis and selection criteria for treatment. In the meantime, apart from the HLA haplotype, no genetic biomarker can contribute to MS diagnosis in the early stage of the disease.

Several studies have found an increased frequency of the *HLA-A2* allele in patients with EOAD and others have demonstrated an association between the *A2* allele and an earlier age of onset of AD [94], with *HLA-A2* homozygotes having an onset of AD 5 years earlier, on average, than either *A2* heterozygotes or those without *A2*, reflecting a gene dosage effect independent of *APOE4* status. It has been suggested that the *HLA-A2* allele may have a role in regulating an immune response in the pathogenesis of AD or that there may be a responsible gene in close linkage to *A2* [94].

The *APOE4* (Cys112Arg) variant of apolipoprotein E is associated with increased risk for late-onset AD (LOAD, after age 60) in a dose-dependent fashion [95, 96]. The *APOE4* allele, by unknown mechanisms, appears to affect age of onset by shifting the onset curve toward an earlier age [97]. Estimates of the total contribution of *APOE* to the variance in onset of LOAD vary widely. As the inheritance of one or more *APOE4* alleles is not deterministic for AD and the mechanism by which inheritance of one or more *APOE4* alleles causes AD is unidentified, the usefulness of *APOE* genotyping in clinical diagnosis and risk assessment remains unclear [98]. *APOE* genotyping may have an adjunct role in the diagnosis of AD because a large proportion of demented individuals with *APOE4* alleles have been found to have neuropathological confirmation of AD at autopsy [31, 32, 99, 100]. Increased risk of LOAD associated with the *APOE4* allele has also been shown in African-Americans [101] and Caribbean Hispanics [102]. In a segregation analysis of LOAD families, Daw et al. [103] found evidence that, in addition to *APOE*, four additional loci make a contribution to the variance in age at onset of LOAD similar to or greater in magnitude than that made by *APOE* alone, with one locus making a contribution several times greater than that of *APOE*. Their results suggested that several genes not yet localized to that time may play a larger role than does *APOE* in LOAD.

No result indicates that homozygosity or heterozygosity for *APOE4* allele causes a greater risk of developing MS, but there are several studies which found an association between the allele *APOE4* and a faster disease progression and also two longitudinal studies which found a relationship between the allele and the progression rate. A recent meta-analysis confirmed such data [104].



Numerous candidate gene studies have been conducted both in MS and AD to identify additional disease modifiers. Some positive findings have been reported, but modestly significant associations were observed, suggesting that the studied genetic variants have only modest effects on risk/protection, age at onset, and disease duration. This observation reinforces the existence of not yet identified additional putative functional variants and demonstrates that none of the investigated single variants taken alone contributed substantially to the development of AD or MS in the respective studied samples. Among the analyzed genes, the following can be cited: *UBQLN1* [105], *APBB2* [106], *TNF alpha* [107], *LRP1*, *ACE*, *A2M*, *BLMH*, *DLST*, *TNFRSF6*, *NOS3*, *PSEN1*, *PSEN2*, *BCHE*, *APBB1*, *ESR1*, *CTSD*, *MTHFR*, and *IL1A* [108]. Croes et al. [109] argued against using genetic testing for AD as a diagnostic tool, suggesting that the contribution of genetic testing to clinical diagnosis is small and does not counterbalance the problems associated either with interpretation or with secondary effects on family members. In contrast to the utility of *APOE* testing as an adjunct diagnostic test in individuals with dementia, there is general agreement that *APOE* testing should not be used for predictive testing for AD in asymptomatic persons.

**Soluble Biomarkers: Analysis of Body Fluids** Biological analyses of body fluids such as CSF, blood, and urine samples have the potential to deliver important biomarkers in the clinic due to the accessibility of these biological materials. However, lumbar puncture is a relatively invasive procedure and repeated collection of CSF needs to be restricted, representing a limitation for its use in the clinical setting and also as a potential surrogate marker in clinical trials. Blood–brain barrier damage, both in MS and AD, allows relatively large amounts of CSF to be absorbed in the blood every day [110], but the main challenge resides in the relevance of brain leakage proteins likely to be present in CSF and/or plasma in poor to very poor levels in comparison to abundant proteins such as albumin. This is due to the fact that peripheral blood represents a traffic compartment for both regulatory and effector lymphomonuclear cells, and it is very unlikely that the processes occurring in the CNS would be reflected and paralleled in the peripheral circulation [111]. Moreover, soluble molecules, which could play a relevant role in the ethiopathogenesis of MS and AD, act in an autocrine or paracrine fashion, which typically show a short half-life [112].

**Common Features between MS and AD** CSF levels of protein tau (total tau) have been shown to be significantly increased in patients with AD, with elevations of two- to threefold CSF total tau, but weak correlations present with changes in cognitive scores [113, 114]. In MS, increased total tau levels in patients with active disease compared to clinically nonactive states indicate axonal pathology in active disease [115], and the report was confirmed in a second study [116]. The association of tau concentrations with AD suggests that the protein could be related to cognitive dysfunction in MS. Over the last 20 years, several studies have confirmed the association of MS with the HLA complex, the most polymorphic gene cluster known in humans with >200 genes, but the MS-associated DR2 haplotype, also known as HLA-DRB1\*1501-DQA1\*0102-DQB1\*0602, extends over a long interval containing multiple genes [93]. Most recent studies based on linkage disequilibrium analysis with highly dense single-nucleotide polymorphism (SNP) maps have demonstrated

that it is the HLA class II genotype patterns that determine susceptibility and resistance to MS, this pattern of susceptibility strongly supporting an autoimmune etiology for MS [117, 118]. *HLA-A2* allele association with earlier age of AD onset has been reported with an additive effect of *HLA-A2* and *APOE4* demonstrating a 3-year shift [94, 119]. Although sensitivity is high, poor specificity limits the diagnostic value of these biomarkers.

**Specific Features of AD**  $A\beta$  is recognized as a normal product of cellular metabolism throughout life and circulates as a soluble peptide in biological fluids. The most reproduced finding is the decrease of CSF  $A\beta_{1-42}$  levels [28, 120, 121] in association with higher levels of total tau in the CSF of AD patients in comparison with the CSF of controls while CSF  $A\beta_{1-40}$  levels (or  $A\beta_{total}$ ) are not different from age-matched controls. The high positive likelihood ratio indicates that combined biomarker tests are useful in confirming the diagnosis of AD, but the low negative likelihood ratio indicates that a negative test result cannot rule out the disease [122]. Changes in CSF biomarkers occur early in the course of AD in most patients. No consistent correlation of  $A\beta_{1-40}$  or  $A\beta_{1-42}$  plasma concentrations with CSF levels has been observed [123]. Based on CSF levels of  $A\beta_{42}$ , tau, and ubiquitin, subgroups presenting different clinical profiles can be distinguished; these different subgroups may benefit from different therapeutic drugs [124].

Analysis of the different CSF phosphorylated tau isoform (p-tau 181, 199, 231, 235, 396, and 404) levels revealed that CSF tau-199 (phosphorylation at serine 199) is significantly increased in AD patients compared to non-AD subjects, exceeding both sensitivity and specificity over 85% as a sole biomarker for AD [125], but with many of the non-AD tauopathy and degenerative dementias also showing increased tau-199 levels. Cross-sectional studies examining three p-tau species as diagnostic measures (tau-181, tau-199, and tau-231) have revealed that tau-231 may have a greater specificity to AD [126]. The observed correlations between levels of CSF phosphorylated tau and hippocampal atrophy are independent of disease duration and severity, suggesting that CSF phosphorylated tau levels may reflect neuronal damage in AD [127].

Combining CSF tau and CSF  $A\beta_{1-42}$  levels improves specificity, allowing differentiation between AD, normal aging, and vascular dementia [128]. Other biological markers such as CSF levels of neurotransmitters, cytokines, or superoxide dismutase were shown to have less diagnostic value [129].

**Specific Features of MS** As regards biomarkers related to the inflammatory phase, it is well known that the detection in the CSF of an intrathecal antibody production, namely the presence of two or more oligoclonal IgG bands (OB) detected by separation of CSF proteins, while not demonstrable in corresponding serum, is supportive of a diagnosis of MS. Several studies suggest that CSF analysis increases diagnostic sensitivity, though perhaps at the cost of specificity and accuracy [130]. However, findings in one large study of PPMS [131] have led to a recommendation to revise the CSF criteria for this population of patients, and it is no longer a requirement for diagnosis [34]. MS patients with and without CSF bands may be clinically indistinguishable but may represent two immunologically distinct populations [132].

Other components of the immune system, and mostly cytokine and chemokine blood levels, have been extensively studied as potential biomarkers with

disappointing results, with only few studies trying to test more than one potential biomarker at the same time in the same subjects [133].

Given the recent evidence that axonal loss occurs in MS since the early stages of the disease, there is a current strong interest in the identification and analysis of biomarkers which reflect the axonal damage, which can be divided into cytoskeleton markers, membrane markers, and other axonal markers [134]. Several studies have shown that concentration of light-chain neurofilament (NfL), one of the major axonal cytoskeleton proteins, is increased in CSF of patients with MS [135], with peak concentrations during relapses. In addition, the correlation of NfL and heavy-chain neurofilament (NfH) autoantibody indices with surrogate markers for atrophy on MRI suggests a relation with disease progression [136]. Combining CSF markers (tau and NfH) can increase sensitivity in predicting the conversion from CIS to MS (40% compared with 34% for MRI), which could be further increased to 60% if CSF and MRI criteria were combined [137]. Similarly, the combination of tau and NfH showed higher specificity (94%) than MRI (82%). The only marker related to neuronal damage that has shown promising results in peripheral blood is 24S-hydroxycholesterol [138], with concentrations decreased in serum of patients with MS, especially if affected by the primary progressive form of MS. Additional markers of axonal loss which have been tested are the ApoE levels, the 14–3–3 protein, and the neuron-specific enolase with, however, conflicting results.

### **14.2.3 CLINICAL DEVELOPMENT APPLICATION: DRUG EFFICACY AND SAFETY**

#### **14.2.3.1 Using Biomarkers to Select Patient Subgroups**

Biomarkers may be applied to drug development in a number of distinct ways [129]. First, additional diagnostic measures in clinically identified populations are provided (high specificity and sensitivity are therefore required). Second, new biomarkers of presymptomatic disease are very important for population enrichment strategies and confirmation of efficacy during the assessment of novel disease-modifying therapies. Third, biomarkers can help to establish a differential therapeutic approach and to treat individuals selectively according to their pathophysiological/genetic subtype and disease status (Table 1). In clinical development, the most important benefit offered by biomarkers is to limit investigational drugs to high-risk groups, who would gain the most of novel therapies and/or in which efficacy is most likely to be observed. Heterogeneity in the patient population can be limited by subgrouping patients according to these additional biomarker criteria. In the absence of subgrouping patients, dilution effect on efficacy may lead to no statistically significant results and arrest of drug development of molecule even if the molecule may be valuable for some patient subgroups. Enrichment trial designs using the same methodology as in cancer drug development can be extremely valuable strategy, especially for unprecedented targets [139]. The improved accuracy in diagnosis with additional diagnostic measurements can lead to better chance of success in proof-of-principle/proof-of-concept (POP/POC) studies, allowing us to concentrate our efforts on successful POP/POC studies to continue to late-phase clinical development and, therefore, to speed up the testing of novel treatments. The reduction in

**TABLE 1 Different Types of Biomarkers**

|                              |                                                                                                                                                                                                                                   |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Predisposition biomarkers    | To use in families where risk of developing disease is high and/or age dependent, such as in AD families with mutations in APP, PSEN1, PSEN2, the penetrance may not be 100%.                                                     |
| Screening biomarkers         | To be used for screening in high-risk presymptomatic populations—mutations in APP, PSEN1, and PSEN2 are too rare to justify the effort and expense, and ApoE is not sufficiently strong to be used in presymptomatic populations. |
| Diagnosis biomarkers         | Neuroimaging biomarkers and oligoclonal bands for MS. In AD, neuroimaging biomarkers are the most promising, ApoE being used as an adjunct.                                                                                       |
| Therapy selection biomarkers | To use to define targeted populations where disease-modifying therapies have been efficacious (targeted therapies).                                                                                                               |
| Prognosis biomarkers         | Once diagnosis is performed, markers defining disease severity and/or rate of progression.                                                                                                                                        |
| Predictive biomarkers        | To anticipate efficacy and/or safety of symptomatic and/or disease-modifying neurotherapies.                                                                                                                                      |

sample size estimates is also very important: Enrollment of fewer subjects is needed if using neuroimaging and soluble biomarkers rather than when using cognitive measurements (for AD) or clinical criteria (for MS) [39, 140].

From a broader clinical standpoint, redefining disease is the challenge of future medicine. Because of heterogeneity in MS and in AD, subtyping patients by genetic, clinical, neuroradiological, and neuroimmunological parameters will be necessary, and biomarkers will be extremely important contributors to the subtyping algorithm.

#### **14.2.3.2 Predicting Changes in Response to Therapy and Disease Course: Predictive Biomarkers**

As new targets have emerged from genomics, it has been put forward that insufficient target validation has caused the increase in failure rates of drug development programs. As a consequence, experimental medicine groups frequently develop and validate biomarkers to enable them to test if the pharmacology can be safely expressed in humans and if the pharmacological activity provides sufficient efficacy and differentiation in the disease indication [141].

A number of strategic considerations are addressed in the course of drug development and questions frequently emerge: Why and how to use biomarkers in clinical drug development? With rare exceptions, biomarkers cannot substitute primary endpoints in drug registration trials and, by definition, a surrogate can only approximate the gold standard. Biomarkers may be applied to drug development in a number of distinct ways [129]. If deployed as early as possible, biomarkers can contribute to the development of three phases of the clinical plan: (1) to confirm if the drug is hitting the target; (2) to test if hitting the target alters the pathophysiological mechanism of disease (POP/POC); and (3) to test if altering the mechanism affects clinical status [142].

Consequently, classes of biomarkers differ not only for each therapeutic area but also according to their objectives. In risk reduction, specific biomarkers may be developed for selecting drug dose, for selecting patient population, for assessing early efficacy, or for predicting toxicity, making some biomarkers suited for one

role and inappropriate for another. Specificity and sensitivity required will depend on the application. Clearly, they are add-ons to facilitate decision making and not a substitute for clinical endpoints requested by regulatory authorities (FDA or EMEA) [143].

Development and validation of biomarkers can be cost effective in drug development programs, aiming for high-risk unprecedented targets (such as biologicals) coming from genomics discoveries and carrying increased risk in early development, as well as in chronic diseases and in pre-phase II studies to separate more quickly and cheaply than traditional phase II studies. In these instances, biomarker studies contribute not necessarily to go–no go drug development decisions but rather to gaining confidence in a novel compound and promoting it for further development, keeping in mind that validation and reliability of biomarkers are frequently difficult points and that safety and efficacy biomarker data are to be weighed against “traditional” safety and efficacy data. In parallel, for the measurement of disease severity, observations of changes in biomarker with disease progression are to be observed in longitudinal studies (high specificity not specially required in this case).

#### 14.2.4 CONCLUSION

The development and validation of diagnostic and predictive biomarkers in diseases with usually extended asymptomatic phases, like MS and AD, are particularly challenging and can take a long time as their validation is by necessity often linked to long-term clinical outcomes. Nonetheless, such biomarkers could have a big impact on health care by facilitating the assessment of new drugs that prevent costly chronic progressive disease [144]. Progress in identifying diagnostic biomarkers has been rate limited due to incomplete understanding of the pathogenesis of these disorders despite the advances in the identification of genes associated with inherited risk. Biomarkers for AD are far from reaching clinical practice while MRI and CSF OB bands are the only validated biomarkers currently used in MS. Yet, if validated diagnostic biomarkers as well as disease-modifying treatments with ability to slow down the disease progression became available—as the first A $\beta$  modulators in AD and interferons in MS—patients would highly benefit from management of the early disease stage.

Biological and neuroimaging biomarkers with strong potential are currently being used in early drug development programs to allow for the exploration of pharmacodynamic effects of targeted therapeutics and to aid early diagnosis, alone or in combination, but require additional validation to be considered for their ability to reveal effect of investigational drugs [11, 12, 145]. CSF A $\beta$  and tau are biological markers of AD pathophysiology, and their measures may have potential clinical utility in the future diagnosis of AD. They might be combined with neuroimaging techniques, and the PET scan with (11)C-PIB tracer is perhaps the most promising candidate [77]. Combinations of biomarkers hold great promise: For example, tau and NfH are valuable biomarkers for axonal damage in CIS patients. Predicting conversion from CIS to MS can be improved if CSF markers are combined with MRI [137]. However, soluble biological biomarkers show considerable variance, with significant overlap between values obtained in disease groups and controls for AD or MS. Meta-analyses of previous studies comparing these markers need to demonstrate similar findings. For most biomarkers, the differences are not sensitive

and specific enough for discrimination between disease and control groups or prediction of disease outcome in individual patients. Few studies on specific markers for neurodegeneration in body fluids have been done. The studies have typically included only small numbers of patients, and, more importantly for monitoring of neurological decline, follow-up needs to be performed on longer time periods. Validation of biomarkers requires longitudinal studies to collect data on disease progression and on neuropathology, especially for AD.

The search for biomarkers in MS and AD has just started and is already yielding promising results. Intensive research efforts are currently being accomplished using mass spectrometry-based proteomics as well as other promising technologies to provide a deep coverage of the body fluid proteomes. In the immediate future, emerging biomarkers for MS and AD are more likely to be used in clinical trials than in clinical practice, but undoubtedly, some of them will bring interesting diagnostic and prognostic tools to molecular medicine [146].

## ACKNOWLEDGMENTS

The authors would like to thank Jose Maria Badenas, MD, Director, Medical & Scientific Services, Quintiles, for his insight and editorial comments.

## REFERENCES

1. Hauser, S. L., and Goodin, D. S. (2005), Multiple sclerosis and other demyelinating diseases, in Kasper, D. L., Braunwald, E., Fauci, A. D., Hauser, S. L., Longo, D. L., and Jameson, J. L., Eds., *Harrison's Principle of Internal Medicine*, 16th ed., Mc-Graw Hill, New York.
2. Khachaturian, Z. S. (1985), Diagnosis of Alzheimer's disease, *Arch. Neurol.*, 42, 1097–1105.
3. Khachaturian, Z. S. (2006), Diagnosis of Alzheimer's disease: Two-decades of progress, *J. Alzheimer's Dis.*, 9, 409–415.
4. McKhann, G., Drachman, D., Folstein, M., et al. (1984), Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, *Neurology*, 34, 939–944.
5. Lue, L. F., Rydel, R., Brigham, E. F., et al. (2001), Inflammatory repertoire of Alzheimer's disease and nondemented elderly microglia in vitro, *Glia*, 35, 72–79.
6. Akiyama, H., Barger, S., Barnum, S., et al. (2000), Inflammation and Alzheimer's disease, *Neurobiol. Aging*, 21, 383–421.
7. Mrak, R. E., and Griffin, W. S. (2001), The role of activated astrocytes and of the neurotrophic cytokine S100B in the pathogenesis of Alzheimer's disease, *Neurobiol. Aging*, 22, 915–922.
8. Zipp, F., and Aktas, O. (2006), The brain as a target of inflammation: common pathways link inflammatory and neurodegenerative diseases, *Trends Neurosci.*, 29, 518–527.
9. Watson, G. S., and Craft, S. (2006), Insulin resistance, inflammation, and cognition in Alzheimer's disease: Lessons for multiple sclerosis, *J. Neurol. Sci.*, 24521–24533.
10. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rate? *Nat. Rev. Drug Discov.*, 3, 711–715.



11. Matthews, P. M., and Wise, R. G. (2006), Noninvasive brain imaging for experimental medicine in drug development, *expert Opin. Drug Discov.*, 1(2), 111–121.
12. Frank, R. A., Galasko, D., Hampel, H., et al. (2003), National Institute on Aging Biological Markers Working Group. Biological markers for therapeutic trials in Alzheimer's disease. Proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease, *Neurobiol. Aging*, 24, 521–536.
13. Rolan, P., Atkinson, A. J. Jr., and Lesko, L. J. (2003), Scientific Organizing Committee; Conference Report Committee. Use of biomarkers from drug discovery through clinical practice: Report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development, *Clin. Pharmacol. Ther.*, 73, 284–291.
14. Tardif, J. C., Heinonen, T., Orloff, D., et al. (2006), Vascular biomarkers and surrogates in cardiovascular disease, *Circulation*, 113, 2936–2942.
15. White, T. J., Clark, A. G., and Broder, S. (2006), Genome-based biomarkers for adverse drug effects, patient enrichment and prediction of drug response, and their incorporation into clinical trial design, *Personalized Med.*, 3(2), 177–185.
16. Atkinson, A. J. Jr., Biomarkers Definitions Working Group (2001), Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
17. Glenner, G. G., and Wong, C. W. (1984), Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein, *Biochem. Biophys. Res. Commun.*, 120, 885–890.
18. Selkoe, D. J. (1991), The molecular pathology of Alzheimer's disease, *Neuron*, 6(4), 487–498.
19. Caramelli, P., Robitaille, Y., Laroche-Cholette, A., et al. (1998), Structural correlates of cognitive deficits in a selected group of patients with Alzheimer's disease, *Neuropsychiatry Neuropsychol. Behav. Neurol.*, 11, 184–190.
20. Knopman, D. S., DeKosky, S. T., Cummings, J. L., et al. (2001), Practice parameter: Diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology, *Neurology*, 56, 1143–1153.
21. De Leon, M. J., George, A. E., Golomb, J., et al. (1997), Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease, *Neurobiol. Aging*, 18, 1–11.
22. Bottino, C. M., Castro, C. C., Gomes, R. L., et al. (2002), Volumetric MRI measurements can differentiate Alzheimer's disease, mild cognitive impairment, and normal aging, *Int. Psychogeriatr.*, 14, 59–72.
23. Gosche, K. M., Mortimer, J. A., Smith, C. D., et al. (2002), Hippocampal volume as an index of Alzheimer neuropathology: Findings from the Nun Study, *Neurology*, 58, 1476–1482.
24. Bresciani, L., Rossi, R., Testa, C., et al. (2005), Visual assessment of medial temporal atrophy on MR films in Alzheimer's disease: Comparison with volumetry, *Aging Clin. Exp. Res.*, 17, 8–13.
25. Wahlund, L. O., Julin, P., Johansson, S. E., et al. (2000), Visual rating and volumetry of the medial temporal lobe on magnetic resonance imaging in dementia: A comparative study, *J. Neurol. Neurosurg. Psychiatry*, 69, 630–635.
26. Herholz, K., Salmon, E., Perani, D., et al. (2002), Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET, *Neuroimage*, 17, 302–316.
27. Mosconi, L. (2005), Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. FDG-PET studies in MCI and AD, *Eur. J. Nucl. Med. Mol. Imaging*, 32, 486–510.



28. Sunderland, T., Linker, G., Mirza, N., et al. (2003), Decreased  $\beta$ -amyloid1-42 and increased tau levels in cerebrospinal fluid of patients with Alzheimer disease, *JAMA*, 289, 2094–2103.
29. Blennow, K. (2004), Cerebrospinal fluid protein biomarkers in Alzheimer's disease, *NeuroRx*, 1, 213–225.
30. Mayeux, R., Honig, L. S., Tang, M. X., et al. (2003), Plasma A[ $\beta$ ]40 and A[ $\beta$ ]42 and Alzheimer's disease: Relation to age, mortality, and risk, *Neurology*, 61, 1185–1190.
31. B  tard, C., Robitaille, Y., Gee, M., et al. (1994), Apo E allele frequencies in Alzheimer's disease, Lewy body dementia, Alzheimer's disease with cerebrovascular disease and vascular dementia, *Neuroreport*, 5, 1893–1896.
32. Mayeux, R., Saunders, A. M., Shea, S., et al. (1998), Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. Alzheimer's Disease Centers Consortium on Apolipoprotein E and Alzheimer's Disease, *N. Engl. J. Med.*, 338, 506–511.
33. McDonald, W. I., Compston, A., Edan, G., et al. (2001), Recommended diagnostic criteria for multiple sclerosis: Guidelines from the International Panel on the diagnosis of multiple sclerosis, *Ann. Neurol.*, 50, 121–127.
34. Polman, C. H., Reingold, S. C., and Edan, G. (2005), Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria," *Ann. Neurol.*, 58, 840–846.
35. Weinshenker, B. G., Bass, B., Rice, G. P., et al. (1989), The natural history of multiple sclerosis: A geographically based study. 2. Predictive value of the early clinical course, *Brain*, 112, 1419–1428.
36. Comi, G., Filippi, M., Martinelli, V., et al. (1995), Brain MRI correlates of cognitive impairment in primary and secondary progressive multiple sclerosis, *J. Neurol. Sci.*, 132, 222–227.
37. Kujala, P., Portin, R., and Ruutiainen, J. (1997), The progress of cognitive decline in multiple sclerosis. A controlled 3-year follow-up, *Brain*, 120, 289–297.
38. Nocentini, U., Pasqualetti, P., Bonavita, S., et al. (2006), Cognitive dysfunction in patients with relapsing-remitting multiple sclerosis, *Mult. Scler.*, 12, 77–87.
39. Brass, S. D., Benedict, R. H., Weinstock-Guttman, B., et al. (2006), Cognitive impairment is associated with subcortical magnetic resonance imaging grey matter T2 hypointensity in multiple sclerosis, *Mult. Scler.*, 12, 437–444.
40. Parmenter, B. A., Denney, D. R., Lynch, S. G., et al. (2007), Cognitive impairment in patients with multiple sclerosis: Association with the APOE gene and promoter polymorphisms, *Mult. Scler.*, 13, 25–32.
41. Knopman, D. S., Parisi, J. E., Salviati, A., et al. (2003), RC2003. Neuropathology of cognitively normal elderly, *J. Neuropathol. Exp. Neurol.*, 62, 1087–1095.
42. Bennett, D. A., Schneider, J. A., Arvanitakis, Z., et al. (2006), Neuropathology of older persons without cognitive impairment from two community-based studies, *Neurology*, 66, 1837–1844.
43. Miller, D. H., Ormerod, I. E., Rudge, P., et al. (1989), The early risk of multiple sclerosis following isolated acute syndromes of the brainstem and spinal cord, *Ann. Neurol.*, 26, 635–639.
44. Petersen, R. C. (2004), Mild cognitive impairment as a diagnostic entity, *J. Intern. Med.*, 256, 183–194.
45. Bielekova, B., and Roland, M. (2004), Development of biomarkers in multiple sclerosis, *Brain*, 127, 1463–1478.
46. Morris, J. C., Kimberly, A., Quaid, K., et al. (2005), Role of biomarkers in studies of pre-symptomatic Alzheimer's disease, *Alzheimer's Dementia*, 145–151.

47. Chong, M. S., Lim, W. S., and Sahadevan, S. (2006), Biomarkers in preclinical Alzheimer's disease, *Curr. Opin. Investig. Drugs*, 7, 600–607.
48. Braak, H., and Braak, E. (1991), Neuropathological staging of Alzheimer-related changes, *Acta Neuropathol.*, 82, 239–259.
49. Jack, C. R. Jr., Shiung, M. M., Weigand, S. D., et al. (2005), Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic MCI, *Neurology*, 65, 1227–1231.
50. De Leon, M. J., DeSanti, S., Zinkowski, R., et al. (2001), Longitudinal CSF and MRI biomarkers improve the diagnosis of mild cognitive impairment, *Neurobiol. Aging*, 27, 394–401.
51. Erten-Lyons, D., Howieson, D., Moore, M. M., et al. (2006), Brain volume loss in MCI predicts dementia, *Neurology*, 66, 233–235.
52. Geroldi, C., Rossi, R., Calvagna, C., et al. (2006), Medial temporal atrophy but not memory deficit predicts progression to dementia in patients with mild cognitive impairment, *J. Neurol. Neurosurg. Psychiatry*, 77, 1219–1222.
53. Chetelat, G., Landeau, B., Eustache, F., et al. (2005), Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *Neuroimage*, 27, 934–946.
54. Bozzali, M., Filippi, M., Magnani, G., et al. (2006), The contribution of voxel-based morphometry in staging patients with mild cognitive impairment, *Neurology*, 67, 453–460.
55. Kantarci, K., Xu, Y., Shiung, M. M., et al. (2002), Comparative diagnostic utility of different MR modalities in mild cognitive impairment and Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.*, 14, 198–207.
56. Falini, A., Bozzali, M., Magnani, G., et al. (2005), A whole brain MR spectroscopy study from patients with Alzheimer's disease and mild cognitive impairment, *Neuroimage*, 26, 1159–1163.
57. Confavreux, C., Vukusic, S., Moreau, T., et al. (2000), Relapses and progression of disability in multiple sclerosis, *N. Engl. J. Med.*, 343, 1430–1438.
58. Poser, C. M., Paty, D. W., Scheinberg, L., et al. (1983), New diagnostic criteria for multiple sclerosis: Guidelines for research protocols, *Ann. Neurol.*, 13, 227–231.
59. Dalton, C. M., Brex, P. A., Miszkiel, K. A., et al. (2002), Application of the new McDonald criteria to patients with clinically isolated syndromes suggestive of multiple sclerosis, *Ann. Neurol.*, 52, 47–53.
60. Tintore, M., Rovira, A., Rio, J., et al. (2003), New diagnostic criteria for multiple sclerosis: Application in first demyelinating episode, *Neurology*, 60, 27–30.
61. Trapp, B. D., Peterson, J., Ransohoff, R. M., et al. (1998), Axonal transection in the lesions of multiple sclerosis, *N. Engl. J. Med.*, 338, 278–285.
62. Miller, D. H., Barkhof, F., Frank, J. A., et al. (2002), Measurement of atrophy in multiple sclerosis: Pathological basis, methodological aspects and clinical relevance, *Brain*, 125, 1676–1695.
63. Dalton, C. M., Brex, P. A., Jenkins, R., et al. (2002), Progressive ventricular enlargement in patients with clinically isolated syndromes is associated with the early development of multiple sclerosis, *J. Neurol. Neurosurg. Psychiatry*, 73, 141–147.
64. Filippi, M., Rovaris, M., Inglese, M., et al. (2004), Interferon beta-1a for brain tissue loss in patients at presentation with syndromes suggestive of multiple sclerosis: A randomised, double-blind, placebo-controlled trial, *Lancet*, 364(9444), 1489–1496.
65. Iannucci, G., Tortorella, C., Rovaris, M., et al. (2000), Prognostic value of MR and magnetization transfer imaging findings in patients with clinically isolated syndromes suggestive of multiple sclerosis at presentation, *Am. J. Neuroradiol.*, 21, 1034–1038.

66. Kaiser, J. S., Grossman, R. I., Polansky, M., et al. (2000), Magnetization transfer histogram analysis of monosymptomatic episodes of neurologic dysfunction: preliminary findings, *Am. J. Neuroradiol.*, 21, 1043–1047.
67. Brex, P. A., Leary, S. M., Plant, G. T., et al. (2001), Magnetization transfer imaging in patients with clinically isolated syndromes suggestive of multiple sclerosis. *Am. J. Neuroradiol.*, 22, 947–951.
68. Fernando, K. T., McLean, M. A., Chard, D. T., et al. (2004), Elevated white matter myo-inositol in clinically isolated syndromes suggestive of multiple sclerosis, *Brain*, 127, 1361–1369.
69. Mondadori, C. R., Buchmann, A., Mustovic, H., et al. (2006), Enhanced brain activity may precede the diagnosis of Alzheimer's disease by 30 years, *Brain*, 129, 2908–2922.
70. Goekoop, R., Scheltens, P., Barkhof, F., et al. (2006), Cholinergic challenge in Alzheimer patients and mild cognitive impairment differentially affects hippocampal activation—a pharmacological fMRI study, *Brain*, 129, 141–157.
71. De Leon, M. J., Convit, A., Wolf, O. T., et al. (2001), Prediction of cognitive decline in normal elderly subjects with 2-[(18)F]fluoro-2-deoxy-D-glucose/positron-emission tomography (FDG/PET), *Proc. Natl. Acad. Sci.*, 98, 10966–10971.
72. Chetelat, G., Desgranges, B., de la Sayette, V., et al. (2003), Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? *Neurology*, 60, 1374–1377.
73. Drzezga, A., Lautenschlager, N., Siebner, H., et al. (2003), Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: A PET follow-up study, *Eur. J. Nucl. Med. Mol. Imaging*, 30, 1104–1113.
74. Mosconi, L., Perani, D., Sorbi, S., et al. (2004), MCI conversion to dementia and the APOE genotype: A prediction study with FDG-PET, *Neurology*, 63, 2332–2340.
75. Anchisi, D., Borroni, B., Franceschi, M., et al. (2005), Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease, *Arch. Neurol.*, 62, 1728–1733.
76. Klunk, W. E., Engler, H., Nordberg, A., et al. (2004), Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B, *Ann. Neurol.*, 55, 306–319.
77. Fagan, A. M., Mintun, M. A., Mach, R. H., et al. (2006), Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans, *Ann. Neurol.*, 59, 512–519.
78. Archer, H. A., Edison, P., Brooks, D. J., et al. (2006), Amyloid load and cerebral atrophy in Alzheimer's disease: An 11C-PIB positron emission tomography study, *Ann. Neurol.*, 60, 145–147.
79. Engler, H., Forsberg, A., Almkvist, O., et al. (2006), Two-year follow-up of amyloid deposition in patients with Alzheimer's disease, *Brain*, 9, 2856–2866.
80. Mintun, M. A., Larossa, G. N., Sheline, Y. I., et al. (2006), [11C]PIB in a nondemented population: Potential antecedent marker of Alzheimer disease, *Neurology*, 67, 446–452.
81. Rocca, M. A., Mezzapesa, D. M., Ghezzi, A., et al. (2005), A widespread pattern of cortical activations in patients at presentation with clinically isolated symptoms is associated with evolution to definite multiple sclerosis, *Am. J. Neuroradiol.*, 26, 1136–1139.
82. Campion, D., Dumanchin, C., Hannequin, D., et al. (1999), Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum, *Am. J. Hum. Genet.*, 65, 664–670.
83. Sherrington, R., Froelich, S., Sorbi, S., et al. (1996), Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant, *Hum. Mol. Genet.*, 5, 985–988.
84. Janssen, J. C., Beck, J. A., Campbell, T. A., et al. (2003), Early onset familial Alzheimer's disease: Mutation frequency in 31 families, *Neurology*, 60, 235–239.

85. Dyment, D. A., Ebers, G. C., and Sadovnick, A. D. (2004), Genetics of multiple sclerosis. *Lancet Neurol.*, 3, 104–110.
86. Lerner, A. J., and Doran, M. (2006), Clinical phenotypic heterogeneity of Alzheimer's disease associated with mutations of the presenilin-1 gene, *J. Neurol.*, 253, 139–158.
87. Campion, D., Flaman, J. M., Brice, A., et al. (1995), Mutations of the presenilin I gene in families with early-onset Alzheimer's disease, *Hum. Mol. Genet.*, 4, 2373–2377.
88. Cruts, M., van Duijn, C. M., Backhovens, H., et al. (1998), Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease, *Hum. Mol. Genet.*, 7, 43–51.
89. Rocca, W. A., Hofman, A., Brayne, C., et al. (1991), Frequency and distribution of Alzheimer's disease in Europe: A collaborative study of 1980–1990 prevalence findings. The EURODEM-Prevalence Research Group, *Ann. Neurol.*, 30, 381–390.
90. Bird, T. D., Sumi, S. M., Nemens, E. J., et al. (1989), Phenotypic heterogeneity in familial Alzheimer's disease: A study of 24 kindreds, *Ann. Neurol.*, 25, 12–25.
91. Jersild, C., Svejgaard, A., and Fog, T. (1972), HLA antigens and multiple sclerosis, *Lancet*, 1(7762), 1240–1241.
92. Peltonen, L., Saarela, J., and Kuokkanen, S. (2002), *Multiple Sclerosis. The Genetic Basis of Common Diseases*, 2nd ed., Oxford Monographs on Medical Genetics, New York: Oxford University Press, pp. 805–817.
93. Haines, J. L., Terwedow, H. A., Burgess, K., et al. (1998), Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. The Multiple Sclerosis Genetics Group, *Hum. Mol. Genet.*, 7, 1229–1234.
94. Zarepari, S., James, D. M., Kaye, J. A., et al. (2002), HLA-A2 homozygosity but not heterozygosity is associated with Alzheimer disease. *Neurology*, 58, 973–975.
95. Corder, E. H., Saunders, A. M., Strittmatter, W. J., et al. (1993), Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families, *Science*, 261, 921–923.
96. Strittmatter, W. J., Saunders, A. M., Schmechel, D., et al. (1993), Apolipoprotein E: High-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease, *Proc. Natl. Acad. Sci. U.S.A.*, 90, 1977–1981.
97. Meyer, M. R., Tschanz, J. T., Norton, M. C., et al. (1998), APOE genotype predicts when—not whether—one is predisposed to develop Alzheimer disease, *Nat. Genet.*, 19, 321–322.
98. Roses, A. D. (1995), Apolipoprotein E genotyping in the differential diagnosis, not prediction, of Alzheimer's disease, *Ann. Neurol.*, 38, 6–14.
99. Saunders, A. M., Hulette, O., Welsh-Bohmer, K. A., et al. (1996), Specificity, sensitivity, and predictive value of apolipoprotein-E genotyping for sporadic Alzheimer's disease, *Lancet*, 348, 90–93.
100. Welsh-Bohmer, K. A., Gearing, M., Saunders, A. M., et al. (1997), Apolipoprotein E genotypes in a neuropathological series from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD). *Annals of Neurology*, 42, 319–325.
101. Green, R. C. (2002), Risk assessment for Alzheimer's disease with genetic susceptibility testing: Has the moment arrived? *Alzheimer's Care Quart.*, 3, 208–214.
102. Romas, S. N., Santana, V., Williamson, J., et al. (2002), Familial Alzheimer disease among Caribbean Hispanics: A reexamination of its association with APOE, *Arch. Neurol.*, 59, 87–91.
103. Daw, E. W., Payami, H., Nemens, E. J., et al. (2000), The number of trait loci in late-onset Alzheimer disease, *Am. J. Hum. Genet.*, 66, 196–204.
104. Pinholt, M., Frederiksen, J. L., and Christiansen, M. (2006), The association between apolipoprotein E and multiple sclerosis, *Eur. J. Neurol.*, 13, 573–580.

105. Kamboh, M. I., Minster, R. L., Feingold, E., et al. (2006), Genetic association of ubiquilin with Alzheimer's disease and related quantitative measures, *Molec. Psychiat.*, 11, 273–279.
106. Li, Y., Hollingworth, P., Moore, P., et al. (2005), Genetic association of the APP binding protein 2 gene (APBB2) with late onset Alzheimer disease, *Hum. Mutat.*, 25, 270–277.
107. Ramos, E. M., Lin, M.-T., Larson, E. B., et al. (2006), Tumor necrosis factor-alpha and interleukin 10 promoter region polymorphisms and risk of late-onset Alzheimer disease, *Arch. Neurol.*, 63, 1165–1169.
108. Prince, J. A., Feuk, L., Sawyer, S. L., et al. (2001), Lack of replication of association findings in complex disease: An analysis of 15 polymorphisms in prior candidate genes for sporadic Alzheimer's disease, *Eur. J. Hum. Genet.*, 9, 437–444.
109. Croes, E. A., Dermaut, B., van der Cammen, T. J. M., et al. (2000), Genetic testing should not be advocated as a diagnostic tool in familial forms of dementia, *Am. J. Hum. Genet.*, 67, 1033–1035.
110. Zipser, B. D., Johanson, C. E., Gonzalez, L., et al. (2007), Microvascular injury and blood-brain barrier leakage in Alzheimer's disease, *Neurobiol. Aging*, 28, 977–986.
111. Saunders, N. R., Habgood, M. D., and Dziegielewska, K. M. (1999), Barrier mechanisms in the brain. I. Adult brain. *Clin. Exp. Pharmacol. Physiol.*, 26, 11–19.
112. Martino, G., Adorini, L., Rieckmann, P., et al. (2002), Inflammation in multiple sclerosis: The good, the bad, and the complex, *Lancet Neurol.*, 1, 499–509.
113. Blennow, K., and Hampel, H. (2003), CSF markers for incipient Alzheimer's disease, *Lancet Neurol.*, 2, 605–613.
114. Wallin, A. K., Blennow, K., Andreasen, N., et al. (2006), CSF biomarkers for Alzheimer's Disease: Levels of beta-amyloid, tau, phosphorylated tau relate to clinical symptoms and survival, *Dement. Geriatr. Cogn. Disord.*, 21, 131–138.
115. Kapaki, E., Paraskevas, G. P., Michalopoulou, M., et al. (2000), Increased cerebrospinal fluid tau protein in multiple sclerosis, *Eur. Neurol.*, 43, 228–232.
116. Sussmuth, S. D., Reiber, H., and Tumani, H. (2001), Tau protein in cerebrospinal fluid (CSF): A blood-CSF barrier related evaluation in patients with various neurological diseases, *Neurosci. Lett.*, 300, 95–98.
117. Dymont, D. A., Herrera, B. M., Cader, M. Z., et al. (2005), Complex interactions among MHC haplotypes in multiple sclerosis: Susceptibility and resistance, *Hum. Mol. Genet.*, 14, 2019–2026.
118. Lincoln, M. R., Montpetit, A., Cader, M. Z., et al. (2005), A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis, *Nat. Genet.*, 37, 1108–1112.
119. Payami, H., Schellenberg, G. D., Zarepari, S., et al. (1997), Evidence for association of HLA-A2 allele with onset age of Alzheimer's disease, *Neurology*, 49, 512–518.
120. Andreasen, N., Minthon, L., Davidsson, P., et al. (2001), Evaluation of CSF-tau and CSF-Abeta42 as diagnostic markers for Alzheimer disease in clinical practice, *Arch. Neurol.*, 58, 373–379.
121. Sunderland, T., Gur, R. E., and Arnold, S. E. (2005), The use of biomarkers in the elderly: Current and future challenges, *Biol. Psychiatry*, 58, 272–276.
122. Herukka, S.-K., Hallikainen, M., Soininen, H., et al. (2005), CSF amyloid-beta-42 and tau or phosphorylated tau and prediction of progressive mild cognitive impairment, *Neurology*, 64, 1294–1297.
123. Mehta, P. D., Pirttila, T., Patrick, B. A., et al. (2001), Amyloid beta protein 1-40 and 1-42 levels in matched cerebrospinal fluid and plasma from patients with Alzheimer disease, *Neurosci. Lett.*, 304102–304106.
124. Iqbal, K., Flory, M., Khatoon, S., et al. (2005), Subgroups of Alzheimer's disease based on cerebrospinal fluid molecular markers, *Ann. Neurol.*, 58, 748–757.

125. Itoh, N., Arai, H., Urakami, K., et al. (2001), Large-scale, multicenter study of cerebrospinal fluid tau protein phosphorylated at serine 199 for the antemortem diagnosis of Alzheimer's disease, *Ann. Neurol.*, 50, 150–156.
126. Hampel, H., Buerger, K., Zinkowski, R., et al. (2004), Measurement of phosphorylated tau epitopes in the differential diagnosis of Alzheimer disease: A comparative cerebrospinal fluid study, *Arch. Gen. Psychiatry*, 61, 95–102.
127. Hampel, H., Burger, K., Pruessner, J. C., et al. (2005), Correlation of cerebrospinal fluid levels of tau protein phosphorylated at threonine 231 with rates of hippocampal atrophy in Alzheimer disease, *Arch. Neurol.*, 62, 770–773.
128. De Jong, D., Jansen, R. W., Kremer, B. P., et al. (2006), Cerebrospinal fluid amyloid beta42/phosphorylated tau ratio discriminates between Alzheimer's disease and vascular dementia, *J. Gerontol. Biol. Sci. Med. Sci.* 61, 755–758.
129. Thal, L. J., Kantarci, K., Reiman, E. M., et al. (2006), The role of biomarkers in clinical trials for Alzheimer disease, *Alzheimer Dis. Assoc. Disord.* 20, 6–15.
130. Link, H., and Huang, Y. M. (2006), Oligoclonal bands in multiple sclerosis cerebrospinal fluid: An update on methodology and clinical usefulness, *J. Neuroimmunol.*, 180(1–2), 17–28.
131. Wolinsky, J. S., the PROMiSe Study Group (2003), The diagnosis of primary progressive MS, *J. Neurol. Sci.*, 206, 145–152.
132. Imrell, K., Landtblom, A. M., Hillert, J., et al. (2006), Multiple sclerosis with and without CSF bands: Clinically indistinguishable but immunogenetically distinct, *Neurology*, 67, 1062–1064.
133. Furlan, R., Rovaris, M., Martinelli Boneschi, F., et al. (2005), Immunological patterns identifying disease course and evolution in multiple sclerosis patients, *J. Neuroimmunol.*, 165, 192–200.
134. Teunissen, C. E., Dijkstra, C., and Polman, C. (2005), Biological markers in CSF and blood for axonal degeneration in multiple sclerosis, *Lancet Neurol.*, 4, 32–41.
135. Haghighi, S., Andersen, O., Oden, A., et al. (2004), Cerebrospinal fluid markers in MS patients and their healthy siblings, *Acta Neurol. Scand.*, 109, 97–99.
136. Eikelenboom, M. J., Petzold, A., Lazeron, R. H., et al. (2003), Multiple sclerosis: Neurofilament light chain antibodies are correlated to cerebral atrophy, *Neurology*, 60, 219–223.
137. Brettschneider, J., Petzold, A., Junker, A., et al. (2006), Axonal damage markers in the cerebrospinal fluid of patients with clinically isolated syndrome improve predicting conversion to definite multiple sclerosis, *Mult. Scler.*, 12, 143–148.
138. Leoni, V., Masterman, T., Diczfalusy, U., et al. (2002), Changes in human plasma levels of the brain specific oxysterol 24S hydroxycholesterol during progression of multiple sclerosis, *Neurosci. Lett.*, 331, 163.
139. Potti, A., Dressman, H. K., Bild, A., et al. (2006), Genomic signatures to guide the use of chemotherapeutics, *Nat. Med.*, 12, 1294–1300.
140. Dickerson, B. C., and Sperling, R. A. (2005), Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease, *NeuroRx*, 2, 348–360.
141. Littman, B. H., and Williams, S. A. (2005), The ultimate model organism: Progress in experimental medicine, *Nat. Rev. Drug Discov.*, 4, 631–638.
142. Patterson, S. D., and DuBose, R. F. (2006), The role of biomarkers in the future of drug development, *Expert Opin. Drug Discov.*, 1, 199–204.
143. Mohs, R. C., Kawas, C., and Carillo, M. C. (2006), Optimal design of clinical trials for drugs designed to slow the course of Alzheimer's disease, *Alzheimer's Dementia*, 2, 131–139.



144. Patterson, S. D., and DuBose, R. F. (2006), The role of biomarkers in the future of drug development, *Expert Opin. Drug Discov.*, 1(3), 199–204.
145. Severino, M. E., Dubose, R. F., and Patterson, S. D. (2006), A strategic view on the use of pharmacodynamic biomarkers in early clinical drug development, *IDrugs*, 9, 849–853.
146. Hye, A., Lynham, S., Thambisetty, M., et al. (2006), Proteome-based plasma biomarkers for Alzheimer's disease, *Brain*, 129, 3042–3050.
147. Simonsen, A. H., McGuire, J., Podust, V. N., et al. (2007), A novel panel of cerebrospinal fluid biomarkers for the differential diagnosis of Alzheimer's disease versus normal aging and frontotemporal dementia, *Dement. Geriatr. Cogn. Disord.*, 24, 434–440.
148. Shi, M., Caudle, W. M., and Zhang, J. (2008), Biomarker discovery in neurodegenerative diseases: A proteomic approach, *Neurobiol. Dis.*, Doi:10.1016/j.nbd.2008.09.004.
149. Akuffo, E. L., Davis, J. B., Fox, S. M., et al. (2008), The discovery and early validation of novel plasma biomarkers in mild-to-moderate Alzheimer's disease patients responding to treatment with rosiglitazone, *Biomarkers*, 13, 618–636.
150. Vellas, B., Andrieu, S., Sampaio, C., et al.; A European Task Force Group (2008), End-points for trials in Alzheimer's disease: A European task force consensus, *Lancet Neurol.*, 7, 436–450.





# 15

---

## *Review Boards*

**Maureen N. Hood,<sup>1</sup> Jason F. Kaar,<sup>2</sup> and Vincent B. Ho<sup>1</sup>**

*<sup>1</sup>Department of Radiology and Radiological Sciences and <sup>2</sup>Office of General Counsel,  
Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814*

### **Contents**

|        |                                                                                                          |     |
|--------|----------------------------------------------------------------------------------------------------------|-----|
| 15.1   | Background                                                                                               | 896 |
| 15.1.1 | Development of Human Research Guidelines                                                                 | 896 |
| 15.1.2 | History and Landmark Doctrines                                                                           | 897 |
| 15.2   | Definitions                                                                                              | 898 |
| 15.2.1 | Institutional Review Board                                                                               | 898 |
| 15.2.2 | Ethics Committee                                                                                         | 898 |
| 15.2.3 | Data Monitoring Safety Boards                                                                            | 898 |
| 15.2.4 | Assurances                                                                                               | 898 |
| 15.2.5 | Human Research                                                                                           | 898 |
| 15.2.6 | Informed Consent                                                                                         | 899 |
| 15.2.7 | Institution                                                                                              | 899 |
| 15.2.8 | IRB Registration                                                                                         | 899 |
| 15.3   | Global Nature of Research Review: International Committee on<br>Harmonisation and Good Clinical Practice | 899 |
| 15.4   | Composition, Roles and Activities of an Institutional Review<br>Board/Institutional Ethics Committee     | 900 |
| 15.4.1 | Composition                                                                                              | 900 |
| 15.4.2 | Oversight                                                                                                | 900 |
| 15.4.3 | Informed Consent                                                                                         | 901 |
| 15.4.4 | Other Concerns                                                                                           | 901 |

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences or the Department of Defense.

|        |                                                   |     |
|--------|---------------------------------------------------|-----|
| 15.5   | Types of IRB Reviews                              | 901 |
| 15.5.1 | Initiating a Clinical Trial and Types of IRB/IECs | 901 |
| 15.5.2 | Exempted Review                                   | 902 |
| 15.5.3 | Expedited Review (45 CFR § 46.110)                | 904 |
| 15.5.4 | Full Review                                       | 904 |
| 15.5.5 | Continuing/Annual Review                          | 904 |
| 15.6   | Adverse Events                                    | 905 |
| 15.7   | IRB Recordkeeping and Administrative Actions      | 905 |
| 15.7.1 | Information That Must Be Retained and Duration    | 905 |
| 15.7.2 | Training and Credentials                          | 905 |
| 15.7.3 | Privacy and Confidentiality                       | 905 |
| 15.8   | Investigator Responsibilities                     | 906 |
| 15.8.1 | Principal Investigator                            | 906 |
| 15.8.2 | Qualifications and Training                       | 906 |
| 15.8.3 | Study Design and Performance                      | 907 |
| 15.8.4 | Logistical Support and Budgetary Planning         | 907 |
| 15.8.5 | Study Security and Maintenance                    | 907 |
| 15.8.6 | Communication with Stakeholders                   | 908 |
| 15.9   | Summary                                           | 909 |
|        | Appendix: Websites                                | 910 |
|        | References                                        | 910 |

The fundamental lesson to be understood by doctors is that when they stray from the traditional concepts of medicine—saving lives, healing, and ameliorating pain and suffering—they enter into a territory that is fraught with danger, not only for their patients, but also for the morality of medicine as a profession.

—Lounsbury DE, Bellamy RF, Beam TE, and Sparacino LR.  
 Editors of *Military Medical Ethics*, Volume 2, 2003,  
 in memory of Sheldon H. Harris, Ph.D.

## 15.1 BACKGROUND

### 15.1.1 Development of Human Research Guidelines

Adherence to the highest ethical and moral standards is important not only for routine clinical practice but also for all research involving human subjects. In the United States, the Food and Drug Administration (FDA) regulates the evaluation of medications and biomedical devices and has the sole responsibility of approval for their commercial sale for use in humans. The Office of Human Research Protections (OHRP) oversees the regulation of federally funded human research [1, 2]. The primary oversight authority of investigators, however, is the responsibility of the institutional review board (IRB) or institutional ethics committee (IEC). The delegation of oversight responsibility to the IRB was made in recognition of the regional and institutional variations that exist in ethical and treatment standards of care [3]. The geographic proximity of the IRB/IEC to the location of the research facilitates timely direct communication between the regulatory body and the investigator. However, there is a growing move toward regional or community review boards, as well as a move toward independent review boards. Open communication between investigators and the IRB/IEC is critical for the timely performance of

audits, reviews, notifications, and amendments and ultimately provides improved safety and welfare for the research participants.

### 15.1.2 History and Landmark Doctrines

The history of IRBs/IECs has not been without significant challenges and problems, however. Prior to the early to mid-twentieth century, human research was governed by individual conscience, professional bodies, societal norms, and government customs of the time [4]. The earliest known legal requirement pertaining to informing human subjects of medical experimentation was the Berlin Code of 1900 [5]. The United States made some initial efforts toward forming human research guidelines and ethics in the early 1900s with studies such as the U.S. Army's Yellow Fever Project; but, unfortunately, most physicians and scientists either knew little of these guidelines or refused to adhere to such strict rules [5, 6].

Formalization of human research standards came as a reaction to the atrocities committed by Nazi medical researchers found during the Nuremberg Military Tribunal [7]. In addition, it is estimated that several hundred thousand deaths occurred from the Japanese biomedical, biological, and chemical warfare experimentation on involuntary subjects from the mid-1920s until 1945 [8]. The atrocities of forced human testing performed by Nazi and Japanese scientists on prisoners and civilians during World War II resulted in a global outcry and the establishment of a variety of ethical guidelines for research involving human subjects. In 1953, the Department of Defense instituted the first directive for an official process for conducting human-use research, called the Wilson Memorandum, which unfortunately, was kept mostly secret [6]. The National Institutes of Health attempted to establish policies for approving research studies on humans, but no consensus could be reached as to which human-use research projects should be subject to regulation [9]. A more substantial and formal systematic protection of human subjects in medical research remained dormant for many years.

In 1964, the World Medical Association issued the Declaration of Helsinki, a groundbreaking treatise for international research protections. The Declaration of Helsinki continues to be followed today and has been revised periodically, most recently in 2000 during the 52nd World Medical Association General Assembly in Edinburgh, Scotland [10]. However, the United States did not formulate human research regulations until after Henry Beecher published his ethics paper in 1966 [11] outlining some of the unethical human research projects being conducted in the United States, which sparked the U.S. Congress to pass the National Research Act in 1974 [12]. The National Research Act established the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, which is best known for formulating the links between ethical principles and human research practices in the Belmont Report of 1979. This legislation formally required the establishment of IRBs to oversee all federally funded research activities [12, 13]. The Belmont Report (available at <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm>) articulates the three main ethical principles upon which all IRBs should operate: respect for persons, beneficence, and justice [13]. Various international and federal regulations that govern human subject research require the oversight of research by an IRB or its equivalent (e.g., ethics committees, institutional ethics committees, health research ethics committees, etc.).

## 15.2 DEFINITIONS

### 15.2.1 Institutional Review Board

An independent group, committee, or similar such entity consisting of a combination of medical/scientific and nonmedical/nonscientific members that is formally organized to oversee biomedical research involving human subjects at an institution. An IRB's primary objective is to protect the rights, safety, and welfare of human subjects participating in biomedical research [14].

### 15.2.2 Ethics Committee

Essentially equivalent to an IRB but usually uses the name ethics committee (EC) or human research ethics committees (HREC) in countries outside the United States. Each country sets specific minimum requirements for the composition of ECs as guided by the International Conference on Harmonisation (ICH) E6 Good Clinical Practice (GCP) guidelines [15].

### 15.2.3 Data Monitoring Safety Boards

An independent, objective group that is charged with monitoring the safety of clinical trials. Clinical trials that are evaluating new drugs, biologics, or devices are required by 21 CFR § 312.50 and 21 CFR § 312.56 to monitor their studies [16]. Data monitoring safety boards (DSMBs) are an additional group that may perform a monitoring function [17]. Although DSMB composition may vary from place to place and country to country, they should be composed of scientists, clinicians, biostatisticians, and bioethicists who have the expertise necessary to be able to evaluate independently an objective safety analysis of the clinical trial [18].

### 15.2.4 Assurances

Documents for acknowledging compliance by an institution of mandated federal policies for human subject research and for the protection of research subjects. Assurances are a written commitment obtained through the Office of Human Research Protections (OHRP) that must be approved before federal funding can be awarded. The most common type of assurance accepted by OHRP is the Federalwide Assurance (FWA). All organizations participating in federal research that have an IRB must obtain an assurance and a written agreement between IRBs and/or organizations outlining their relationships and commitments to the recognized assurance [19].

### 15.2.5 Human Research

A systematic investigation conducted by an investigator involving living individuals or identifiable information of the human subjects that is conducted in a manner that is designed to contribute new knowledge. All research activities that fall under the regulation of Title 45 CFR part 46 are required to abide by IRB/IEC oversight [3].

### 15.2.6 Informed Consent

The process by which human subjects are informed of all aspects of a research study including potential risks and benefits that may influence the subject's decision on whether or not to volunteer to participate in the study. The informed consent document must be reviewed and approved by the IRB/IEC. In the event that a research subject is not legally competent to consent, as in the case with a minor or a person who lacks mental capacity (such as a person in a coma), the law normally allows a third person, typically a family member, to consent on the person's behalf. Check the requirements of the jurisdiction you are in to determine whether it is appropriate and the procedure to be followed. Care should be taken to ensure the IRB/IEC is acting in the best interests of the research subjects. In at least one instance an IRB was cited for erroneously putting the interests of investigators over the interests of subjects [20]. Informed consent may be oral or written depending on the risk level. More than minimal risk studies typically require formal written consent. Documentation of informed consent is done through a document that is signed and dated [15].

### 15.2.7 Institution

Any medical or dental entity or facility, public or private, where clinical trials can be conducted.

### 15.2.8 IRB Registration

Registration is required by any IRB/IEC that reviews U.S. federally supported human research or any institution that is subject to the Common Rule [3]. IRBs/IECs must be registered with OHRP before they may be designated on an FWA. IRBs/IECs that are registered are listed on the OHRP website at <http://ohrp.cit.nih.gov/search/asearch.asp#ASUR>. An OHRP assurance is also required.

## 15.3 GLOBAL NATURE OF RESEARCH REVIEW: INTERNATIONAL COMMITTEE ON HARMONISATION AND GOOD CLINICAL PRACTICE

All human subject research, regardless of the country in which the research is taking place, should be conducted under the International Conference on Harmonisation (ICH), E6 Good Clinical Practice (GCP) guidance (available at <http://www.fda.gov/cder/guidance/959fnl.pdf>). The International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use grew out of a meeting hosted by the European Federation of Pharmaceutical Industries and Associations in Brussels in 1990 that brought together the regulatory authorities and pharmaceutical industries in the European Union, United States, and Japan to discuss joint regulation of the pharmaceutical industry. The early intent of the ICH was to establish an international process for the development and regulation of new medical products in a more efficient manner.

The ICH has continued to expand its guidance through harmonization of therapeutic advances, technical developments, and other specific medical products based

on scientific consensus and the protection of public health and individual safety on an international perspective. The World Health Organization (WHO) has endorsed these ICH GCP international standards as well as the Declaration of Helsinki and the Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines for Biomedical Research Involving Human Subjects as the ethical and scientific standards for conducting biomedical human subject research [21]. These guidelines are intended to help facilitate biomedical research and clinical trials worldwide. The Office for Human Research Protections has been working on compiling an international listing of countries that have laws, regulations, or guidelines to cover human research. The 2005 international compilation listing includes 72 countries with links to standards and international organizations (available at <http://www.hhs.gov/ohrp/international/HSPCompilation.pdf>).

## **15.4 COMPOSITION, ROLES, AND ACTIVITIES OF INSTITUTIONAL REVIEW BOARD/INSTITUTIONAL ETHICS COMMITTEE**

### **15.4.1 Composition**

The composition of an IRB/IEC will vary but should consist of a reasonable number of members with qualifications and expertise in a variety of aspects so as to ensure the rights, safety, and welfare of subjects involved in research are protected. In the United States, 45 CFR § 46.107 strongly encourages diversity, not only as to race, creed, sex, and national origin, but also professional and social experience [3]. If a vulnerable population, such as prisoners are used, there must be someone on the committee representing their interests (45 CFR § 46.304). Minimum requirements: at least five members, at least one member who is from a nonscientific background, and at least one member who is independent of the site or institution. Both genders must be represented, including making quorum, which must be at least 50% of the primary membership (45 CFR § 46.108) [14]. Many institutions are moving to a two-tiered system that adds a scientific review component step before the protocol is presented to the IRB/IEC. The scientific review is intended to aid in promoting research excellence. This two-step system is intended to help streamline the IRB/IEC review process by having a scientific review by experts on the subject matter first address the study design aspects of the protocol, which then allows the IRB/IEC to focus on the human use issues. However, the IRB/IEC still needs to review the report from the scientific review as well as review the protocol as a whole in the human use context.

### **15.4.2 Oversight**

The main goals of an IRB/IEC are to ensure that the rights of the research subject especially in regards to privacy and informed consent are maintained, that investigators conduct their research in a professional and ethical manner, and that the potential benefits of the study outcomes outweigh the potential risks to the patient. IRBs/IECs are also tasked with the responsibilities to ensure that the studies that fall under their purview are methodologically sound in their design, whereby the objectives of the research will provide meaningful data and the risks and benefits



for study involvement are suitable for human participation. If an institution uses a scientific review committee, the findings from the scientific review should be reviewed by the IRB/IEC. These roles for the IRB are not exclusive to the initial protocol approval but are continuous throughout the duration of the study and for at least 3 years after the study has completed, and even longer for pharmaceutical clinical trials according to 45 CFR § 46.115 [3].

### 15.4.3 Informed Consent

The informed consent process and documentation is a critical element the IRB/IEC must review. The information contained within the informed consent document must be at a language level appropriate for the study population (usually sixth grade level) and may need to be made available in more than one language in order to support all participants. The information must be clear to the potential volunteer so that the decision of whether to participate can be made freely by the subject. When the research involves a medical product such as a device or pharmaceutical, an adequate summary of the product must be communicated to the potential subject that includes the safety information and clinical experience or the product. Any compensation for subject participation should be contained in the informed consent document as well. Compensation may not be at a level that is seen as improper inducement to participate in the research. The process of recruiting the potential subjects as well as a description of the informed consent process must be approved by the IRB/IEC in order to reduce potential coercion.

### 15.4.4 Other Concerns

There are other important concerns that an IRB/IEC must also consider for each research protocol such as legal issues and intellectual property rights. Members of the IRB must remember that research protocols may contain sensitive or proprietary information. It is imperative that members of the IRBs/IECs respect the confidentiality of the research protocols. “Products of human ingenuity” such as a new medication or device should be considered intellectual property as guided by *Diamond v. Chakrabarty*, 447 U.S. 303, 1980, [22, 23]. The IRB will also be assessing a number of other issues as outlined in Table 1, all in accordance with respect for persons, beneficence, and justice [24].

## 15.5 TYPES OF IRB REVIEWS

### 15.5.1 Initiating Clinical Trial and Types of IRBs/IECs

Before starting any human research or clinical trial, an investigator must attain written approval from his or her IRB/IEC prior to initiation of the study. This approval process includes approval of informed consent forms, advertisements, and other written documents as well as specific procedures, data collection, analysis, and storage. When an institution does not have an IRB/IEC, there are independent IRBs/IECs that can be contracted to serve as an IRB/IEC to review research under your institution’s OHRP accepted assurance. For multicenter clinical trials, central IRBs/IECs can be used [25]. Clinical trials that fall under the investigative new drug

**TABLE 1 Overview of Basic Items in an IRB Review**

| Purpose of Study                            | Background, Summary, and Aims                                                                                                              |
|---------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| Sponsor                                     | The name of the organization that takes the responsibility for and initiates a clinical trial. It may be a private or public organization. |
| Investigator                                | Qualifications, financial disclosures, conflicts of interest.                                                                              |
| Study population                            | Inclusion/exclusion criteria, subject recruitment, appropriateness of subject selection, advertisement.                                    |
| Scientific design                           | Objectives, design description adequate to answer question, research procedures.                                                           |
| Informed consent                            | Written documentation of process is essential.                                                                                             |
| Potential risks and discomforts to subjects | Risk relative to alternatives, risk–benefit analysis.                                                                                      |
| Confidentiality                             | Are procedures adequate to protect the privacy and confidentiality of the subject?                                                         |
| Data oversight                              | Stopping rules, data monitoring board.                                                                                                     |
| Data analysis                               | Appropriate data analysis and statistical analysis.                                                                                        |
| Compensation and costs                      | Is compensation or reimbursement of subject reasonable?                                                                                    |
| Other                                       | Depending on the type of clinical trial, other checks may be considered by the IRB.                                                        |

(IND) regulations must be approved by an IRB/IEC. Central IRBs/IECs help facilitate the IRB/IEC process by reducing redundancy in effort, and for small institutions, can provide a more diverse perspective for reviewing the research protocol. Centralized IRBs/IECs are consistent with 21 CFR § 56.114 that provides for institutions involved in multi-institutional studies and joint reviews. Centralized IRBs/IECs can be used by institutions in whole or in part for the review of a multi-center clinical trial [25]. In addition, if there is a registered IRB/IEC in your geographic region, you can negotiate having them conduct reviews for your institution, or an institution may also consider establishing its own IRB/IEC.

In all cases in which research is conducted using funds from the U.S. government each institution must obtain its own assurance that is accepted by OHRP (45 CFR § 46.103). Regardless of the type of review system, there are a number of elements in common that must be addressed in the protocol submission (Table 1). Local IRBs/IECs have the flexibility to specify the format and detailed content required for the review as long as the information meets federal requirements. The type of review a protocol receives is up to the IRB/IEC, not the investigator. It is prudent for an investigator to discuss a potential protocol with the IRB/IEC for guidance before submitting a protocol. However, the IRB/IEC or its properly established subcommittee must review the protocol before deciding which type of review is most appropriate. This is normally determined by local guidelines.

### 15.5.2 Exempted Review

Certain types of research can be exempted from the full detailed review of the IRB. The research must meet all of the criteria in 45 CFR § 46.101(b) (1-6) and involve no greater than minimal risk to be eligible for exempted review (e.g., retrospective review of last 100 abdominal CT studies) (Table 2). The IRB typically does not conduct continuing reviews of exempted protocols, but the investigator will be

**TABLE 2 Exempted Research Criteria as Found in 45 CFR§ 46.101 (b)**

| Category                                                                                                                                                                                                                        | Conditions                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as:                                                                                               | (a) Research on regular and special education instructional strategies;<br>OR<br>(b) Research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.                                                                                                                                                                                                                                                                                                                                                                                                         |
| (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, if:                                              | (a) Information obtained is recorded in such a manner that human subjects cannot be identified, directly or through identifiers linked to the subjects;<br>AND<br>(b) Any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subject's financial standing, employability, or reputation;<br>AND<br>(c) Subjects are not under the age of 18 or members of a vulnerable class, including prisoners, pregnant women, individuals who are mentally disabled or economically or educationally disadvantaged. |
| NOTE: all three conditions must apply.                                                                                                                                                                                          | (a) The human subjects are elected or appointed public officials or candidates for public office;<br>OR<br>(b) Federal statutes(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.                                                                                                                                                                                                                                                                                                                              |
| (3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (2) above if: | (a) These sources are publicly available;<br>OR<br>(b) If the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.                                                                                                                                                                                                                                                                                                                                                                                                      |
| (4) Research, involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if:                                                                                      | (a) Public benefit or service programs;<br>OR<br>(b) Procedures for obtaining benefits or services under those programs;<br>OR<br>(c) Possible changes in or alternatives to those programs or procedures;<br>OR<br>(d) Possible changes in methods or levels of payment for benefits or services under those programs.                                                                                                                                                                                                                                                                                                     |
| (5) Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine:                                       | (a) Wholesome foods without additives are consumed;<br>OR<br>(b) If a food is consumed that contains a food ingredient at or below the level <b>and</b> for a use found to be safe, <b>or</b> agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the USDA.                                                                                                                                                                                      |
| (6) Taste and food quality evaluation and consumer acceptance studies, if:                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

*Note:* In order to qualify for exemption, all of a project's activities must qualify as one or more categories from the left column. (To qualify for a category, a project must meet the conditions to the right of the category). The IRB/IEC makes the decision regarding exempted status. This is presented as a guideline only.  
*Source:* Adapted from Ref. [3].

required to complete an annual report. The investigator is also still bound by all the federal ethical requirements and regulations for good clinical practice (GCP) [15]. Informed consent is typically not required by law for exempted protocols, however, the sponsor or local IRB/IEC may require it. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) authorization or waiver may also be needed. The IRB/IEC will make the determination as to which protocols fit exempted review status and the human use requirements for the investigator.

### **15.5.3 Expedited Review (45 CFR § 46.110)**

Certain kinds of research can be conducted through an expedited review process. Protocols that may be subject to expedited review need to have certain characteristics that may include: (1) must be no more than minimal risk; (2) may be exempted research; (3) must fit within the categories outlined by federal regulations 45 CFR § 46.110; (4) may be a review of minor revisions for a previously approved research study; and (5) may still require informed consent. The local IRB/IEC has the authority to make the decision for expedited reviews, not the investigator.

### **15.5.4 Full Review**

Most prospective studies and some retrospective studies will require a full review. Consult with the local IRB/IEC when in doubt. The review must be conducted at a convened IRB meeting that has a quorum of at least half of its primary members. The IRB must determine that all of the federal requirements specified in 45 CFR § 46.111 and 21 CFR § 56.111 are satisfied [3, 14]. A protocol must receive the approval of the majority of the IRB members present at the meeting to be approved. The IRB must inform the investigators and the institution in writing of its decision to approve, modify, or disapprove the research. Most protocols go through at least one round of revisions prior to final approval.

### **15.5.5 Continuing/Annual Review**

Continuing reviews of each protocol will occur at intervals appropriate for the degree of risk, but must not be less than once a year. If 365 days elapse without continuing review (or a shorter period of time specified by the IRB), the study is suspended [26]. There is NO grace period. The purpose of the continuing review is to review the progress of the entire study, not just the changes to it. The review process not only protects the study subjects, but also exists to make sure enrollment is being done reasonably, and that the study is being conducted as outline by the research protocol. Studies with higher risk levels are reviewed more frequently (e.g., every 3 months or after every 3 subjects). The IRB/IEC will make the determination for the frequency, plus the IRB/IEC has the authority to change the frequency of the reviews based on what it considers appropriate for the safety of the subjects and the integrity of the study.

The IRB/IEC reviews all amendment requests by the investigator/sponsor. The amendments must not deviate significantly from the original proposal and need to be critically evaluated for potential risk to the subjects and to scientific validity of the study design.

## 15.6 ADVERSE EVENTS

Adverse events are taken very seriously by IRBs/IECs. Serious and unexpected adverse events must be reported expediently to the IRB [24]. A serious event is considered to be any experience that results in death or a life-threatening situation where the subject is admitted to the hospital or the outcome is a persistent or significant disability, incapacity, or congenital anomaly in a progeny. An unexpected event is any adverse experience not consistent with the current description of the IRB-approved protocol and informed consent form. Not all adverse events need to be evaluated individually by the IRB/IEC or DSMB. Events that are not significant or not clinically meaningful to the safety of the subjects should be documented, but normally do not require special review [24]. The IRB/IEC and/or DSMB will review protocols at least annually to assess the protocol for safety and documentation. Sometimes, unexpected frequencies of adverse events may cause the IRB/IEC or DSMB to either require a modification to the risk–benefit section of the protocol and informed consent document, or in rare cases, terminate the study.

## 15.7 IRB RECORDKEEPING AND ADMINISTRATIVE ACTIONS

### 15.7.1 Information That Must Be Retained and Duration

Institutional review boards are required to maintain accurate records of every protocol and investigator under their authority [1]. In the United States, the FDA requires IRBs/IECs to maintain accurate records of all research activities of the institution and store them for at least 3 years after completion or termination of a protocol [1]. The international guidelines similarly recommend the same careful documentation and 3-year minimum storage of records [27]. Some sponsored clinical trials and institutions may keep their research records longer.

### 15.7.2 Training and Credentials

The main functions of the IRB/IEC are to thoroughly review each research protocol, perform continuing reviews, and review adverse events. The breadth of an IRB/IEC's responsibilities is truly great and even includes the responsibility for ensuring that investigators have adequate training to perform the research procedures and have the necessary credentials to adequately care for the participants. In addition, the members of the IRB/IEC need to have periodic continuing education and training for their roles in the IRB/IEC process.

### 15.7.3 Privacy and Confidentiality

One additional aspect of the IRB/IEC's responsibility that cannot be overlooked concerns the protection of the participants' privacy and confidentiality rights. The IRB/IEC must assure that the investigators maintain the participants' privacy in all aspects of the research protocol. In the United States, there are added security requirements (HIPAA), which went into effect April 14, 2003 [28]. Table 3 outlines the protected health information.

**TABLE 3 Health Information Protected by 45 CFR§ 164.514(b)**

|                                                                   |                                                                      |
|-------------------------------------------------------------------|----------------------------------------------------------------------|
| Names                                                             | Biometric identifiers                                                |
| Dates that directly relate to an individual                       | Persons over the age of 89                                           |
| Telephone numbers                                                 | Fax numbers                                                          |
| Medical record number                                             | Social Security number                                               |
| Health plan beneficiary number                                    | Electronic e-mail address                                            |
| Certificate/license numbers                                       | Account numbers                                                      |
| Web universal resource locators (URLs)                            | Internet protocol (IP) address                                       |
| Full-face photograph                                              | Geocodes                                                             |
| Any copyrighted/unique identifying number, characteristic or code | Geographic subdivisions smaller than a state (some exceptions apply) |

*Source:* Adapted from 45 CFR § 164.514(b) [29].

The IRBs/IECs must also make sure the intellectual property rights of each protocol and the investigators are protected as well. IRB/IEC research protocols are considered confidential material for the participants and the investigators and must be kept in a secure or locked file. Each institution must also maintain accurate and adequate files and databases of their research that are kept secure.

## **15.8 INVESTIGATOR RESPONSIBILITIES**

### **15.8.1 Principal Investigator**

Investigators that participate in human research have a variety of responsibilities that include qualifications and expectations, study design and performance, and communication with research stakeholders. The principle investigator, or PI, has the ultimate responsibility to ensure that all investigators (i.e., associate investigators) and staff on the study have the appropriate qualifications to perform their roles within the study design. The PI bears the primary responsibility for study adherence to regulatory and ethical standards and the integrity of both the scientific and clinical aspects of the research.

### **15.8.2 Qualifications and Training**

A primary responsibility of any investigator is that he or she has the adequate education, training, and experience to participate in the proposed research. If an investigator is a novice, it is generally expected that additional experience under the tutelage of an expert is obtained prior to the individual's performance of the task independently. If a formal certification process exists for that procedure or task, the investigator is expected to have obtained that certification. Furthermore, it is expected that investigators adhere to the laws, regulations, and ethics of the country and locality that pertain to the tasks or research that are being performed. Financial disclosure is required when the investigator has a monetary relationship with sponsors. This is to avoid possible conflicts of interest in the clinical trial [30].

In addition to being qualified to perform their roles within a project, all investigators must also receive proper human research protections training. This can be

achieved various ways such as via a traditional classroom instruction, interactive CD ROM program, or Web-based tutorial. In special cases, if an IRB/IEC feels that a researcher has not met full qualifications to participate in research, the IRB/IEC can require additional training or the involvement of a more senior investigator to oversee the work and mentor the less seasoned investigator [26]. The investigators' responsibilities, regardless of the type of study or type of scientific or medical discipline, are based on the ICH and GCP documents. In the United States, in addition to human research protection training, investigators should be given a copy of the Belmont Report (<http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm>) as an ethical guide.

### **15.8.3 Study Design and Performance**

Investigators are responsible for designing studies that have not only scientific merit but also do not unduly jeopardize the well-being of the participants. Accordingly, human research studies must be designed such that the study can be validated or replicated, the study participants are not exposed to unnecessary risks, and the benefits of the research outweigh the risks to the individuals volunteering their time for the research. The investigator and the IRB should note that "a poorly designed human research protocol is unethical" [31, p. 109]. Inherent to this concept is that the study design includes proper surveillance methods for detecting potential risks and/or qualified individuals to identify abnormalities.

### **15.8.4 Logistical Support and Budgetary Planning**

Prior to the performance of the study, the investigators must ensure that there is adequate logistical support (e.g., personnel, equipment, space, etc.) and funding for the performance of the proposed study. Should resources lapse (e.g., key investigator leaves the institution or funding abruptly be terminated) during the conduct of a study, it is the responsibility of the investigators to notify the IRB and potentially suspend or amend the protocol to more modest goals should the resource not be replenished.

### **15.8.5 Study Security and Maintenance**

Investigators are responsible for the integrity of the data, the human subject recruitment, informed consent process, and maintenance not only of the regulatory documentation but also of subject data files. Investigators are also responsible for maintenance of subject privacy and confidentiality, which includes development of the processes for the assurance of these elements. It is common practice for the secure storage of study records and materials. Secure storage of research documents is usually required to be no less than 2 years following FDA market application approval or discontinuation (21 CFR § 312.62) [16], but some sponsors and institutions require much longer storage periods, so investigators are responsible for finding out specific requirements. The records that must be maintained by the investigator are: investigational drug disposition, dates, quantities of drugs given, unused supplies, case histories, all recorded observations, supporting data, signed informed consent documents, progress notes, and any other pertinent medical records or data



collected for the trial (21 CFR § 312.62) [16]. These recordkeeping requirements may also be required by the institution. Investigators have the responsibility to ensure that records for the research study are being properly completed and maintained.

### 15.8.6 Communication with Stakeholders

Investigators are expected to communicate honestly, openly, and freely with not only the subjects but also the IRB/IEC or other outside regulatory (e.g., FDA) or interested bodies (e.g., sponsor or funding agency)—that is all persons with a “stake” in the research. Investigators, thus, are responsible not only for the preparation of the research proposal document but also for the presentation and defense of the proposed study to the IRB/IEC and any other applicable regulatory committee. A clinical study which entails the use of chest X rays, for example, typically must also receive approval from the institution’s radiation safety committee. Investigators also have a responsibility to maintain the communication with the IRB/IEC and other stakeholders even after the study has been initially approved.

**New Findings** Investigators are expected to notify the IRB/IEC of any new findings such as unanticipated risks or benefits. There is a similar expectation that subjects also be notified. Investigators are expected to incorporate their new information into the study and request amendment to the study design or documentation (e.g., informed consent documentation) in accordance to the new information. Note that this pertains not only to new information that the investigator may have learned from the specific research undertaken but also from the reported literature or available information from studies performed by other investigators. If the study is an investigational new drug trial, the pharmaceutical company (i.e., sponsor) typically provides regular updates on new findings, namely serious adverse events, that arise from all sites such that individual sites can benefit from the experience of other investigators. In response to any new study finding, investigators have the responsibility of determining if the protocol needs amendment and the responsibility of filing the appropriate regulatory paperwork for protocol amendment.

**Amendment Requests and Protocol Deviation Reporting** Any changes or amendment requests to the protocol require IRB/IEC approval. If, after the fact, a deviation for the protocol is noted, the deviation must be documented and explained to the IRB/IEC and the sponsor of the study. Deviations are considered protocol violations and are taken seriously by an IRB/IEC. Deviations made for patient safety may be done prior to IRB/IEC approval but must be reported and an amendment request made immediately to the IRB/IEC, the sponsor, and any regulatory authorities, or the trial may be suspended [15]. The protection of the human subject is the primary responsibility of the investigator. The investigator must not coerce subjects into volunteering for trials, and the investigator must be careful to fully disclose the potential risks and benefits to the subjects prior to enrollment (45 CFR § 46.116) [3]. Only investigators on the IRB/IEC-approved protocol are allowed to administer investigation drugs and with disposing of or returning unused supplies to the sponsor.

**Complying with Audits** Investigators also are responsible of complying with regulatory audits, which can be performed as a scheduled audit or impromptu inspection by the IRB or on occasion, the FDA or other regulatory body. IRBs/IECs are required to review human subject research no less than annually. In studies associated with higher risk, IRBs/IECs may choose to review the protocol more frequently. In any case, the investigators are responsible for adhering to these regulatory responsibilities which typically include an interview with an IRB/IEC representative, a reporting of study progress, and preparation of data files for review.

**Safety and Data Safety Monitoring Boards** Investigators must also report regularly to the sponsor any data or information specified by the agreement between the sponsor and investigator for the conduction of the trial. Types of reports are IRB/IEC approvals and rosters, progress reports, safety reports, and final reports. In addition, the investigators may be monitored by study monitors that are supplied by the sponsor, in addition to any monitoring requirement made by the IRB/IEC. DSMBs are growing in use by IRBs/IECs. However, there is some confusion and overlap of responsibilities between the IRBs/IECs and the DSMBs because the FDA requirements (21 CFR § 312.66) for reporting unanticipated problems are vague [16]. The overall goal should be to adequately protect the research participants through a process that can reliably detect harm to participants and then respond quickly and appropriately to the particular instances [32]. DSMBs are meant to help the research process, but more standardization of roles and adverse event reporting need to be established on an international basis [32]. Nonetheless, investigators must comply with approved protocol procedures and safety regulations. Noncompliance by investigators such as failure to report changes in the protocol, failure to report safety problems, or misuse of informed consent documents may result in the IRB/IEC halting the protocol and investigating the issues [26]. Serious lapses or continued lapses by the investigator may lead to termination of the protocol. IRBs/IECs want to ensure that the research being performed with human subjects is being done in a manner that protects the rights and welfare of the participants. Similarly, sponsors and researchers hope to bring safe and beneficial health care products, techniques, and devices to clinical practice.

## 15.9 SUMMARY

Institutional review boards or institutional ethics committees are an essential part of the clinical trial system. IRBs/IECs should help the researchers conduct research in an ethical manner that protects the rights and welfare of the human participants. There are many international regulations and guidelines such as the ICH/GCP that help guide IRBs/IECs in clinical trials. These boards are in place because of the atrocities that have occurred in past medical research, especially during World War II. This chapter presented an overview of the IRB/IEC process along with resources for researchers who want to be involved in clinical trial research. The bottom line is to perform clinical research that is safe and ethical to the human participant such that new knowledge can be gained for the advancement of health care.

## APPENDIX: WEBSITES

### General Regulations and Guidance

<http://www.fda.gov/oc/gcp/default.htm>  
<http://www.fda.gov/oc/ohrt/irbs/default.htm>  
<http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm>  
<http://www.fda.gov/cder/guidance/959fnl.pdf>  
<http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm>  
<http://www.hhs.gov/ohrp/international/HSPCompilation.pdf>  
[http://www.cioms.ch/frame\\_guidelines\\_nov\\_2002.htm](http://www.cioms.ch/frame_guidelines_nov_2002.htm)  
<http://www.pre.ethics.gc.ca/english/policystatement/policystatement.cfm>  
<http://www.dh.gov.uk/assetRoot/04/01/86/39/04018639.pdf>  
<http://eudract.emea.europa.eu>  
[http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/eudralex\\_en.htm](http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/eudralex_en.htm)  
<http://www.nihs.go.jp/hse/drug/drug-e.html>  
<http://www.tga.gov.au/index.htm>  
<http://www.hhs.gov/ohrp/documents/OHRPRegulations.pdf>

### Clinical Trial Guidance

<http://www.fda.gov/cber/gdlns/clindatmon.pdf>  
<http://www.emea.europa.eu>  
[http://www.hc-sc.gc.ca/dhp-mpps/prodpharma/applic-demande/guide-ld/clini/qual\\_cta\\_dec\\_e.html](http://www.hc-sc.gc.ca/dhp-mpps/prodpharma/applic-demande/guide-ld/clini/qual_cta_dec_e.html)

### Ethics

<http://www.ich.org/cache/compo/276-254-1.html>  
<http://www.wma.net/e/policy/b3.htm>  
<http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm>

### Investigator Agreement

<http://www.hhs.gov/ohrp/humansubjects/assurance/spa-aii.htm>

## REFERENCES

1. U.S. Department of Health and Human Services Food and Drug Administration. Information Sheets: Guidance for Institutional Review Boards and Clinical Investigators 1998 Update; available at: <http://www.fda.gov/oc/ohrt/irbs/appendixc.html>; accessed April 28, 2002.
2. Department of Health and Human Services, Office of the Secretary (2000), Office of Public Health and Science, and National Institutes of Health, Office of the Director, Statement of organization, functions and delegations of authority, *Fed. Reg.*, 65(114), 37136–37137.

3. National Archives and Records Administration (2006), Electronic Code of Federal Regulations (e-CFR), Title 45, Public Welfare, Part 46, Protection of Human Subjects (45 CFR §46), Government Printing Office, Washington, D.C; available at: <http://www.hhs.gov/ohrp/documents/OHRPRegulations.pdf>.
4. Schneider, W. H. (2002), The establishment of institutional review boards in the U.S. background history; available at: <http://www.iupui.edu/~histwhs/G504.dir/irbhist.html>; accessed November 2, 2002.
5. Nicholson R. H. (2003), The regulation of medical research: A historical overview, in Eckstein S., Ed., *Manual for Research Ethics Committees*, 6th ed., Cambridge University Press, New York.
6. Advisory Committee on Human Radiation Experiments (1996), *Human Radiation Experiments: Final Report of the Presidents Advisory Committee*, Oxford University Press, New York.
7. *United States v. Karl Brandt*, et al. (1949), The Medical Case: Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law 10. *Nuremberg. Volumes I–II*, Government Printing Office, Washington, D.C.
8. Harris, S. H. (2003), Japanese biomedical experimentation during the World-War-II era, in Lounsbury D. E., Bellamy, R. F., Beam, T. E., and Sparacino, L. R., Eds., *Military Medical Ethics*, Vol. 2, Office of the Surgeon General, Department of the Army, Washington, D. C.
9. National Institutes of Health (2002), A short history of the National Institutes of Health—exhibits and galleries; available at: <http://history.nih.gov/exhibits/history/main.html>; accessed December 31, 2003.
10. World Medical Association General Assembly (2002), World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Humans; available at: <http://www.wma.net/e/policy/pdf/17c.pdf>; accessed December 31, 2003.
11. Beecher, H. K. (1966), Ethics and clinical research, *N. Engl. J. Med.*, 274(24), 1354–1360.
12. *National Research Act of 1974*. P.L. 348, 93d Cong., 2d Sess (12 July 1974). A bill to amend the Public Health Service Act to establish a national program of biomedical research fellowships, traineeships, and training to assure the continued excellence of biomedical research in the United States, and for other purposes, 1974.
13. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979), The Belmont Report Ethical Principles and Guidelines for the Protection of Human Subjects of Research; available at: <http://ohsr.od.nih.gov/mpa/belmont.php3>; accessed October 27, 2002.
14. National Archives and Records Administration (2002), *Code of Federal Regulations, Title 21, Chapter I—Food and Drug Administration*, Department of Health and Human Services, Part 56, Institutional Review Boards (21 CFR §56), Government Printing Office, Washington, D.C.
15. ICH (2006), International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Guideline for good clinical practice, E6 (R1); available at: <http://www.ich.org/LOB/media/MEDIA482.pdf>; accessed April 6, 2006.
16. National Archives and Records Administration (2002), *Code of Federal Regulations, Title 21, Chapter I—Food And Drug Administration*, Department of Health and Human Services, Part 312, Investigational New Drug Application (21 CFR §312), Government Printing Office, Washington, D.C.
17. FDA (2001), U. S. Department of Health and Human Services, Food and Drug Administration, Guidance for Clinical Trial Sponsors. Establishment and Operation of Clinical

- Trial Data Monitoring Committees; available at: <http://www.fda.gov/OHRMS/DOCKETS/98fr/01d-0489-gdl0003.pdf>; accessed April 6, 2006.
18. Slutsky, A. S., and Lavery, J. V. (2004), Data safety and monitoring boards, *N. Engl. J. Med.*, 350(11), 1143–1147.
  19. HHS (2006), U.S. Department of Health & Human Services, Office for Human Research Protections (OHRP), available at: <http://www.hhs.gov/ohrp/>; accessed April 6, 2006.
  20. *Grimes v. Kennedy Krieger Institute* (2001), Case 129, Maryland Court of Appeals.
  21. Council for International Organizations of Medical Sciences (2006), International Ethical Guidelines for Biomedical Research Involving Human Subjects; available at: [http://www.cioms.ch/frame\\_guidelines\\_nov\\_2002.htm](http://www.cioms.ch/frame_guidelines_nov_2002.htm); accessed April 6, 2006.
  22. *Diamond v. Chakrabarty* (1980), 447 US 303–310. Diamond, Commissioner of Patents and Trademarks V. Chakrabarty. Certiorari to the United States Court of Customs and Patent Appeals. No. 79–136.
  23. Resnick, D. (2001), DNA patents and scientific discovery and innovation: Assessing benefits and risk, *Sci. Eng. Ethics*, 7(1), 29–62.
  24. Amdur, R., Kornetsky, and Bankert, L. (2002), Reviewing a research proposal, in Amdur, R. Ed., *The Institutional Review Board Member Handbook*, Jones and Bartlett, Sudbury, MA.
  25. HSS (2006), U.S. Department of Health and Human Services, Food and Drug Administration, Guidance for industry. Using a centralized IRB review process in multicenter clinical trials; available at: <http://www.fda.gov/cder/guidance/index.htm>; accessed April 6, 2006.
  26. U.S. Department of Health & Human Services (1993), Protecting Human Research Subjects: Institutional Review Board Guidebook; available at: [http://www.hhs.gov/ohrp/irb/irb\\_guidebook.htm](http://www.hhs.gov/ohrp/irb/irb_guidebook.htm); accessed April 6, 2006.
  27. World Health Organization (2000), Operations Guidelines for Ethics Committees That Review Biomedical Research; available at: <http://www.who.int/tdr/publications/publications/pdf/ethics.pdf>; accessed April 6, 2006.
  28. Steinbrook, R. (2002), Health Policy Report: Improving protection for research subjects, *N. Engl. J. Med.*, 346(18), 1425–1430.
  29. National Archives and Records Administration (2006), *Electronic Code of Federal Regulations (e-CFR)*, Title 45, Public Welfare, Part 162.514, Administrative Requirements (45 CFR §162.514), Government Printing Office, Washington, D.C.; available at: <http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=%2Findex.tpl>; accessed April 6, 2006.
  30. National Archives and Records Administration. (2002), *Code of Federal Regulations, Title 21, Chapter I—Food And Drug Administration*, Department of Health and Human Services, Part 54.4, Financial Disclosure by Clinical Investigators (21 CFR §54.4), Government Printing Office, Washington, D.C.
  31. Federman, D. D., Hanna, K. E., and Rodriguez, L. L. (2003), *Responsible Research: A Systems Approach to Protecting Research Participants*, National Academies Press, Washington, D.C., p. 109.
  32. Califf, R. M., Morse, M. A., Wittes, J., et al. (2003), Toward protecting the safety of participants in clinical trials, *Controlled Clin. Trials*, 24, 256–271.

# 16

---

## *Size of Clinical Trials*

**Jitendra Ganju**

*Amgen Inc., South San Francisco, California*

### **Contents**

|        |                                                                |     |
|--------|----------------------------------------------------------------|-----|
| 16.1   | Introduction                                                   | 913 |
| 16.1.1 | Challenges Faced in Practice                                   | 914 |
| 16.2   | Statistical Methods for Determining Trial Size                 | 917 |
| 16.2.1 | Difference in Means                                            | 918 |
| 16.2.2 | Differences in Means Adjusted for Baseline Data                | 920 |
| 16.2.3 | Differences in Proportions                                     | 920 |
| 16.2.4 | Time to Event                                                  | 921 |
| 16.3   | Realistic Assessment of Trial Size                             | 922 |
| 16.3.1 | Misspecification of Rates                                      | 922 |
| 16.3.2 | Misrepresentation of Estimate of Treatment Effect and Variance | 923 |
| 16.3.3 | Sensitivity to Projected Number of Events                      | 925 |
| 16.3.4 | Deviations from Assumptions and from Protocol Procedures       | 925 |
| 16.4   | Sample Size Reestimation                                       | 926 |
| 16.5   | Points to Consider When Planning Size of Trial                 | 928 |
| 16.5.1 | Noninferiority/Equivalence Trials                              | 928 |
| 16.5.2 | Unequal Allocation                                             | 929 |
| 16.6   | Conclusion                                                     | 929 |
|        | References                                                     | 930 |

### **16.1 INTRODUCTION**

On the face of it, the number of subjects required for a clinical trial is obtained from a mechanical application of a formula. A number representing the trial size pops out after presumed values of efficacy and variance, and specified levels of false-

positive and false-negative rates are plugged into the formula. However, arriving at a well-planned trial size is much more than a routine calculation. Because the derivations of formulas for different analysis methods are amply and adequately covered in the literature, they will be kept to a minimum here. The emphasis instead will be on the many challenges that arise when practical aspects of the size of a trial are considered. For purposes of keeping the focus on certain practical aspects, the chapter will be limited to randomized trials comparing two treatments and unless stated otherwise, a 1:1 allocation ratio.

Scientific, strategic, operational, and ethical deliberations exert a push and pull on the appropriate number of subjects needed for a trial. The tension this creates is described later in the section. Section 16.2 will give a limited overview of the technical details. In Sections 16.3 and 16.4 methods to avoid certain pitfalls will be described. Section 16.5 will briefly discuss noninferiority/equivalence testing and unequal allocation.

The minimum inputs to trial size (better known as sample size) assessment are the presumed values of efficacy and variability, the type I error (false-positive) rate or  $\alpha$  and the type II error (false-negative) rate or  $\beta$ . Power is  $1 - \beta$ , or in words, the probability of obtaining a statistically significant result assuming the presumed value of efficacy to be true. When power is held fixed, the number of subjects and  $\alpha$  have an inverse relationship. Determining the sample size at fixed power is equivalent to determining the power at fixed sample size. Typically, for confirmatory phase III trials, the one-sided  $\alpha$  is set at 0.025 and  $\beta$  at either 0.1 or 0.2. The fact that  $\alpha$  is chosen to be less than  $\beta$  is a reflection of how we, the clinical trials community, including regulatory bodies such as the U.S. Food and Drug Administration (FDA), are more concerned about the risk of declaring an ineffective treatment effective than we are of the risk of failing to declare an effective treatment effective. We have agreed to reach our conclusions by risking on average, for every 40 trials, no more than 1 false positive or no more than 4–8 false negatives.

The presumed value of efficacy used in sample size calculations is stated under the alternative hypothesis for the primary variable (also referred to as the response or outcome variable or endpoint). Its complement, the null hypothesis, states that the presumed value of efficacy is 0—that is, the efficacy of the experimental treatment is no different from control. (But see noninferiority/equivalence testing in Section 16.5 for different formulations of the null and alternative hypotheses.) For example, for a trial comparing cholesterol-lowering drugs, the null hypothesis may state that the average low-density lipoprotein, LDL or “bad cholesterol,” in the two groups is the same, and the alternative may state that the experimental treatment, compared to the control, lowers the average LDL by 15 mg/dL.

We now get into how the scientific, strategic, operational, and ethical matters exert a push and pull on how many subjects are appropriate for addressing the trial’s objectives.

### 16.1.1 Challenges Faced in Practice

**Scientific** Let  $\Delta$  denote the presumed value of efficacy. (For our cholesterol example, we would write  $\Delta = 15$  mg/dL.) Since the sample size formula relies heavily on the value specified for efficacy (the sample size is a function of  $\Delta^2$ ), it is important



to investigate where  $\Delta$  came from. In assessing its reliability, the following sorts of questions need to be investigated.

1. Did  $\Delta$  come from an earlier trial conducted by the same institution/sponsor? If so, how is the trial being planned different from the earlier one? Was the earlier one a small-sized trial conducted in a few centers? Will the upcoming trial include a more heterogeneous population? Are there differences in the lengths of follow-up between the two trials? Is the control group the same?
2. If  $\Delta$  was provided to a statistician by a clinician, what was their source of information?
3. Was  $\Delta$  inferred from a published article that reported results on a recently approved treatment with a similar mechanism of action?
4. Could medical advances in the field influence the magnitude of  $\Delta$  estimated from earlier trials? For example, organ transplant patients receive immunosuppressants to weaken the body's immune system to prevent it from attacking the transplanted organ. Advances in surgical methods for organ transplant may influence  $\Delta$  independent of the drug effect.
5. If  $\Delta$  was obtained from a peer-reviewed journal article, then taking that  $\Delta$  at face value will likely be an optimistic value. Positive trials tend to get published, whereas negative trials tend not to. If we had results from positive and negative trials, then our presumed value of efficacy would weigh the evidence from all trials. Overlooking the fact that negative results get censored leads to an overly favorable  $\Delta$ .

All things considered, if  $\Delta$  seems optimistic, it would be advisable to err on the side of being conservative. Even in the simple case of  $\Delta$  defined as the difference in means, the statistician, if only as a watchdog, has an important role to play in checking whether the disparate bits of information contribute a well-inferred assessment of the presumed value of efficacy.

If the primary variable is instead time to some event (e.g., time to heart attack in our cholesterol-lowering example), then other challenges are present. Here  $\Delta$  is the ratio of hazards, a concept generally not readily understood by nonstatisticians when adjustments to it are made depending on the subject accrual rate, the follow-up duration, nonproportional hazards, censoring due to loss to follow-up, and competing risks. For such an endpoint the statistician needs to be even more involved in carefully projecting the number of events and in ensuring that  $\Delta$  has been carefully selected.

Some of the same questions also arise when choosing a value for the variance. Let  $s^2$  denote the estimate of variability to be used in sample size calculations that is associated with the estimate of  $\Delta$ . Often in a covariate-adjusted analysis  $s^2$  is obtained by conditioning on marginal totals (e.g., the number of subjects per center). The estimate is valid only when there is no interaction between treatment and the stratification factor. When interaction exists, the estimated treatment effect is biased and the estimated variance is an underestimate (Section 16.3).

Additionally,  $s^2$  obtained from an earlier trial does not represent, in the statistical repeated sampling sense, an estimate of variability for the upcoming trial. No two trials are ever conducted under identical conditions subject only to random variation. For instance, the inclusion/exclusion criteria or the centers may be different or

the definition of the disease may have been refined over time. It is more appropriate, therefore, to refer to  $s^2$  as a “guesstimate” of the true variance. The estimate–guesstimate delineation is a reminder that the upcoming trial will not be identical, except for random variation, to the completed one from which  $s^2$  was calculated. This realization prepares us for questioning or challenging our assumptions while the trial is ongoing and possibly even reestimating the sample size from interim trial data.

Values of  $\Delta$  and  $s^2$  that get plugged into the sample size formula often assume compliance to the protocol. But the assumption of compliance itself is one to guard against. As Efron [1] notes: “There could not be worse experimental animals on earth than human beings; they complain, they go on vacations, they take things they are not supposed to take, they lead incredibly complicated lives, and, sometimes, they do not take their medicine.” Too often little attention is paid at the *planning* stage about the consequences of protocol deviations on the power of the trial. Standard sample size formulas do not automatically adjust for imperfect compliance. When subjects prematurely terminate from the trial, their efficacy assessment at the final visit will be missing. If the intention is to analyze per the intent-to-treat principle—a principle that mandates inclusion of all randomized subjects—then the results depend on the imputation method. Careful planning of sample size should acknowledge the pervasiveness of missing data, make realistic projections on the proportion likely to be missing, and calculate the sample size consistent with the proposed imputation method.

Other challenges to sample size calculation arise when subjects midstream change their trial conduct. In blinded trials with long-term follow-up, some subjects may modify their behavior after they become unblinded to the blinded treatment. Unblinding can occur when subjects are able to infer the treatment administered by monitoring the trajectories of their outcomes. Upon becoming unblinded, subjects may then, for example, start taking prohibited study medications, thus confounding the causal assessment of efficacy. Adjusting for such a deviation in the sample size is difficult. The combination of even slightly unrealistic values for  $\Delta$  and  $s^2$ , and deviations from protocol procedures can drastically impact the chance of achieving the trial’s objectives. Section 16.3 discusses this in more detail.

**Strategic** The size of a trial is generally governed by its primary objective. Sometimes, however, trialists have strategic objectives in mind that go beyond the trial’s primary objective. These concern finding expeditious ways to license a drug or terminate the trial early for futility. This situation is best described with some examples. (1) Suppose a sponsor is initiating its phase II trial around the time news on the approval status of its competitor’s drug is forthcoming. The sponsor adopts the following strategy. If the competitor receives approval, then the sponsor will amend the protocol to add another cohort of subjects who would be randomized to the competitor’s newly approved drug or to the sponsor’s experimental treatment. Results from the additional cohort would facilitate planning of the sponsor’s phase III trials by giving them an “early read” of the efficacy of their experimental treatment relative to the newly approved one. If the competitor does not receive approval, then the trial size remains unchanged. (2) In a phase II trial with long-term follow-up, if enrollment turns out to be much slower than expected, the sponsor may no longer wish to wait for trial completion. Instead the sponsor may decide to test for superiority (or noninferiority) on a surrogate endpoint that is measured before the primary clinical endpoint, while continuing enrollment. If superiority (or noninferi-

ority) is established, then enrollment and follow-up for evaluation of the clinical endpoint continues, otherwise the trial terminates for futility. Because the trial may be underpowered with interim data, the sponsor may choose to lower the threshold for demonstration of its objective on the surrogate endpoint.

Such strategic objectives can influence the trial's size, but they may not get written into the trial protocol, particularly when the sponsor's thinking is still evolving at the time of initiating the trial.

**Operational** Trials can be very expensive, especially late-phase trials. The National Heart, Lung, Blood Institute–funded Coronary Drug Project Trial, designed to assess lipid-influencing regimens, randomized 8341 subjects and was reported to cost \$42 million [2], amounting to an average cost of approximately \$5000 per subject. In 2010 this amount is about \$11,000 per subject after adjusting for inflation, a cost by no means out of the ordinary for a typical trial. More expensive trials, particularly those sponsored by nascent pharmaceutical or biotechnology companies, may simply not get off the ground. (The possibility of joint sponsorship with institutions or larger companies is not always viable.) The solution, however, is not necessarily to reduce drastically the trial size because that would also lower the chance of success. Indeed, the failure to demonstrate success has been attributed to an inadequate number of subjects [3]. As aptly put by Lachin [4]: “Clinical trials with inadequate sample size are thus doomed to failure before they begin and serve only to confuse the issue of determining the most effective therapy for a given condition.” A trial that has failed to demonstrate efficacy because it was not sufficiently sized, may simply not get repeated, risking termination of not just the trial but the development of a potentially important therapy. Innovative alternatives to reducing the trial cost need to be considered without substantially compromising statistical power.

**Ethical** If trials could be done on plots of land, many of the challenges noted above would not arise. There would be full adherence to the protocol and ample data, none of it missing. But ethical concerns with human participation bring the size of trials into glare. Trials with mortality or irreversible morbidity as the primary variable lend themselves to extra scrutiny if they are too small or too large. Small-sized trials have a low probability of success and a high probability of equivocal results, whereas excessively large trials expose more subjects than necessary to at least one suboptimal or detrimental therapy. Ethical concerns sharply bring to focus the point that subjects are, after all, volunteers who provide data to help trialists decide fairly conclusively whether or not the experimental treatment (a) can proceed to the next stage for further testing or (b) is safe and effective for licensure.

## 16.2 STATISTICAL METHODS FOR DETERMINING TRIAL SIZE

In this section an outline of the methods for calculating the sample size for different analysis methods will be described. The methods will be described in some detail for differences in means and proportions, but sketched for the time-to-event endpoint. The level of detail in this section is intended to provide a basic understanding. For an in-depth understanding the references in this section and the next should be consulted. The excellent paper by Lachin [4] belongs to the “must read” category.

### 16.2.1 Difference in Means

Let  $Z_\alpha$  and  $Z_\beta$  denote the  $\alpha$  and  $\beta$  percentiles of the normal distribution, respectively. Power is mathematically expressed as

$$1 - \beta = P(Z > Z_{1-\alpha} | \Delta) \quad (1)$$

Un-normalizing the  $Z$  test gives

$$1 - \beta = P\left(\frac{\hat{\Delta}}{\sqrt{\text{Var}(\hat{\Delta})}} > Z_{1-\alpha} | \Delta\right) \quad (2)$$

where  $\hat{\Delta}$  is the estimated treatment effect and where  $\text{Var}(\hat{\Delta}) = 4\sigma^2/N$ , with  $\sigma^2$  denoting the variance of each observation and  $N$  denoting the *total* trial size. The above equation simplifies to

$$1 - \beta = P\left(Z > \frac{Z_{1-\alpha}\sqrt{\text{Var}(\hat{\Delta})} - \Delta}{\sqrt{\text{Var}(\hat{\Delta})}}\right) \quad (3)$$

If the two treatments are identical (i.e.,  $\Delta = 0$ ), then regardless of the size of the trial, power equals the false-positive rate  $\alpha$ . From (3)

$$Z_\beta = \frac{Z_{1-\alpha}\sqrt{\text{Var}(\hat{\Delta})} - \Delta}{\sqrt{\text{Var}(\hat{\Delta})}}$$

We almost always choose power to exceed 50%, so  $Z_\beta$  is negative. Noting that  $-Z_\beta = Z_{1-\beta}$  we solve for  $N$  to get the total sample size:

$$N = \frac{4\sigma^2 (Z_{1-\alpha} + Z_{1-\beta})^2}{\Delta^2} \quad (4)$$

For  $\alpha = 0.025$  it is simpler to remember the total sample size as  $N = 42/(\text{signal-to-noise ratio})^2$  for 90% power or  $N = 32/(\text{signal-to-noise ratio})^2$  for 80% power where  $\Delta/\sigma$  denotes the “signal-to-noise ratio.” When testing on the difference scale a trial with 90% power requires about 30% more subjects than a trial with 80% power (and  $\alpha = 0.025$ ).

The sample size calculation relies on the  $Z$  test, which assumes that the variance is known. The data are analyzed, however, using the  $t$  test instead, which is like the  $Z$  test, except that the variance is estimated from the data. There is very little loss in power caused by the discrepancy in using one statistic ( $Z$  test) for calculating power and another ( $t$  test) for analyzing data. The  $Z$  and  $t$  tests make assumptions that (a) the variances associated with each treatment are equal, and (b) the data are normally distributed. Do violations of any, or both, of these assumptions make the sample size formula unreliable? There may be some loss in power due to violations of assumptions but (4) is still reliable. The  $t$  test is remarkably robust to violations in assumptions. Posten, Yeh and Owen [5] study robustness when the equal variance

assumption is violated. A powerful conclusion they reach in the case of normal data and a sample size of 15 per group is that “no matter how much  $\sigma_1^2$  varies from  $\sigma_2^2$ , the true significance level [of the  $t$  test] is within 0.01 units of the desired level  $\alpha = 0.05$ .” Heeren and D’Agostino [6] demonstrate the robustness of the  $t$  test for data on 3, 4, or 5-point ordinal scales. Aside from invoking normality to analytically calculate power, power may also be empirically calculated by simulating data under the alternative for the type of nonnormal data at hand (e.g., ordinal or skewed data).

Some investigators [7] have argued that the sample size calculated by the above method is “really large.” They say that this method “permits an illogical interpretation that tempts investigators and plagues readers. The large sample size can lead one to declare an unimportant difference important.” What they mean by this is that an unimportant but statistically significant difference in means may be interpreted as clinically meaningful. It would be inappropriate to conclude, as the authors assume readers are wont to do, that a statistically significant result is clinically meaningful. They instead advocate the following sample size that was proposed by Feinstein [8]:

$$N \geq \frac{4\sigma^2 Z_{1-\alpha}^2}{\Delta^2} \quad (5)$$

According to them power is not the issue—instead one needs to pick an “intellectually honest”  $\Delta$  before data collection and then draw inferences by reference to  $\Delta$  and 0. The following explains their recommendation (wherein positive values of  $\Delta$  signify improvement, LL and UL denote the lower and upper confidence limits around  $\hat{\Delta}$ ):

If  $\hat{\Delta} > \Delta$  and  $LL > 0$ , conclude an important difference.

If  $\hat{\Delta} < \Delta$  and  $LL < 0$ , concede the null.

If  $\hat{\Delta} < \Delta$  but  $LL > 0$  and  $UL > \Delta$ , conclude an important difference may exist.

If  $LL > 0$  and  $UL < \Delta$ , conclude result is unimportant even though statistically significant.

Their approach, however, is not feasible. First, as an inequality (5) does not indicate how large  $N$  should be relative to  $4\sigma^2 Z_{1-\alpha}^2 / \Delta^2$ . Second, power is paramount, although the authors do not believe in it. In the end it is necessary to know what chance of success a sponsor has of meeting the trial’s objectives. For example, with  $\Delta = \sigma = 1$ , an  $N$  equal to  $4\sigma^2 Z_{1-\alpha}^2 / \Delta^2$  would require 16 subjects ( $\alpha = 0.025$ ). The power of declaring that  $\hat{\Delta}$  is statistically significantly different from 0 is approximately 50%; the power of declaring that an important difference exists (as they have defined it) would be even less. Why would a sponsor undertake a trial that has less than a 50% chance of finding an important difference? Third, presupposing an “intellectually honest”  $\Delta$  belies the reality that proposing a carefully thought-through  $\Delta$  may involve the weighing of evidence from fragmentary and conflicting data. It would be unrealistic to expect a sponsor, a regulatory agency, or a committee advising the agency to all agree on the same value of  $\Delta$ . If clinical rather than statistical significance is of interest, then it would be better to power the trial so that the LL

exceeds the clinically important difference. However, the power required for a clinically important difference would result in an even *larger* sample size than (4).

### 16.2.2 Differences in Means Adjusted for Baseline Data

When baseline data on the same endpoint are available, variability can be reduced by taking the change from baseline or by fitting an analysis of covariance (ANCOVA) model. Under standard assumptions ANCOVA improves variability by nearly  $1 - r^2$  [9] and  $0.5(1 + r)$  [10] compared to the unadjusted and change from baseline models, respectively, where  $r$ , which takes values between  $-1$  and  $1$  [11], denotes the correlation of the baseline and posttreatment data. (For pre/post data,  $r$  usually takes values between  $0$  and  $1$ .) The sample size for the ANCOVA model is obtained as  $N_{\text{adj}} = N(1 - r^2)$  where  $N$  is as shown in (4). Since ANCOVA results in a smaller variance than change from baseline, there is no reason to continue to use change from baseline as the endpoint.

### 16.2.3 Difference in Proportions

A simple method for calculating the sample size for comparing the difference in proportions makes use of the fact that the proportions are approximately normally distributed. The derivation proceeds along the same lines as for the difference in means with one important distinction explained below.

The null and alternative hypotheses are:  $H_0: p_1 - p_2 = 0$  and  $H_1: p_1 - p_2 = \Delta$ , where  $p_1$  and  $p_2$  denote population proportions. The estimate of the treatment effect is  $\hat{\Delta} = \hat{p}_1 - \hat{p}_2$ . The important distinction is that there are two formulas for the variance of  $\hat{\Delta}$  that feature in the sample size formula: one under the null hypothesis and the other under the alternative. We start by rewriting (2) as

$$1 - \beta = P\left(\frac{\hat{\Delta}}{\sqrt{\text{Var}_{H_0}(\hat{\Delta})}} > Z_{1-\alpha} | \Delta\right)$$

where the variance under the null is  $\text{Var}_{H_0}(\hat{\Delta}) = (4/N) \bar{p}(1 - \bar{p})$  and  $\bar{p} = (p_1 + p_2)/2$  is the average of the proportions specified under the alternative. Recognizing that power is calculated under the alternative, this can be reexpressed as

$$1 - \beta = P\left(Z > \frac{Z_{1-\alpha} \sqrt{\text{Var}_{H_0}(\hat{\Delta})} - \Delta}{\sqrt{\text{Var}_{H_1}(\hat{\Delta})}}\right) \quad (6)$$

where the variance under the alternative is

$$\text{Var}_{H_1}(\hat{\Delta}) = \frac{2}{N} \{p_1(1 - p_1) + p_2(1 - p_2)\}$$

The total sample size derived from (6) is

$$N = \frac{[Z_{1-\alpha} \sqrt{4\bar{p}(1 - \bar{p})} + Z_{1-\beta} \sqrt{2p_1(1 - p_1) + 2p_2(1 - p_2)}]^2}{\Delta^2} \quad (7)$$

Noting that  $\text{Var}_{H_0}(\hat{\Delta}) \geq \text{Var}_{H_1}(\hat{\Delta})$  a more conservative sample size formula is obtained by replacing  $\text{Var}_{H_1}(\hat{\Delta})$  in (6) with  $\text{Var}_{H_0}(\hat{\Delta})$  giving

$$N = \frac{4\bar{p}(1-\bar{p})(Z_{1-\alpha} + Z_{1-\beta})^2}{\Delta^2} \quad (8)$$

Formulas (7) and (8) yield approximate but not exact power. This is because the difference in proportions has a discrete distribution that is only approximately normal. Adjustments have been proposed so that power of the test gets closer to the stated value of  $1 - \beta$  [12]. The formula implemented in the sample size software nQuery Advisor [13] is the one proposed by Fleiss, Tytun, and Ury [14]; their formula is

$$N' = N + \frac{4}{|p_1 - p_2|} \quad (9)$$

where  $N$  is shown in (7).

When either proportion is expected to be very small ( $\leq 0.05$ ) or very large ( $\geq 0.95$ ), Fisher's or Barnard's [15] exact test is used for analysis because tests that rely on the normal approximation are less reliable. The sample size calculation should then reflect the power of the exact test and not one based on the normal approximation. It is beyond the level of technical detail intended for this chapter to describe the exact methods, but the interested reader is referred to [16]. Gordon [17] provides an excellent review of the sample size methods for comparing proportions.

#### 16.2.4 Time to Event

Instead of analyzing data on a binary scale, the time to the occurrence of the event (i.e., time to event) is often a more powerful analysis [18]. The sample size calculation is sketched for one of the more commonly used time-to-event test statistics, the log-rank test [19].

The concept of a hazard is central in time to event analysis. Loosely put, the hazard at time  $t$  is the instantaneous risk that the event will occur at time  $t$ . A large portion of the sample size literature rests on the assumption that the hazard of an event at any given time for one treatment group is proportional to the hazard at that time for the other treatment group. The assumption of proportional hazards enables the derivation of the sample size formula for the log-rank statistic, although once the data become available, the log-rank test remains valid even if that assumption is unmet.

The null and alternative hypotheses are:  $H_0: \lambda_1(t)/\lambda_2(t) = 1$  and  $H_1: \lambda_1(t)/\lambda_2(t) = \Delta$ , where  $\lambda(t)$  denotes the hazard of an event at time  $t$  for an individual. On a log scale we test  $H_0: \log_e \lambda_1(t) - \log_e \lambda_2(t) = 0$  against  $H_1: \log_e \lambda_1(t) - \log_e \lambda_2(t) = \log_e \Delta$ . For large samples, the log-rank statistic has an approximate normal distribution with mean  $\log_e \Delta$  and, under a 1:1 allocation ratio, variance  $4/d$ . Following the same steps in this section in going from Equations (1)–(3), we see that the variance of the log-rank statistic can also be written as  $(Z_{1-\alpha} + Z_{1-\beta})^2/(\log_e \Delta)^2$ . Therefore, the



number of events required to achieve power  $1 - \beta$  at a type I error rate of  $\alpha$  is  $d = [4(Z_{1-\alpha} + Z_{1-\beta})^2] / (\log_e \Delta)^2$ .

The required number of subjects is obtained from  $N = d/P(\text{event})$  where  $P(\text{event})$  is the probability of the event and is a function of efficacy, the accrual rate, and follow-up period. For randomized trials  $P(\text{event})$  is the average of  $P(\text{event in group 1})$  and  $P(\text{event in group 2})$ . To obtain this probability for each treatment group, let  $a$  denote the accrual period,  $f$  denote the follow-up period going from the end of accrual to a fixed period (in the same time unit), and assume that enrollment occurs at a uniform rate. Then it can be shown that  $P(\text{event in group } i)$  equals

$$1 - 1/6 \{P_i(T > f) + 4P_i(T > 0.5a + f) + P_i(T > a + f)\} \quad (10)$$

for  $i = 1, 2$  [20]. In the equation above,  $T$  is a random variable denoting the time to event, and the notation  $P_i(T > x)$  denotes the probability that the time to event in group  $i$  exceeds  $x$ . When enrollment is simultaneous,  $a = 0$ , and  $P(\text{event in group } i)$  simplifies to  $1 - P_i(T > f)$ . When enrollment is staggered  $P_i(T > f)$ ,  $P_i(T > 0.5a + f)$ , and  $P_i(T > a + f)$  are all decreased as the accrual period increases, and (10) shows how this increases the probability of the occurrence of an event.

Often a terminal event such as death or irreversible morbidity is the “event” in time-to-event trials. (Such trials also get the most media attention.) It is therefore important to understand the concept of hazards, the derivation of the formula, and the assumptions that the formula depends upon beyond the silhouette presented above. A highly accessible account of the sample size derivation is in Collett [20]. Other useful references include George and Desu [21], Schoenfeld [22], and Lachin and Foulkes [23]. A novel method to calculate the sample size that does not make any of the restrictive assumptions is discussed in Section 16.3.

## 16.3 REALISTIC ASSESSMENT OF TRIAL SIZE

PhRMA (the Pharmaceutical Research and Manufacturers of America) says that of drugs completing Phase II trials, about 50% fail in Phase III, often for lack of efficacy (quoted in [24]). But Temple [24] points out that phase II trials are meant to demonstrate efficacy, so it is startling that the phase III failure rate is as high as 50%. It may well be that when sizing phase III trials misjudgments in  $\Delta$  or  $s^2$  contribute to the high failure rate. This section will discuss the sensitivity of the sample size to various assumptions regarding  $\Delta$ ,  $s^2$ , and protocol compliance.

### 16.3.1 Misspecification of Rates

A simple hypothetical example is presented to help understand better why the sample size (or, equivalently, power) is so sensitive to the initial assumptions. Table 1 displays the sample sizes for 80 and 90% power for testing  $H_0: p_1 - p_2 = 0$  against  $H_1: p_1 - p_2 = 0.07$ . As explained in Section 16.2, to calculate the sample size or power we also need the presumed population proportions. Suppose we believe that  $p_1 = 0.12$  and  $p_2 = 0.05$  are the true population proportions. Table 1 shows that we need a total of 552 subjects to have 80% power and 720 subjects for 90% power. If in reality the population proportions are instead  $p_1 = 0.13$  and  $p_2 = 0.06$ , a slight change from our

**TABLE 1 Sample Sizes under Various Assumptions ( $\alpha = 0.025$ )**

|                   |      |      |      |      |
|-------------------|------|------|------|------|
| $p_1$             | 0.12 | 0.13 | 0.14 | 0.13 |
| $p_2$             | 0.05 | 0.06 | 0.07 | 0.07 |
| $p_1 - p_2$       | 0.07 | 0.07 | 0.07 | 0.06 |
| $N$ (power = 80%) | 552  | 606  | 656  | 848  |
| $N$ (power = 90%) | 720  | 790  | 860  | 1114 |

Note: Samples sizes obtained from nQuery Advisor Version 5.0 or equivalently from (9).

**TABLE 2 Outcome Is Measure of Neurological Functioning**

| Stratum | Social Class and Gender | Treatment A |             |                  | Treatment 2 |             |                  |
|---------|-------------------------|-------------|-------------|------------------|-------------|-------------|------------------|
|         |                         | $n_A$       | $\bar{y}_A$ | $\hat{\sigma}_A$ | $n_B$       | $\bar{y}_B$ | $\hat{\sigma}_B$ |
| 1       | Low, female             | 41          | 1.38        | 0.22             | 40          | 1.36        | 0.28             |
| 2       | Low, male               | 41          | 1.26        | 0.25             | 38          | 1.28        | 0.19             |
| 3       | Medium, female          | 33          | 1.51        | 0.31             | 35          | 1.41        | 0.27             |
| 4       | Medium, male            | 45          | 1.46        | 0.28             | 46          | 1.39        | 0.33             |
| 5       | High, female            | 18          | 1.61        | 0.34             | 20          | 1.51        | 0.41             |
| 6       | High, male              | 23          | 1.59        | 0.46             | 23          | 1.44        | 0.30             |

Source: From Fleiss [25].

assumptions although the difference of 0.07 is maintained, we need 606 and 790 subjects for 80 and 90% power, respectively. The sample size increases even though  $\Delta$  is unchanged because  $\text{Var}(\hat{\Delta})$  increases. The next column shows a similar increase in sample sizes (656 and 860 subjects for 80 and 90% power) for a similar increase in the proportions. The last column shows how significant the impact can be if both the proportions and the difference in proportions are slightly off target. For  $p_1 = 0.13$  and  $p_2 = 0.07$  we need 848 and 1114 subjects, respectively, for 80 and 90%. (Note that in each case the sample size for 90% power is roughly 30% more than the sample size for 80% power.)

### 16.3.2 Misrepresentation of Estimate of Treatment Effect and Variance

Consider the data in Table 2 on 403 subjects displayed by social class and gender for two treatments [25]. We assume the data were obtained from a trial that employed the method of simple (permuted block) randomization. Further assume that the design of the upcoming trial is similar to the completed one. When data become available for the upcoming trial, the analysis will adjust for the six stratification factors to reduce variability. Although the data in Table 2 suggest that there is no evidence of treatment by stratum (social class or gender) interaction, we will assume for the sake of illustration that such an interaction exists. We use this information to estimate the variability of  $\hat{\Delta}$ .

It is common to estimate the treatment effect as  $\hat{\Delta} \sum_{i=1}^6 w_i \hat{\Delta}_i / (\sum_{i=1}^6 w_i)$  where  $\hat{\Delta}_i$  is the estimated effect for the  $i$ th stratification level, and where  $w_i$  is a weight that is a function of the number of subjects in each cell in the  $i$ th stratum. For example, for  $i = 1$ ,  $\hat{\Delta}_1 = 1.38 - 1.36 \equiv 0.2$  and  $w_1 = 41 \times 40 / (41 + 40) \equiv 20.25$ . When interaction exists, which is another way of saying that  $\Delta_i$ 's are not all equal, then  $\hat{\Delta}$

is biased [26]. The reason  $\hat{\Delta}$  is biased is because the weights,  $w_i$ 's, are treated as fixed constants, whereas according to the trial design the weights are random. (If randomization was stratified within each stratum, then  $\hat{\Delta}$  is unbiased [27]). Moreover, the pooled estimate of variance is as an underestimate. The pooled variance estimate,  $s^2$ , is calculated as a weighted average of the variance in each cell in Table 2:

$$\left\{ \frac{(41-1) \times 0.22^2 + (40-1) \times 0.28^2 + \dots}{+(23-1) \times 0.30^2} \right\} / (40 + 39 + \dots + 22) = 0.0883$$

Typically, this is the value of  $s^2$  that gets plugged in (4) to calculate the sample size. This variance is known as the conditional variance where the condition is the observed number of subjects in each of the 12 cells. Because  $\hat{\Delta} = \sum_{i=1}^6 w_i \hat{\Delta}_i / (\sum_{i=1}^6 w_i)$  is biased when interaction exists, the unweighted and unbiased estimate of the treatment effect may be preferred. Accordingly, we need to use the appropriate variance formula. It has been shown that [26] the unconditional variance is

$$\text{Var}(\hat{\Delta}) \approx \frac{4}{N} \left( \sigma^2 + \frac{K}{2} \right)$$

where  $K = \sum_i \pi_i \{(\mu_{Ai} - \mu_A)^2 + (\mu_{Bi} - \mu_B)^2\}$ ;  $\pi_i$  denotes the proportion of subjects who belong to the  $i$ th stratification level ( $i = 1, 2, \dots, 6$ ),  $\mu_{Ai}$  and  $\mu_{Bi}$  denote the population means of treatments A and B for the  $i$ th strata with  $\mu_A = \sum \pi_i \mu_{Ai}$  and  $\mu_B = \sum \pi_i \mu_{Bi}$  denoting the population means. The unconditional variance adds  $K/2$  to the pooled variance  $\sigma^2$  to free the conditional variance from requiring inferences limited to fixed cell sizes. Just as the data in Table 2 provided an estimate of  $\sigma^2$ , so too we use the data to estimate  $\pi_i$ ,  $\mu_{Ai}$ ,  $\mu_{Bi}$ ,  $\mu_A$ , and  $\mu_B$ . For example, the estimate of the proportion of subjects who belong to the "medium, female" level,  $\pi_3$ , is  $68/403 = 0.17$  and the estimates of  $\mu_{A3}$  and  $\mu_{B3}$  are 1.51 and 1.41, respectively. Substituting estimates for population parameters, we get  $K/2 = 0.0065$ , so the pooled estimate of variance goes from 0.0883 to  $0.0883 + 0.0065 = 0.0968$ .

The sample size formula corresponding to the unconditional variance becomes

$$N = \frac{(4\sigma^2 + 2K)(Z_{1-\alpha} + Z_{1-\beta})^2}{\Delta^2} \quad (11)$$

Using the conventional sample size formula (4) gives a total sample size  $N$  of 772 subjects to achieve 80% power, whereas (11) requires 964 subjects to achieve the same power, an increase of 25%. Put another way, the power that will be achieved with 772 subjects is 70% and not, as claimed, 80%. The tradition of using an unconditional variance for sample size is common in survey sampling but not, however, in the clinical trials literature. [Because the population proportions are known in survey sampling, the unconditional variance expression is different from (11)]. For survey sampling the unconditional variance has been recommended by none other than Deming [28] who says that it is the "formula which one will use at the planning stage."

### 16.3.3 Sensitivity to Projected Number of Events

A large vaccines trial was designed to reduce the burden of illness due to herpes zoster in a population at least 60 years of age with a minimum of 6 months of follow-up [29]. Approximately 38,000 subjects included in the analysis received either an investigational vaccine or placebo, in a 1:1, randomized, double-blind manner. Although the primary analysis variable for the trial was the relative reduction in the herpes zoster burden of illness score, we will consider the relative reduction in the proportion of subjects with herpes zoster.

There were a total of 957 confirmed cases of herpes zoster included in the analysis, 315 among vaccine recipients and 642 among placebo recipients. The relative risk is 0.49 and the 95% confidence interval (CI) is (0.43, 0.56). Because the interval excludes 1, the result is statistically significant. (In fact, since the upper limit is much below 1, the investigational vaccine substantially reduced the rate of herpes zoster.) Table 3 shows the observed result in the first row and hypothetical ones in the next two rows. Comparing the first and second rows, we see that width of the CI is the same even though  $N$  in the second row is less by 28,000 subjects. The massive reduction in sample size has no effect on the width of the interval because the variance of ratios of estimates of event rates is driven by the number of events (and only trivially by the total sample size). Comparing the second row with the third, we see that  $N$  is the same but the number of cases are halved and that this only slightly increases the width of the interval. Although the reduction in number of cases from 957 to 450 is large, the result is, in effect, unchanged. This is because the relative risk  $Z$  values for both are very large, raising questions whether the trial was overpowered. An interim analysis allowing for early termination upon demonstration of a significant reduction in herpes zoster cases may have markedly reduced the trial size. Indeed, if mortality were the outcome variable, an interim analysis would be mandatory so that the trial not enroll more subjects than necessary.

### 16.3.4 Deviations from Assumptions and from Protocol Procedures

Here we consider sample size assessment when the outcome variable is binary or time to some event and when there is loss to follow-up, noncompliance, “drop-in,” nonproportional, or nonconstant hazards. (Drop-in refers to subjects changing their randomized treatment in violation of protocol procedures.) For binary outcomes Lakatos [30] has presented a method for calculating the sample size under such realistic conditions. His method uses a Markov model for adjusting the proportions,  $p_1$  and  $p_2$ , for losses due to noncompliance, drop-in, and so forth. Each treatment group is modeled separately. In Lakatos’ Markov model, a transition matrix is

**TABLE 3** Number of Cases of Herpes Zoster,<sup>a</sup> Relative Risk Estimates and 95% CI

| Sample Size        | Vaccine | Placebo | Relative Risk | 95% CI       |
|--------------------|---------|---------|---------------|--------------|
| $N \approx 38,000$ | 315     | 642     | 0.49          | (0.43, 0.56) |
| $N^* = 10,000$     | 315     | 642     | 0.49          | (0.43, 0.56) |
| $N^* = 10,000$     | 150     | 300     | 0.50          | (0.41, 0.61) |

*Note:* The number randomized to each group is roughly  $N/2$ . The first row shows real data, the next two rows (indicated by  $N^*$ ) show hypothetical data.

<sup>a</sup>From [29].

created for each time interval (the length of the time interval is user specified). The rows and columns of the transition matrices are states subjects belong to and will transition to with probabilities that are also specified by the user. Typical examples of states that subjects belong to include: a loss to follow-up state (no further information available for that subject), an event state (subject has had the event), an at-risk state for a subject who is a complier (subject is at risk of experiencing the event assuming compliance), and an at-risk state for a subject who is a noncomplier (subject is at risk of experiencing the event assuming noncompliance). Needless to say, subjects would transition from at-risk states to the same or other states, but not from a loss to follow-up or an event state to an at-risk state. At the beginning of the trial the probability of belonging to the at-risk state is 1 for compliers and is 0 for all other states. At the end of the trial the probability of belonging to any particular state is the Markov process, obtained as the product of the transition matrices. The adjusted proportions,  $p_1$  and  $p_2$ , obtained from the Markov process for the end of the trial get plugged in the sample size formula.

Lakatos compares the sample size for the SHEP (systolic hypertension in the elderly program) trial calculated the traditional unadjusted way and upon application of the Markov model. The outcome variable for the trial was fatal or nonfatal stroke. Subjects were to be followed for a minimum of 4 years. The experimental treatment was assumed to lower the rate of stroke relative to control by 40%. Without adjusting for losses, noncompliance, and drop-ins the control and treatment rates were assumed to be 0.0775 and 0.0471, respectively. Application of (9) results in a sample size of 2784 to achieve 90% power ( $\alpha = 0.025$ ). After making reasonable adjustments for noncompliance, losses to follow-up, and the like, the Markov model gave rates of 0.0677 and 0.0463. With these rates, the required sample size is 5116.

Lakatos later [31] extended his method for the log-rank statistic enabling calculation of the sample size under unrestrictive conditions. Unlike other methods in the literature, Lakatos' method does not rely on the assumption of proportional hazards. Indeed, the advantage of this method is that it can accommodate any arbitrary pattern of data projected for the future. The difference in sample sizes calculated with and without adjustments for loss to follow-up, noncompliance, and drop-in for nonproportional hazards data is striking. For example, in Table 2 of Lakatos [30], assuming constant recruitment the sample size to achieve 90% power is 4397 ( $\alpha = 0.025$ ). For the same constant recruitment assumption but adjusting for a lag effect (i.e., a certain type of nonproportional hazard), modest loss to follow-up, noncompliance, and drop-in rates the required sample size is 8009.

These examples exemplify the point that trialists should pay careful consideration at the planning stage of the impact on power when rates are misspecified or when protocol compliance is less than perfect.

## 16.4 SAMPLE SIZE REESTIMATION

Methods have been proposed to estimate the variance while the blinded trial is still ongoing to allow for reestimation of the sample size. One option is to break the treatment code after data on some fraction of subjects has become available

and estimate the variance. However, because this step unblinds the trial, it has to be done with much forethought. One has to be concerned about protecting the integrity of trial results at its scheduled termination due to interim unblinding. Here we discuss a method proposed for continuous outcome variables that does not require breaking the blind [32, 33]. To apply the method it suffices to know the enrollment order and randomization block size. The method is simple to apply and works reasonably well for small randomization block sizes when compared to its unblinded counterpart. It works as follows. At an interim stage we have data on the outcome variable for, say,  $\tilde{N}$  subjects. Suppose the randomization block size is  $n$  and there are  $k$  blocks, so that  $\tilde{N} = nk$ . Denote the data on the outcome variable as

$$Y_{11}, Y_{21}, \dots, Y_{n1} | Y_{12}, Y_{22}, \dots, Y_{n2} | \dots | Y_{1k}, Y_{2k}, \dots, Y_{nk}$$

The data structure written this way means that of the first set of observations,  $Y_{11}, Y_{21}, \dots, Y_{n1}$ , half of those belong to one treatment group and the other half to the other treatment group. Of the second set of  $n$  observations,  $Y_{12}, Y_{22}, \dots, Y_{n2}$ , half belong to one group and half belong to the other, and so on until the last set  $Y_{1k}, Y_{2k}, \dots, Y_{nk}$ .

Take the sum of the  $Y$ 's within each set. For the  $j$ th set denote this sum by  $T_j$ . The blinded variance estimator is (variance of  $T_j$ )/block size or:

$$\tilde{\sigma}^2 = \frac{1}{n} \frac{\sum_{j=1}^k (T_j - \bar{T})^2}{k-1} \quad (12)$$

This blinded variance estimator is unbiased and achieves minimum variance—that is, the variation in  $\tilde{\sigma}^2$  is smallest—when  $k$  equals the number of treatments. The following example demonstrates how the method works.

Normally distributed data of size 20 with seed 1234 were generated in SAS version 8.2 with mean 0 and variance 1 for group 1 and mean 2 and variance 1 for group 2. If the means are equal, then blinded data estimation of variability would amount to unblinded estimation. For this reason the means chosen were sufficiently different. The data are shown in Table 4. Assume the randomization block size is 2. Data that belonged to the blocks were taken to be those generated sequentially by SAS for the chosen seed. The first two observations, one from each group, were assigned to block 1, the next two to block 2, and so on. The unblinded estimates of the variances of the data in groups 1 and 2 are 1.19 and 1.81, respectively. Since they are estimating the same variance, we take their average, 1.24, as the unblinded variance estimate. From (12) the blinded variance estimate is 1.28, a result similar to the unblinded estimate. As the block size gets larger, the performance of the blinded variance estimator is less impressive. For normally distributed data, the ratio of the standard deviation of the blinded variance estimator relative to its unblinded counterpart is  $\sqrt{n(\tilde{N}-2)/(\tilde{N}-n)}$ ; for  $n = 2$  the ratio equals  $\sqrt{2}$  [32]. This method based on block sums has been extended to estimate the variance after adjusting for covariates [33]. Other simple methods are also available for estimating the within-group variance [34, 35]. These methods depend, however, on guessing the true treatment effect and are limited to two-treatment trials.

**TABLE 4 Blinded Variance Reestimation Example**

| Block | Group 1 | Group 2 | Block Sum |
|-------|---------|---------|-----------|
| 1     | 0.277   | 3.462   | 3.74      |
| 2     | 0.299   | 0.293   | 0.59      |
| 3     | 0.895   | 1.396   | 2.29      |
| 4     | 0.591   | 2.857   | 3.45      |
| 5     | -0.331  | 2.874   | 2.54      |
| 6     | -1.839  | 1.990   | 0.15      |
| 7     | -1.261  | 1.321   | 0.06      |
| 8     | 0.902   | 0.867   | 1.77      |
| 9     | -0.701  | 1.378   | 0.68      |
| 10    | -0.829  | 3.049   | 2.22      |
| 11    | -1.21   | 0.141   | -0.98     |
| 12    | 0.657   | 4.126   | 4.78      |
| 13    | -1.487  | 4.192   | 2.71      |
| 14    | -1.409  | 1.880   | 0.47      |
| 15    | 0.648   | 1.491   | 2.14      |
| 16    | 1.246   | 2.351   | 3.60      |
| 17    | -1.072  | 2.978   | 1.91      |
| 18    | -0.067  | 1.913   | 1.85      |
| 19    | 0.925   | 2.208   | 3.13      |
| 20    | -2.490  | 1.393   | -1.10     |

Note: Group 1 is  $N(0, 1)$  and group 2 is  $N(2, 1)$ . Data generated in SAS (version 8.2) using seed 1234. The unblinded pooled variance is 1.24, the blinded variance, for block size 2, is 1.28.

## 16.5 POINTS TO CONSIDER WHEN PLANNING SIZE OF TRIAL

The topics discussed above will be summarized in this section after briefly discussing noninferiority/equivalence trials and unequal allocation.

### 16.5.1 Noninferiority/Equivalence Trials

When the control is a marketed product, it is not uncommon to design a noninferiority trial in which the objective is to demonstrate that the experimental treatment is not inferior to control. For example, a sponsor may wish to demonstrate that the once-a-day dosing regimen of its experimental treatment is not inferior to the twice-a-day dosing regimen of an approved drug. The critical issue is the choice of a “margin” that defines noninferiority. For superiority trials, the margin is unambiguously defined to be 0; in other words, to establish superiority, the objective is to demonstrate that  $\hat{\Delta}$  is statistically significantly different from 0. For noninferiority trials the objective is to demonstrate that  $\hat{\Delta}$  is statistically significantly different from some margin  $\delta$ . For example, to demonstrate whether the experimental treatment is not inferior to a marketed control, the sponsor may wish to test whether the difference (experimental minus control) in the proportion of subjects surviving at the end of the trial is greater than  $\delta = -10\%$ . A regulatory agency, such as the FDA, may consider it more appropriate if instead  $\delta = -5\%$ . The choice of  $\delta = -5\%$  is a more difficult objective to meet than  $\delta = -10\%$  and requires a larger sample size.

Equivalence trials involve a lower and an upper margin,  $(\delta_L, \delta_U)$ . An example of an equivalence trial is a vaccine lot consistency trial whose goal is to demonstrate



that the vaccine lots (as assessed, e.g., by the titer values obtained in a clinical trial across the different vaccine lots) are equivalent.

The null hypothesis for noninferiority is  $H_0: \Delta \leq \delta$  and for equivalence is  $H_0: \Delta \leq \delta_L$  or  $H_0: \Delta \geq \delta_U$  with alternative hypotheses, respectively,  $H_1: \Delta > \delta$  and  $H_1: \delta_L < \Delta < \delta_U$ . The sample size derivation for two-treatment noninferiority/equivalence trials is conceptually similar to that for superiority. However, since the sample size is calculated under the alternative hypothesis, a value of  $\Delta$  needs to be specified. In some equivalence trials it has been common to let  $\Delta = 0$ ; this is often a mistake. For vaccine lot consistency (or equivalence) trials, it is demonstrated why the  $\Delta = 0$  assumption may lead to a severely underpowered trial [36]. Following the steps in Section 16.2 for calculating the sample size for noninferiority/equivalence will make evident why the choice of the margin(s) influences the trial size. In particular, the closer the margin(s) is to 0, the larger the sample size. Before planning the size of a noninferiority or equivalence trial, the sponsor needs to be confident in its choice of margin(s). Useful references for determining the sample size in noninferiority/equivalence settings include Farrington and Manning [37] and Nam [38].

### 16.5.2 Unequal Allocation

Although the variance is minimized when the allocation to the two treatments is equal, in large placebo-controlled trials it is not uncommon to allocate more subjects to the experimental treatment. Such allocation provides more safety data for the experimental treatment in subjects who are representative of the population to be treated. In small- to midsized dose-ranging trials where the objective is often to compare increasing doses of the drug to control, it is statistically preferable to allocate more subjects to control, yet it is not uncommon to allocate an equal number of subjects to each group. Such an allocation provides more data on various doses of the drug at a stage in the drug's development when there is little prior evidence of its dose-ranging activity or efficacy. In the large or the dose-ranging trial pragmatism was elevated over statistical idealism. Before determining the sample size, first an allocation ratio deemed to be pragmatically (but not necessarily statistically) appropriate should be determined.

## 16.6 CONCLUSION

In conclusion, the gist of the chapter can be organized around a few assertions. Determining the number of subjects needed is much more than a calculation. Unfortunately, too often the sample size assessment is done mechanically. The inputs to the sample size formula require carefully combining pieces of information. This is where collaboration between statisticians and nonstatisticians is most important. In general, the presumed values of efficacy and variance should be on the conservative side. The impact of missing data should be considered upfront. As stated by Lavori [39]: "Do not expect nature to be kind . . . Many power calculations are based on expected causal differences (what would happen under complete control), and not on expected practical differences (when uncontrolled extra treatments or undelivered study treatments intervene)." It is very helpful to calculate sample sizes under various values of presumed efficacy and variance after adjusting those values by the

intended imputation method for missing data expected in the trial. When the trial is ongoing, examine the blinded data for noncompliance. The impact of disproportionate noncompliance should be assessed. If possible, estimate the variance without breaking the blind after applying the imputation method that was stated in the trial protocol. If the blinded variance estimate is very different from the assumed variance, then recalculate the power of the trial and evaluate the options. Monitor the number of events for time-to-event or relative risk analyses, recalling that the variance is a function of the number of events not the total number of subjects. It will be informative to compare the total number of events observed during the blinded portion with the number that was projected in the sample size calculation.

The calculation of the size of a trial, like any other learning process, is iterative, and not a one-step solution. A more reliable sample size will be obtained if the factors that influence it are given due attention and subject to debate, the assumptions challenged, the blinded data evaluated than if the calculation is performed mechanically.

## REFERENCES

1. Efron, B. (1998), Foreword: Limburg compliance symposium, *Stat. Med.*, 17, 249–250.
2. Canner, P. L. (1984), How much data should be collected in a clinical trial? Experience of the coronary drug project, *Stat. Med.*, 3, 423–432.
3. Friedman, J., Chalmers, T., Smith, H., and Kuebler, R. (1978), The importance of Beta, the Type II error and sample size in the design and interpretation of the randomized controlled trial, *N. Engl. J. Med.*, 299, 690–694.
4. Lachin, J. (1981), Introduction to sample size determination and power analysis for clinical trials, *Controlled Clin. Trials*, 2, 93–113.
5. Posten, H. O., Yeh, H. C., and Owen, D. B. (1982), Robustness of the two-sample t-test under violations of the homogeneity of variance assumption, *Commun. Stat. Theory Methods*, 11, 109–126.
6. Heeren, T., and D'Agostino, R. B. (1983), Robustness of the two independent samples t-test when applied to ordinal scaled data, *Stat. Med.*, 6, 79–90.
7. Jacobson, R., and Poland, G. A. (2005), Sample sizes and negative studies in clinical vaccine research, *Vaccine*, 23, 2318–2321.
8. Feinstein, A. (1984), *Principles of Medical Statistics*, Chapman and Hall/CRC, Boca Raton, FL.
9. Cox, D. R. (1958), *Planning of Experiments*, Wiley, New York, pp. 53–58.
10. Ganju, J. (2004), Some unexamined aspects of analysis of covariance in pretest–posttest studies, *Biometrics*, 60, 829–833.
11. Cramer, M. R. (1987), Analysis of covariance in parallel-group clinical trials with pretreatment baselines, *Biometrics*, 43, 895–901.
12. Casagrande, J. T., Pike, M. C., and Smith, P. G. (1978), An improved approximate formula for calculating sample sizes for comparing two binomial distributions, *Biometrics*, 34, 483–496.
13. Statistical Solutions (2005), nQuery Advisor, version 5.0, MA.
14. Fleiss, J. L., Tytun, A., and Ury, H. K. (1980), A simple approximation for calculating sample sizes for comparing independent proportions, *Biometrics*, 36, 343–346.
15. Barnard, G. A. (1947), A new test for  $2 \times 2$  tables, *Nature*, 156, 177.

16. Mehta, C. R., and Hilton, J. F. (1993), Exact power of conditional and unconditional tests: Going beyond the  $2 \times 2$  contingency table, *Am. Statist.*, 47, 91–98.
17. Gordon, I. (1994), Sample size for two independent proportions: A review, *Australian J. Stat.*, 36, 199–209.
18. Cuzick, J. (1982), The efficiency of the proportions test and the log-rank test for censored survival data, *Biometrics*, 38, 1033–1039.
19. Mantel, N. (1966), Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemother. Rept.*, 50, 163–170.
20. Collett, D. (1994), *Modelling Survival Data in Medical Research*, Chapman and Hall, London, pp. 255–264.
21. George, S., and Desu, M. (1974), Planning the size and duration of a clinical trial studying time to some critical event, *J. Chronic Dis.*, 27, 15–24.
22. Schoenfeld, D. A. (1983), Sample size formula for the proportional-hazards regression model, *Biometrics*, 39, 499–503.
23. Lachin, J. M., and Foulkes, M. A. (1986), Evaluation of sample size and power for analyses of survival with allowance for non-uniform patient entry, losses to follow-up, non-compliance, and stratification, *Biometrics*, 42, 507–519.
24. Temple, R. J. (2004). *The Critical Path Opportunities for Efficiency in Development*. FDA Science Board Advisory Committee Meeting Maryland, April 22.
25. Fleiss, J. L. (1986), *The Design and Analysis of Clinical Experiments*, Wiley, New York, p. 152.
26. Ganju, J. (2008), Post stratified analysis of clinical trial data, under preparation.
27. Ganju, J., and Mehrotra, D. V. (2003), Stratified experiments re-examined with emphasis on multicenter trials, *Controlled Clin. Trials*, 24, 167–181. Correction: 24, 830.
28. Deming, W. E. (1960), *Sample Design in Business Research*, McGraw-Hill, New York.
29. Oxman, M. N., Levin, M. J., Johnson, G. R., et al. (2005), A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults, *N. Engl. J. Med.*, 352, 2271–2284.
30. Lakatos, E. (1986), Sample size determination in clinical trials with time-dependent rates of losses and noncompliance, *Controlled Clin. Trials*, 7, 189–199.
31. Lakatos, E. (1988), Samples sizes based on the log-rank statistic in complex clinical trials, *Biometrics*, 44, 229–241.
32. Xing, B., and Ganju, J. (2005), A method to estimate the variance of an endpoint from an on-going blinded trial, *Stat. Med.*, 24, 1808–1814.
33. Ganju, J., and Xing, B. (2009), Re-estimating the sample size of an on-going blinded trial based on the method of randomization block sums, *Stat. Med.*, 28, 24–38.
34. Gould, A. L., and Shih, W. J. (1992), Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance, *Commun. Stat. Theory Methods*, 21, 2833–2853.
35. Zucker, D. M., Wittes, J. T., Schabenberger, O., and Brittain, E. (1999), Internal pilot studies II: Comparison of various procedures, *Stat. Med.*, 18, 3493–3509.
36. Ganju, J., Izu, A., and Anemona, A. (2008), Sample size for equivalence trials: A case study from a vaccine lot consistency trial, *Stat. Med.*, 27, 3743–3754.
37. Farrington, C. P., and Manning, G. (1990), Test statistics and sample size formulae for comparative binomial trials with null hypotheses of non-zero risk difference or non-unity relative risk, *Stat. Med.*, 9, 1447–1454.
38. Nam, J. (1995), Sample size determination in stratified trials to establish the equivalence of two treatments, *Stat. Med.*, 14, 2037–2049.
39. Lavori, P. W. (1992), Clinical trials in psychiatry: Should protocol deviation censor patient data, *Neuropsychopharmacology*, 6, 39–48.



## *Blinding and Placebo*

**Artur Bauhofer**

*Institute of Theoretical Surgery, Philipps-University Marburg, Marburg, Germany*

### **Contents**

|       |                                                              |     |
|-------|--------------------------------------------------------------|-----|
| 17.1  | Introduction                                                 | 933 |
| 17.2  | Need for Placebo                                             | 934 |
| 17.3  | Features of Placebo                                          | 936 |
| 17.4  | Coding and Randomization                                     | 936 |
| 17.5  | Blinding More Than Patients and Physicians                   | 938 |
| 17.6  | Blinding Trials Other Than Oral Drug Comparisons             | 940 |
| 17.7  | Waiving Blindness                                            | 941 |
| 17.8  | Safety Mechanisms: Breaking Code in Case of Emergency        | 941 |
| 17.9  | Assessment of Blinding and Expectation                       | 942 |
| 17.10 | Indices for Assessment of Blinding                           | 943 |
| 17.11 | Breaking Code at End of Trial                                | 944 |
| 17.12 | Conclusion                                                   | 944 |
|       | Appendix: Critical Questions for Blinded, Placebo-Controlled |     |
|       | Randomized Trials                                            | 945 |
|       | References                                                   | 945 |

### **17.1 INTRODUCTION**

Treatment recommendations in guidelines are mainly based on the results of randomized controlled trials (RCTs) [1]. Results of RCTs should provide the closest possible approximation to the truth. Bias-reducing safeguards (e.g., concealments of randomization and blinding) are important because their omission can exaggerate a treatment effect by 20–45% relative to the true treatment effect [2]. Overestimation in this order may be important since most RCTs seek to detect treatment

**TABLE 1** Quality of Different Allocation Methods<sup>a</sup>

| Allocation by | Constant Probability | Equal Probability | Independent Probability |
|---------------|----------------------|-------------------|-------------------------|
| Randomization | Yes                  | Yes               | Yes                     |
| Alternation   | Yes                  | Yes               | No                      |
| Birthday      | (Yes)                | No                | Yes                     |
| Initials      | (Yes)                | No                | Yes                     |

<sup>a</sup>Different allocation methods with limitations compared to randomization.

Source: Adapted from Lorenz et al. [10].

effects of moderate size of about 20–35% [3]. For this reason, moderate degrees of bias can lead to important distortions in treatment effect estimates; and authorities [4] encourage the use of RCT methodology and researchers conducting systematic reviews of RCTs. In systematic reviews heterogeneity in trial results are often explained by the possibility of differences in trial methodology. Sometimes the effect size is weight based on the methodology used in the RCT [5]. Such corrections are problematic since a trial that is not double blinded must per se not be of inferior quality than a trial in which only data collectors and outcome assessors were blinded or even where an open trial methodology was used.

Often only wordings without common sense such as double blind were used without exactly reporting who was blinded and who was not. Despite efforts to improve the reporting of RCTs, such as the Consolidated Standards of Reporting Trials (CONSORT), studies have documented in a low number concealment of randomization and blinding [2, 6].

Most researchers appreciate the meaning of blinding, but beyond general appreciation there is confusion. Terms such as single blind, double blind, and triple blind mean different things to different people (Fig. 1) [7]. Moreover, many medical researchers confuse the term blinding with allocation concealment. Allocation concealment is primarily used to prevent selection bias and to protect an assignment sequence before and until allocation. In contrast, blinding refers to keeping persons (participants, investigators, and data assessors) unaware of the allocated treatment or therapy, so that they are not influenced psychologically or physically by that knowledge. In well-blinded trials one can have more certainty that any differential effect between groups is due to the treatment rather than the subjects' or researchers' bias [8].

About 44–93% of publications of randomized controlled trials lack a clear description of allocation concealment [9]. This is really a poor rate since a clear description without quantitative information needs no additional trial effort and inadequate information exaggerates the treatment effect [6].

There is agreement that the allocation should be performed by chance, but randomization is not always used as it should be. There are still other allocation procedures in use (Table 1), such as allocation by alternation, by birthday, and by initials. These methods do not fulfill all three criteria: constant, equal, and independent probability [10].

## 17.2 NEED FOR PLACEBO

In contrast to several other study types, randomized controlled trials always have a control group. Having control patients who are completely without treatment does



(A)



(B)



(C)

**FIGURE 1** An exact description of who is blinded when is preferable to stating the trial was “single” (A), “double” (B), or (C) “triple” blinded. In the pictures the author is (A) “single”, (B) “double”, and (C) “triple” blinded.

not allow to decipher whether any response improvement in the treatment group is due to therapy or due to the fact of being treated in some way. Even if therapy is irrelevant to the patients condition, the patient’s attitude to his or her illness and indeed the illness itself may be improved by a feeling that something is being done to improve his or her condition [11].



Gribbin [12] argues that many patients could be effectively treated by placebo, especially by the use of attractive pills and a convincing statement by the physician as to their value. Hence, in any randomized drug trial versus untreated controls, it is worth considering giving the latter a placebo. The use of a placebo allows to eliminate the “placebo effect” from the therapeutic comparison. For this reason placebos are commonly used now in many disease trials.

The decision whether some standard active drug therapy should be a control and the new drug is given as an add-on is an entirely separate issue. One basic principle is that patients cannot ethically be assigned to only a placebo if there exists an alternative standard therapy of established efficacy [11]. The problem is that in many areas of medicine the standard therapy was never tested by sound evidence-based trials. Without supportive evidence the physician has to rely on clinical experience and opinion in deciding whether it is ethical to withhold what has become accepted as standard therapy.

Sometimes two active components should be compared. In this scenario often a “double-dummy” method is used. For example, if we want to compare two medicines, one presented as a blue tablet and one as a red capsule, we could also supply blue placebo tablets and red placebo capsules. Both patient groups have to take one blue tablet and one red capsule [13].

Placebos are used to make patients’ and observers’ attitudes to the trial as similar as possible in treatment and control groups. Under many circumstances the use of placebos is a prerequisite to perform a double-blinded trial.

### 17.3 FEATURES OF PLACEBO

Placebos have to be identical in all respects to the active drug, except that the active drug is absent. This means for oral placebos that they are not distinguishable by color, size, shape, taste, or texture. For oral therapy there are several modes of application available: capsules, tablets, or liquids. Since many active drugs have a distinctive taste, the use of capsules is most often feasible.

When using liquids for oral therapy or for injections, in addition to the color the viscosity and the potential to generate foam by shaking should also be considered. For example, liquids containing protein generate foam but buffer solutions do not. In this case, the ideal placebo will contain also an inactive protein in low concentrations such as human albumin. Some clinicians are very clever in breaking the code; they not only shake or identify the drugs by taste and smell but even photometers were used to determine differences between active drug and placebo. For such reasons it is sometimes very difficult to guarantee blinding.

### 17.4 CODING AND RANDOMIZATION

Double-blinded randomized trials require most careful organization in the allocation of treatments. The allocation to different treatments should have similar probabilities (see above and Table 1).

Before the generation of a randomization list is performed, principal considerations have to be made. Should the randomization list be generated by simple

random permutation or is there a need for stratification in the trial. In some trials, for example, stratification for different operations and age are useful. To allocate approximately the same number of patients to the placebo and the treatment group in each center, a block randomization is often used. This means, for example, in a trial with 100 patients it could have a block size of 20 patients randomizing 10 placebo patients and 10 treatment patients to one center. If one block is assigned, the next block will be used by the individual study center.

Another topic to be discussed is the use of an unequal randomization to placebo and treatment group. Driven by the enthusiasm for a new treatment, sometimes more patients were randomized to the treatment group, even though it would involve some loss of statistical efficiency [14]. Thus randomization in a 2:1 or 3:1 ratio for the new:standard treatment is a realistic consideration; however, there is an even greater loss in statistical power in the 3:1 approach.

There are several methods for the generation of randomization lists available. A detailed description of which method was used is given in Table 2. In general, the preparation of random lists and drug packages should be performed by persons not otherwise involved in conducting the trial. Most times random lists are generated by the study statistician. In all cases it is important to have a simple coding system linking the drug packages to the randomization list. Each package must have a unique trial code number that is also written on the randomization list and patient record form to safeguard a traceable link between, patients, packs, and the list.

Today, more and more central computers are used for randomization and for distribution and allocation of medication packages to the different study centers over a interface with the telephone (interactive voice response, IVR) or the Internet (interactive Web response, IWR). These methods led to significant savings in the amount of medication needed, but also to an increased risk of breaking the code [15]. A deduction from the dispensing order is in some cases possible. An example of such a situation is given by McEntegart et al. [15]. Four packs—two of each group—were delivered to the site in a trial with two treatments (A and B) stratified by gender. The first two patients randomized into the study were male (numbers

**TABLE 2 Generation of Randomization Lists and Allocation Concealment<sup>a</sup>**

| Method        | Adequate Descriptions                                                                                                   | Inadequate, Incomplete Descriptions                                                                                                                             |
|---------------|-------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Computer      | Random lists generated by a computer were used by the physician to sequentially allocate the patients to a treatment.   | Just mentioning the use of a computer is not enough information. Further details should be given since computers are rarely involved in the allocation process. |
| Envelopes     | For allocation serially numbered, sealed and opaque envelopes were used.                                                | Allocation was done by the use of opaque envelopes                                                                                                              |
| Vehicles      | Vehicles were indistinguishable, sequentially numbered, and sequentially administered without knowledge of the content. | For allocation indistinguishable vehicles were used.                                                                                                            |
| Pharmacy unit | Drug preparations from the pharmacy unit, using indistinguishable vials and content were used.                          | Allocation was performed by the pharmacist.                                                                                                                     |

<sup>a</sup>Criteria for adequate description of allocation concealment were taken in part from Schulz [34].

001 and 002). They were assigned to treatments A and B. The next patient, a female, was assigned to treatment A, as defined by the random list and received a matching pack, which was in this example number 004. From this numbering can be concluded that pack 003 must contain different medication from pack 004 because 003 was not given to the patient. Since pack 003 and 004 are different from each other also 001 and 002 must be different from each other.

Another problem is called pack separation, which occurs when the medication is supplied or used in a different ratio to the one used for packaging. More packs from one type will be left, for example, due to a higher withdrawal rate on active treatment. Repeated resupply of packs with equal numbers of packs will allow a division of packs into two distinct groups (one group with more and one with fewer packs), which allows assumptions on their difference. This clearly demonstrates also that sophisticated blinding and drug supply procedures with IVR and IWR technology are susceptible to bias unblinding.

To overcome these problems two strategies are possible. In the first approach the number of delivered packs in each block is increased, but then a part of the drug savings obtained by the individual delivery of packs is lost. Another approach is a double randomization [15]. In this case packs are randomly shuffled around so that there is no longer any association between the order in the file and the pack numbers. This method probably increases the complexity of labeling and distribution, and care has to be taken that no confusion occurs.

## 17.5 BLINDING MORE THAN PATIENTS AND PHYSICIANS

For the term *blinding* sometimes the synonym *masking* is used but blinding is used more often. Blinding, however, is used to reduce bias, but it is not always easy to obtain. Before an *open (nonblinded) trial* is performed, the question should be asked as to who can be blinded of the persons involved in the trial. A list of the individual groups is defined in Table 3 in accordance with Montori et al. [16].

*Single-blind trials* (where either only the investigator or the patient is blind to the allocation) are most times preferable to *open trials*. In *double-blinded trials* it

**TABLE 3 Individual Groups That Could Be Blinded**

| Group                 | Definition <sup>a</sup>                                                                                                                                                                 |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Participants          | Individuals who are randomly assigned to the interventions under evaluation.                                                                                                            |
| Health care providers | The physicians, nurses, or other personnel that actually care for the participants during the study period and/or administer the interventions.                                         |
| Data collectors       | The individuals who actually collect data for the study outcomes. Data collection could include a quality-of-life questionnaire, talking and/or recording a blood pressure measurement. |
| Outcome assessors     | The individuals who ultimately decide if a patient has suffered the outcome of interest.                                                                                                |
| Data analysts         | The individual who conducts the data analysis.                                                                                                                                          |
| Manuscript writers    | The individuals who write alternative versions of the manuscript before breaking the randomization code.                                                                                |

<sup>a</sup>Definitions were taken from Montori et al. [16].

is implicated that the assessment of patient outcome is done without knowing the treatment received [13]. Such blind assessment of outcome can often also be achieved in trials that are open. For example, pathological findings can be assessed by someone else who was not involved in running the trial. Blinded assessment of patient outcome is also valuable in epidemiological studies, such as cohort studies. In diagnostic tests and other trials, persons evaluating the performance of those performing the test should be unaware of the true diagnosis. Or in studies for the evaluation of the reproducibility of a measurement technique, the observers should be unaware of previous measurements on the same individuals [13].

One particular advantage of double-blind trials is that they allow the objective evaluation of side effects, both by the patient and by the physician. For instance, side effects (usually minor effects such as headache, fatigue, nausea) were reported by patients on placebo. This enables to correct for the overreporting of side effects on active therapy to get an unbiased estimate of adverse reactions attributed to the treatment itself [11].

Beyond the widely used term *double blinding* sometimes the term *triple blinding* is also used. This term usually means a double-blind trial that also maintains a blinded data analysis [11], but this is not always the case. Some people think it denotes that investigators and assessors as well as the participants are all unaware of the assignment.

Since there is some confusion on the use of the terminology *single blind*, *double blind*, and *triple blind* in the report of trials, clear statements should be given on who was blinded at what part of the study (Fig. 1). In general, the quality of the reports of concealment and blinding is poor (Table 4), but there seems to be an improvement in the reports from 2002 to 2004, although the studies from Montori et al. [16] and from Devereaux et al. [2] did not use identical criteria for trial selection. Montori et al. [16] analyzed RCT reports from five leading journals, and Devereaux et al. [2] used reports from internal medicine patients. Poor reporting does not mean automatically poor conduction of the trials. Some authors stated that they performed concealing randomization and blinding in adequate ways [2]. From this observation it can be concluded that readers should not assume at least in general that bias-reducing procedures not reported in RCTs did not occur.

However, on average, randomized trials that have not used appropriate levels of blinding show larger treatment effects than blinded studies [17]. In parallel,

**TABLE 4 Reporting Allocation Concealment and Blinding in Randomized Trials<sup>a</sup>**

|                       | Montori et al. [16] 2002 | Devereaux et al. [2] 2004 |
|-----------------------|--------------------------|---------------------------|
| Number of trials      | 191                      | 105                       |
| Report of concealment | n.a.                     | 45%                       |
| Blinding status of    |                          |                           |
| Participants          | 15%                      | 74%                       |
| Health care providers | 5%                       | 36%                       |
| Data collectors       | 12%                      | 16%                       |
| Outcome assessors     | 23%                      | 17%                       |
| Data analysts         | 5%                       | 4%                        |
| Manuscript writers    | 0%                       | n.a.                      |

<sup>a</sup>Information given on allocation concealment and blinding in randomized controlled trials. n.a. = not assessed in the study.

diagnostic test performance is often overestimated when the reference test is interpreted with knowledge of the test results [18]. Quantification of the blinding bias by adjusting the results for a single trial or a meta-analysis of several trials for trial quality is problematic. However, without quantification blinding makes it difficult to bias results intentionally or unintentionally and so helps to ensure the credibility of study results [13].

Beyond the blinding of participants, health care providers, data collectors, outcome assessors, and data analysts (Table 3) some trialists postulate even the blinding of writers [19]. In this approach two or more manuscripts were prepared before the code of the trial is broken. In one manuscript the test drug is significantly better than placebo and in the other there is no benefit of the new treatment. In my opinion this approach is too much directed by statistical rigidity, aside of clinical reality. There should be a clear hypothesis published in advance for the primary outcome of the trial, but the manuscripts will differ gradually. For example, the new treatment can be statistically better than the standard, as assumed beforehand, or even better at a lower significance level, or unchanged or even significantly adverse. However, hardly anyone would write five different manuscripts and interpretations of the results in advance.

## 17.6 BLINDING TRIALS OTHER THAN ORAL DRUG COMPARISONS

Blinding in oral drug trials is most times the easiest part, but it can be also more complex, for example, in the comparison of repetitive applications with a single application. In this case multiple placebo applications are needed to safeguard blinding.

Placebo injections present a greater practical and ethical problem than oral placebos. Additionally, subcutaneous and peripheral injections, for instance, are more easily accepted than more invasive procedures such as insertions of central venous lines.

The problem of blinding is most prominent in surgical trials. Blinding of surgeons is in most cases impossible, and a placebo operation (sham operation) will be in most cases unethical, but it is under some circumstances possible. In a trial by Freed et al. [20] patients received a sham surgery by drilling holes in the skull without penetration of the dura mater. This trial was accepted by the local ethics committee and published in the *New England Journal of Medicine*. A sham treatment such as this one is probably not possible in a lot of other centers. This example demonstrates that ethical considerations may be influenced by local and cultural differences.

Even if blinding of the surgeon is impossible, very elegant designs were used to reduce bias in surgical trials. For instance, in several trials comparing laparoscopic cholecystectomy with conventional surgery, an improved outcome for the new technique was reported [21]. Expectation bias was excluded for the first time in a randomized controlled trial from Majeed et al. [22] comparing laparoscopic and small-incision cholecystectomy. In this trial identical wound dressings in both groups were used so that nurses and trial personnel were blinded to the type of operation. Investigators found no difference between the groups with regard to hospital stay, time back to work, and time to full activity. In this case blinding has destroyed the illusion of a significant improvement by the use of laparoscopy.

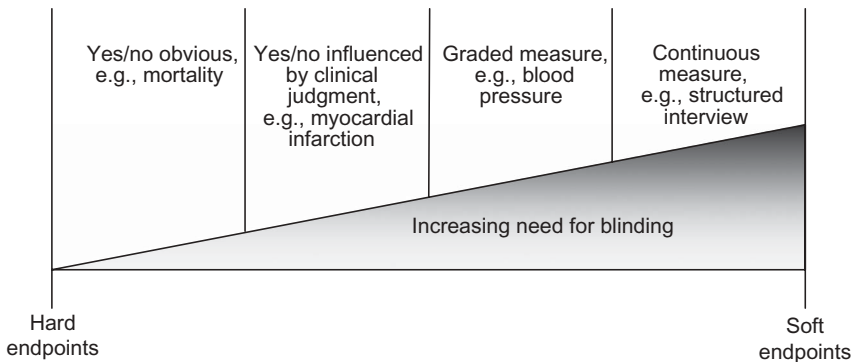


**FIGURE 2** Need for blinding: In parallel with the decrease of objectivity from “hard” endpoints to “soft” endpoints, the influence of expectation increases. Blinding is more important for the assessment of outcomes with high expectation influence.

## 17.7 WAIVING BLINDNESS

Ethical considerations often rule out a double-blind trial design. As already mentioned, in most cases of surgical trials it would be unethical to subject a control group to incisions under anesthesia, mimicking genuine surgery. But, except for the surgeon, all others involved in the trial can be blinded (Table 3).

For some treatments it is totally impossible to arrange a double-blind design. For instance, the evaluation of cytotoxic drugs in cancer therapy is often not double blinded because of complicated dose schedules, the likelihood of serious side effects, and dose modifications to suit each patient’s needs. All these points make it necessary for the treating physician to know the patient’s therapy [11].

It should be considered how serious the bias might be without blinding. In general, for more subjective endpoints a double-blind design is more important than for more objective ones such as mortality (Fig. 2). In each trial organizers have to weigh the pros and cons when planning the trial and writing the trial protocol. Both, ethical and practical problems with blinding have to be considered.

## 17.8 SAFETY MECHANISMS: BREAKING CODE IN CASE OF EMERGENCY

In the case of an emergency a rapid decoding must be possible. For this reason the sponsor and investigators at the different centers must be immediately able to break the code of each individual patient. For this event in most randomized, blinded trials envelopes are prepared for decoding. With each drug package an opaque, sealed, unique coded envelope is delivered to the responsible physician that allows decoding. An additional set of envelopes are left with the sponsor. Another method is the preparation of lists that allow one to scratch off the number code to identify the drug assigned to the participant. In no case should the entire randomization list be transferred to the investigator, which would allow decoding all patients in the trial. Decoding should be allowed only when necessary for the care of the individual patient. Each time the code is broken, the reasons why, the date, and person who broke the code should be recorded. The condition of envelopes will be checked at

each monitoring visit. At the end of the trial the envelopes will be collected, inspected, and mentioned in the final study report [23].

## 17.9 ASSESSMENT OF BLINDING AND EXPECTATION

Ideally, investigators state in the report if blinding was successful. The success of blinding can be easily determined by asking participants, health care providers, data collectors, and outcome assessors to guess which intervention was provided. Table 5 demonstrates an example from a randomized, placebo-controlled trial. Patients were nicely blinded, but health care providers and data collectors were not. In this trial [24, 25] the study drug was G-CSF (granulocyte colony-stimulating factor), which increases the number of granulocytes. Unblinding was probably performed by the knowledge of the leukocyte count. Furthermore, in our trial there was no change over time in the different groups of persons asked. Others have found a change over time, so a repeated measurement of blinding is favored by them [26].

In general, the number of trials providing evidence on the success of blinding is poor. In a analysis by Fergusson et al. [27] it was shown that only 8% (15 of 191 trials) provided qualitative or quantitative information about blinding. Only in 5 of these 15 trials was blinding successful, and only two presented qualitative data. Due to the poor information given, the reporting of randomized controlled trials in regard to this topic should be improved.

In surgery, independent of the treatment, whether the new test drug or placebo is applied, patients believed that the disease will be cured and symptoms reduced by the operation (Table 6). Surgeons had a much more realistic expectation knowing the current literature and their own results. Positive expectations by the patient have a strong influence on the outcome as demonstrated by Koller et al. [28]. Furthermore, a comparison of patients included in the randomized controlled trial and patients treated under routine conditions in the same institution demonstrated an improved outcome of the patients under trial conditions [29]. This observation is not new but often neglected at the time of interpretation of trial efficacy data and extrapolation in regard to clinical efficiency in routine.

In many trials participants showed a strong tendency to believe they had been assigned to the active intervention (Table 6). Expectation can be analyzed also with questionnaires. In our G-CSF trial [24, 25] 68/75 patients thought immediately before the operation that they got the active substance and not the placebo. This belief may be influenced by the time of asking (beginning or end of the trial) and

**TABLE 5** Guess of Treatment (Blindness) of Patients, Surgeons, Ward Assistants, and Data Collectors at Day 3 after Operation (and at Discharge)<sup>a</sup>

| Group           | Correct Guess (%) | Phi Coefficient | <i>P</i> value  |
|-----------------|-------------------|-----------------|-----------------|
| Patients        | 41 (50)           | −0.16 (0.10)    | 0.155 (0.37)    |
| Surgeons        | 73 (73)           | 0.46 (0.49)     | <0.001 (<0.001) |
| Ward assistants | 67 (76)           | 0.33 (0.54)     | <0.005 (<0.001) |
| Data collectors | 73 (76)           | 0.46 (0.54)     | <0.001 (<0.001) |

<sup>a</sup>Blinding was evaluated in a randomized controlled trial with filgrastim prophylaxis versus placebo in patients with increased risk (ASA class 3 and 4) for an adverse outcome after colorectal cancer surgery [24, 25]. Most patients guessed (68/75) that they received the active study drug and not the placebo. *P* values were determined with the Chi-square test.



**TABLE 6** Expectation of Patients and Surgeons before Operation<sup>a</sup>

| Do You Expect:                   | Patient |         |                | Surgeon |         |                |
|----------------------------------|---------|---------|----------------|---------|---------|----------------|
|                                  | G-CSF   | Placebo | <i>P</i> Value | G-CSF   | Placebo | <i>P</i> Value |
| Stop tumor growth                | 36/36   | 39/40   | 1              | 33/36   | 35/40   | 0.71           |
| Reduction of tumor size          | 36/36   | 35/40   | 0.06           | 31/36   | 37/40   | 0.47           |
| Healing                          | 36/36   | 39/40   | 1              | 32/36   | 30/40   | 0.15           |
| Prevention of metastasis         | 35/36   | 39/40   | 1              | 32/36   | 30/40   | 0.15           |
| Prevention of tumor relapse      | 36/36   | 40/40   | 1              | 31/36   | 34/40   | 0.58           |
| Prevention of pain increase      | 36/36   | 39/40   | 1              | 34/36   | 33/40   | 0.16           |
| Pain relief                      | 34/36   | 38/40   | 1              | 20/36   | 25/40   | 0.49           |
| Free of pain without medication  | 34/36   | 38/40   | 1              | 31/36   | 29/40   | 0.17           |
| Relief of tumor-related symptoms | 34/35   | 37/40   | 0.61           | 31/36   | 35/40   | 1              |
| Psychological stabilization      | 35/36   | 39/40   | 1              | 32/36   | 38/40   | 0.41           |

<sup>a</sup>Expectation of patients and surgeons was evaluated in a randomized controlled trial with filgrastim prophylaxis versus placebo in patients with increased risk (ASA class 3 and 4) for an adverse outcome after colorectal cancer surgery. *P* values were determined with the Chi-square test.

by repeated asking itself to guess what treatment they got. Differentiated outcome analysis of blinded and unblinded patients revealed that correct guess of treatment by the study personnel had no influence on patients' outcomes in all major end-points like: quality of life, the McPeck recovery index, length of stay, mortality, complication rate, and re-operation rate.

Expectation bias can be further reduced by using different persons for study coordination and psychometric rating [30]. Since during the trial patients become increasingly familiar and comfortable with the staff and trial office, this often triggers a therapeutic response to the patients [31].

To reduce expectation effects in randomized trials with psychotropic drugs, often a placebo run-in phase is included, which is a single-blind placebo period that lasts about 7–14 days. The placebo run-in phase occurs before randomization, and all study-eligible subjects are given the placebo treatment and are withdrawn most times from any psychoactive drugs during this interval. Responders to the placebo in this preliminary phase were excluded from the trial to reduce expectation and to increase the size of the treatment effect. A recent meta-analysis showed that trials with a placebo run-in phase and elimination of the placebo responders before randomization tended to result in a larger drug effect size, but without statistical significance [32].

Another approach to reduce expectation is to orientate patient expectations with educational statements. This may include statements about the possibility of the study drug working as intended; they may get a placebo and consequently may not improve, and they may receive admonitions to report the actual symptoms honestly without attempting to “help” in accentuating the positive [30]. For instance, site personnel would read the educational statement to the patient during the consent and review it at each succeeding visit. Information about these complex interactions between patients and staff is also important and will further reduce expectation.

## 17.10 INDICES FOR ASSESSMENT OF BLINDING

The success of blinding can be assessed not only by questionnaires but also by validated indices that are available [8, 33]. In the simple and easy-to-calculate index of

James et al. [8] the value increases from 0 to 1 with the success of blinding, while 0.0 is the complete lack of blinding, and 1.0 perfect blinding. In contrast to this index the more complex index of Bang et al. [33] includes also the important opinion of uncertainty. The index is scaled to an interval of  $-1$  to  $1$ ;  $1$  being completely lack of blinding,  $0$  being consistent with perfect blinding, and  $-1$  indicating opposite guessing, which may be related to unblinding.

It is the decision of the persons responsible for the particular trial to use a score or questionnaire. They should be aware that these tools can help to assess the rate of unblinding and the potential influence on the outcome.

### 17.11 BREAKING CODE AT END OF TRIAL

Before the code is broken, all parameters and endpoints listed in the trial protocol should be determined and saved in a log file. This procedure should guarantee as much as possible blindness of data collectors, outcome assessors, and data analysis. At that time point when all steps of data analysis were performed without group comparison, it is time to break the code. Unblinding is a milestone in randomized controlled trials and is for this reason often performed as a ceremonial act.

When the whole trial is completed and the results are interpreted, then it is desirable to inform investigators of each patient treatment. This allows one to get feedback from the experience of the investigators and understanding of the treatments. This procedure allows further the improvement of future trials. Some patients also want to know to which treatment group they were assigned. At this time the information should be accessible for them.

### 17.12 CONCLUSION

There is no uniform methodology for all randomized controlled blinded trials. The need for blinding is increasing in parallel with the decrease of objectivity from “hard” endpoints to “soft” endpoints as the influence of expectation increases. Blinding is more important for the assessment of outcomes with high expectation influence. Ethical and practical problems with blinding have to be considered early in trial development. Reducing bias by blinding is also possible in trials where health care providers (e.g., surgeons) cannot be blinded; but data collectors, outcome assessors, and data analysts can effectively be blinded. Some critical questions to be asked are listed in the Appendix. The success of blinding can be determined at various time points by asking participants, health care providers, data collectors, and outcome assessors to guess which intervention was provided. Blinding can be also assessed by the use of a scoring system. The decision of which system of quantification is used is not so important, but the trial report should include a clear statement of how randomization and allocation were performed and who was at which time point blinded. Following the general consideration mentioned here, bias can be reduced and the plausibility of trial results can be increased.

## APPENDIX CRITICAL QUESTIONS FOR BLINDED, PLACEBO-CONTROLLED RANDOMIZED TRIALS

| Item                     | Questions to Be Answered                                                                                                                                                                    |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Use of placebo           | What is the current standard for the treatment of the specific disease under investigation?<br>Is the use of a placebo ethically sound?                                                     |
| Preparing the placebo    | Which pharmaceutical form is the best to hide the identity of the drug in regard to smell, taste, color, and structure?<br>Is the placebo identical to the active medication?               |
| Coding and randomization | Which method is used for coding and randomization?                                                                                                                                          |
| Blinding                 | Does the study allow to blind: participants, health care providers, data collectors, outcome assessors, data analysts?<br>Does the blinding result in any harm or undo risk to the patient? |
| Breaking the code        | Which safety mechanisms are provided to break the code in an emergency?                                                                                                                     |
| Testing blinding         | Should the study include a test of blinding of the different individuals involved in the trial?                                                                                             |
| Reporting                | Is the generation of the randomization code, allocation process, and blinding adequately described?                                                                                         |

## REFERENCES

1. Ellis, J., Mulligan, I., Rowe, J., et al. (1995), Inpatient general medicine is evidence based. A-Team, Nuffield Department of Clinical Medicine, *Lancet*, 346, 407–410.
2. Devereaux, P. J., Choi, P. T., El-Dika, S., et al. (2004), An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods, *J. Clin. Epidemiol.*, 57, 1232–1236.
3. Yusuf, S., Collins, R., and Peto, R. (1984), Why do we need some large, simple randomized trials? *Stat. Med.*, 3, 409–422.
4. The European Agency for the Evaluation of Medical Products (2001), ICH Topic E10—Choice of Control Group in Clinical Trials: Note for Guidance on Choice of Control Group in Clinical Trials. CPMP/ICH364/96, 1–29.
5. Moher, D., Jadad, A. R., Nichol, G., et al. (1995), Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists, *Controlled Clin. Trials*, 16, 62–73.
6. Pildal, J., Chan, A. W., Hrobjartsson, A., et al. (2005), Comparison of descriptions of allocation concealment in trial protocols and the published reports: Cohort study, *BMJ*, 330, 1049.
7. Devereaux, P. J., Manns, B. J., Ghali, W. A., et al. (2001), Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials, *JAMA*, 285, 2000–2003.
8. James, K. E., Lee, K. K., Kraemer, H. C., et al. (1990), An index for assessing blindness in a multicenter clinical trial: Disulfiram for alcohol cessation—a VA cooperative study, *Stat. Med.*, 15, 1421–1434.
9. Moher, D., Schulz, K. F., and Altman, D. G. (2001), The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials, *Lancet*, 357, 1191–1194.

10. Lorenz, W., Ohmann, C., Immich, H., et al. (1982), Patientenzuteilung bei kontrollierten klinischen Studien, *Chirurg*, 53, 514–519.
11. Pocock, S. J. (1983), Blinding and placebo, in Pocock, S. J., Ed., *Clinical Trials—A Practical Approach*, Wiley, Chichester, pp. 90–99.
12. Gribbin, M. (1981), Placebos: Cheapest medicine in the world, *New Sci.*, 89, 64–65.
13. Day, S. J., and Altman, D. G. (2000) Statistics notes: Blinding in clinical trials and other studies, *BMJ*, 321, 504.
14. Pocock, S. J. (1979), Allocation of patients to treatment in clinical trials, *Biometrics*, 35, 183–197.
15. McEntegart, D., Lang, M., and Wood, R. (2005), Protecting the blind, *Good Clin. Pract. J.*, 11, 10–13.
16. Montori, V. M., Bhandari, M., Devereaux, P. J., et al. (2002), In the dark: The reporting of blinding status in randomized controlled trials, *J. Clin. Epidemiol.*, 55, 787–790.
17. Schulz, K. F., Chalmers, I., Hayes, R. J., et al. (1995), Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA*, 273, 408–412.
18. Lijmer, J. G., Mol, B. W., Heisterkamp, S., et al. (1999), Empirical evidence of design-related bias in studies of diagnostic tests, *JAMA*, 282, 1061–1066.
19. Gotzsche, P. C. (1996), Blinding during data analysis and writing of manuscripts, *Control Clin. Trials*, 17, 285–290.
20. Freed, C. R., Greene, P. E., Breeze, R. E., et al. (2001), Transplantation of embryonic dopamine neurons for severe Parkinson's disease, *N. Engl. J. Med.*, 344, 710–719.
21. McGinn, F. P., Miles, A. J., Uglow, M., et al. (1995), Randomized trial of laparoscopic cholecystectomy and mini-cholecystectomy, *Br. J. Surg.*, 82, 1374–1377.
22. Majeed, A. W., Troy, G., Nicholl, J. P., et al. (1996), Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy, *Lancet*, 347, 989–994.
23. Spriet, A., and Dupin-Spriet, T. (1994), *Good Practice of Clinical Drug Trials*, Karger, Basel.
24. Bauhofer, A., Lorenz, W., Stinner, B., et al. (2001), Granulocyte-colony stimulating factor in the prevention of postoperative infectious complications and sub-optimum recovery from operation in patients with colorectal cancer and increased preoperative risk (ASA 3 and 4). Protocol of a controlled clinical trial developed by consensus of an international study group. Part two: Design of the study, *Inflamm. Res.*, 50, 187–205.
25. Bauhofer, A., Plaul, U., Torossian, A., et al. (2007), Perioperative prophylaxis with granulocyte colony-stimulating factor (G-CSF) in high-risk colorectal cancer patients for an improved recovery: A randomized controlled trial, *Surgery*, 141, 501–510.
26. Rees, J. R., Wade, T. J., Levy, D. A., et al. (2005), Changes in beliefs identify unblinding in randomized controlled trials: A method to meet CONSORT guidelines, *Contemp. Clin. Trials*, 26, 25–37.
27. Fergusson, D., Glass, K. C., Waring, D., et al. (2004), Turning a blind eye: The success of blinding reported in a random sample of randomised, placebo controlled trials, *BMJ*, 328, 432–437.
28. Koller, M., Lorenz, W., Wagner, K., et al. (2000), Expectations and quality of life of cancer patients undergoing radiotherapy, *J. R. Soc. Med.*, 93, 621–628.
29. Kopp, I., Bauhofer, A., and Koller, M. (2004), Understanding quality of life in patients with colorectal cancer: Comparison of data from a randomised controlled trial, a population based cohort study and the norm reference population, *Inflamm. Res.*, 53 (Suppl. 2), S130–S135.

30. Kirby, L., Borwege, S., Christensen, J., et al. (2005), Reducing placebo response: Triple blinding & setting expectations, *Apl. Clin. Trials*, November, 48–52.
31. Fritze, J., and Moller, H. J. (2001), Design of clinical trials of antidepressants: Should a placebo control arm be included? *CNS. Drugs*, 15, 755–764.
32. Lee, S., Walker, J. R., Jakul, L., et al. (2004), Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation, *Depress. Anxiety*, 19, 10–19.
33. Bang, H., Ni, L., and Davis, C. E. (2004), Assessment of blinding in clinical trials, *Control Clin. Trials*, 25, 143–156.
34. Schulz, K. F., Altman, D. G., Moher, D., et al. (2002), Allocation concealment in clinical trials, *JAMA*, 288, 2406–2408.



# 18

---

## *Pharmacology*

**Thierry Buclin**

*Division of Clinical Pharmacology and Toxicology, University Hospital of Lausanne  
(CHUV), Lausanne, Switzerland*

### **Contents**

|        |                                                                    |     |
|--------|--------------------------------------------------------------------|-----|
| 18.1   | Introduction                                                       | 949 |
| 18.1.1 | Ethical Issues in Human Pharmacology Trials                        | 950 |
| 18.1.2 | Roles and Importance of Clinical Pharmacology in Drug Development  | 951 |
| 18.1.3 | Classical Study Questions and Designs in Clinical Pharmacology     | 953 |
| 18.2   | Pharmacokinetic Characterization of Drugs                          | 956 |
| 18.2.1 | Drug Assay Methods                                                 | 956 |
| 18.2.2 | Sample-Rich Trials                                                 | 959 |
| 18.2.3 | Population PK Studies                                              | 973 |
| 18.3   | Pharmacokinetic–Pharmacodynamic Evaluations                        | 977 |
| 18.3.1 | Pharmacodynamic Variables                                          | 978 |
| 18.3.2 | Concentration–Response Relationships                               | 980 |
| 18.3.3 | Design and Analysis of PKPD Trials                                 | 983 |
| 18.4   | Conclusion                                                         | 984 |
|        | Appendix: Note on Geometric Averages and Coefficients of Variation | 985 |
|        | References                                                         | 986 |

### **18.1 INTRODUCTION**

A vast majority of clinical trials are designed to scientifically assess the efficacy, safety, effectiveness, or usefulness of a given medical procedure, most often a therapeutic intervention (drug treatment, surgery, physical therapy, etc.) and sometimes a diagnostic test or a prognostic marker. Such clinical trials challenge a small



segment of actual or upcoming medical practice. Their goal is to improve the welfare of the enrolled patients, which represents a determinant condition for the acceptance of new medical techniques. Proper conduct of clinical research is necessary to ensure the scientific validity of such claims for improvement. However, besides this important body of medically oriented trials, one can identify a minority of clinical experiments not strictly directed toward the improvement of the participants condition. Such trials are typically aimed at studying a given scientific question, which is to be tested during a clinical investigation, however, without therapeutic purpose. Depending on the type of question, one may distinguish between clinical studies in *physiology*, which address normal functions of the organism, *pathophysiology*, dealing with disease mechanisms, and *pharmacology*, oriented toward characterization of drug properties in humans. This chapter is devoted to the latter type of trials. Clinical pharmacology trials play an important role both in drug development and in optimization of patient care, and their specific methodological characteristics will be reviewed.

### 18.1.1 Ethical Issues in Human Pharmacology Trials

Even if scientific progress has indirect repercussions beneficial for patient care, clinical trials in pharmacology are most often without clear benefit to the subjects included. Those subjects may be either healthy volunteers or patients presenting with a given condition to investigate, such as renal or hepatic insufficiency susceptible to alter the disposition of a drug. Even in the latter case, the drug tested is generally given without medical indication (e.g., a single dose of an antibiotic may be given to hemodialysis patients not currently suffering from infection). This has definite ethical implications [1]:

- Obtaining a fully informed consent is absolutely necessary, and the inclusion of subjects not able to understand extensively the study conditions is unacceptable (while it may be considered feasible to some extent in therapeutically oriented trials).
- Not the slightest foreseeable risk is acceptable for study participants (while an expectedly favorable benefit–risk ratio is required in therapeutic trials, even when potentially hazardous treatments are administered).
- Study subjects will usually receive significant financial compensation for their participation (which is usually not the case in therapeutic trials).
- No pressure is to be exerted on the candidates to stimulate their participation in the study, making it problematic to include economically disadvantaged people, who may depend too strongly on financial compensation, or collaborators of the investigator, bound by professional dependency.
- Insofar as altruistic motivations play some role in the decision of subjects to participate in a trial, the investigators have a moral responsibility of exploiting the study results with the best consideration to public health.
- Eventually, particular attention must be paid to psychological and relational aspects of the study: Most people will not easily accept unnecessary medical investigations and drug treatments, even with compensation. Participating subjects may experience irrational concerns and anxiety or unpleasant feelings of

“hiring one’s body.” Thus, it is important that investigators in clinical pharmacology be capable of psychological attention, receptive attitude, recognition of valuable motivations, and thankfulness toward study subjects.

A detailed discussion of ethical issues raised by the inclusion of healthy volunteers in clinical trials will be found elsewhere in this textbook (see Chapters 24.1 and 24.2). It is worth noting that the ethical standards applied to the conduct of human pharmacology trials have changed over the time and do not apply equally all over the world. The preferential involvement of prisoners as study subjects had been advocated for centuries; it culminated under the Nazi regime in Germany, and it was still practiced in the United States up to the 1970s [2]. It is now considered incompatible with human rights. The systematic inclusion of homeless people is still current practice in some places. Similarly, numerous academic centers would regularly ask their employees to participate in clinical investigations requiring healthy subjects. On the other hand, the importance of financial aspects in pharmacological research becomes of increasing public awareness. For-profit contract research organizations (CROs) play nowadays a leading role in the conduct of phase I trials, a vast majority of them being exclusively supported by industrial sponsors. In addition, the general trust of western populations in pharmaceutical companies is decreasing [3]. In this changing world, a strong commitment is required from investigators to apply rigorous ethical standards and protect human pharmacology research from any suspicion of implication in postmodern forms of slavery.

### **18.1.2 Roles and Importance of Clinical Pharmacology in Drug Development**

Clinical pharmacologists working in pharmaceutical companies are traditionally in charge of phase I trials, mostly devoted to the characterization of the pharmacological profile of new compounds. Issues related to the organization and performance of phase I trials are discussed elsewhere in this textbook (see Chapter 9.2). However, human pharmacology studies cannot be merely equated with phase I trials. Pharmacological questions are to be investigated at every step of drug development [4]. Pharmacokinetic analyses are increasingly included in phase II and III trials. Many drug interaction studies are conducted during phase IV. Moreover, it is often only after the commercialization of a drug that the suitability of therapeutic drug monitoring (TDM) for dosage individualization is examined, justifying further pharmacological assessments. In other instances, a biomarker useful for patient follow-up is identified and requires proper pharmacodynamic validation before large-scale evaluation and implementation.

The pharmacological characterization of new drugs has been considered by many authors as not entirely satisfactory at the time of product registration. Numerous drugs have been launched at suboptimal dosage, either too high or (less frequently) too low, because of a poor description of their dose–response relationship. A systematic analysis of 499 new prescription drugs approved by the Food and Drug Administration (FDA) over 20 years revealed that dosage changes had occurred in 73 of the drugs over the following years, the vast majority of which were decreases in dosage due to safety reasons [5]. Similar findings were reported in Europe [6]. Recommended dosing schedules happen to prove inadequate, either globally or in

specific subgroups of patients. For example, while most intravenous  $\beta$ -lactam antibiotics should probably be proposed as 24-hour infusions [7], they are still prescribed as injections on a t.i.d. (three times a day) or q.i.d. (four times a day) basis. Significant drug interactions frequently remain unknown at launch time, which may cause serious adverse outcomes in patients and even the withdrawal of products from the market [8, 9]. This is also the case for pharmacogenetic influences, which should receive more consideration [10]. Lastly, treatment monitoring and dosage adaptation issues are paid insufficient attention from drug developers, preventing the patients from obtaining the best possible benefit from new drugs until rational individualization strategies have been elaborated. Typical examples can be found in HIV (human immunodeficiency virus) treatment with protease and reverse transcriptase inhibitors [11], immunosuppression [12], and probably cancer treatment with new signal transduction inhibitors.

Clinical pharmacology has much to bring to drug development not only during the clinical phases I–IV but may even complement or replace some parts of the preclinical development. The sensitivity of drug assay methods is continuously increasing, with accelerator mass spectrometry techniques currently enabling determination of minute concentrations of radiolabeled compounds in biological fluids [13]. The administration of tiny amounts of new drug candidates to human volunteers, well below efficacy and toxicity dose thresholds, has been advocated for screening purposes, before the initiation of costly preclinical pharmacology and toxicology programs. This so-called phase-zero approach puts emphasis on the human pharmacokinetic profile as a relevant criterion for early drug development decisions [14]. It must be recalled that less than one in five drug candidates entering phase I eventually reach the market; this attrition plays a significant role in increasing the global costs of pharmaceutical research [15]. Clinical pharmacologists employed in the pharmaceutical industry are those who regularly bring the elements leading to a decision to kill a drug candidate during development. An earlier acquisition of such elements might therefore allow significant savings in research expenditures.

Finally, clinical pharmacology is still a science under development, and presently unknown concepts probably remain to be discovered. The whole story of cytochrome P450 enzymes has occupied many scientists since the 1970s, with numerous clinical trials devoted to elucidate their metabolic roles, genetic variations, and modulation by interacting drugs [16]. More recently, drug transporters went into scrutiny and justified a further amount of clinical investigations, which are still currently underway [17]. Other mechanisms of drug disposition are still poorly characterized and deserve definite research efforts (e.g., degradation pathways for peptide products). Advances in biopharmacy and new drug delivery techniques raise further issues that require to be tested at the clinical level. Pharmacodynamic mechanisms, besides their detailed dissection through *in vitro* experiments, also warrant to be better described through clinical experiments. Much remains to be discovered, too, in the field of adverse effect causation. Pharmacogenetics and ethnopharmacology represent further areas for intensive scientific exploration. Such “knowledge-oriented” research has been traditionally ascribed to academic groups, in opposition to “product-oriented” development. It is, however, important that both types of research conform to similar quality standards. Noticeably, a part of the scientific findings produced by universities lead to patent protection and industrial develop-

ment. Conversely, phase I trials devoted to a specific product development can identify new pharmacological concepts and produce generalizable findings. Thus, not only is clinical pharmacology important for drug development, but drug development itself invigorates clinical pharmacology in various scientific aspects.

### 18.1.3 Classical Study Questions and Designs in Clinical Pharmacology

Clinical pharmacology aims to describe, explain, and predict the fate and the effects of drugs in the human organism and to provide scientific foundations for the rational use of drugs in therapeutics. It relies on established knowledge in physiology and pathophysiology, and it uses specific methodological approaches to derive useful conclusions from clinical measurements, obtained during drug administration trials designed for that purpose. The general concepts governing clinical pharmacology can be found in dedicated textbooks [18, 19], and only their practical investigation through human trials will be covered here.

**Main Types of Study Questions** Traditionally, the scientific questions addressed during clinical pharmacology studies have been classified into two main groups:

- *Pharmacokinetic* (PK) issues are those dealing with the fate of a drug in the human organism, that is, *absorption* from administration site, *distribution*, *metabolism*, and *excretion* (metabolism and excretion are regarded together as *elimination*, and distribution and elimination are referred to collectively as *disposition*) [19]. The variables to be measured in a pharmacokinetic trial are mainly circulating concentrations or excreted amounts of the drug and/or its metabolites.
- *Pharmacodynamic* (PD) issues are those related to the effects of a drug on the human organism, that is, dose and time dependency of the response, effect modulation over repeated administration, such as *sensitization* or *tolerance*, and interference with other endogenous or exogenous agents, such as *synergism* or *antagonism*. The variables to be assessed during a pharmacodynamic trial are biological effect markers, or *biomarkers*.

During the past decade, the advantages of simultaneously addressing both groups of questions in clinical pharmacology trials have been recognized. Such *pharmacokinetic–pharmacodynamic* (PKPD) studies are increasingly performed [20]. They require an elaborated methodology for the analysis of their results, which take specifically into account the mechanisms of translation of concentration profiles into effects, as they can be observed clinically [21].

**Usual Study Objectives** The aim of a clinical pharmacology study can be a mere *description* of the average trends that characterize dose–concentration (PK), or dose–response or concentration–response (PD) relationships. However, many investigators will also propose an *interpretation* of study observations, with reference to a theoretical model (e.g., compartmental modeling in PK trials, or  $E_{\max}$  modeling in PD trials). This approach provides clinical estimates of a set of parameters describing the drug PK, PD, or PKPD profile, according to the model chosen. Those parameters can then be used to explain or predict the influence of *specific factors* on the

pharmacological profile. The type of factors entering into consideration delineate another way to classify clinical pharmacology investigations, based on the comparisons planned in the study protocol:

- *Dose–range trials* investigate the influence of the dose administered on drug PK descriptors (e.g., pharmacokinetic *linearity* issues), or PD or PKPD profile (*dose–response* issues).
- *Demographic characteristics trials* examine the influence of gender, age, race, or body weight on a drug PK, PD, or PKPD parameters. This basic approach is applied, for example, to the display of results in most PK trials.
- *Drug formulation trials* examine how the route of administration and galenical conditioning of an active ingredient influences its PK, PD, or PKPD profile. *Bioavailability* and *bioequivalence* trials enter into this category, the latter being oriented toward the demonstration of an acceptable degree of similitude between two different formulations of a given pharmacological agent, for example, an original and a generic brand (which raises specific statistical issues).
- *Food effect trials* investigate the influence of food intake, mainly on the PK parameters of the study drug.
- *Pregnancy trials* aim to determine how childbearing alters the pharmacological profile of a drug. Some investigations also include an evaluation of the transplacental passage of the drug agent to the fetus (e.g., through measure of the drug in cord blood at delivery). The secretion of a drug agent into breast milk can be determined in women subjects during *lactation*.
- *Drug interaction trials* examine the modifications induced by a specific co-medication in the PK, PD, or PKPD characteristics of a given drug.
- *Organ insufficiency trials* assess the impact of a given degree of failure affecting a given physiological system on the PK, PD, or PKPD profile of a drug. While most such trials investigate alterations of drug disposition in subjects suffering from kidney or liver failure, other types of *drug–disease interaction* are sometimes conducted. Drug *dializability* studies are to be cited here as well.
- *Pharmacogenetic trials* try to demonstrate the influence of specific individual genetic traits on the pharmacological profile of a given drug; among classical examples for such investigations are those addressing genetic polymorphisms of cytochrome P450 enzymes or drug receptors.

The combination of those study objectives with the PK, PD, or PKPD principal aims outlines a general typology of human pharmacology trials. Not infrequently, several objectives are combined in a single trial (it is, however, recommended to hierarchize the study questions, and to predefine one or a few main questions for confirmatory analysis, while keeping an exploratory status for all other comparisons).

***Classes of Study Designs*** A further axis for study classification relies on the *trial design*. While many variations can be imagined with regard to the questions addressed and the context of the study, trials can be roughly classified into one of the following main categories:

- *Sample-rich trials* correspond to the classical pharmacological experiments in human subjects, as they were performed by innumerable investigators over more than one century. Such trials continue to be required for new drug registration and represent the most straightforward way to bring answers to well-defined scientific questions of clinical pharmacology [22]. Traditionally, they include a limited number of subjects (between 5 and 50), while each subject provides a large amount of data (repeated concentration and/or effect measurements).
- (a) *Single-dose trials* apply PK and/or PD measurements to the follow-up of one drug administration. Such trials are often built up in *cross-over*, that is, the study subjects receive several single doses (using different dose levels, formulations, interacting medications, etc.) according to a given sequence, the administrations being separated by wash-out periods. This design enables the subtraction of difficult-to-control individual determinants of drug exposure and response, which cancel out during within-subject comparisons. Ideally, the sequence would be randomized to minimize confounding influences such as time-related trends. However, during early development of new products, a *dose-rising sequence* is often preferable for safety reasons; the inclusion of interspersed placebo periods allows randomization and blinding. Placebo control and blinding are usually considered of little relevance in PK trials but are recommended in principle when PD measurements are performed, to evaluate and subtract the baseline response profile observed after a zero dose of the test drug. Factors such as demographic or genetic characteristics can only be tested according to a *parallel* design.
- (b) *Repeated-dose trials* imply a prolonged period of observation and therefore often follow a parallel design. Such trials with *longitudinal* follow-up and analysis are necessary to observe time-related trends in exposure and response, for example, concentration accumulation, autoinduction of metabolism, or tolerance to drug effects. Sometimes, a *transversal* snapshot over one dosing interval is considered sufficient to assess some aspects of the drug PK or PD profile in patients treated on the long term.
- *Population trials* have been proposed since about the last three decades and rely upon data analysis techniques based on pharmacostatistical modeling. Such techniques allow for the estimation of pharmacologically relevant PK and PD parameters from observations drawn in sparse sample conditions, that is, when a large number of subjects provide few data each. Observations can thus be collected on a large scale directly among the target population of patients receiving the treatment, under less stringent study conditions than in classical trials. Population modeling techniques can accommodate varying dosing levels and schedules, unbalanced designs, and fragmentary data. They enable the estimation of both the mean values and the variability of PK and PD parameters among the population. Moreover, the influence of demographic factors, genetic markers, food effects, organ insufficiency, and co-medication can be assessed, provided that the trial population covers a sufficient range of individual differences. While subject heterogeneity is considered as a nuisance in traditional sample-rich trials, it is rather an advantage in population trials. Thus, population approaches have received much attention from both drug



developers and registration authorities [23], and they contribute currently to extend the scope of pharmacological characterization of drugs throughout their clinical development [24–26].

## 18.2 PHARMACOKINETIC CHARACTERIZATION OF DRUGS

During clinical PK trials, drug assay techniques are applied to a series of biological samples to determine the concentration profile that develops following drug administration; the results are analyzed to describe the fate of the drug, and possibly its derivatives, in the organism. This analysis, interpreted with reference to the classical laws governing the absorption and disposition of drugs, provides estimates of the *pharmacokinetic parameters* that characterize the drug studied. Such parameters represent specific values appearing in the pharmacokinetic equations used to model the drug PK profile. The parameter values are specific not only for a given drug but also for a given type of individual in a given condition. Many pharmacokinetic investigations will include a search for important factors that influence the parameters to a clinically significant extent. Knowing a drug PK profile is a prerequisite for designing rational therapeutic regimens and individualization strategies.

### 18.2.1 Drug Assay Methods

**Key Steps of Drug Assay** A PK trial has to rely on efficient and valid laboratory methods. The measurement of drugs and their metabolites in plasma, urine, or other fluids is most often achieved using chromatographic methods. Without entering into details of analytical chemistry, we just summarize here the main steps of a drug assay:

- *Preanalytical processing* covers the operations done on the sample between its collection and assay. Blood will usually be separated into either plasma or serum by centrifugation, with or without anticoagulant, respectively (although certain assay methods are developed specifically for whole blood). For poorly stable compounds, the sample may have to be added with preservatives (pH or oxidation modifiers, chelators, or enzyme inhibitors). In general, a series of samples will be stored under refrigeration until analysis. The influence of storage conditions on sample stability must be checked. Not infrequently, the samples will have to be sent frozen (e.g., on carbonic ice) to a distant lab. Preanalytical processing can be a source of serious bias in PK results: The tested drug may undergo spontaneous degradation or adsorption on the walls of sample tubes, or additives such as anticoagulants may interfere with the assay. Another important issue is *sample labeling*, which has to be thoroughly designed and organized to avoid any confusion between samples. A proper sample label should indicate the study name, the study site, the patient identification number, and the date and time of its collection; it must survive in storage conditions. All sequential tubes used in preanalytical processing must be labeled, working conditions must be organized to minimize the risk of confusion, and the technicians must be taught to “always read the tube they hold.”



- *Extraction* is often necessary to avoid slandering of analytical devices by protein-rich fluids. Protein denaturation, for example, by trichloroacetic acid, followed by centrifugation is often sufficient. *Solid-phase extraction* on micro-columns or *liquid-phase extraction* in an organic solvent may achieve a rough preliminary separation and compensate for variations in the sample matrix (e.g., urine concentration). It is important to check the degree of completeness of extraction through recovery experiments at various concentration levels. *Microdialysis* or *ultracentrifugation* techniques will enable the specific extraction of the unbound moiety of a drug in plasma or serum and the determination of free circulating drug concentrations. Such techniques are delicate to implement and highly sensitive to preanalytical processing conditions.
- *Separation* represents the central step in a chromatographic assay. It may be achieved in liquid phase, with a solvent eluting the components of the sample extract along a solid column under high pressure (*high-performance liquid chromatography*, HPLC) or in gaseous phase along a tiny capillary tube (*gas chromatography*, GC). Numerous choices are available to optimize the separation process, which depends on the relative physicochemical affinity of the analytes for the mobile and fixed phases. It is frequently possible to develop a method enabling the simultaneous separation of a drug and its metabolites, or of several drugs in a single run. On the other hand, different components of the sample happen to come out of the separation column at the same time, which represents an important risk of interference when nondiscriminative detection methods are used.
- *Detection* of the analytes at the exit of the separation column enables their quantification. Traditional HPLC methods use spectrometric, fluorimetric, or amperometric detection, with a modest discriminative power. More recently, mass spectrometry methods have been implemented, enabling detection of both high specificity and sensitivity, and thus decreasing the requirements on separation. *Gas chromatography and liquid chromatography with mass spectrometry* (GCMS and LCMS) assays have imposed a new standard on PK analytical methods, while both increasing the sensitivity and specificity of drug determination and decreasing sample processing time. The addition of ion fragmentation and mass spectrum analysis of fragments adds a further degree of specificity; it even enables the qualitative identification of chemical components in the sample, with reference to a database of fragmentation spectra.
- *Quantification* uses the signal issued by the detection apparatus. It usually consists of integrating the electronic signal over the duration of the peak corresponding to the analyte of interest. The value is then transformed into a concentration level according to a calibration curve. Many choices are available for the calibration method and may significantly impact on the performance of the assay (curve shape, regression method, coverage of the dynamic range by calibrators). It is frequently recommended to use an *internal standard*, that is, a chemical compound different from the analyte, which is added in fixed amount into the sample before extraction and separation, and detected and quantified on its own. Normalizing the measurement signal of the analyte by the signal of the internal standard will compensate for variations in the whole assay process and can significantly improve its reproducibility.

- *Quality control* determines the acceptance of a series of results. It is usually done by including a few *internal control* samples of known concentration and checking whether their results lie in a predefined acceptance range (e.g.,  $\pm 10\%$ ). In case of problematic precision, one may decide to perform all measurements in duplicate or triplicate. For drug measurements widely performed, programs of *external quality control* may be available, with one center sending samples of undisclosed concentration to all participating labs, receiving the results and analyzing the respective performance of each lab. If only one lab performs a given drug assay, it can be a good idea to provide *blinded duplicates*, that is, to separate certain study samples into two tubes and send one with a coded label.

Other analytical methods are used as well in clinical PK besides chromatographic assays and can be roughly described according to similar steps. For example, biotech products are often poorly characterized at the molecular level, as they represent a mixture of numerous chemical species (e.g., proteins with different stages of glycosylation). For such types of drugs, immunoassays are often preferable, with antibodies ensuring the step of analyte separation; the detection is ensured by radioactivity counting (*radioimmunoassay*, RIA), enzymatic reactions followed by spectrometry (*enzyme immunoassay*, EIA), or *fluorescence polarization* (FPIA). One frequent problem is the cross-reactivity of drug metabolites, either active or inactive, or other sample constituents with the assay method. Another interesting family of separation techniques uses a preparation of drug receptors instead of antibodies: such *receptor assays* are able to specifically measure the sum of chemical species showing affinity for the pharmacological drug target. Going one further step in the direction of pharmacodynamics, *bioassay* techniques directly measure the pharmacological activity contained in the sample (e.g., inhibition of bacterial growth by antibiotics). Such assay methods are often of high relevance, however, of limited performance.

**Elements of Assay Validation** Whatever the assay technique considered for a PK trial, it must be set up methodically, and its performance and reliability must be controlled in a preliminary validation step, before it can be applied to study samples. The elements to be collected during this step have been listed in the recommendations published following the Washington Conference on Analytical Methods Validation [27], considered as a basic document for good laboratory practice in the field of PK assay methodology. Similar elements are mentioned in the Q2 guideline of the International Conference on Harmonisation (ICH), which, however formally, address pharmaceutical quality control rather than PK determinations [28]. Briefly, the following pieces of information are required to characterize the performance of an assay method:

- *Accuracy* is the extent to which results generated by the method agree with the “true value.” It is usually determined by assaying control samples of known concentration (e.g., serum spiked with precise amounts of the analyte). Accuracy is expressed as a *relative bias*, which should not depart statistically from zero.
- *Precision* is the extent to which the individual results of multiple measurements of a series of standards agree together. It is expressed as a *coefficient of varia-*

tion. Intraassay (also termed *repeatability*), interassay and possibly interlab (also termed *reproducibility*) components of variability can be identified using appropriate tests.

- *Sensitivity* is given by the *limit of detection*, that is, the point at which a measured signal is significantly larger than the background noise (signals from blank samples), allowing to conclude that some analyte is definitely present in the sample; and by the *limit of quantification*, that is, the minimum concentration that allows precise measurement. The latter limit, together with the upper level of concentrations still measurable with linearity and precision, define the *dynamic range* of the analytical method.
- *Linearity* is the ability of the assay to elicit results that are (either directly or by means of a given mathematical transformation), proportional to the concentration of analyte in the samples within a given range. It can be determined by serial dilutions of a highly concentrated sample.
- *Specificity* (or *selectivity*) is the ability of the assay to measure unequivocally an analyte in the presence of chemical interferences, such as drug excipients, degradation products, co-medications, impurities, or other components of the sample matrix. Its assessment requires extensive search for cross-reactive substances.
- *Robustness* is the resistance of the assay methods toward changes in technical conditions of operation (e.g., temperature, processing time, batch of reactants, etc.).
- Finally, the period of assay validation is a suitable time to check *sample stability* issues under various collection and storage conditions (type of tubes, matrix effects, and length and temperature of storage).

### 18.2.2 Sample-Rich Trials

In traditional pharmacokinetic experiments, a panel of individuals are administered one or several standard doses of the drug studied and provide a series of blood samples (and possibly urine collections) for the determination of drug concentrations [29]. Practical aspects of subject selection, drug administration, sample collection, and data analysis are presented in this section.

**Selection of Subjects** Many PK trials are performed in healthy young volunteers, considered as reference subjects for the assessment of PK parameters of a given drug in “normal conditions.” Any departure from normality would raise a risk of biased evaluation of the drug PK profile. Therefore, rather stringent inclusion criteria are usually applied to ensure that the study population will actually correspond to this goal. The list below enumerates inclusion criteria frequently applied in PK study protocols on healthy subjects.

- As a principle, both male and female subjects should be included so that the results reflect possible sex-related differences and avoid selection bias [30]. Exceptions are for treatments indicated only in either gender or expected to seriously interfere with reproduction (impact on menstrual cycle, potential for teratogenic activity). Pregnant women should not be included, as pregnancy

significantly influences the absorption and disposition of numerous drugs [31]. Contraceptive use is usually permitted, however, potential interaction of the study drug with contraceptive agents should be taken into account. A pregnancy test may be included in lab screen.

- “Young volunteers” are usually considered between 18 and 45 years. Beyond age 30, a progressive decline of many physiological functions by roughly 1% per year is reported in the average population [19]. It probably results less from programmed physiologic changes than from a trivial increase in the probability of facing sequels of diseases and traumas. In any case, there must be some predefined cutoff to avoid significant impact of such changes on the PK study results.
- Significant underweight or overweight subjects should be avoided as well. Usual weight limits range between 55 and 95 kg for men and from 45 to 80 kg for women, with a body mass index (BMI, i.e., body weight divided by squared height) between 18 and 28 kg/m<sup>2</sup>.
- The medical history and physical examination should not reveal “significant findings,” that is, any condition susceptible to bias the evaluation of normal PK parameters or to impair the safety of exposure to the study drug. A history of severe drug allergy or atopic condition is frequently considered as an exclusion criterion, as are hypertension, cardiac disease history, or significant electrocardiographic abnormalities (including prolonged QT interval).
- Routine laboratory tests should be ordered (blood count, electrolytes, proteins, kidney, and liver tests), and subjects having clinically significant laboratory abnormalities should not be included. Noticeably, athletic young men can have rather high serum creatinine values without any renal impairment due to abundant production by the muscular mass. Gilbert’s syndrome (a rather frequent condition with increased total and unconjugated bilirubin on fasting, due to impaired glucuroconjugation) can be accepted if the study drug is not a substrate for glucuroconjugation reactions.
- Urine drug screens for common street drugs are usually requested, to refuse candidate volunteers with current substance use disorder, in which unwanted interactions might happen. The existence of false-positive results for those tests is to be remembered (e.g., poppy seed bread consumption producing positive opiate screen).
- In general, there should be no limitation of race in PK studies. However, it is to be remembered that ethnic factors play a definite role in drug disposition, mainly because the frequency of genetic polymorphisms affecting drug metabolism or transport varies between the different populations across the world (e.g., higher number of CYP2C19-poor metabolizers in Asians, of CYP2D6 ultrarapid metabolizers in Ethiopians [32]). Genotyping of the study subjects for important pharmacogenetic polymorphisms is increasingly requested in PK trials and is now recommended by registration authorities to explore the influence of given genes on the PK profile of new drugs.
- The use of any medication is to be forbidden 2 weeks prior to study drug administration and during the follow-up, including aspirin or other over-the-counter preparations. The use of long half-life drugs with induction or inhibition

potential may affect drug metabolism even more durably (e.g., phenobarbital, amiodarone). Acetaminophen may be permitted before and during the study as a rescue medication with investigator permission.

- The subjects must generally be asked not to consume quinine or grapefruit beverages, as they inhibit CYP2D6 and CYP3A4/5, respectively. Tobacco smoking or exposure to hydrocarbons induce CYP1A2, and alcohol abuse induces CYP2E1. More generally, the study subjects should be asked to refrain from excessive food consumption, prolonged fasting, strenuous exercise, dehydration, and blood donation before and during the whole study period.
- Finally, ability to understand the study conditions, willingness to participate, provision of written informed consent, and appropriate compliance toward study restrictions represent obvious inclusion criteria.

**Dose Determination** The pharmacokinetic literature reveals a striking contrast between the marked attention devoted to analytical and mathematical aspects of the exploitation of biological samples and the poor consideration given to the pharmaceutical aspects of drug administration. A recent literature survey reported that the assessment of doses actually administered was mentioned only in a small minority of pharmacokinetic trials published in clinical pharmacology journals, while about one quarter of those trials met criteria indicating a theoretical relevance for such an assessment [33]. No mention of procedures for actual dose assessment is found in the recommendations of registration authorities for the performance of pharmacokinetic studies. However, current pharmacopeia requirements set acceptance criteria to  $\pm 10\%$  around the nominal strength for the pharmaceutical quality control of most manufactured drugs; moreover, analytical methods used in quality control are usually not identical to assays used for pharmacokinetic determinations, which can quantify similar drug amounts fairly differently (especially when based on immunological reactivity and/or applied to complex agents or biologicals); salt factors may be neglected; and finally, the preparation of solutions for infusion introduces a supplementary risk of dilution errors, leakage, residues, adsorption on syringes or lines, or inaccuracy in infusion rate and volume. All those factors can be expected to introduce significant bias in the determination of pharmacokinetic or pharmacodynamic parameters if the calculations rely upon the labeled dose instead of the best estimate of the dose actually administered. The following pragmatic steps are thus recommended to implement actual dose assessment when necessary in pharmacological trials involving human subjects:

- While dose determination is certainly not necessary in all kinds of trials, the question of the relevance of such assessment should be raised for each study at the time of protocol elaboration. Various criteria should suggest investigators to plan such procedures in a given trial: use of intravenous injections or infusions, experimental drug manufactured on a small scale, especially peptides, proteins, or other complex agents (e.g., phytopharmaceuticals), immunologically based analytical methods (e.g., RIA or EIA), parallel-group design with qualitatively different treatment assigned to each group, and estimation of a drug absolute clearance, volume, or effective dose parameters (the calculation of which include a dose term).

- For trials using a ready-for-use drug formulation, such as an oral tablet, a global evaluation of the average actual dose present in the trial drug is usually sufficient. Pharmaceutical quality controls should be checked, and their results should be used if they indicate any significant departure from the nominal dose of the formulation. The identity and actual titer of the substance used to prepare the calibrators for drug assay in biological fluids should be checked as well (taking a possible salt factor into account). For complex preparations such as biological agents, one may consider determining actual dose amounts using the assay method developed for plasma samples.
- For trials using a drug in solution, typically for intravenous infusion, the actual dose delivered to each study subject should be assessed individually. First, the volume of solution actually infused will be carefully measured and recorded (use the indications given by an electronic pump after a preliminary validation of its flow rate; alternatively, weight precisely the solution vials or bags connected with lines before and after drug administration, and divide the difference by the specific gravity of the solution measured on a sample). Second, the concentration of active drug in the infusion solution will be determined using the assay applied to plasma samples (which must have been validated for aqueous matrix). Whenever feasible, prepare more infusion solution than required, and retain samples from each treatment at the outlet of the line (after postinfusion weighting).
- The differences between assigned dose levels and the amounts actually administered should be reported, and the cause for significant departures should be elucidated. Attention should be paid to the accuracy and precision of the measurement and assay methods used for dose assessment, to avoid the unnecessary introduction of variability in PK or PD calculations (the reproducibility should be reported as standard deviation or coefficient of variation).
- Whatever the approach, the procedures applied to actual dose assessment should be described in study reports and publications, and the results used in the calculation of PK or PD parameters.

In PK experiments, it is important to properly standardize the conditions of drug administration. Unless specified by the protocol, oral dose intake should occur on an empty stomach, hence the subjects are to remain fasted over several hours before. A predefined amount of water is to be given with the oral dose. Even the posture of the subjects should be standardized, as gastrointestinal absorption, hepatic first pass, and renal excretions are influenced by body position. Similarly, the site of a subcutaneous drug injection is to be standardized across the whole trial. Further precautions are necessary for topical drug administration. Finally, it is important to hold a comprehensive drug accountability throughout the study, and to preserve all drug containers and packages until the end of the study.

**Sample Collection** Study protocols of PK trials usually define a series of sequential sampling times. The goal is to cover with sufficient thoroughness the concentration curves to be described. Ideally, the duration of PK follow-up should last over at least three to four half-lives of the drug. Sampling should be more frequent immediately after drug administration and around the peak time or immediately



after infusion cessation; thereafter, the samples can be separated by larger time intervals. However, the terminal phase is often the most relevant part of the PK profile to be described, as it determines effect duration and accumulation.

Several pragmatic recommendations deserve to be mentioned with regard to sample collection:

- Prepare a gallery of sampling tubes with appropriate labels, along with a worksheet to be filled in during the experiment. The worksheet comprises a sample record log (Fig. 1). Ensure that the labels correspond to worksheet designations for each tube and that consistent designation rules apply to all subsequent procedures.
- Record on the worksheet the precise clock time of drug administration(s) and blood samplings. While the protocol defines relative sampling time after dose administration, it is convenient to write down on the worksheet the absolute times scheduled for sampling (which depend on the actual dosage time). In case of an infusion, the worksheet will mention start and end times, infusion flow rate, weight of the pouch before and after infusion, and collection of a retain sample.
- For repetitive blood samplings, it is suitable that the subjects are equipped with a venous cannula with cap and lateral valve. Large forearm veins are best suited to draw blood (avoid antecubital and dorsal hand veins for subjects convenience). In case of a drug infusion, use different venous accesses for drug administration and blood sampling, as sampling through the same line has every chance to contaminate the samples with the infusion solution (which is about 1000 times more concentrated).
- Before drawing samples, discard 0.5 mL of blood (to flush the cannula dead volume); immediately after sampling, recap properly the cannula and rinse it with 2 mL normal saline through the lateral plug. Normal saline is sufficient to main the permeability of venous access over several hours; anticoagulants are unnecessary and may only cause problems (interference with assays, sensitization). Tell the subjects to ask for supplemental rinsing in case of blood reflux observable in the tip of the cannula.
- When collecting a blood sample takes a significant amount of time, record the midsample time on the worksheet and mention this difficulty in a remark. Whatever the system used for blood collection, drawing blood gently rather than trying to pull it out with a strong suction pressure will always do it better. Blood collection may be facilitated by warming the arm bearing the cannula (as a cold skin associated with increased venous tone and decreased blood flux through superficial veins). If it is impossible to obtain blood, report the problem and keep the empty tube in the series.
- If urine collections are scheduled in the study protocol, record precisely each micturition time, and measure the total collection volume before taking samples (it is advisable to weight precisely the recipient and subtract the tare; as urine density varies between 1000 and 1030 g/L, the measurement error is not expected to exceed 3%, which is often more precise than direct measurement of the volume in a column). Offer the subjects a quiet place and ask them to empty their bladder as completely as possible. Ensure sufficient provision of water



| Fluoxetine Bioequivalence Study – Clinical Investigation Unit<br>Worksheet for Sample Collection |                   |                        |                     |      |                                         |
|--------------------------------------------------------------------------------------------------|-------------------|------------------------|---------------------|------|-----------------------------------------|
| Subject                                                                                          | LAR               | AN                     | 101                 |      |                                         |
| Period                                                                                           | 1                 | Date                   | Apr 14, 05          |      |                                         |
| Sample label                                                                                     | Relative time (h) | Scheduled time (hh:mm) | Actual time (hh:mm) | Visa | Remarks                                 |
| -30 min                                                                                          | -0.5              | 07:30                  | 7:43                | TB   |                                         |
| (dose intake)                                                                                    | 0                 | 08:00                  | 8:12                | TB   | Dose intake with 200 ml water; OK       |
| 1.0 h                                                                                            | 1                 | 9:12                   | 9:12                | TB   |                                         |
| 2.0 h                                                                                            | 2                 | 10:12                  | 10:12               | TB   |                                         |
| 3.0 h                                                                                            | 3                 | 11:12                  | 11:17               | TB   | Delay due to difficult blood collection |
| 4.0 h                                                                                            | 4                 | 12:12                  | 12:12               | TB   | Subject mentions nausea (see AE form)   |
| 5.0 h                                                                                            | 5                 | 13:12                  | 13:12               | TB   |                                         |
| 6.0 h                                                                                            | 6                 | 14:12                  | 14:12               | TB   |                                         |
| 7.0 h                                                                                            | 7                 | 15:12                  | 15:12               | TB   |                                         |
| 8.0 h                                                                                            | 8                 | 16:12                  | 16:12               | MA   | Difficult blood collection              |
| 9.0 h                                                                                            | 9                 | 17:12                  | 17:12               | MA   |                                         |
| 10.0 h                                                                                           | 10                | 18:12                  | 18:18               | MA   | Inadvertent delay                       |
| 12.0 h                                                                                           | 12                | 20:12                  | 20:05               | MA   | Subject on a hurry to leave             |
| 24.0 h                                                                                           | 24                | Apr 15, 8:12           | 8:22                | TB   | Inadvertent delay                       |
| 48.0 h                                                                                           | 48                | Apr 16, 8:12           | 8:12                | TB   |                                         |
| 72.0 h                                                                                           | 72                | Apr 17, 8:12           | 8:12                | TB   |                                         |
| day 5                                                                                            | 96                | Apr 18, 8:12           | 8:18                | MA   |                                         |
| day 7                                                                                            | 144               | Apr 20, 8:12           | void                | TB   | Subject forgot the visit; unreachable   |
| day 14                                                                                           | 312               | Apr 27, 8:12           | 9:25                | TB   | Subject late on visit; warned           |
| day 21                                                                                           | 480               | May 4, 8:12            | 8:12                | MA   |                                         |
| day 28                                                                                           | 648               | May 11, 8:12           | 8:17                | JB   |                                         |
| day 35                                                                                           | 816               | May 18, 8:12           | 8:12                | MA   | Subject late on visit                   |
| day 42                                                                                           | 984               | May 25, 8:12           | 9:02                | MA   | Subject late on visit                   |

**FIGURE 1** Example of worksheet for sample recording over one period in a PK trial, where 30 subjects received twice 20 mg of fluoxetine orally (Prozac versus generic) at 2-month interval.

over the investigation. Calculate the duration of each collection as the difference between end of collection time and initial micturition time; recall that the initial urine does not belong to the collection and is to be discarded. If a subject needs to urinate before the scheduled collection end, save that urine in the container. If a subject forgets to give his or her urine, ask for an evaluation of the missing volume and record the deviation on the worksheet.

- Process the sample tubes as soon as possible and store them in structured racks and boxes, at the appropriate temperature. It makes sense to dispatch the plasma or serum of each blood sample into two tubes, so as to obtain two parallel collections, including one for backup. Ensure that the storage conditions are respected up to sending the samples to the lab. Temperature “spy” recorders may help in monitoring the storage conditions. Refrigerators should be connected to an alarm system to prevent unrecognized malfunction.
- Have all samples shipped to the laboratory, along with the list of tubes. The study samples will most often need to be shipped frozen on dry ice, and packaging issues are of critical importance. The lab should immediately check the completeness of the sample series. If possible, all samples of one subject should be analyzed on the same day.

**Description of PK Results** Once the lab has produced measurement data for the series of samples, after careful quality check, the first step is to produce graphics of *individual time–concentration profiles* (Fig. 2). Not infrequently, the graphics will identify remaining errors in data handling or suspect data points requiring reanalysis. The actual sampling times are to be used to draw individual PK curves. In most situations, the use of a log *Y* scale is advisable, as it will linearize the exponential components of the time–concentration relationship and indicate the most suitable model to describe the data. The issue of results below quantification limit (BQL) has been abundantly discussed in the literature, and many authors have issued recommendations on that specific question. When sufficient amounts of data are

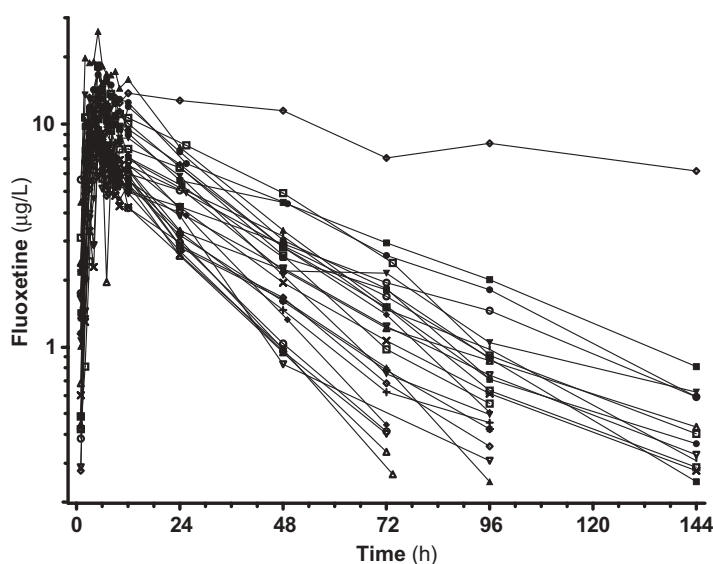


**FIGURE 2** Individual graphic of raw data in a PK trial, where 30 subjects received 20 mg of fluoxetine orally (generic formulation). A log *Y* scale is used. Observe several delayed samples, one suspect point at 6 hours, the fairly log-linear shape of the terminal phase, the decrease in relative measurement precision at low concentration levels, and the presence of one slow eliminator among the subjects.

available (e.g., in Fig. 2), the best solution is to merely drop such data points from individual curves and calculations. For the calculation of means and during population analyses, a simple approach is to equate the first BQL point to half the value of the quantification limit, drop all subsequent points, and force the residual error to include at least the range between zero and the quantification limit. This will somewhat correct a selection bias of data points, which would otherwise lead to overrepresentation of higher values over BQL, and thus to overestimation of terminal half-life.

In a second step, an *average time–concentration profile* can be computed and plotted (Fig. 3). It is highly recommended to compute geometric means with corresponding coefficients of variation and to plot them along a log Y scale (see Appendix ). Several average curves can be compared on a common graph when the PK experiment involves different doses, formulations, co-medications or subgroups of subjects (according to gender, age, genotype, clinical condition, etc.).

Most sample-rich PK trials will apply a “two-step” approach in data analysis, that is, the characteristic values describing the PK profile are first calculated for each subject individually, and then submitted to statistical tests. The classical descriptors of PK profile curves are as follows:

- The *maximum concentration* ( $C_{\max}$ ) corresponds to the highest observed value in the curve. This descriptor is of significance mainly for drugs administered through an extravascular route (oral, subcutaneous, etc.). It is clearly dose dependent, and it represents for some drugs the best predictor of pharmacodynamic efficacy.

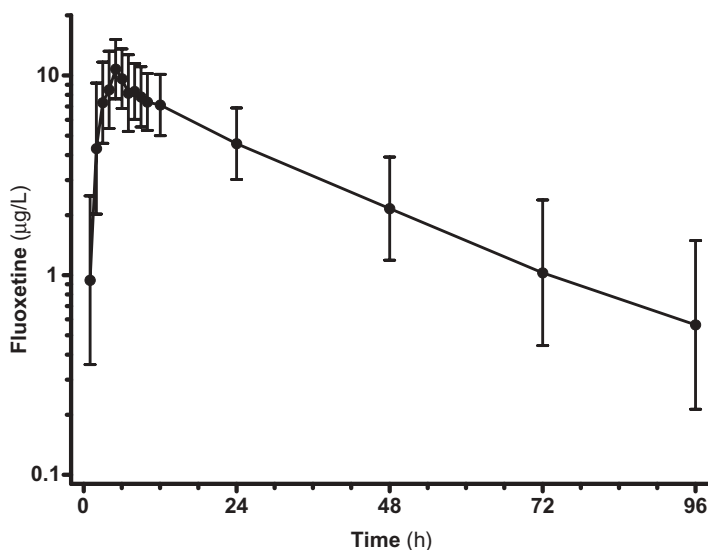


**FIGURE 3** Individual graphic of average data in a PK trial, where 30 subjects received 20mg of fluoxetine orally (generic formulation). A log Y scale with geometric mean and standard deviations is used. Based on the graphic of individual values (Fig. 2), the characterization of average curve appears appropriate up to 96 hours after dose. Observe the relative homogeneity of the dispersion, however, slightly increasing toward low concentration levels.

- The *time of maximum concentration* ( $t_{\max}$ ) is merely the time corresponding to the observation of  $C_{\max}$ . It is affected by both the rate of absorption and the rate of disposition of the drug, according to the pharmaceutical formulation studied. It is usually determined with less precision than other PK descriptors, due to the sparse distribution of sampling times and to the sensitivity of the concentration plateau to slight measurement errors.
- The *rate constant of the terminal phase* ( $\lambda_z$ ) is the absolute value of the slope of log-transformed concentration values versus time (only natural logarithms must be used). It is computed by linear regression over the terminal part of the curve, including all points that appear sufficiently well aligned on the graph (e.g., in Fig. 1, the data points would be included beyond 10 hours after dose intake). Various algorithms have been devised to select formally the points to be included, but visual inspection is often sufficient. When a drug is given at different doses, the quantification limit of the assay often restricts the possibility of estimating  $\lambda_z$  at the lowest doses, producing artificially high  $\lambda_z$  values; this must be differentiated from the case of true PK nonlinearity, where increase in dosage actually slows down drug elimination.
- A distribution phase is sometimes identifiable on the log concentration versus time curve, translating into a steeper initial segment; this is mainly the case after intravenous drug injection. A *rate constant of the initial phase* ( $\lambda_1$ ) can be computed in such situations.
- After administration of an intravenous bolus dose, a *concentration at time zero* ( $C_{(0)}$ ) can be back-extrapolated as the intercept of the log-linear regression line fitted over the points of the initial phase.
- The *area under the curve* (AUC) is the most important descriptor to calculate after single-dose administration. It is estimated by numerical integration over all segments of the concentration curve. It is dose dependent, and it represents for many drugs the best exposure marker correlated with pharmacodynamic efficacy. The *trapezoidal rule* is often used to calculate AUC segments: if a concentration  $C_i$  is measured on time  $t_i$ , followed by a concentration  $C_j$  on time  $t_j$ , the corresponding segment has the area

$$\text{AUC}_{i-j} = (t_j - t_i) \left[ \frac{1}{2} (C_j + C_i) \right]$$

During the decay phase, especially when samples are taken at sparse intervals, the trapezoidal rule tends to overestimate segment areas, as their upper boundary is in fact curvilinear; using the *log-trapezoidal rule* is thus considered as preferable:

$$\text{AUC}_{i-j} = (t_j - t_i) \left[ \frac{C_i - C_j}{\log(C_i) - \log(C_j)} \right]$$

The sum of all segments is then calculated from the first to the last observed concentration ( $\text{AUC}_{\text{last}}$ ). However, this sum still underestimates the true AUC, and it must be complemented with extrapolations. An extrapolation between dosage time and the first measured concentration will use  $C_{(0)}$  for the estimation

of the first segment area in case of an intravenous bolus injection; for other routes of administration, the concentration at dosage time can be considered to be zero, and the first segment of AUC can be equated to a triangle having an area of  $AUC_{0-1} = 0.5C_1/t_1$ . At the other end of the time–concentration curve, an extrapolation of the AUC toward infinity can be based on the last measurable point ( $C_{last}$ ) and the rate constant of the terminal phase:

$$AUC_{last-\infty} = \frac{C_{last}}{\lambda_z}$$

Some authors recommend to replace the observed  $C_{last}$  by a fitted value based on all points of the terminal phase. Both extrapolations  $AUC_{0-1}$  and  $AUC_{last-\infty}$  are added to the sum of segments to compute  $AUC_{0-\infty}$ , the best available estimate of total AUC. As the extrapolations represent the least robust part of AUC estimation, it is often required that they do not represent more than 10% of the total AUC, to ensure an accurate estimation of the concentration exposure. In repeated dose trials, when concentrations are followed over a dosing interval at steady state (i.e., once accumulation is complete), no extrapolation to infinity must be applied, as the AUC limited to one dosing interval is equivalent to the AUC to infinity, which would have been observed after a single dose.

- The *area under the first moment curve* (AUMC) is another descriptor of PK curves used sometimes. The first moment curve is the curve of concentration values multiplied by corresponding sampling times. Its integration can be interpreted as an exposure marker assigning bonus to concentration levels according to their duration. The trapezoidal rule for AUMC estimation is

$$AUMC_{i-j} = (t_j - t_i) \left( \frac{t_i C_j + t_j C_i}{2} \right)$$

A log-trapezoidal rule has also been proposed but it is far more complicated and rarely used. The estimation of AUMC has also to be extrapolated toward infinity:

$$AUMC_{last-\infty} = \frac{t_{last} C_{last}}{\lambda_z} + \frac{C_{last}}{\lambda_z^2}$$

As this descriptor weights concentrations by sampling times, the percent extrapolated is often fairly high, which limits the confidence that can be placed on AUMC estimation.

**Noncompartmental PK Parameters** Once the descriptors mentioned above have been computed for each individual PK curve, they can serve to estimate various dose-independent parameters, taking into account the actual dose administered ( $D$ ), the administration route, and the study design [34]. The following PK parameters are usually calculated:

- The *terminal half-life* is taken as the inverse of the rate constant of the terminal phase multiplied by the natural logarithm of 2 (i.e., 0.69):

$$t_{1/2z} = \frac{\log(2)}{\lambda_z}$$

In case of intravenous injection or rapid extravascular absorption,  $t_{1/2z}$  reflects the drug *elimination half-life*; however, for slow-absorption formulation, it may reflect the *absorption half-life*, which becomes the factor limiting drug disposition. Similarly, the terminal half-life of a drug metabolite measured in blood can reflect the kinetics of either its formation or its elimination. When an initial log-linear segment can be identified on the time–concentration curve, its rate constant can be used to estimate the *initial half-life*, which most often represents a *distribution half-life*.

- The *total drug clearance* is calculated as the ratio of the dose over the resulting AUC:

$$CL = \frac{D}{AUC}$$

Notice that after extravascular administration, this formula gives the *apparent clearance* ( $CL/F$ ) rather than the absolute clearance, as the concentration exposure results only from the fraction of the dose able to reach the circulation, which is the *bioavailability* ( $F$ ). When urinary determinations are performed, the amount of drug excreted into the urine ( $A_e$ ), which may require extrapolation toward infinity, can be used to determine the *renal clearance* of the drug:

$$CL_R = \frac{A_e}{AUC}$$

The *fraction of dose undergoing renal excretion* is calculated as the ration  $A_e/D$ . When a drug metabolite is measured in plasma or serum, the AUC of this metabolite is usually presented along with its half-life. However, to determine the fraction of the dose eliminated through the corresponding metabolic pathway requires that the absolute clearance of the metabolite is evaluated after intravenous administration of the metabolite itself. Alternatively, the fraction of the dose recovered in urine as metabolite can provide some indication. Otherwise, the metabolite AUC cannot be assigned a precise meaning, as it results from both the drug transformation into metabolite and the disposition kinetics of the metabolite, which are unidentifiable.

Finally, it is to be recalled that any estimation of a drug clearance refers always to a specific medium where drug concentrations are measured: for most drugs, the plasma clearance will significantly differ from the whole blood clearance. By default, total plasma is considered as the reference medium. There is growing interest for the determination of free drug concentrations in the plasma; for drugs highly bound to plasma carriers (albumin,  $\alpha$ -1-glycoprotein, lipoproteins, etc.), as free concentrations are very low, *free drug clearance* can

reach impressive values, especially in case of high liver extraction ratio, when the metabolic activity of the liver is able to displace the binding equilibrium in plasma during blood transit through the liver. Free drug concentrations are considered to be more tightly related to pharmacodynamic activity than total plasma levels.

- The determination of the *absolute bioavailability* of an extravascular formulation requires to compare the AUC after its administration with the AUC after intravenous injection of the active ingredient, accounting for possible dose differences:

$$F_{\text{extravascular}} = \frac{\text{AUC}_{\text{extravascular}} D_{\text{intravenous}}}{\text{AUC}_{\text{intravenous}} D_{\text{extravascular}}}$$

When two extravascular formulations are compared, this formula gives the *relative bioavailability* of one formulation with respect to the other (taken as reference). This is the key parameter estimated in bioequivalence trials.

- Finally, several types of *volume of distribution* can be defined. The *volume of the terminal phase* is simply derived from the total clearance and the terminal constant rate:

$$V_z = \frac{\text{CL}}{\lambda_z}$$

Notice that after extravascular administration, this formula gives the *apparent terminal volume* ( $V_z/F$ ) rather than the absolute one. After intravenous bolus injection, an *initial* or *central volume* can be deduced from the concentration extrapolated at time zero:

$$V_1 = \frac{D}{C_{(0)}}$$

When an AUMC value has been estimated after intravenous injection, it can be used to calculate the *mean residence time*:  $\text{MRT} = \text{AUMC}/\text{AUC}$ , a weighted estimator of the average duration of sojourn of drug molecules in the organism before their elimination. After intravenous infusion, the ratio  $\text{AUMC}/\text{AUC}$  is equal to the MRT plus one half the infusion duration. After extravascular administration, it is equal to the MRT plus the *mean absorption time* (MAT), which can thus be determined only by comparison with the intravenous administration. Knowing the MRT of a drug enables one to calculate its *volume at steady state* (i.e., when concentrations in the different compartments are at equilibrium):

$$V_{\text{ss}} = (\text{CL})(\text{MRT})$$

Like clearances, volumes of distribution always refer to a given measurement medium, and for agents highly bound to plasma carriers, the distribution volume of the free moiety can exceed many times that of the total plasma drug.



**Pharmacokinetic Modeling** Beyond the basic calculations outlined above, many more sophisticated calculations can be devoted to the analysis of PK data, depending on the study aims and conditions. A detailed description of all possible approaches is beyond the scope of this chapter [19]. Briefly, the models that are regularly applied belong to the following classes:

- *Exponential models* use nonlinear regression techniques to fit the data with a general equation, usually a sum of exponential terms; for example, a two-exponential concentration curve after intravenous injection will be described according to the equation:

$$y = Ae^{-\alpha t} + Be^{-\beta t}$$

As PK data are characterized by noticeable heteroscedasticity (i.e., heterogeneity in variability), the issue of appropriate weighting schemes for the regression is of relevance. The coefficients estimated ( $A$ ,  $\alpha$ ,  $B$ ,  $\beta$ , termed PK *macroconstants*) enable the calculation of classical PK parameters ( $\lambda_1$ ,  $\lambda_z$ , CL,  $V_1$ ,  $V_z$ ). Notice that the model equation can also be built up using directly clearance and volume parameters. Parameter values estimated this way not infrequently show fair differences with their noncompartmental counterparts (with the exception of CL, which is the most robust regarding estimation methods). Exponential models are ideally suited to draw continuous curves across the data points and calculate various simple extrapolations (e.g., dose repetition).

- *Compartmental models* aim to fit the data with a set of differential equations. The equations characteristically describe the evolution of drug amounts throughout body compartments. For example, the bicompartmental disposition of a drug after intravenous injection will be described according to the set of differential equations:

$$\frac{dY_1}{dt} = k_{21}Y_2 - k_{12}Y_1 - k_{10}Y_1$$

$$\frac{dY_2}{dt} = k_{12}Y_1 - k_{21}Y_2$$

where  $Y_1$  is the amount of drug in the central compartment (related to the concentration through the volume  $V_1$  of this compartment), and  $Y_2$  the amount of drug in the peripheral compartment. The coefficients ( $k_{12}$ ,  $k_{21}$ ,  $k_{10}$ ,  $V_1$ , termed PK *microconstants*) enable as well the derivation of macroconstants and classical PK parameters. Compartmental models can accommodate a wider spectrum of conditions, such as time-varying infusion rates, first-pass effect and enterohepatic cycle, multiple metabolites production, or situations where some PK processes are nonlinear, for example, saturable elimination pathways following Michaelis–Menten kinetics. The results are more susceptible to be interpreted mechanistically and can be used in simulations where PK mechanisms are selectively modified (e.g., prediction of the effects of a change in formulation or organ function).

- *Physiological models* try to incorporate physiological variables, such as anatomical volumes of tissues and organs, or circulation flow rates. While those models are mainly used in animal experiments, they can be useful in selected clinical situations, such as the determination of drug kinetics through an organ, either natural or artificial (e.g., renal replacement machines), especially when it is accessible to measurement (e.g., cerebral imaging of drug distribution into the brain).

While the basic PK calculations described above can be achieved using a simple spreadsheet, fitting exponential, compartmental, or physiological models requires specific computer facilities. Numerous pieces of software are available to perform such advanced PK data analyses, most of them being represented on the Internet (see, e.g., the online catalog maintained by Bourne [35]). Among such programs, WinNonLin and Kinetica currently experience a high level of recognition among pharmacokineticists. In addition, several computer packages for general statistics have now implemented specific routines for PK data analysis (e.g., S-Plus, R, and Stata). Nowadays, the limiting factor for a sound exploitation of PK data using computer resources is only the degree of knowledge and experience required from users.

**Statistical Analysis of PK Results** Once the PK parameters have been deduced from the raw concentration data, they are ready for description, analysis, and comparisons according to the study objectives and design. It is standard practice to compare PK parameters rather than the concentration results themselves. Many specialized PK computer programs incorporate statistical tools to perform tests on the parameter estimates that have been calculated. Alternatively, the parameters may be copied into a standard statistical computer package.

If the main study objective is a mere *description* of a drug PK profile, the importance will be put on the summary table of PK parameters. Several choices are available for statistical description. The calculation of geometric means and associated percent variations should be encouraged, as it is often the most appropriate alternative (see Appendix). Nevertheless, arithmetic means and standard deviations are still often encountered. When parameter values appear to follow an irregular distribution pattern, nonparametric descriptors should be preferred (i.e., median and interquartile range). The descriptive presentation of PK results is often complemented with some statistical comparisons of the parameters according to various factors, such as gender, age, or the genotype regarding given polymorphisms. Such tests must be viewed as *exploratory*, that is, the *P* values have only an indicative interest, and no Bonferroni-like correction is applied for multiple testing.

When a PK trial is built up toward the assessment of a given hypothesis, the main study objective must be specified in terms of one or a few PK parameters to be tested regarding the influence of relevant factors, and *confirmatory* statistical tests must be predefined in the protocol. A preference should be given to parametric tests after preliminary log transformation of the parameters. For example, in a metabolic interaction study following a cross-over design, the effect of a given treatment on the log of the clearance of the tested drug is to be assessed through an analysis of variance; in a renal failure study including patients with various degrees of renal impairment, the correlation of the log of the clearance of the tested drug with the log of the glomerular filtration rate is to be evaluated. When the distribution of PK

parameters departs significantly from a log-normal, nonparametric tests should be applied instead.

A particular case is represented by *bioequivalence* trials, which are intended to establish the absence of clinically relevant difference between two formulations of the same agent. In this situation, it is not sufficient to show an absence of statistically significant differences. The ratio of the parameters of interest ( $AUC$ ,  $C_{\max}$ ,  $T_{\max}$ ) between both formulations is to be calculated in each study subject, and the bioequivalence can be claimed only if the statistical confidence interval at the 90% level around the geometric average of this ratio lies entirely inside a predefined range. Registration authorities set this range between 0.8 and 1.25 for  $AUC$  and  $C_{\max}$  and between 0.7 and 1.4 for  $T_{\max}$  [36, 37]. This is roughly equivalent to performing two one-sided  $t$  tests, and to conclude that the probabilities that the “true” ratio lies either below the lower limit or above the upper limit are less than 5% each [38]. A similar approach can be used to establish the absence of influence of food or of a potentially interacting drug on the PK profile of a given agent. Another investigation oriented toward showing an absence of difference is the assessment of PK linearity, usually performed by comparing dose-corrected  $AUC$  values. Here again, all the statistical calculations are to be done on log-transformed values (as suggested by the asymmetrical acceptance range for the confidence interval around the ratio of parameters of interest).

### 18.2.3 Population PK Studies

Population PK studies differ in many aspects from the classical sample-rich trials, while their general aims remain similar, namely to characterize the absorption and disposition profile of drugs in humans. They must rely as well on efficient, reliable, and validated analytical methods. Their importance is increasingly recognized in both academic circles and registration authorities, and they are progressively invading drug development [39]. They are not aimed to replace, but rather to complement, the small-sized trials performed during early drug development in healthy volunteers. They can be easily engrafted, indeed, on other types of studies, such as phase II or phase III therapeutic trials.

***Specificities of Population PK Study Protocols*** While classical PK trials mainly derive from the tradition of physiological experimentation, population PK studies owe something to epidemiological concerns: drug characteristics should be established from observations gathered in field conditions, where the drug is employed for therapeutic purposes. This explains substantial differences in study protocols and data analysis techniques [40].

The *study subjects* will thus preferably be representative patients having to receive the drug for a medical condition, rather than healthy subjects selected according to strict normality criteria. This does not mean that population approaches cannot be used in phase I trials: They can indeed, only with less relevance. Subject heterogeneity, which is considered as a nuisance in experimental trials, is regarded in population studies as a valuable opportunity to assess the variability of PK parameters and to identify determinant factors affecting their value. A condition, however, is to carefully record what the relevant subject characteristics are (gender, age, body weight, medication, morbidity, renal and liver function tests, genotype results for

polymorphic enzymes or transporters, etc.), so that they can be used as explanatory *covariates* during data analysis.

Population *study designs* are also less stringent: While healthy subjects are able to undergo intensive investigations, involving numerous blood samplings over a short time, medical patients should not be exposed to cumbersome study protocols. Specific population data analysis techniques can accommodate sparse-sample designs. Therefore, instead of obtaining a lot of measures in a restricted group of subjects, population studies will tend to include a larger number of patients (several dozen at least), with each patient providing only a limited number of observations (usually between 3 and 12). This makes, for instance, population studies represent the only way to address pharmacokinetic issues in pediatrics. Moreover, neither the dosing regimen nor the blood sampling schedule need to be rigorously standardized, as population techniques will accommodate unevenness in doses, intervals, blood sampling times, and even administration routes. So the study patients may take their drug and undergo venipuncture on convenient times (this may be initially accepted with perplexity by health professionals, who often associate clinical research with the requirement for strict timetables). Of prime importance is, however, the thorough recording of all actual dosing and sampling times. Dose determination issues need also to be considered (see above). In addition, compliance may represent a noteworthy problem, while it is rarely an issue in classical PK experiments. Compliance determination using electronic devices may thus be suitable in some population PK trials.

This flexibility makes population PK studies relatively easy to append to a variety of clinical trials. In particular, phase II and phase III therapeutic trials may clearly benefit from this kind of extension (which can be applied only to a subgroup of patients, e.g., those in the hospital). This will help to assess concentration–efficacy and concentration–toxicity relationships in the target patient population, to uncover causes of treatment failure or toxicity, to design dosing adaptation strategies for specific patient subgroups, and to optimize rationally the medical utilization of new drugs. A further step will be to pool the PK data gathered during all phase I, II, and III trials of a new drug into a large database, and run a systematic population analysis with proper exploration strategies and confirmation cycles, which will extract all clinically relevant knowledge from concentration data. This has been achieved for some newly developed drugs; however, it represents nonnegligible technical requirements and expenses, in terms of data management, computer facilities, and specialist time [24].

In summary, population PK studies are remarkably flexible toward patient inclusion criteria, drug administration, and sampling time schedules, but this implies that each blood sample tube comes associated with a set of carefully recorded clinical data. The data to record vary from one study to another, and the following tentative list is only indicative:

- Patient identity and unique number in the study
- Study group and investigation center, when relevant
- Usual dosing regimen regarding the study drug
- Last dose administered, along with precise dosing time and other relevant details (e.g., route, formulation, infusion rate, compliance assessments, etc.)

Accurate sampling time

Demographic covariates, such as gender, age, body weight, height, and ethnicity  
Scores of organ insufficiency (heart, kidney, liver), global functional scores (in chronic disease patients)

Biochemical markers of organ function, such as typically creatinine or bilirubin  
Concentration of plasma carriers (albumin,  $\alpha_1$ -acid glycoprotein), for drugs highly bound

Genotype results regarding relevant polymorphisms susceptible to affect drug disposition (whenever tested, which is increasingly the case)

Concomitant medications, to be classified at the end of data collection according to their potential for interaction with the study drug (e.g., inhibitors of inducers of a given cytochrome P540 isoenzyme)

**Analysis of Population PK Data** The analysis of a set of observations collected during a population PK study represents a nontrivial assignment. It requires both specific computer facilities and specialized skills [40]. An analysis will include in a single run the PK data obtained in all study subjects. The system will fit them with a family of PK models, based on a common equation (or system of equations), with variable parameters allowed to take different values between each subject.

Formally, if we call  $f$  the common equation, without specifying it, the model for the  $j$ th concentration value observed in the  $i$ th subject will correspond to the general expression:

$$y_{ij} = f(x_{ij}, \phi_i) + \varepsilon_{ij}$$

where the  $x_{ij}$  are the *fixed effects* that apply to sample  $y_{ij}$ , such as dose and relative sampling time, the  $\phi_i$  are the *individual PK parameters* values that characterize subject  $i$ , and  $\varepsilon_{ij}$  is a *residual error* (i.e., the difference between model prediction and actual observation). The errors  $\varepsilon_{ij}$  are assumed to be randomly distributed around zero with a variance  $\Sigma$ , which quantifies the residual or *intrasubject* variability, a mixture of measurement errors, biological oscillations, and model misfit. More complex error structures may be used, for example, involving both addition and multiplication of error terms and necessitating to define  $\Sigma$  as a matrix of variance components. This model for observations is complemented with a second layer of models, which express the variability of the subjects PK parameters and their dependency toward individual factors:

$$\phi_i = g(z_i, \theta) + \eta_i$$

where the  $z_i$  are the *covariates* characterizing subject  $i$  (which may even vary along time:  $z_{ij}$ ), the  $\theta$  are the average *population PK parameters* including constants that correlate individual parameter values with covariates, and  $\eta_i$  is a *random effect* affecting the parameter value in subject  $i$  (i.e., the difference between its “actual” value and the value predicted equation  $g$ ). The random effects  $\eta_i$  are distributed around zero with a variance  $\Omega$ , which quantifies the unexplained *intersubject* variability, while equation  $g$  captures the part of variability that is explained by the covariates. Here again, additive and multiplicative random effect models may be

used. Moreover, there may exist some degree of correlation between the random effects affecting various PK parameters, necessitating to define  $\Omega$  as a variance–covariance matrix.

Thus, fitting a population model represents a problem of hierarchical, or *mixed-effect* regression, which involves the presence of two levels of errors (a third level is sometimes introduced, leading to distinguish between intersubject, interoccasion, and residual variability terms). The individual PK parameters  $\phi_i$  cannot be directly measured and represent latent variables. In general, the mathematical expressions for the functions  $f$  and  $g$  above are not reducible to linear equations. The population PK parameters to be found are not only the values of the constants  $\theta$  but also the components of the variance matrices  $\Sigma$  and  $\Omega$  (hence the designation of *pharmacostatistical model*). Their estimation requires one to decide on appropriate fitting criteria and to apply efficient numerical algorithms. Many choices must be done regarding the equations  $f$  and  $g$ , the inclusion of influent covariates among the available  $z_i$ , and the acceptability of various underlying assumptions. That is why a series of sequential models are usually adapted to a given set of population PK data, requiring the adoption of appropriate decision rules, and giving utmost importance to a rational and clinically guided model building strategy.

Several software tools have been developed for this type of calculations [35]. NonMEM, developed by L. B. Sheiner and S. L. Beal at UCSF, remains the most widely accepted program in the field. While depending on a whole set of assumptions often difficult to validate, it provides a highly flexible framework for population data analysis, which can accommodate a unrivaled variety of study conditions. Several computer tools have been devised to facilitate NonMEM operation, lacking notably from user-friendliness (Wings, PDx-POP, Xpose). Other tools are currently available for population PK analysis, which use different algorithms and will suit a more restricted range of study problems (WinNonMix, P-Pharm, Kinetica, WinBUGS, Monolix, PopKinetics, USC\*Pack, etc.). Extension modules for population PK analysis are also available in all-purpose statistical packages (NLME for R and S-Plus, NLMIX for SAS, etc.).

Without going into much detail, we just summarize here the main concrete steps of a population data analysis [40]:

- The preparation of the data set may take a while, as stringent requirements of the computer program must be followed. The drug measurement results must be entered in a structured way, along with all corresponding clinical data (subject identification, dosing schedule, sampling times, and covariate values). Automatized extraction from a database is tricky to set up and requires careful checking.
- The simplest possible population model, chosen according to preliminary knowledge of the drug PK profile, should be adapted initially, without including any covariate. The resulting set of population PK parameters will serve as a baseline for further model building.
- A set of decision rules for the acceptance of refined models should be chosen. It is usual to consider indices of goodness-of-fit such as the improvement in log-likelihood or Akaike information criterion, diagnostic graphics such as scatter-plots of predictions versus observations or of residuals versus predictions (taking

both the individual and the population predictions), and commonsense arguments such as the clinical relevance of refinements introduced in the model.

- Usually, model building will begin with the search of an appropriate structural PK model, that is, deciding on the number of exponentials or compartments. The measurement of metabolites or urinary excretion may further complicate the problem. A second issue is the choice of the most appropriate error model, both for intraindividual and interindividual variability. Thereafter, the sequential inclusion of covariates will be tested. *Covariate selection* may be valuably assisted by graphical explorations, showing possible trends between individual parameter estimates and known subject characteristics. It is important as well to be aware of the meaning of each sophistication proposed for the model and to judge its clinical pertinence. Covariate centering is usually recommended; for example, when testing a dependency of distribution volume toward body weight (BW), the expression

$$V = \theta_1 + \theta_2 \frac{BW - 70}{70} + \eta_V$$

will be more satisfactory than

$$V = \theta_1 + \theta_2(BW) + \eta_V$$

As in the first case,  $\theta_1$  represents the volume of a typical 70-kg individual and  $\theta_2$  the change predicted a doubling in body weight, while in the second case, no particular meaning can be attributed to  $\theta_1$  and  $\theta_2$ . Moreover, convergence issues are better dealt with using the first type of expression.

- Once the final model is established, it must undergo a formal validation. Internal validation will be based on careful examination of diagnostic graphics, standard errors and cross-correlations of each parameter, sequential removal of each covariate to test its marginal significance level, estimation of the lack of fit, and the like. Whenever possible, external validation on a different set of data is recommended. Overparametrization is often an issue, and the principle of parsimony should be kept in mind during data modeling.
- The equations and population parameters of the final model should be presented, along with the model building history. A graphical representation of the model is often useful to show the typical concentration curve expected in an average subject receiving a usual dosage; a 90% prediction interval should ideally be drawn around the typical curve, using either simulations or error propagation calculations. Sophisticated visual predictive check techniques are increasingly recommended.

### 18.3 PHARMACOKINETIC–PHARMACODYNAMIC EVALUATIONS

Most drugs induce measurable effects, which mediate their therapeutic efficacy, and at least in part the adverse reactions that they can cause. Ultimately, human pharmacology trials aim to describe, explain, and predict not so much the concentration



profile than the effect profile of pharmaceutical agents, taking into account the administration scheme and the clinical characteristics of recipients. It is suitable that clinical pharmacodynamic trials include drug concentration measurements whenever possible, as the pharmacokinetic phase of drug action in humans represents usually the major source of variability in pharmacological response [19]. PKPD trials thus tend to break down the analysis of dose–response relationships into the sequential assessment of dose–concentration (PK) and concentration–response (PD) relationships. This section presents some considerations on the latter part of the work.

### 18.3.1 Pharmacodynamic Variables

Whereas a single type of data is encountered in PK modeling, namely drug concentration measures, PD trials happen to deal with a variety of response modalities. The global aims of both PK and PD investigations in humans remain, however, to record the profile of the respective variables along time. Hence many methodological considerations discussed above for PK studies apply similarly to PD and PKPD trials: necessity to develop and validate accurate, sensitive, specific, and reproducible measurement methods, importance of dose determination, accurate recording of dosage and measurement times, choice of an appropriate study design with respect to the scientific questions addressed, rational approaches for data description, interpretation and modeling, and the like.

All kinds of clinical, physiological, or biochemical measures can be employed as PD variables, as soon as they are shown to change in response to a given drug treatment. Their clinical relevance may, however, vary considerably. Certain variables correspond themselves directly to the *clinically relevant outcome* of a treatment, for example, the level of pain after administration of an analgesic. Other variables are used as *surrogate markers* for treatment outcome, based on sufficient preliminary evidence showing that they play a significant role in the pathophysiology of the disease under treatment and that their correction correlates with clinical improvement in the patient, for example, the levels of viremia and CD4-positive lymphocytes in chronic HIV infection. The evidence allowing a marker to be recognized as a valid surrogate for clinical outcome may be difficult to establish. For example, it has taken decades to confirm that blood pressure or plasma cholesterol levels, beyond their predictive potential regarding the occurrence of cardiovascular diseases, were amenable to pharmacological correction with an actual benefit for the patients; a famous counterexample is the case of ventricular arrhythmia, a condition apparently improved by antiarrhythmic agents, however, at the expense of increased patient mortality. Finally, mere *biomarkers* reflecting drug action are frequently measured during PKPD trials, however, without claim for a strong predictive value with regard to clinical outcome. Such markers may, however, help to understand the clinical pharmacology of new agents, to build up rational dosage strategies, and to develop useful treatment monitoring methods. Noticeably enough, some biomarkers are so tightly correlated with circulating drug concentrations that they can even be used for PK assessments: for example, circulating angiotensin-converting-enzyme (ACE) activity closely reflects blood concentration of ACE inhibitors.

The pharmacometric characteristics of PD variables are important to take into account when designing a PKPD trial. Whenever possible, *graded response markers* are preferable to accurately assess concentration–response relationships using

regression techniques and formal PKPD modeling. This kind of measures are therefore most frequently encountered in clinical pharmacology studies and include, for example, biochemical markers (e.g., glucose, lipids, electrolytes, hormones, cytokines, coagulation tests, etc.) and functional variables (e.g., electrocardiographic or electroencephalographic variables, spirometric flows, psychometric tests, etc.). The measurement of subjective symptoms can also be performed using continuous variables: This is the principal merit of *visual analog scales* (VAS), such as those used to assess pain or sleepiness. The subjects are presented with a 100-mm horizontal line printed on a paper, labeled with “none at all” on the left end and “worst imaginable” on the right end. They are instructed to place a mark on the line to report the intensity or quality of the sensation being experienced. The distance between the left end and the mark is recorded precisely and provides a suitable quantification of the subjective response. Symptoms and clinical signs may also be recorded using validated scores, typically on a 1–4 scale corresponding to clear definitions, for example, for the assessment of tolerability and adverse events (1, mild; 2, moderate; 3, severe; 4, life threatening). Such *ordinal variables* deserve a different treatment during data analysis. Formal validation is not less important for scoring methods than for graded response markers, however, slightly different approaches are used, including test–retest and interinvestigator concordance assessments. Lastly, *binary variables* can intervene in PD studies, such as typically success/failure assessment, for example, for infection cure.

Most effect variables measured during PD and PKPD trials are less precise and less specific than PK measurements, thus making effect profiles more irregular and noisy than concentration profiles. Evaluating the *signal-to-noise ratio* of a response marker is one of the objectives of the procedure of validation and may help to choose the most efficient measure among several available ones. Highly variable markers further increase the requirement for appropriate data analysis approaches, able to deal with imprecision and to extract pharmacologically pertinent trends. The choice of an appropriate *scale* for marker measurement is of importance, as it may dramatically influence data analysis and results interpretation (e.g., expressing the response to antacids as a change in either pH or  $H^+$  concentration in gastric juice). Frequently, PD response measures start from a certain *baseline level* that is to be thoroughly assessed before drug administration. Moreover, the level of the response variable may display systematic *fluctuations* even in the absence of an active substance, for example, due to cyclic modifications (circadian, circannual), spontaneous disease evolution, psychological influences, and the like. Therefore, PD and PKPD trials will often include a placebo arm, while this is useless in purely PK trials. Effect variables may also be distinguished according to their degree of *reactivity* or inertia. Some markers quickly respond to drug exposure and return to baseline thereafter, while others integrate the response over prolonged times. A striking example can be found in osteoporosis treatment: A marked response to bone resorption inhibitors can be recorded over a few hours in healthy volunteers while following specific bone resorption markers such as plasma or urine collagen degradation products; on the other hand, a radiological follow-up of bone density in patients takes years to reveal significant changes in this marker, which seems, however, better correlated with the fracture risk. In addition, the sensitivity of a response variable may decrease over time, due to a *tolerance* phenomenon, such as a decrease in number and/or excitability of cell receptors (e.g., with opioids), or the induction of physiological

counteraction mechanisms (e.g., with  $\alpha$ -blocker antihypertensives). Inversely, some responses display a phenomenon of *sensitization*, meaning an increase in amplitude over prolonged exposure (e.g., drug addiction or allergy). Appropriate PKPD models are thus required to account for such time-related changes in concentration–response relationships. Finally, the description of a drug effect over fairly long time periods may necessitate to take into account the proper kinetics of pathophysiological events that characterize the disease and that modulate its response to treatment. Such *disease evolution modeling* represents an important research area, leading clinical pharmacology to meet the traditional questions of therapeutics.

### 18.3.2 Concentration–Response Relationships

**Pharmacodynamic Data Description** Effect measurement results gathered during sample-rich trials can first be described using simple calculations fairly similar to those developed for PK data description (see above). This basic approach will be useful to summarize the results, apply statistical comparisons, and reveal important trends. The following calculations are applicable even when the effect curve shows some retardation compared with the concentration curve, provided the effect has been followed over a sufficiently long time to capture the major part of its development and resolution:

- The *maximum effect observed* ( $E_{\max}$ ) corresponds to the highest marker value in the curve.
- The *time of maximum effect* ( $t_{E_{\max}}$ ) is merely the time corresponding to the observation of  $E_{\max}$ . It is rarely determined with good precision, as it is measured on effect plateau.
- The *area under the effect curve* (AUEC) can be estimated by numerical integration over all segments of the effect curve. The *trapezoidal rule* is generally used to calculate AUEC segments. As it is exceptional that a model is proposed to fit the terminal part of the effect curve, no extrapolation to infinity is usually added to the sum of AUEC segments.
- Even if some usefulness could be imagined for an area under the moment of the effect curve, no such calculations are usually performed.

When sufficient amounts of data are available, those PD descriptors can be submitted to simple statistical tests to assess the effects of dose, demographic factors, medical condition, co-medication, or genotype and to explore their correlation with the elementary PK descriptors. Among various questions in consideration, the identification of the best PK predictor for the effect is of relevance: For example, among the antibiotics, one can distinguish agents with an efficacy depending on peak concentration (e.g., aminoglycosides), on the time of concentrations exceeding a threshold (e.g., penicillins), or on the AUC (e.g., macrolides) [41]. When different effects depend on diverse PK descriptors, manipulating the dosing regimen will help to enhance the specificity of the treatment for a given effect. This is a way to optimize the efficacy/toxicity ratio for anticancer agents, for example. Finally, such basic explorations are useful as an orientation for appropriate model selection in further PKPD analyses.

**Pharmacokinetic–Pharmacodynamic Models** The simplest representation of the dependency between circulating drug concentrations and effect levels postulates merely an instantaneous correlation between both variables. This view characterizes the *immediate PKPD models*, where the development and the resolution of the effect are directly driven by the shape of the PK curve. In the situations satisfying the assumption of immediacy, a simple correlation analysis between simultaneous effect and concentration measures is sufficient to describe the effect profile. Only exceptionally is a *linear* correlation appropriate for an accurate description. In most conditions, the level of response tends rather to increase proportionally to the logarithm of the circulating concentration, making relative concentration changes more accurately predict the effect level than absolute concentration values. *Log-linear* relationships are thus sometimes used in immediate PD models to relate an effect to a circulating drug concentration. This model is, however, not entirely satisfactory, as it predicts negative effects at very small concentration values. Moreover, in many pharmacological situations, the effect is observed to saturate at the highest values of concentration. Thus, a hyperbolic equation derived from the receptor theory is most often used to describe clinical concentration–effect relationships. This is the  $E_{\max}$  model relating an effect  $E$  to the drug concentration  $C$ :

$$E = E_0 + \frac{E_{\max}C}{EC_{50} + C}$$

where  $E_0$  is the baseline response level predicted for a zero concentration,  $E_{\max}$  the maximal increase in response predicted at infinite concentration, and  $EC_{50}$  the concentration associated with one-half of the asymptotic maximal effect. This expression corresponds to a sigmoid curve when represented on a log  $X$  axis. Inhibition of a response variable may require subtraction of the right-sided term from the baseline  $E_0$ . The value of  $E_{\max}$  is traditionally considered to reflect the *efficacy* of the drug, while  $EC_{50}$  expresses its *potency*. These parameters can serve to compare various drugs among a pharmacological class. For a given drug, different  $EC_{50}$  values characterize various types of responses (e.g., therapeutic and adverse effects), which may help to assess the *therapeutic margin* and define a suitable therapeutic range. Notice that the parameters of the clinical  $E_{\max}$  model have little to do with their counterparts found during in vitro pharmacological experiments. A supplemental coefficient of steepness  $\gamma$  may be added in the concentration–response curve, giving *Hill's* or *sigmoid  $E_{\max}$  model* (initially developed to describe the affinity of oxygen for hemoglobin):

$$E = E_0 + \frac{E_{\max}C^{\gamma}}{EC_{50}^{\gamma} + C^{\gamma}}$$

Further PD models have been developed to account for less usual shapes of concentration–effect curves.

For binary response variables, a convenient description of the PK dependency uses the *logistic regression* model, which associates to each concentration level  $C$  a probability  $\pi(E)$  of successful achievement of the effect:

$$\pi(E) = \frac{e^{\alpha + \beta C}}{1 + e^{\alpha + \beta C}}$$

where the logit of  $\alpha$  gives the baseline probability in the absence of drug, and  $\beta$  indicates the log of the odds associated with a unit increase in concentration. A less sophisticated model sometimes used for binary responses is the step function curve, characterized by a single threshold concentration value, above which the effect is assumed to be achieved (e.g., the minimum inhibitory concentration for antibiotics).

Not infrequently, the expansion of the PD curve does not parallel the PK curve, and the effect appears with retardation, while still depending on the circulating concentration profile. In such situations, a concentration–effect scatterplot will not reveal a steady correlation but rather separate lines over the periods of concentration ascent and descent, respectively. Such a looping appearance of the concentration–effect curve calls for the application of *hysteresis PKPD models*. Several approaches have been devised to handle this phenomenon in modeling. The *link model* derives from the compartment theory and assumes that the circulating drug distributes in an effect compartment, where it induces the effect through an immediate relation as shown above [42]. Thus, the drug concentration at effect site is constructed as a latent variable. The retardation between plasma concentration and effect curves is explained by the time constant that determines drug accumulation and persistence in this peripheral effect compartment (this is in fact the rate constant of elimination from this compartment,  $k_{e0}$ ). This model is in accordance with physiological concepts, with receptors sometimes actually located in peripheral tissues, or receptors themselves found to progressively accumulate and slowly release ligands. Another class of models deserving mention are the *indirect models* [43]. These models, based on a compartmental approach, consider the effect marker as an endogenous substance, which is produced and eliminated with its proper kinetics. The drug administered is assumed to act either on the production or on the elimination rate of the endogenous marker, either to stimulate or to inhibit the process, through one of the direct concentration–effect relationships introduced above. While the drug concentration is supposed to influence immediately the rate of change of the endogenous marker, the resulting absolute marker level integrates this response over time with a definite inertia, thus accounting for the retardation between concentration and response curves. This model finds physiological justifications for a series of drugs known to act through second messengers, sometimes amenable to measurement (e.g., oral anticoagulants, which inhibit the synthesis of coagulation factors having specific turnover rates). It can also be applied to other situations where the turnover of a latent endogenous marker is nothing but a virtual representation.

A number of sophisticated PKPD models have been developed to describe more complex dose–concentration–effect relationships [44]. The production of a virtual antagonistic metabolite can be invoked to describe a phenomenon of tolerance, which is characterized by an antihysteresis loop. Feedback stimulation or inhibition loops can be added, either on PK or on PD processes. Further intermediary effect compartments can be designed. A complex interplay between the drug and living

pharmacological targets (e.g., viruses, cancer cells) can be modeled. Nonlinearity in PKPD models explain the schedule dependency shown by some drugs, meaning that the effect of a given total amount of drugs strongly depends on the way it is administered over time. Even paradoxical responses can be accounted for, with different administration schemes resulting in opposite effects (e.g., gonadotrophin-releasing-hormone analogs, which stimulate ovulation and hormone production when given in pulses but achieve a so-called chemical castration when released slowly from a depot formulation [45]).

### 18.3.3 Design and Analysis of PKPD Trials

As outlined above, PKPD studies are much more versatile than traditional PK trials due to the variety of measurable response markers, each of them having its own clinical relevance, pharmacometric properties, and endogenous profile. Hence PKPD study protocols have to be somewhat creative, but care must be given to have them produce good science anyway. The following recommendations deserve consideration for the design, execution, and analysis of PKPD trials:

- Precise study objectives should always be defined in advance, and one or a few of them should be specified as primary. The study design should be decided in accordance with the study aims, with consideration for preliminary knowledge of the characteristics of the drug tested. PKPD trials may involve single or multiple dosages, parallel or cross-over designs, and sample-rich or sparse-sample measurement schedules.
- The inclusion of a *placebo* arm is often useful in PKPD trials to assess the evolution of response markers in the absence of pharmacological agents. For trials testing an experimental treatment, the inclusion of a *verum* arm, that is, a reference treatment of known impact on the response markers, may be useful as positive control.
- A restricted choice of effect variables should be done, based on a selection of the most appropriate response markers. When a large number of response markers are available, the selection should be oriented to minimize the redundancy and to mainly retain variables that complement each other. For example, a rapidly reactive variable will be chosen along with a slowly integrative one. Marker measurement methods should undergo formal validation. A preliminary pilot study may be useful to check the feasibility of the methods before undertaking a large-scale trial.
- Whenever possible, basic robust descriptors should be calculated separately on both PK and PD results data, before initiating any formal PKPD modeling. In rich-sample trials, the questions corresponding to main study objectives are frequently addressed only with reference to such simple descriptors, while the fitting of a formal PKPD model is left among the secondary objectives. Individual and average response curves should be graphed for all study variables.
- In sparse-sample trials, population approaches are perfectly suited for data analysis and may work even better in this context than in purely PK trials. Most



computer facilities developed for this purpose allow the inclusion of both concentration and effect data, even though it makes the problem technically more demanding. *Population PKPD analysis* represents probably the best available tool for a rational exploitation of the results of human pharmacology trials [46]. Sometimes, only the PK data deserve a population analysis, while the PD data and PKPD associations are merely analyzed by simple statistical approaches applied to basic descriptors outlined above (e.g., AUEC), which can be correlated to the individual PK parameters produced by the population analysis. Another alternative is to first fit the data from the PK part, then fix all PK parameters, and then rerun the population analysis to estimate the PD parameters.

## 18.4 CONCLUSION

Clinical pharmacologists have the task to gather, produce and summarize relevant knowledge on the pharmacological profile of therapeutic drugs in order to optimize their medical utilization. Rational dosing regimens must be established, and consistent recommendations for dosage individualization must be issued. Many choices are to be discussed during early clinical drug development: selection of the best candidate among a family of compounds, route of administration, pharmaceutical formulation, dose level and dosing frequency, duration of treatment, dosage adaptation strategies (according to body weight, age, organ failure, coexistent diseases, interacting drugs or pharmacogenetic tests), requirement for some form of monitoring based on either biomarkers or drug concentrations, and the like. The answers to such questions should ideally be found before a candidate drug enters into large phase III confirmatory trials, where it should be ready for use as in clinical practice [4]. Considering all those questions, and the number of effect variables susceptible to be examined, one may envision vast drug development projects with hundreds of studies to explore any imaginable combination of choices. However, this is neither feasible nor really suitable. The aim of early clinical drug development is to find the most efficient pathway reaching the optimal combination of choices, across the tortuous hills and valleys of a multidimensional space of possible decisions. This requires to arrange a series of efficient, small-sized clinical experiments, each of them being designed according to the results of the previous ones. Exploration–confirmation cycles are continued until sufficient information has been obtained to devise rational dosing and monitoring recommendations. PKPD modeling and simulation represent highly valuable tools to improve this process [47, 48]. This effort does not end up at authorization to market. It is to be continued not only by drug manufacturers but also by academic researchers taking up the responsibility to raise critical questions and to challenge the therapeutic tools, so that the patients and the public obtain the best service from the drugs they buy. This is how clinical drug development definitely resembles the art of “response surface analysis,” to quote Sheiner [4] referring to an elegant theoretical framework developed in engineering [49]. This amazing vision must not be restricted to the sole domain of clinical pharmacology applied to drug development but will probably prove relevant in many conditions where human trials are planned, performed, and analyzed.



## APPENDIX NOTE ON GEOMETRIC AVERAGES AND COEFFICIENTS OF VARIATION

Both drug concentration data and pharmacokinetic parameters usually follow a skewed distribution among individuals, which is best described according to a *log-normal* rather than a Gaussian statistical model. This is because most pharmacokinetic processes are governed by *first-order rates* (i.e., transfer or reaction rates are proportional to the amount of substrate available), giving importance to *relative* rather than absolute changes.

In this situation, it is not appropriate to summarize results using ordinary mean and standard deviation (SD), as the thin and tall upper tail of the distribution will drag the mean well above the bulk of typical values and increase the magnitude of SD up to exceeding the size of the mean. A more suitable summary descriptor is the *geometric mean*, which is defined as the  $n$ th root of the product of  $n$  individual values. Practically, it is obtained by transforming the individual values into logarithms, computing the usual (arithmetic) mean of log transforms, and calculating back the antilog of this mean. For example, the three values 80, 100, and 125 correspond to a geometric mean of 100.

Computing the SD of those log-transformed data, and then taking its antilog, provides an estimate of the relative spread of log-normally distributed values: It can be interpreted as the typical factor either multiplying or dividing the geometric mean, which encompasses about two thirds of the data. For our three values 80, 100, and 125, this “geometric SD” represents a factor of 1.25. (Remember that the classical SD around an arithmetic mean is the typical difference either added to or subtracted from the mean, which encompasses about two thirds of a set of data that follow a Gaussian distribution). It may be more convenient to express this relative spread as a *percentage of variability*, by simply subtracting 1 from the geometric SD and multiplying by 100. In our example, this leads to a value of 25%. This value is higher than the “arithmetic” coefficient of variation, defined as the (arithmetic) SD divided by the mean (which would give  $22.6/101.7 = 22\%$  in our example). Nevertheless, it is sometimes called a *geometric coefficient of variation* (a more proper denomination would be the “coefficient of variation of the multiplicative model”). A more exact estimate for this coefficient can be calculated as the square root of the antilog of the variance of log-transformed data minus one (giving  $\sqrt{0.051} = 22.6\%$  in our example) [50].

For graphic purposes, it is recommended to represent geometric means associated with error bars showing their multiplication and division by the corresponding geometric SD. This is the easiest way to compute error bars that appear symmetric along a log Y axis (Fig. 3). Likewise, when parametric statistical tests are to be used for treatment or group comparisons on log-normally distributed data, a preliminary log transformation of the data is clearly indicated. Subsequently, it is possible to translate a difference between average log values back into a ratio of relative change. Study factors are thus assumed to impact multiplicatively rather than additively on the variables of interest, which they do most of the time (e.g., an enzyme inhibitor will be said to increase the concentrations of an interacting drug by a given percentage, rather than by a fixed amount in milligrams/liter). While this log-transform approach is standard practice in clinical pharmacokinetics, it would probably be suitable in many other areas of data analysis for clinical research.

## REFERENCES

1. Miller, F. G. (2003), Ethical issues in research with healthy volunteers: Risk-benefit assessment, *Clin. Pharmacol. Ther.*, 74, 513–515.
2. Hornblum, A. M. (1997), They were cheap and available: Prisoners as research subjects in twentieth century America, *BMJ*, 315, 1437–1441.
3. Lenzer, J. (2004), Scandals have eroded U.S. public's confidence in drug industry, *BMJ*, 329, 247.
4. Sheiner, L. B. (1997), Learning versus confirming in clinical drug development, *Clin. Pharmacol. Ther.*, 61, 275–291.
5. Cross, J., Lee, H., Westelinck, A., et al. (2002), Postmarketing drug dosage changes of 499 FDA-approved new molecular entities, 1980–1999, *Pharmacoepidemiol. Drug Safety*, 11, 439–446.
6. Heerdink, E. R., Urquhart, J., and Leufkens, H. G. (2002), Changes in prescribed drug doses after market introduction, *Pharmacoepidemiol. Drug Safety*, 11, 447–453.
7. Mouton, J. V., and Vinks, A. A. T. M. M. (1996), Is continuous infusion of  $\beta$ -lactam antibiotics worthwhile?—efficacy and pharmacokinetic considerations, *J. Antimicrob. Chemother.*, 38, 5–15.
8. Krayenbuhl, J. C., Vozeh, S., Kondo-Oestreicher, M., et al. (1999), Drug-drug interactions of new active substances: Mibefradil example. *Euro. J. Clin. Pharmacol.*, 55, 559–565.
9. Williams, D., and Feely, J. (2002), Pharmacokinetic-pharmacodynamic drug interactions with HMG-CoA reductase inhibitors, *Clin. Pharmacokinet.*, 41, 343–370.
10. Shah, R. R. (2005), Pharmacogenetics in drug regulation: Promise, potential and pitfalls, *Philos. Trans. Roy. Soc. London*, 360, 1617–1638.
11. Rendon, A., Nunez, M., Jimenez-Nacher, I., et al. (2005), Clinical benefit of interventions driven by therapeutic drug monitoring, *HIV Med.*, 6, 360–365.
12. Van Gelder, T., Meur, Y. L., Shaw, L. M., et al. (2006), Therapeutic drug monitoring of mycophenolate mofetil in transplantation, *Ther. Drug Monitor.*, 28, 145–154.
13. Brown, K., Tompkins, E. M., and White, I. N. (2006), Applications of accelerator mass spectrometry for pharmacological and toxicological research, *Mass Spectrom. Rev.*, 25, 127–145.
14. Garner, R. C., and Lappin, G. (2006), The phase 0 microdosing concept, *Br. J. Clin. Pharmacol.*, 61, 367–370.
15. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Economy*, 22, 151–185.
16. Ortiz de Montellano, P. R. (2005), *Cytochrome P450: Structure, Mechanism, and Biochemistry*, Kluwer Academic/Plenum, New York.
17. Ho, R. H., and Kim, R. B. (2005), Transporters and drug therapy: Implications for drug disposition and disease, *Clin. Pharmacol. Ther.*, 78, 260–277.
18. Atkinson, A. J. Jr., Daniels, C. E., Dedrick, R. L., et al. (2001), *Principles of Clinical Pharmacology*, Academic, San Diego.
19. Tozer, T. N., and Rowland, M. (2006), *Introduction to Pharmacokinetics and Pharmacodynamics: The Quantitative Basis of Drug Therapy*, Lippincott Williams & Wilkins, Philadelphia.
20. Colburn, W. A., and Lee, J. W. (2003), Biomarkers, validation and pharmacokinetic-pharmacodynamic modelling, *Clin. Pharmacokinet.*, 42, 997–1022.

21. Gabrielsson, J., and Weiner, D. (2001), *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*, 3rd ed., Taylor & Francis, London.
22. FDA Guidance for Industry (2003), *Exposure-Response Relationships—Study Design, Data Analysis, and Regulatory Applications*; available at: [www.fda.gov/cber/guidelines.htm](http://www.fda.gov/cber/guidelines.htm).
23. FDA Guidance for Industry (1999), *Population Pharmacokinetics*; available at: [www.fda.gov/cber/guidelines.htm](http://www.fda.gov/cber/guidelines.htm).
24. Williams, P. J., and Ette, E. I. (2000), The role of population pharmacokinetics in drug development in light of the Food and Drug Administration's "Guidance for Industry: Population pharmacokinetics," *Clin. Pharmacokinet.*, 39, 385–395.
25. Minto, C., and Schnider, T. (1998), Expanding clinical applications of population pharmacodynamic modelling, *Br. J. Clin. Pharmacol.*, 46, 321–333.
26. Vozech, S., Steimer, J. L., Rowland, M., et al. (1997), The use of population pharmacokinetics in drug development, *Clin. Pharmacokinet.*, 30, 81–93.
27. Shah, V. P., Midha, K. K., Dighe, S., et al. (1991), Analytical methods validation: Bioavailability, bioequivalence and pharmacokinetic studies. Conference report. *Euro. J. Drug Metabolism Pharmacokinet.*, 16, 249–255.
28. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1994, rev. 1996, 2005), *Validation of Analytical Procedures*; available at: [www.ich.org](http://www.ich.org).
29. Directive 75/318/EEC as amended (1987), *Pharmacokinetic Studies in Man*; available at: [ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-3/pdfs-en/3cc3aen.pdf](http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-3/pdfs-en/3cc3aen.pdf).
30. Schwartz, J. B. (2003), The influence of sex on pharmacokinetics, *Clin. Pharmacokinet.*, 42, 107–121.
31. Loebstein, R., Lalkin, A., and Koren, G. (1997), Pharmacokinetic changes during pregnancy and their clinical relevance, *Clin. Pharmacokinet.*, 33, 328–343.
32. Mizutani, T. (2003), PM frequencies of major CYPs in Asians and Caucasians, *Drug Metabolism Rev.*, 35, 99–106.
33. Buclin, T., Perrottet, N., and Biollaz, J. (2005), The importance of assessing the dose actually administered in pharmacokinetic trials, *Clin. Pharmacol. Ther.*, 77, 235–240.
34. Gillespie, W. R. (1991), Noncompartmental versus compartmental modeling in clinical pharmacokinetics, *Clin. Pharmacokinet.*, 20, 253–262.
35. Boomer, D. (2006), Pharmacokinetic and Pharmacodynamic Resources; accessible at: <http://www.boomer.org/pkin/soft.html>.
36. FDA Guidance for Industry (2001), *Statistical Approaches to Establishing Bioequivalence*; available at: [www.fda.gov/cber/guidelines.htm](http://www.fda.gov/cber/guidelines.htm).
37. EU Directive 65/65/EEC and 75/318/EEC as amended (1991), *Investigation of Bioavailability and Bioequivalence*; available at: [ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-3/pdfs-en/3cc15aen.pdf](http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-3/pdfs-en/3cc15aen.pdf).
38. Senn, S. (2001), Statistical issues in bioequivalence, *Stat. Med.*, 20, 2785–2799.
39. Williams, P. J., and Ette, E. I. (2000), The role of population pharmacokinetics in drug development in light of the Food and Drug Administration's "Guidance for Industry: Population Pharmacokinetics," *Clin. Pharmacokinet.*, 39, 385–395.
40. Ette, E. I., and Williams, P. J. (2004), Population Pharmacokinetics. I: Background, Concepts, and Models; II: Estimation Methods; III: Design, Analysis, and Application of Population Pharmacokinetic Studies, *Ann. Pharmacother.*, 38, 1702–1706; 1907–1915; 2136–2144.

41. Craig, W. A. (2002), Pharmacodynamics of antimicrobials: General concepts and applications, in Nightingale, C. H., Murakawa, T., and Ambrose, P. G. Eds., *Antimicrobial Pharmacodynamics in Theory and Clinical Practice*, Marcel Dekker, New York, pp. 1–22.
42. Sheiner, L. B., Stanski, D. R., Vozeh, S., et al. (1979), Simultaneous modeling of pharmacokinetics and pharmacodynamics: Application to d-tubocurarine, *Clin. Pharmacol. Ther.*, 25, 358–371.
43. Jusko, W. J., and Ko, H. C. (1994), Physiologic indirect response models characterize diverse types of pharmacodynamic effects, *Clin. Pharmacol. Ther.*, 56, 406–419.
44. Csajka, C., and Verotta, D. (2006), Pharmacokinetic-pharmacodynamic modelling: History and perspectives, *J. Pharmacokinet. Pharmacodyn.*, 33, 227–279.
45. Stanislaus, D., Pinter, J. H., Janovick, J. A., et al. (1998), Mechanisms mediating multiple physiological responses to gonadotropin-releasing hormone, *Mol. Cell. Endocrinol.*, 144, 1–10.
46. Aarons, L., Karlsson, M. O., Mentre, F., et al. (2001), Role of modelling and simulation in Phase I drug development, *Euro. J. Pharma. Sci.*, 13, 115–122.
47. Meibohm, B., and Derendorf, H. (2002), Pharmacokinetic/pharmacodynamic studies in drug product development, *J. Pharma. Sci.*, 91, 18–31.
48. Chien, J. Y., Friedrich, S., Heathman, M. A., et al. (2005), Pharmacokinetics/pharmacodynamics and the stages of drug development: Role of modeling and simulation, *AAPS J.*, 7, 544–559.
49. Myers, R. H., and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley, Hoboken, NJ.
50. Limpert, E., Stahel, W. A., and Abbt, M. (2001), Log-normal distributions across the sciences: Keys and clues. *Bioscience*, 51, 341–352.

# 19

---

## *Modeling and Simulation in Clinical Drug Development*

**Jerry Nedelman,<sup>1</sup> Frank Bretz,<sup>2</sup> Roland Fisch,<sup>3</sup> Anna Georgieva,<sup>1</sup>  
Chyi-Hung Hsu,<sup>2</sup> Joseph Kahn,<sup>1</sup> Ryosei Kawai,<sup>4</sup> Phil Lowe,<sup>3</sup> Jeff Maca,<sup>2</sup>  
José Pinheiro,<sup>2</sup> Anthony Rossini,<sup>3</sup> Heinz Schmidli,<sup>5</sup> Jean-Louis Steimer,<sup>3</sup>  
and Jing Yu<sup>4</sup>**

<sup>1</sup>*Modeling and Simulation and* <sup>2</sup>*Clinical Information Sciences, Novartis Pharmaceuticals  
Corp., East Hanover, New Jersey*

<sup>3</sup>*Modeling and Simulation and* <sup>5</sup>*Clinical Information Sciences, Novartis Pharma AG,  
Basel, Switzerland*

<sup>4</sup>*Modeling and Simulation, Novartis Institutes for BioMedical Research, Inc.,  
Cambridge, Massachusetts*

### **Contents**

|        |                                      |     |
|--------|--------------------------------------|-----|
| 19.1   | Introduction                         | 990 |
| 19.1.1 | What Is a Model?                     | 990 |
| 19.1.2 | What Is Simulation?                  | 992 |
| 19.1.3 | Model Building                       | 992 |
| 19.1.4 | Bayesian Methods                     | 993 |
| 19.1.5 | What Lies Ahead                      | 993 |
| 19.2   | Types of Modeling and Simulation     | 994 |
| 19.2.1 | Regression Models                    | 994 |
| 19.2.2 | Longitudinal Models                  | 995 |
| 19.2.3 | Exploratory Modeling and Data Mining | 996 |
| 19.2.4 | Dose–Exposure–Response Modeling      | 997 |

|                                                           |      |
|-----------------------------------------------------------|------|
| 19.2.5 Pathways Modeling                                  | 1002 |
| 19.2.6 Physiological Modeling                             | 1003 |
| 19.2.7 Disease Progression Modeling                       | 1004 |
| 19.2.8 Clinical Trial Simulation                          | 1005 |
| 19.2.9 Decision Analysis                                  | 1006 |
| 19.3 M&S-Based Clinical Trials                            | 1007 |
| 19.3.1 First-in-Man Studies                               | 1007 |
| 19.3.2 Oncology Phase I Studies                           | 1009 |
| 19.3.3 Proof-of-Concept Studies                           | 1010 |
| 19.3.4 MCP-Mod: Unified Strategy for Dose-Finding Studies | 1011 |
| 19.3.5 Adaptive Trials                                    | 1012 |
| 19.4 Discussion                                           | 1013 |
| References                                                | 1013 |

## 19.1 INTRODUCTION

This chapter describes modeling and simulation (M&S) applied to clinical drug development within a pharmaceutical company. In that context, M&S serves the two goals of drug development: learning and confirming [1].

From the learning perspective, M&S is a tool to learn how the drug works. M&S helps us design and interpret clinical trials, the learning experiences of clinical drug development. M&S is also a framework for representing what we learn in the precise language of mathematics.

For confirming, M&S provides a rigorous context in which to make statistical judgments about relationships, such as whether increases in dose are significantly associated with increases in response.

As you can see from the previous paragraphs, it is conventional to use the phrase “modeling and simulation” as a singular noun. But modeling and simulation are different things. Let us begin with a separate consideration of each. We will then discuss some general notions about how models are built and briefly introduce Bayesian methodology, which plays a large role in M&S. Then we will pause to outline how the remainder of the chapter describes M&S in clinical drug development.

### 19.1.1 What Is a Model?

A model is a way of predicting outputs from inputs.

For example, consider a model to predict the steady-state concentration  $C_{ss}$  obtained by infusing a drug at rate  $R$  into a subject who has clearance  $CL$ :

$$C_{ss} = \frac{R}{CL} \quad (1)$$

The output is  $C_{ss}$  and the inputs are  $R$  and  $CL$ . Note that different inputs may play different roles. The infusion rate is at the discretion of the experimenter (physician, pharmacologist, . . .). The clearance is a characteristic of the subject and drug. Both might be called variables, or we might refer to the clearance, which may be unknown, as a parameter.

Now consider the systolic blood pressures at baseline, the start of the trial, among patients enrolled in a clinical trial for an antihypertensive drug. The protocol will specify some severity of hypertension for included patients, but there will still be variability among patients at baseline. We use the language of probability to create a model for such variability. Suppose the average baseline blood pressure among all eligible patients is  $\mu$  and the variance is  $\sigma^2$ . It is often the case that a collection of blood pressures follows a bell-shaped curve that we might characterize as a normal probability distribution. We then say the output—baseline blood pressure  $B$ —is predicted from the inputs— $\mu$  and  $\sigma^2$ —according to the Normal probability model:

$$B \sim N(\mu, \sigma^2) \quad (2)$$

In (2),  $B$  is called a random variable;  $\mu$  and  $\sigma^2$  are both considered parameters. Note that to represent a random variable as an output from its distributional model we use  $\sim$  instead of  $=$ .

Imagine observing steady-state concentrations by taking blood samples and assaying them. Even if model (1) is reasonable, and even if we know  $CL$  as well as  $R$ , our observed concentration will probably not exactly equal  $C_{ss}$  because of measurement error and because of physiological fluctuations that render the model incomplete. We account for this discrepancy by introducing a new variable, usually denoted  $\varepsilon$ , called the residual error. It is modeled as a random variable, analogous to the baseline blood pressures (2), and we often use the same Normal probability model. We usually assume a mean of zero for residual errors, so the only input is the variance,  $\sigma^2$ . We then expand (1) to

$$C_{ss} = \frac{R}{CL} + \varepsilon \quad (3)$$

$$\varepsilon \sim N(0, \sigma^2) \quad (4)$$

Note that in (3)  $\varepsilon$  plays the role of an input, but in (4) it plays the role of an output. Chaining models together in this way gives modeling great power while maintaining conceptual simplicity.

Suppose we administer a single intravenous bolus dose  $D$  of a drug at time 0, and suppose the systemic circulation behaves as if it is a single well-mixed compartment. Then the systemic concentration  $C$  of drug may be predicted as a function of the inputs dose ( $D$ ), time post dose ( $t$ ), clearance ( $CL$ ), and compartmental volume ( $V$ ) as

$$C(t) = D \exp\left(\frac{-CL t}{V}\right) \quad (5)$$

Models with time as one of the inputs play a very important role in clinical trials (and throughout science), so it is common to highlight the input time by using it as an argument in a representation of the output as a function, as  $C(t)$  in (5).

The one-compartment model (5) is somewhat special because we can write the dependence of concentration on time algebraically. Frequently, models with time as input can only be expressed by describing how the rate of change of the output



depends on time and other inputs, that is, as a differential equation. A differential equation corresponding to (5) is

$$\frac{dC}{dt} = -\left(\frac{CL}{V}\right)C \quad C(0) = D \quad (6)$$

The algebraic representation (5) is the solution to the differential equation (6). But often no algebraic solution can be found, and we must solve the differential equation numerically, which takes us to simulation.

### 19.1.2 What Is Simulation?

Simulation is the act of computing values of the output for given values of the input. In (1), if we plug in values for  $R$  and  $CL$  and compute the resulting value of  $C_{ss}$ , we have accomplished a very simple simulation.

Usually, however, simulation refers to repeating such calculations many times for different values of the inputs to see how changes in the inputs translate into changes in the outputs. The three most common situations are:

1. One of the inputs is time, and the computations are done to see how the output evolves as time progresses.
2. One or more of the inputs is a random variable, and the computations are done to see how the output varies due to the random process(es) represented by the random variable(s).
3. Both 1 and 2 together.

### 19.1.3 Model Building

Models seldom spring fully formed from the heads of modelers, at least if “fully formed” includes specification of the values of the model’s parameters. Usually, the parameters must be estimated by fitting the model to data. Often the structure of the model (e.g., algebraic functions, differential equations, or probability distribution) and the inputs that best predict the outputs are determined partly if not wholly by fitting different models to the data and choosing the one that seems best, a process called model building.

Fitting the model to data means selecting parameter values of the model to minimize some measure of discrepancy or maximize some measure of similarity between observed values of the output variable (the data) and predictions of the output variable from the model. The sum of squared differences between data and predictions is a measure of discrepancy; minimizing it yields a least-squares fit. A common similarity measure is the likelihood, which, as its name implies, quantifies how probable the observed data are according to the model; maximizing it yields a maximum-likelihood fit.

The more complex the model the better the fit in terms of least squares or likelihood. To decide when to stop building, called model selection, criteria may be used that penalize models for being too complex, both for the aesthetic appeal of parsimony and because theory shows that these criteria effectively home in on the best model. Two such criteria are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [2].

### 19.1.4 Bayesian Methods

In the Bayesian approach to modeling, parameters are considered random variables not unknown constants. A parameter's probability distribution represents our uncertainty regarding its value. To learn is to reduce that uncertainty, corresponding to a reduction in the variance of the parameter's probability distribution. At the start of a new learning step, the probability distribution is called the prior. When we acquire new data and thereby learn something about the parameter, the updated probability distribution is called the posterior. This process of updating our knowledge, called Bayesian inference, is accomplished via Bayes rule, a mathematical formula. Berry [3] provides an introduction to Bayesian methods and their application in clinical trials.

### 19.1.5 What Lies Ahead

The remainder of this chapter will describe types of M&S used in clinical drug development and types of clinical trials that strongly depend on M&S. Each section will begin with a statement of purpose for the type of M&S covered or for the use of M&S in the clinical trial.

The sections will frequently cross reference one another. Curly braces { } will denote section number cross references, whereas square brackets [ ] will denote bibliographical citations and parentheses ( ) will denote equation numbers.

Although the chapter's parts must be sequentially ordered in a print medium such as this, there is no perfect sequencing of the material. Figure 1 displays one possible configuration of the dependencies among the sections.

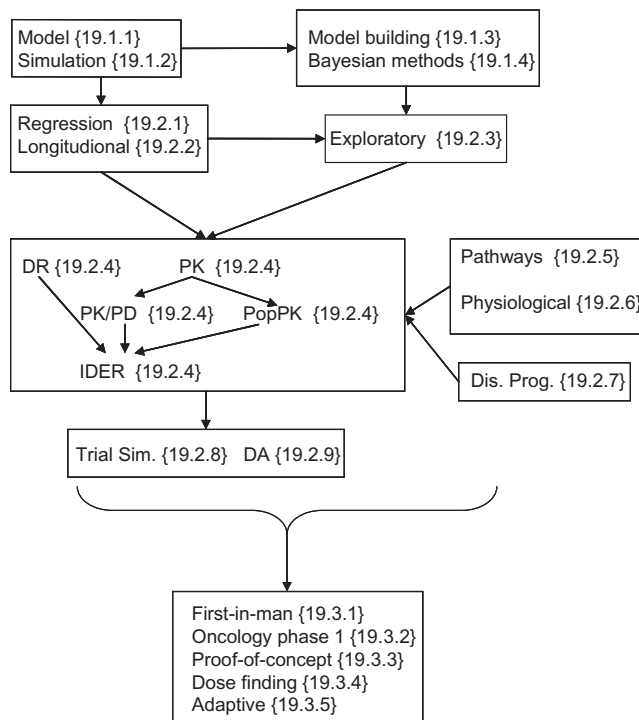


**FIGURE 1** Dependencies among chapter sections.

The heart of the chapter is {19.2.4}, which covers dose–exposure–response (DER) modeling, the fundamental clinical characterization of how a drug works and how M&S helps determine what dose to give to which patients. {19.1.1}–{19.2.3} provide the fundamentals of M&S needed to understand how DER models are built and used. {19.2.5}–{19.2.6} introduce exciting new areas of biological modeling that support DER modeling with mechanistic underpinnings. {19.2.7} describes methods for characterizing higher level disease progression. Within the process of drug development, an important application of models is to support decision making; {19.2.8}–{19.2.9} cover that. Finally {19.3} showcases five types of clinical trials whose design and interpretation benefit from M&S; they are presented roughly in the order they might appear in a drug development program.

## 19.2 TYPES OF MODELING AND SIMULATION

### 19.2.1 Regression Models

**Purpose:** *To infer how one variable may be predicted empirically from one or more others using independent observations of those variables in a sample of subjects.*

One variable, such as blood pressure (BP) at the end of a hypertensive trial, is the output, called the response; and one or more others, such as baseline blood pressure, treatment, dose, age, and sex, are inputs, called predictors. The goal of relating the response to the predictors may be learning—for example, to devise a rule to predict future observations of the response—or confirming—for example, to test if one or more of the predictors, such as treatment, is statistically significant. The model is fitted to data comprising one record per subject, each record containing observations of the response and the predictors, with records from different subjects considered independent.

Instead of being derived from any scientific theory, the form of a regression model is selected as a compromise between mathematical convenience and the empirical representation of the data's shape. Often a linear model (7) suffices because relationships tend to be approximately linear over limited ranges of the data.

Simple linear regression has one continuous response, for example, BP, that depends on a continuous predictor, for example, dose, and residual error:

$$\text{BP}_i = \beta_0 + \beta_1 \text{dose}_i + \varepsilon_i \quad (7)$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad (8)$$

where  $\beta_0$ ,  $\beta_1$ , and  $\sigma^2$  are parameters that are estimated by fitting the model to data and where the subscript  $i$  indexes the data, that is, here the subjects. Presumably, if the drug were an antihypertensive,  $\beta_1 < 0$ .

The adjective “linear” refers to the linearity of (7) with respect to the parameters  $\beta_0$  and  $\beta_1$ , not dose. If in (7) dose were replaced by, for example, the logarithm of dose, it would still be a linear model even though a plot of  $\beta_0 + \beta_1 \log(\text{dose})$  versus dose would be curved.

If there are several predictors, such a model becomes a multiple regression model.

Nonlinear regression is when some parameters appear nonlinearly, such as

$$BP_i = \beta_0 + \frac{\beta_1 \text{dose}_i}{\beta_2 + \text{dose}_i} + \varepsilon_i \quad (9)$$

Suppose we create a new response variable,  $Y$ , by considering only whether BP dropped by 3 mmHg or not, for which we assign  $Y = 1$  or 0, respectively. Logistic regression is where the logit transformation of the probability that  $Y = 1$ , say  $p$ , is modeled linearly as

$$\text{logit}(p) = \log \left[ \frac{p}{(1-p)} \right] = \beta_0 + \beta_1 \text{dose} \quad (10)$$

Logistic regression is an example of a broader class of models called generalized linear models [4], where some function of the mean of the response is related linearly to the predictors.

A clinical outcome may be the time to some event, such as death. Regression models for such a response time are a major component of survival analysis. The chance of the event occurring in the next instant of time, given that it has not happened yet, is called the hazard. In the Cox proportional hazards model, the logarithm of the hazard is related linearly to the predictors.

Many statistics textbooks cover regression modeling. See [5] for a nontechnical introduction or [6] for a technical one.

### 19.2.2 Longitudinal Models

**Purpose:** *To infer how one variable may be predicted empirically from one or more others using observations measured at more than one time per subject in a sample of subjects.*

Whereas regression models {19.2.1} are appropriate when subjects independently contribute one observation each, longitudinal models describe multiple observations collected over time for each subject. The several observations of the same individual tend to be correlated, which must be accounted for by the model. This additional complication buys, for confirming purposes, additional precision and power, and, for learning purposes, additional insight into how the response evolves as conditions change.

Regression models generalize to longitudinal models in two ways: marginal models and mixed-effect models.

The focus of a marginal model is to predict the population's average response, generally for confirming purposes. As in generalized linear models {19.2.1}, some transformation of the average response is assumed to admit a linear structure in the predictors. The correlation structure is handled somewhat empirically using methods that are known to be robust with sufficiently large samples; see [7].

Marginal models have two main limitations. First, they require that all subjects have the same number of observations under the same design. Second, they generally are not informative about individual responses.

Mixed-effect models (MEMs) overcome these limitations. An MEM is constructed hierarchically, by chaining models together as described in {19.1.1}, starting from a model for the response of an individual. The parameters of the individuals' separate models are themselves outputs of regression-type models. This accounts for the necessary correlation structure of the data and allows the models to be fitted to data even where subjects have different numbers of observations at different times.

Let us consider an MEM generalizing (7–8) to describe a study where each subject receives successively higher doses at each of which BP is measured. Let  $BP_{ij}$  and  $dose_{ij}$  be the  $j$ th observed BP and dose for the  $i$ th subject. The MEM is

$$BP_{ij} = b_{0i} + b_{1i} dose_{ij} + \varepsilon_{ij} \quad (11)$$

$$b_{0i} = \beta_0 + \eta_{0i} \quad b_{1i} = \beta_1 + \eta_{1i} \quad (12)$$

$$\eta_{0i} \sim N(0, \omega_0^2) \quad \eta_{1i} \sim N(0, \omega_1^2) \quad \text{corr}(\eta_{0i}, \eta_{1i}) = \rho \quad (13)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2) \quad (14)$$

Each subject has his or her own intercept and slope  $b_{0i}$  and  $b_{1i}$ . These inputs to model (11) are outputs of the models (12). The random variables  $\eta_{0i}$  and  $\eta_{1i}$  are called random effects. They characterize the variation among subjects of the intercepts and slopes of the BP-vs.-dose relationships, and their presence accounts for the correlations among the several  $BP_{ij}$  values for the  $i$ th subject.

Suppose we believe that some of the variation in the dependence of BP on dose is due to age. We need only make age an input to the model for  $b_{1i}$  in (12):

$$b_{1i} = \beta_1 + \beta_2 \text{age}_i + \eta_{1i} \quad (15)$$

If any of the parameters in an MEM besides the variances appears nonlinearly, then the model is called a nonlinear MEM or NLMEM, which will figure prominently in {19.2.4}.

Longitudinal models for clinical trials are reviewed in [8]. See also [9].

### 19.2.3 Exploratory Modeling and Data Mining

**Purpose:** *In large data sets, to find patterns that can be used as hypotheses to guide further experimentation or as retrospective explanations of surprising signals.*

Exploratory modeling in clinical drug development has two main applications. The first is part of the guided knowledge generation that is common to all science, when some general goals of discovery are pursued in the hopes of generating hypotheses that might be confirmed with further experimentation. Examples are mining genomic data to find potential biomarkers, or mining clinical data to find possible new indications for marketed drugs.

An alternative setting occurs when a clinical trial produces some surprising signal, such as a worrisome toxicity. Sometimes called fire-fighting, exploratory modeling then seeks to propose a plausible story for the observed results. Such a story would ultimately require confirmation, and thus this second application is really one of

hypothesis generation like the first; but if strong enough, such modeling results might be proposed as tentatively confirmatory or at least satisfactorily descriptive for, say, a regulatory submission.

The steps of exploratory modeling are (a) defining the problem, (b) exploring the data with graphical, tabular, and numerical summaries, (c) building models, and (d) validating models.

In the first two steps, the scope of potential models is clarified by learning the structure and content of the data and identifying any important substructures such as grouping in study centers, or local laboratories with different reference ranges, or time profiles of observations.

The nomenclature of data mining, derived largely from computer science, often refers to modeling as the application of various algorithms. Commonly used algorithms are listed below. To link with our view of modeling {19.1.1}, for each algorithm we identify the kind of outputs. In every case, the inputs might be anything else in the database; finding the right inputs and their relationship with the outputs is the model building {19.1.3}.

*Classification Algorithms* The output is one or more discrete variables, such as responder versus nonresponder. Methods include logistic regression {19.2.1}, rule induction, decision trees, neural networks,  $K$ -nearest neighbors, and support vector machines.

*Regression Algorithms* The output is one or more continuous variables, such as blood pressure. Methods include linear regression {19.2.1}, regression trees, and neural networks.

*Segmentation Algorithms* The output is the assignment of a patient into one of several groups, or clusters, of patients that have similar properties. Methods include clustering techniques, neural networks, and visualization methods.

*Dependency Algorithms* The output is any variable in the data set, and often outputs are inputs for other relations in a chain or network that represents associations among different variables. Methods include correlation analysis, association rules, and Bayesian networks.

Finally, after a model is built, its prediction performance is evaluated. Cross-validation, bootstrapping, and bagging are the techniques usually used. Sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and receiver operating characteristic (ROC) curve are the mostly used evaluation metrics. For a comprehensive introduction to these ideas and data mining generally, see [10].

#### 19.2.4 Dose–Exposure–Response Modeling

A fundamental goal of drug development is to learn what dose of the drug to give to which patients. Most drugs reach their site of action via the systemic circulation, so drug concentrations in the blood or plasma are often more predictive of response than dose. Hence, one way of building a dose–response model is to concatenate a dose–exposure (or pharmacokinetic) model with an exposure–response (or pharmacokinetic/pharmacodynamic) model. But direct modeling of dose–response is also done. This section describes these various models.

**Dose–Response Modeling**

**Purpose:** *To characterize the relationship between dose and any measurable response.*

Dose–response modeling is regression modeling {19.2.1} or longitudinal modeling {19.2.2} with dose as input and response as output. Examples are models (7–14). Responses may be biomarkers, clinical outcomes, adverse events, or even pharmacokinetics, although the latter merits separate consideration (see below). See {19.3.4} for use of such models in dose-finding trials.

**Pharmacokinetic (PK) Modeling**

**Purpose:** *To describe measured drug concentrations as a function of dosing regimens and, usually, time.*

Pharmacokinetics is “what the body does to the drug,” as in absorption, distribution, metabolism, and elimination (ADME). Systemic drug concentrations are measured in repeated blood (or plasma) samples over time after one or multiple doses administered to healthy volunteers or patients. PK models characterize such concentrations by mathematically quantifying the ADME processes as parameters such as clearance (the flow at which the drug leaves the body), volume of distribution (the apparent volume in which the drug is distributing), half-life (the rate at which the drug is declining, once transient events in the body have taken place, a parameter that combines influences of elimination and distribution). Examples of such models are (1, 3–6).

Models (5–6) are appropriate when the PK behaves as if drug distributes within a single, well-mixed compartment. Quite often, however, the PK data require assuming the drug distributes among multiple compartments; usually two or three are adequate to describe a PK profile. All such models are referred to as compartment models.

A comprehensive reference on PK models is [11].

Ultimately, the number of compartments in the body is (at least) the number of homogeneous tissues. This is recognized by physiologically based PK (PBPK) modeling, which, in the situation of transitioning from animal studies to first-study-in-man, attempts to predict the key human pharmacokinetic parameters in advance before doing the actual experiments {19.3.1}. A PBPK model is based on literature data in human and animal models about blood flows and sizes of respective organs, and on actual compound-specific PK data in animals (e.g., the rat, mouse, dog, and less often the monkey). For animal-to-human extrapolation, the approach is extremely resource intensive and problematic, despite recent advances. In rare situations, retrospective attempts to develop PBPK models have been largely successful, for instance, for well-known drugs (oncology, morphine) or in other applications, like extrapolations from adults to children [12].

Population PK is another extension of PK modeling; see discussion below for details.

Pharmacokinetics results provide instrumental knowledge along the development of a compound and are essential for the drug submission dossier to help characterize the clinical pharmacology of the new chemical entity [13]. The classical method for data analysis initially avoids models and is therefore sometimes called “noncompartmental analysis.” Metrics of exposure to the drug like  $C_{\max}$  (the



maximum concentration value) and AUC (the area under the measured concentration time points) are determined through simple calculations. This approach, combined with cross-over experiments assessed by analysis of variance, is extremely powerful for descriptive and comparative purposes (e.g., across formulations or dosing regimens).

Quite often, the exposure metrics  $C_{\max}$  and AUC are considered as outputs in dose–response models (see discussion above). A commonly used model is the so-called power model:

$$\text{AUC} = \beta_0 \text{dose}^{\beta_1} \quad (16)$$

Of interest is whether or not  $\beta_1 = 1$ , in which case the PK is said to be dose proportional or linear. Testing whether  $\beta_1 = 1$  is a confirming role for (16).

### **Pharmacokinetic/Pharmacodynamic (PK/PD) Modeling**

**Purpose:** *To characterize the mutual variation of drug concentrations and biomarkers or clinical responses to the drug.*

Pharmacodynamics is “what the drug does to the body.” PD data comprises observations of a marker of drug effect in a subject, be it a systemic marker (e.g., a hormone or a pathway intermediate), a diagnostic assessment [e.g., a QT interval within the electrocardiogram (ECG)], a surrogate marker [e.g., HbA1c for diabetes or CD4 counts for human immunodeficiency virus (HIV)], or a clinical outcome (e.g., fractures in osteoporosis, death in oncology). Usually, the observations are repeated over time, or, for single events such as death, the time to event is relevant.

A very simple class of PK/PD models, commonly used for clinical outcomes in chronic diseases, uses the same empirical forms as dose–response models (see discussion above) but replaces the input dose with a summary measure of systemic exposure such as AUC.

The objective of PK/PD modeling is often to describe temporal delays of the PD response relative to the PK. A classical type of PK/PD modeling for this purpose is the effect-compartment model, when the drug exhibits its effect in a compartment that is not in rapid equilibrium with the systemic circulation [14]. Such a compartment may represent a real physiological space, such as the brain for anesthetics, or it may be empirically imputed solely to account for the observed delay. The drug concentration in such a compartment is an output of the PK part of the model and an input for the effect of the drug on the PD. In some cases, when the distribution to the site of action is extremely slow, the maximal effect can be achieved late after the administration of the drug, when the systemic concentrations are no longer detectable (have reached levels below the limit of assay quantification).

Another more physiological approach for accounting for delays between the time course of the drug effect relative to drug concentration is the turnover model (initially introduced as the “indirect-response model”) [15]. Under normal physiological conditions, the rate of production,  $k_{\text{in}}$ , of the marker of drug effect is equal to the rate of elimination,  $k_{\text{out}}$ ; “nothing happens”—the level of the marker remains constant at  $k_{\text{in}}/k_{\text{out}}$ . Depending on the compound and the marker, the drug effect is described as a stimulation or inhibition of the rate of production or of the rate of elimination. For instance, a single turnover model is adequate to describe, across a broad range of bisphosphonate doses, the kinetics of selected markers of bone

degradation which are inhibited by 0–80% even in healthy subjects. These markers are fast, reacting over hours/days. Bone mineral density is a slow bone marker, where changes are detectable over months/years.

It is important that the PD assessment is part of the causal chain linking the dose of drug to the ultimate clinical effect (e.g., reduction of fracture in case of a drug for osteoporosis, diminishing the risk of heart failure for a cardiovascular drug). Biomarkers are important tools to make drug development more efficient. Reacting more quickly to therapeutic interventions than clinical responses, sometimes even in healthy subjects, they may be signals to guide rapid and inexpensive decision making. The number and type of links in the causal chain will vary; for example, a biomarker may be a parallel process to the clinical effect (as shown in Fig. 2) or it may be directly in the causal chain. Nonetheless, the biomarker should clearly indicate that the drug has bound to the intended target.

In the effect-compartment model and in the turnover model, concentration serves as the stimulus to effect. In contrast, there are situations where PK and PD are intimately linked and cannot be separated, for example, monoclonal antibody kinetics [16], where the PD, the interaction of the drug with its receptor, plays an important role in the actual disposition and elimination of the drug, the PK. With such target-mediated drug disposition, further complexities arise; for example, the system may become nonlinear, with overproportional increase in AUC with dose. PK/PD models are necessary; noncompartmental approaches for PK analysis are largely insufficient.

Further mechanism-based PK/PD models have been reviewed recently [17]. These authors highlight that the principles are well known and illustrate a large amount of experience with these models in a wide variety of therapeutic areas. Much of this experience has accumulated presently in academic settings. However, experience within the pharmaceutical industry is increasing dramatically as the potential of the approaches reaches the attention of senior management and regulators.

In early-phase drug development, PK and PK/PD models are often used to predict the concentration (and response) after multiple dosing based on the first human trial after single-dose administration. Such models may be used in clinical trial simulation {19.2.8} to help in planning future studies, such as proof-of-concept studies {19.3.3}.

Since the uncertainty under such circumstances is great, one cannot expect a model to be always right, even if it is based on the most accurate quantitative data and knowledge available at this particular point. Learning in the form of ongoing revision to the model ought to occur all along the development of the compound and should incorporate insights from the development of similar pharmacological entities or different compounds addressing the same therapeutic indication.

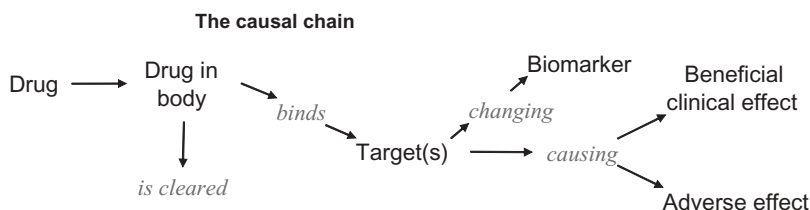


**FIGURE 2** Causal chain.

In terms of regulatory submission, when a model is used for extrapolation, it must be validated beforehand or agreed upon with the regulators that the approach taken is convincing [18].

As an integrative tool later in development, population PK/PD modeling is more and more used for dose–exposure–response assessment [19]; see discussion below.

### **Population PK Modeling**

**Purpose:** *To explain and predict how systemic drug exposure varies among subjects.*

Population pharmacokinetics (PopPK) originated as an attempt to infer the patterns of PK response within a patient population from sparse PK data collected under routine clinical care [20]. Nonlinear mixed-effect modeling {19.2.2} was employed, according to which the parameters of compartmental PK models are themselves considered outputs of models where patient covariates (and random effects) comprise inputs. For example, clearance might depend on body size, age, and the presence of some concomitant therapy that induces a drug–drug interaction.

PopPK has evolved into a widely used tool in clinical drug development for characterizing the variability of PK in patient populations by means of nonlinear mixed-effect models [21, 22].

PopPK is valuable when “a reasonable a priori expectation exists that intersubject kinetic variation may warrant altered dosing for some subgroups in the target population” [23]. It can be used strategically to help assess a need for altered dosing, as in ethnic bridging [24]. It is also useful whenever PK understanding is needed but only sparse PK data can be collected, for example, pediatric trials [25]. Regulators frequently request a PopPK analysis as part of a submission dossier, and many drug development scientists believe that a PopPK characterization of variability is an essential component of a development program.

The steps of using PopPK for clinical trials are design, data collection, model building, model validation, and model application.

The design goal for PopPK is usually to extract useful PK information from few blood samples per patient. Optimal design methods have been implemented in software for such purposes [26].

Data collection can be problematic because the clinical sites are not always used to it. Care must be taken that dates and times of dosing and blood sampling are accurately recorded because errors in that data can lead to biased and imprecise estimates of the model’s important PK parameters.

Building the submodels that describe how the compartmental PK parameters depend on covariates is an intensively data-driven task. Optimal strategies for this step are the subject of much current research [27].

Model validation is demonstrating that the model is credible for its intended purpose. If the only purpose is quantification and explanation of variability, then validation might consist of checking stability of the parameter estimates when refitting the model to random subsets of the data [28]. Ever more frequently, concentration outputs from PopPK models are used as inputs into models for clinical responses as part of integrated dose–exposure–response models (see next section) and for clinical trial simulation {19.2.8}. To be used for such a purpose, the model should be validated by demonstrating that its predictions are consistent with observed data

(predictive checking [29]), preferably new data that was not used as part of the model building (external validation [30]).

### ***Integrated Dose–Exposure–Response Modeling***

**Purpose:** *To comprehensively describe relationships among dose, exposure, and clinical response (efficacy and safety), often by pooling data from several clinical studies.*

The clinical response (effectiveness, undesirable effects) to a specific dose of a drug varies from patient to patient. With an integrative dose–exposure–response (IDER) model, one tries to understand which factors are responsible for individually differing responses. This information is important for the safe and effective use of a drug in individual patients. For example, it can help to specify an appropriate individualized starting dose, or the best way to adjust doses for individual patients [31].

The IDER analysis is typically done in two steps: First, the relationship between dose and drug concentration in blood (exposure) is described, and factors responsible for variations in exposure are identified (such as body weight or renal function); a PopPK analysis (see above) can provide this information. Second, the PK/PD relationship between drug concentration in blood and clinical response is investigated, and factors responsible for individual differences (such as gender or ethnicity) are identified. The PK and PK/PD models are cumulatively learned throughout the drug development process. Both mechanistic models (e.g., pathway models {19.2.5} or physiological models {19.2.6}) and empirical models (e.g., regression models {19.2.1} or longitudinal models {19.2.2}) are used for the development of the IDER relationships.

An IDER model can have a powerful impact in getting the drug approved and providing guidance on dosing. An example is the development of Certican (everolimus) in combination with Neoral (cyclosporine) for prophylaxis of acute organ rejection in renal transplant recipients. Studies in about 1200 patients showed good efficacy but unsatisfactory renal function when everolimus was administered as 1.5 or 3.0 mg daily together with full dose cyclosporine [32, 33]. An IDER model [34] revealed that improved efficacy was associated with everolimus exposures above 3 ng/mL, but that efficacy did not vary with observed cyclosporine exposures. On the other hand, better renal function was associated with lower cyclosporine exposures, whereas renal function did not vary with observed everolimus exposures. The IDER model thus suggested that therapeutic drug monitoring (TDM) should be used for everolimus to ensure that everolimus exposure is above 3 ng/mL in each patient, and that cyclosporine exposure should be individually reduced using TDM. This regimen was then prospectively tested in an additional study with about 250 patients that confirmed the predicted good efficacy and renal function [35, 36].

### **19.2.5 Pathways Modeling**

**Purpose:** *To describe and quantify the dynamic behavior of biological signaling pathways.*

As discussed in {19.2.4}, clinical response to a given drug varies from patient to patient. This variability can be due to routinely collected demographic factors such as gender and race, but it can also be due to different protein and gene expression

levels, as well as mutations in relevant biological entities, such as the drug target or the drug metabolizing enzymes.

Mathematical models of signaling pathways elucidate the dynamical behavior of such relevant biological entities. In terms of target discovery and validation, such models may suggest what are the optimal ways that the system can be manipulated by drugs [37]. In targeted therapies used in oncology, such models can help elucidate possible combinations of drugs that would together achieve efficacy without interfering with the physiological function of the drugs' targets in normal tissue, thus limiting toxicity [38, 39]. In combination with molecular dynamics simulations, such models can address the role that a mutation in the drug target has in the overall clinical efficacy seen with a drug [40]. Schoeberl et al. [41] have used mathematical modeling to show that different protein expression levels from different cell lines and tissues can lead to largely different behavior of the same signaling network and thus helped to reconcile seemingly contradictory literature results.

Signaling pathways are usually modeled by a set of differential equations {19.1.1}, where the state variables are proteins or protein complexes within the signaling pathway. The kinetic parameters in the system include the Michaelis–Menten constant, turnover numbers, and rate constants of association and dissociation [42].

Development of highly predictive pathway models is difficult because such models require (1) complete characterization of the molecular interactions in the pathway and (2) numerical values for the parameters associated with the model equations. Parameter values are often determined via numerical optimization against experimentally measured quantities (model fitting, {19.1.3}). This, in turn, creates an additional complexity for models of realistic size since the number of the unknown parameters is too large as compared to the available experimental data [37]. Sensitivity analysis can determine the dominant parameters, which helps both by pointing to data that can best inform and constrain the model, as well as by suggesting possible biomarkers [43]. In addition, specialized techniques such as extended metabolic control analysis [44], determining the components that control the dynamics of signaling pathways, facilitate the extraction of the relevant pathway modules to be included in a larger disease model {19.2.7} for clinical development.

### 19.2.6 Physiological Modeling

**Purpose:** *To describe mathematical models of tissues and organs and outline possible utility in clinical drug development.*

Mathematical models of human physiology have been of interest for a long time [45]. The IUPS Physiome project (<http://www.physiome.org.nz>) represents a worldwide public domain to build a computational framework and quantitative understanding of human physiology and pathophysiology. As new data accumulates, these models have become increasingly more sophisticated and biologically plausible. Organs that have been described by such mathematical models include kidney [46], respiratory tract [47], heart [48], and even the whole human physiology [49]. Because constructing these models is complex, data intensive, and computationally expensive, such modeling efforts have been largely the subject of academic research with somewhat limited applications to human health. However, the situation appears to be changing with the accumulation of data, government funding, and establishment of consortia and collaborative efforts that pool data across different organizations.

One example for a tissue-level model that is being used to guide clinical decision making is that of cardiac electrophysiology. The last decade has been marked by the withdrawal of several medicines whose use in patients has been associated with the development of torsade de pointe (TdP), a potentially fatal polymorphic tachycardia [50]. Regulatory guidelines [51] propose the prolongation of the QT interval on the electrocardiogram as a biomarker for torsadogenic risk of a given compound. However, TdP is triggered by a dynamic combination of multiple factors, and so, in order to reduce the number of false positives and false negatives, pharmaceutical companies are looking into combining the strengths of different methodologies (both experimental and modeling) to make the best informed decision. A detailed description of the cardiac system in humans allows inclusion of experimental data on the effects of different drugs in a relatively simple manner [52] and prediction of the effect of the molecules both in reference to healthy subjects as well as high-risk individuals. It is worthwhile to note that as the models become more predictive and computation becomes cheaper, such prospective simulations can be used routinely in the drug development process and can serve as data integration tools.

### 19.2.7 Disease Progression Modeling

**Purpose:** *To improve the assessment and predictability of a drug's impact on a disease by distinguishing drug-specific properties from disease-specific properties.*

Many chronic diseases that are targets of drug development, such as Alzheimer's, diabetes, and osteoporosis, are degenerative. The durations of clinical trials may not be insignificant on the time scales of disease progression. So disease progression is a source of variability, accounting for which can improve the detection of drug effect. On the other hand, clinical trials are almost always short relative to the expected duration of a drug's use by a chronic-disease patient. So a reliable forecast of a drug's long-term impact is important for developers, prescribers, patients, and payors.

Disease progression models first quantify the evolving status of a disease absent a drug and then characterize alterations induced by the drug. This hierarchical approach facilitates clarity, extrapolation to different dose regimens, and comparison of different drugs.

Although models for different diseases might be expected to have little in common, certain simple model forms have been found practical as well as conceptually enlightening. Chan and Holford [53, 54] considered the constant progression model,

$$\frac{dS}{dt} = \alpha \quad S(t) = S_0 \quad (17)$$

where  $S(t)$  is the disease status at time  $t$ ,  $S_0$  is the initial status, and  $\alpha$  is the rate of disease progression. They cited use of such a model for Alzheimer's, Parkinson's, chronic obstructive pulmonary disease (COPD), and diabetic nephropathy.

A drug may alter  $S(t)$  by affecting  $S_0$ ,  $\alpha$ , or both. A drug effect on  $S_0$  is considered a symptomatic effect; if drug treatment stops,  $S(t)$  reverts to its original trajectory. A drug effect on  $\alpha$  is considered a protective effect; if drug treatment stops,  $S(t)$

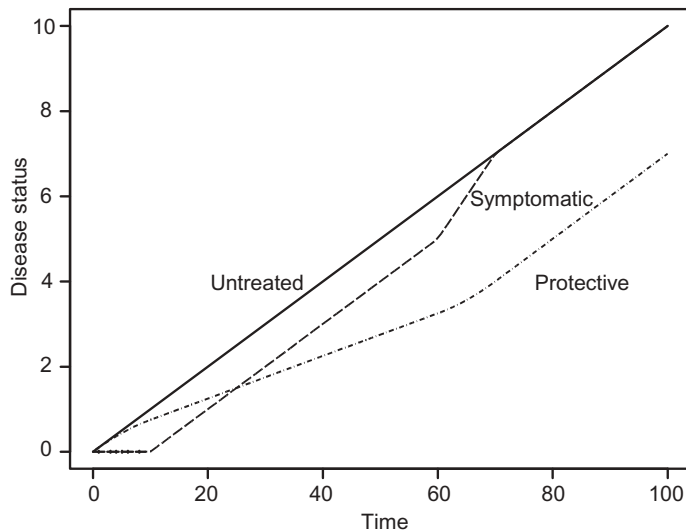


**FIGURE 3** Untreated disease progression and symptomatic and protective drug effects.

continues on a trajectory displaced from the original at the position where treatment stopped. See Figure 3.

Post et al. [55] introduced a class of models based on the disease process being a perturbation of a turnover model {19.2.4}. The disease alters  $k_{in}$  or  $k_{out}$  in a progressive way. Symptomatic drug effects add to or multiply the perturbed  $k_{in}$  or  $k_{out}$  without affecting the process of perturbation. Protective drug effects affect the process of perturbation directly. Models of this sort were applied to type 2 diabetes in [56].

### 19.2.8 Clinical Trial Simulation

**Purpose:** *To assess the expected performance of clinical trial designs and analysis methods.*

Clinical trials often have one or both of two objectives related to the two goals of drug development {19.1}: (i) learning—estimate some characteristic of the drug's behavior, such as the maximum tolerated dose; and (ii) confirming—test some hypothesis about the drug, such as whether the drug is superior to placebo. Modelers have long studied how different experimental designs, estimates, and tests affect the accuracy and precision of such inferences [57]. When their algebra and calculus are not powerful enough to quantify these properties, modelers resort to simulating {19.1.2} their experiments on the computer, for example, [58, 59].

Holford et al. [60] identified three broad classes of questions for which simulation experiments might be useful: (1) To investigate the properties of new model-based methods, such as nonlinear mixed-effect modeling for PopPK {19.2.4} [61]. (2) To study the properties of novel trial designs, such as concentration-controlled trials [62]. (3) Applications to specific clinical trials, for example, [63, 64]. For the latter, different designs can be input to the model and the best design selected for actual implementation based on comparison of the outputs. Simulating the different designs



may cost tens of dollars; actually conducting the trial may cost millions; the small investment of model-based forethought can have a big impact.

A trial simulation model may be thought of as one large model, the key input to which is the design of a clinical trial, and the key output of which is some performance measure of the trial. However, it is useful to consider component models. Commonly, these include an integrated dose–exposure–response (IDER) model [19.2.4], a covariate distribution model, and a trial execution model [60].

The covariate distribution model describes the trial’s target patient population in terms of the distribution of relevant covariates, such as the distribution of blood pressures in (2).

The trial execution model is a computer algorithm for simulating the trial. It generates simulated patients according to the covariate distribution model who experience outcomes according to the IDER model. The trial execution model may allow deviations from the trial protocol through noncompliance, missing observations, and the like. The trial’s simulated data are statistically analyzed according to the protocol. This is done repeatedly for many simulated replications of the trial. Then the statistical outcomes of all of the trials are summarized as a performance measure that addresses the question that motivated the simulation experiment.

An excellent compendium on clinical trial simulation is [65].

### 19.2.9 Decision Analysis

**Purpose:** *To help clinical teams create, compare, and evaluate strategies for clinical development, and to achieve clarity of action in choosing among strategies in the face of uncertainty.*

Modeling and simulation is a powerful tool for decision making due to its role in knowledge integration and its ability to quantify “what-if” scenarios via trial simulation [19.2.8]. Decision analysis is a structured process designed to facilitate effective communication between decision makers and a project team (including modelers and experts), driving toward clarity of action [66, 67]. Matheson and Matheson [68] characterize six components of a decision analysis: (1) framing, (2) strategy generation, (3) meaningful, reliable information, (4) clear values and trade-offs, (5) logically correct reasoning, and (6) commitment to action. Decision quality is achieved when further improvement in each component is not worth the additional time and resources that would be required. Each of these six components is now explained:

1. *Framing* consists of understanding the key questions. A decision hierarchy is developed to distinguish between those decisions already made, those decisions that are to be made now, and those that can be settled at a later time.
2. In *strategy generation*, creative doable strategies are assembled from different alternatives associated with the decisions to be made now.
3. *Meaningful, reliable information*. An integral feature of decision-analytic thinking is to accept the existence of uncertainties and to quantify them by means of probability distributions. Probabilities over clinical (and sometimes also market) possibilities may be developed via modeling from data, but also may be carefully assessed from experts in a manner designed to reduce various biases [69].

4. *Clear values and trade-offs* are selected to allow meaningful, quantitative comparisons of strategies. For-profit companies will typically be concerned with expected (i.e., probability averaged) net present (dollar) value. However, some useful proxies may also be evaluated, and then explicitly balanced against resources (i.e., cost and time). In early-phase “proof-of-concept” trials {19.3.3}, the proxy focus is typically the chance of a correct “go/no-go” after completion of the trial. In later phase (dose finding) trials {19.3.4}, the focus shifts toward the chance for correctly choosing the best dose regimen(s). For phase III trials, the focus is the chance of correctly confirming a compound’s safety and efficacy, leading to successful registration. Beyond the clinical development of a single compound, larger decisions may be analyzed, such as allocation of resources between many development projects in a portfolio [70].
5. *Logically correct reasoning* proceeds comparing the performance of selected strategies against the chosen evaluation criteria. Typically, trial simulation and probabilistic (Bayesian) inference {19.1.4} must be used. Sensitivity analysis can also be performed and the value of information computed [67], allowing better focus on the most salient information.
6. *Commitment to action* among stakeholders is achieved by involving the right people in the above steps. This includes a planned sequence of opportunities for the decision maker to provide input into the frame, the strategies chosen for consideration, and the evaluation criteria.

Other authors might use the term “decision analysis” to refer to steps 3–5. Examples of such decision-theoretic M&S applied to clinical trial designs can be found in [71, 72]. However, the fuller decision analysis process that includes all of steps 1–6 often generates insights leading to improved strategies; teams who engage in such a systematic process frequently report that the open, guided discussion of assumptions, uncertainties, alternatives, and values along with emergent insights provide as much or more benefit compared with the mathematical model outputs.

## 19.3 M&S-BASED CLINICAL TRIALS

### 19.3.1 First-in-Man Studies

**Purpose of M&S:** *To estimate doses for first-in-man studies.*

A hallmark of a good development candidate is to have a high probability of surviving through the clinical proof-of-concept PoC study {19.3.3}, showing pharmacologically useful activity. To assess this one answers three questions: (1) Is the site of action exposed to sufficiently efficacious drug concentration for a suitable duration? (2) Do these concentrations have a suitable safety profile? (3) Can formulations that achieve these concentrations be developed? Several research and development functions generate data, each utilizing different methodologies. M&S should play a key role to fully integrate these data, anticipate doses likely to elicit pharmacological and adverse effects in humans, and then optimize the early clinical development program.

The dose for a first-in-man (FIM) study is generally determined by following international, public guidances of the International Conference on Harmonisation (ICH) E8 [73] and ICH M3 [74]. These provide the rationale for estimating a safe starting dose by specifying target organs, dose dependency, relationships to exposure, and potential reversibility in FIM trials. Further, ICH S6 [75] (for biologicals) defines the principle of relative safety exposure between test animals and humans. The no-observed-adverse-effect level (NOAEL) and the toxic dose in animals are the bases of anticipating the FIM dose, and uncertainty of estimation is accounted for by setting safety margins. The Food and Drug Administration (FDA) [76] gives guidance on the maximum recommended starting dose (MRSD) for FIM studies by  $\text{mg}/\text{m}^2$ -based scaling of Good Laboratory Practice (GLP) toxicology data with a typical 10-fold safety margin.

That said, the guidance [76] does not cover how to determine the pharmacologically active dose, that is, the lowest dose with the intended pharmacological activity, making MRSD essentially a safety secured dose. However, an MRSD far below the pharmacologically active dose requires a long dose escalation scheme, while an MRSD far exceeding the right dose range will not allow us to detect the complete dose–effect relationship [77]. In addition, the guidance-compliant approach does not necessarily fully secure the safety, as highlighted in 2006 by TGN1412, a monoclonal antibody whose MRSD was calculated by conventional approaches with a 500-fold safety margin, but six healthy volunteers were seriously injured in the FIM study.

Collectively, a rational approach is demanded to efficiently anticipate FIM doses, in addition to the guidances, following the core scientific principle that drug effects are a function of target exposure. A three-step process is practical to anticipate the human dose–response [78] (Fig. 4): (1) Experimentally, determine exposure–response relationship in preclinical (animal) studies. Blood (plasma) concentration is often considered as a surrogate to the target exposure, assuming drug distribution equilibrium between circulating blood and drug target. Summary measures such as AUC or  $C_{\text{max}}$  [19.2.4] may be used if appropriate. (2) Adjust the exposure response curve from preclinical species to human. This will involve correcting for interspecies differences (e.g., known binding affinities to blood components and/or the drug target). (3) Convert the anticipated exposure–response to dose–response in humans using

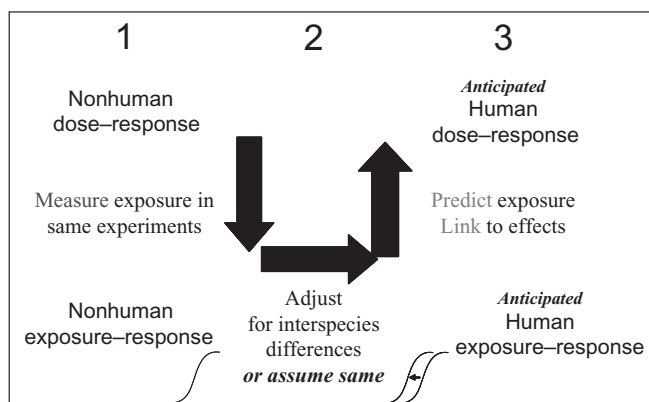


**FIGURE 4** Anticipating the first-in-man dose.

pharmacokinetic scaling, such as allometry [79] or physiologically based pharmacokinetic modeling {19.2.4} [80].

### 19.3.2 Oncology Phase I Studies

**Purpose of M&S:** *To optimally design and analyze first-in-man trials for oncology drugs.*

First-in-man trials for oncology trials differ from other therapeutic areas because oncology drugs often cannot be tested in healthy volunteers due to their toxicity at therapeutic doses. The typical design for such trials is therefore a dose escalation scheme in patients, with interim analysis and dose escalation/stopping decisions after each cohort. The primary endpoint is the occurrence of a dose-limiting toxicity (DLT), and the primary objective is the assessment of a maximally tolerated dose (MTD), corresponding to an acceptable target rate of DLTs. In order to optimally design such trials, the M&S challenge is to define decision rules for dose escalation and MTD selection (specifically, model-based rules), and to evaluate their operating characteristics through simulation.

Heuristic rules to decide upon dose escalation and MTD declaration have been proposed [81, 82], the most popular being the so-called “3+3” method, which bases decisions on whether to escalate, keep, or deescalate doses on how many DLTs are observed in the previous cohorts of size 3. Starting with the so-called continual reassessment method (CRM) [83], many model-based methods have been introduced, where decision rules and MTD selection are based on a model fitted to the available data. In its simplest form, where DLT is binary (i.e., occurrence or nonoccurrence), a model for DLT has the rate of DLTs as output and the dose as input along with parameters: rate of DLTs =  $f(\text{dose, parameters})$ . Due to the sparseness of the data (binary data from few patients), complicated models with many parameters cannot be fitted; thus, the proposed models typically have one or two parameters. An example is the two-parameter logistic function:  $\text{logit}(\text{rate of DLTs}) = \alpha + \beta \log(\text{dose})$ , where  $\text{logit}(p) = p/(1 - p)$  {19.2.1}. After each cohort, the model is fitted to the data, either through maximum likelihood {19.1.3} or using a Bayesian approach {19.1.4}. The latter has some distinct advantages since it allows to base decisions on posterior probabilities of the quantities of interest [84]. For example, one may want to specify the optimal target rate of DLTs to be in the range [0.25, 0.33]; anything beyond 0.33 would be deemed unacceptable toxicity. In a Bayesian framework, this is easily transformed to a decision rule, by choosing a dose maximizing the posterior probability of the rate of DLTs being in the interval [0.25, 0.33], while keeping the posterior probability that the rate of DLTs exceeds 0.33 below a threshold, for example, 0.05; this may be complemented with restrictions on how many dose levels to escalate.

The evaluation of operating characteristics is done through clinical trial simulation {19.2.8}. Generic features can be evaluated through the simulation of typical situations (see, e.g., [84]). However, specific trials with a unique set of prior assumptions and logistic restrictions will very often require specific simulation-based evaluation.

For a discussion of different approaches, with an emphasis on Bayesian methods, see [84] and the references given therein. For a nontechnical introduction see [85]. For references to extensions such as time-to-event and continuous outcomes, covari-

ates, combination trials, and combined safety–efficacy outcome, see references given in [84–87].

### 19.3.3 Proof-of-Concept Studies

**Purpose of M&S:** *To improve the information value of proof-of-concept trials with regard to efficacy and safety evaluation.*

Early development studies for establishing proof-of-concept (PoC) generally use small patient cohorts (typically between 10 and 20) observed for a relatively short period of time (several weeks) in a dose escalation design to evaluate early efficacy and safety signals. Cohorts are assigned, in sequence, to increasing doses until the maximum tolerated dose, generally determined in a previous study, is reached, or unacceptable safety is observed for a given cohort. A new cohort is only allowed to start once acceptable safety signals are verified for all previous doses. At the end of the study, one hopes either to determine a dose range for further exploration in phase IIb or to conclude that no PoC can be established based on the efficacy/safety trade-off.

Because of the small cohort sizes, only safety problems occurring in a relatively large percentage of patients can be reliably detected. Likewise, only relatively strong efficacy signals can be detected with reasonable statistical power using traditional pairwise hypothesis tests.

Safety and efficacy variables are often measured on a continuous scale and observed several times over the duration of the study. However, typically, the end-points for the go/no-go decision are based on a single time point (e.g., change from baseline at end of study) and on dichotomized versions of the original variables to characterize responder/nonresponder behavior. An example of the latter is the transformation of continuous liver function test measurements [e.g., alanine aminotransferase (ALT) and aspartate aminotransferase (AST)] into binary indicators of exceeding three times the upper limit of normal. There are, therefore, two types of information loss often present in PoC studies: (1) not using of all available longitudinal measurements collected in the study and (2) the dichotomization of continuous endpoints.

Safety and efficacy signal detection can be made more efficient in a variety ways: by using information external to the trial and longitudinal modeling approaches {19.2.2} to exploit all available information. Furthermore, the utility of PoC studies within the drug development program can be enhanced by incorporating the information obtained in them directly into later phase trials. Bayesian modeling techniques {19.1.4} are particularly useful in implementing these different approaches.

The key idea is to use a Bayesian mixed-effect model (typically nonlinear) {19.2.2} to characterize the longitudinal dose–response behavior of efficacy and safety variables, measured on a change from baseline scale (e.g., log of measurement divided by baseline). Preclinical or other external data can be used in the elicitation of priors for the Bayesian model. The estimated model is then used to simulate longitudinal profiles of change from baseline for different doses, which are combined with a sample of baseline values, obtained from a database of previous trials with a similar population, to produce dose–response longitudinal profiles. Dichotomization and other transformations can be applied to the simulated profiles to evaluate efficacy and safety more efficiently compared to using a single dichotomized time point. An

additional benefit of the approach is that it allows the use of external information, frequently available in pharmaceutical companies.

### 19.3.4 MCP-Mod: Unified Strategy for Dose-Finding Studies

**Purpose of M&S:** *To combine multiple comparison procedures and modeling in dose finding.*

Understanding and adequately representing the dose response (DR) profile of a compound, for both efficacy and safety, is a fundamental objective of clinical drug development. Proper understanding of this relationship has both a confirming and a learning component: confirming whether there is an overall DR effect (proof-of-concept {19.3.3}); and if so learning which doses to select for further development (dose finding). Selecting too low a dose decreases the chance of showing efficacy in later studies, while selecting too high a dose may result in tolerability or safety problems.

The analysis of dose-finding studies can be classified into two major strategies: modeling [88,89] and multiple comparison procedures (MCP) [90,91]. Dose–response (DR) modeling {19.2.4} assumes a functional relationship between dose (taken as a quantitative factor) and the response, according to a prespecified parametric model (defined in the study protocol). While this provides flexibility in investigating the effect of doses not used in the actual study, the validity of its conclusions highly depends on the correct choice of the DR model, which is typically unknown.

A multiple comparison procedure regards dose as a qualitative factor and makes very few, if any, assumptions about the underlying DR model. It can either be used for detecting an overall dose signal by means of trend tests or for the estimation of target doses by stepwise testing strategies, while preserving the overall type I error rate at a prespecified level. Such procedures are relatively robust to the underlying DR shape but are not designed for extrapolation of information beyond the observed dose levels.

MCP-Mod is a hybrid methodology combining features of MCP and modeling to provide a flexible and principled approach for designing and analyzing dose-ranging studies. Its central idea is the use of a set of candidate DR models to cover the possible shapes anticipated for the DR relationship. MCP are applied to a set of test statistics, determined optimally for the models in the candidate set, to decide which shapes give statistically significant signals. If no candidate model is statistically significant, the procedure stops and declares that no DR relationship can be established from the observed data (i.e., no proof of concept). Otherwise, out of the statistically significant models a best model is selected for dose estimation in the last stage of the procedure. The selection of the dose estimation model can be based on the minimum  $P$  value (of the candidate models) or some other relevant model selection criteria such as AIC or BIC {19.1.3}. The selected dose–response model is then employed to estimate target doses using inverse regression techniques [where the output (response) and input (dose) exchange roles] and possibly incorporating information on clinically relevant effects.

Simulation results suggest that MCP-Mod is as powerful as standard trend tests, while allowing more precise estimation of target doses than MCP, due to its modeling component [88, 92]. It is worth pointing out that this method can be seen as a



seamless design that combines proof of concept (phase Ib/IIa) with dose finding (phase IIb) into one single study.

### 19.3.5 Adaptive Trials

**Purpose of M&S:** *To determine the operating characteristics of an adaptive design, and ensure that the optimal design will be conducted.*

Adaptive designs are a relatively new type of clinical trial in which data from the trial can be used to modify some design aspect of the trial while the trial is still ongoing. This greater flexibility allows for design aspects to be modified in the case where some original trial design assumptions were incorrect, without having to wait for the conclusion of the trial. Such flexibility can save a considerable amount of development time for a novel drug. Adaptive trials can also combine what traditionally might be two separate studies into a single trial, thereby eliminating the time between the two trials. However, in order to design these trials, the decision process must be carefully thought out in advance and specified in the protocol. Often, clinical trial simulation {19.2.8} is the best way to understand and modify the operating characteristics of an adaptive trial.

Although there are many types of adaptation, one common type, sample size reestimation, is very appealing because often the parameters needed for sample size calculations are completely unknown at the beginning of the trial. Therefore, the idea is to try to estimate these needed parameters during the trial and modify the sample size accordingly. However, it has to be understood that at this interim time, the estimates for these parameters have considerable variability in them as well. Therefore, it might not be optimal to use such estimates directly, but instead use them to guide the calculation for the final sample size. A decision rule can be defined for how the sample size will be changed, and then computer simulation can be used to determine how the sample size would change under various scenarios.

Another type of adaptive design is a seamless adaptive design in which two trials are combined together. For example, a trial could be designed such that multiple doses are started in the beginning of the trial, and then the optimal dose is selected during an interim analysis. Additional patients would then only be randomized to the selected dosage group (and control), and all the data would be used at the conclusion of the trial to determine if the selected dose was indeed efficacious. If proper statistical testing procedures are in place at the end of the trial, then this type of design could cover objectives of both a phase II dose-finding study {19.3.4} and a confirmatory phase III trial.

For this type of seamless adaptive trial, simulation must also be used to determine the operating characteristics of the trial. For example, once a dose–response model {19.2.4} is assumed, and the dose selection algorithm is defined, the trial could be simulated to determine how often the selected dose would have a significant statistical test for efficacy at the conclusion of the trial (the statistical power for the trial). Another important characteristic of the trial that could be determined via simulation is the selection probabilities for each of the doses. Simulations should then be performed using different assumed dose responses, selection rules, and sample sizes to ensure that the trial has robust operating characteristics under many different scenarios.

For an introduction to adaptive trials, see [93].



## 19.4 DISCUSSION

This chapter introduces M&S in clinical drug development. Just a chapter cannot do the topic full justice, and indeed a thick book has recently been published that surveys the field [94]. The references herein provide some historical markers as well as some general introductions to specific methodologies. Readers interested in following trends would do well to monitor the journals *Clinical Pharmacology and Therapeutics*, *Journal of Pharmacokinetics and Pharmacodynamics*, *Statistics in Medicine*, and others cited in the references.

## REFERENCES

1. Sheiner, L. B. (1997), Learning versus confirming in clinical drug development, *Clin. Pharmacol. Therap.*, 61, 275–291.
2. Linhart, H., and Zucchini, W. (1986), *Model Selection*, Wiley, New York.
3. Berry, D. A. (2006), Bayesian clinical trials, *Nature Rev. Drug Dis.*, 5, 27–36.
4. McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, Chapman and Hall, New York.
5. Katz, M. H. (2006), *Multivariable Analysis: A Practical Guide for Clinicians*, 2nd ed., Cambridge University Press, Cambridge, UK.
6. Harrell, F. E., Jr. (2001), *Regression Modeling Strategies*, Springer, New York.
7. Liang, K.-Y., and Zeger, S. L. (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13–22.
8. Albert, P. S. (1999), Longitudinal data analysis (repeated measures) in clinical trials, *Stat. Med.*, 18, 1707–1732.
9. Diggle, P. J., Heagerty, P., Liang, K.-Y., et al. (2002), *Analysis of Longitudinal Data*, 2nd ed., Oxford University Press, Oxford, UK.
10. Witten, I. H., and Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, San Francisco.
11. Gibaldi, M., and Perrier, D. (1982), *Pharmacokinetics*, Marcel Dekker, New York.
12. Edgington, A. N., Schmitt, W., Voith, B., et al. (2006), A mechanistic approach for the scaling of clearance in children, *Clin. Pharmacokin.*, 45, 683–704.
13. Welling, P. E., and Balant, L. P., eds. (1994), *Handbook of Experimental Pharmacology, Vol. 110, Pharmacokinetics of Drugs*, Springer, Heidelberg.
14. Sheiner, L. B., Stanski, D. R., Vozeh, S., et al. (1979), Simultaneous modeling of pharmacokinetics and pharmacodynamics: Application to d-tubocurarine, *Clin. Pharmacol. Therapeut.*, 25, 358–371.
15. Dayneka, N. L., Garg, V., and Jusko, W. J. (1993), Comparison of four basic models of indirect pharmacodynamic response, *J. Pharmacokinetic. Biopharmaceut.*, 21, 457–478.
16. Mager, D. E., Wyska, E., and Jusko, W. J. (2003), Diversity of mechanism-based pharmacodynamic models, *Drug Metabolism Disposition*, 31, 510–519.
17. Danhof, M., de Jongh, J., De Lange, E. C. M., et al. (2007), Mechanism-based pharmacokinetic-pharmacodynamic modeling: Biophase distribution, receptor theory, and dynamical systems analysis, *Ann. Rev. Pharmacol. Toxicol.*, 47, 357–400.
18. Stanski, D. R., and Orloff, J. J. (2008), Communicating with the FDA: The “third rail” of a new model for drug development, *J. Clin. Pharmacol.*, 48, 144–145.
19. Sheiner, L. B., and Steimer, J. L. (2000), Pharmacokinetic/pharmacodynamic modeling in drug development, *Ann. Rev. Pharmacol. Toxicol.*, 40, 67–95.

20. Sheiner, L. B., Rosenberg, B., and Marathe, V. V. (1977), Estimation of population characteristics of pharmacokinetic parameters from routine clinical data, *J. Pharmacokinet. Pharmacodynam.*, 5, 445–479.
21. Yuh, L., Beal, S., Davidian, M., et al. (1994), Population pharmacokinetic/pharmacodynamic methodology and applications: A bibliography, *Biometrics*, 50, 566–575.
22. Bonate, P. (2005), Recommended reading in population pharmacokinetic pharmacodynamics, *AAPS J.*, 7, E363–E373.
23. Food and Drug Administration (1999), *Guidance for Industry: Population Pharmacokinetics*; available at: <http://www.fda.gov/cder/guidance/index.htm>.
24. Uyama, Y., Shibata, T., Nagai, N., et al. (2005), Successful bridging strategy based on ICH E5 guideline for drugs approved in Japan, *Clin. Pharmacol. Therapeut.*, 78, 102–113.
25. Meibohm, P., Laer, S., Panetta, J. C., et al. (2005), Population pharmacokinetics studies in pediatrics: Issues in design and analysis, *AAPS J.*, 7, E475–E487.
26. Duffull, S., Waterhouse, T., and Eccleston, J. (2005), Some considerations on the design of population pharmacokinetic studies, *J. Pharmacokinet. Pharmacodynam.*, 32, 441–457.
27. Ribbing, J., and Jonsson, E. N. (2004), Power, selection bias and predictive performance of the population pharmacokinetics covariate model, *J. Pharmacokinet. Pharmacodynam.*, 31, 109–134.
28. Ette, E. I. (1997), Stability and performance of a population pharmacokinetic model, *J. Clin. Pharmacol.*, 37, 486–495.
29. Yano, Y., Beal, S. L., and Sheiner, L. B. (2001), Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J. Pharmacokinet. Pharmacodynam.*, 28, 171–192.
30. Brendel, K., Comets, E., Laffont, C., et al. (2006), Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide, *Pharma. Res.*, 23, 2036–2049.
31. International Conference on Harmonisation (1994), *Dose Response Information to Support Drug Registration*; available at: <http://www.emea.europa.eu/pdfs/human/ich/037895en.pdf>.
32. Vítko, S., Margreiter, R., Weimar, W., et al. (2004), Everolimus (Certican) 12-month safety and efficacy versus mycophenolate mofetil in de novo renal transplant recipients, *Transplantation*, 78, 1532–1540.
33. Lorber, M. I., Mulgaonkar, S., Butt, K. M., et al. (2005), Everolimus versus mycophenolate mofetil in the prevention of rejection in de novo renal transplant recipients: A 3-year randomized, multicenter, phase III study, *Transplantation*, 80, 244–252.
34. Lorber, M. I., Ponticelli, C., Whelchel, J., et al. (2005), Therapeutic drug monitoring for everolimus in kidney transplantation using 12-month exposure, efficacy, and safety data, *Clin. Transplant.*, 19, 145–152.
35. Vitko, S., Tedesco, H., Eris, J., et al. (2004), Everolimus with optimized cyclosporine dosing in renal transplant recipients: 6-month safety and efficacy results of two randomized studies, *Am. J. Transplant.*, 4, 626–635.
36. Kovarik, J. M., Tedesco, H., Pascual, J., et al. (2004), Everolimus therapeutic concentration range defined from a prospective trial with reduced-exposure cyclosporine in de novo kidney transplantation, *Therap. Drug Monit.*, 26, 499–505.
37. Aksenov, S., Church, B., Dhiman, A., et al. (2005), An integrated approach for inference and mechanistic modeling for advancing drug development, *FEBS Lett.*, 579, 1878–1883.

38. Christopher, R., Dhiman, A., Fox, J., et al. (2004), Data-driven computer simulation of human cancer cell, *Ann. NY Acad. Sci.*, 1020, 132–153.
39. Fitzgerald, J., Schoeberl, B., Nielsen, U., et al. (2006), Systems biology and combination therapy in the quest of clinical efficacy, *Nature Chem. Biol.*, 2, 458–466.
40. Liu, Y., Purvis, J., Shih, A., et al. (2007), A multiscale computational approach to dissect early events in the Erb family receptor mediated activation, differential signaling and relevance to oncogenic transformations, *Ann. Biomed. Eng.*, 35, 1012–1025.
41. Schoeberl, B., Pace, E., Shavonne, H., et al. (2006), A data-driven computational model of the ErbB receptor signaling network, *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1, 53–54.
42. Schoeberl, B., Eichler-Johnson, C., Gilles, E., et al. (2002), Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized receptors, *Nature Biotechnol.*, 20, 370–375.
43. Nielsen, U., and Schoeberl, B. (2005), Using computational modeling to drive the development of targeted therapeutics, *IDrugs*, 8, 822–826.
44. Hornberg, J., Binder, B., Bruggeman, F., et al. (2002), Control of MAPK signalling: From complexity to what really matters, *Oncogene*, 24, 5533–5542.
45. Hoppensteadt, F. C., and Peskin, C. S. (2002), *Modeling and Simulation in Medicine and the Life Sciences*, Springer, New York.
46. Thomas, S. R., Layton, A. T., Layton, H. E., et al. (2006), Kidney modeling: Status and perspectives, *Proc. IEEE*, 94, 740–752.
47. Subramaniam, R., Ashagarian, B., Freijer, J., et al. (2003), Analysis of lobar differences in particle deposition in the human lung, *Inhal. Toxicol.*, 15, 1–21.
48. McCulloch, A., Bassingthwaite, J. B., Hunter, P. J., et al. (1998), Computational biology of the heart: From structure to function. *Prog. Biophys. Mol. Biol.*, 69, 151–572.
49. Abram, S., Hodnett, B., Summers, R., et al. (2007), Quantitative circulatory physiology: An integrative mathematical model of human physiology for medical education, *Adv. Physiol. Ed.*, 31, 202–210.
50. Dumotier, B., and Georgieva, A. (2007), Preclinical cardio-safety assessment of torsadogenic risk and alternative methods to animal experimentation: The inseparable twins, *Cell Biol. Toxicol.*, 23, 293–302.
51. International Conference on Harmonisation. (2005), *The Non-Clinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals*; available at: <http://www.emea.europa.eu/pdfs/human/ich/042302en.pdf>.
52. Bottino, D., Penland, C., Stamps, A., et al. (2006), Preclinical cardiac safety assessment of pharmaceutical compounds using an integrated systems-based computer model of the heart, *Prog. Biophys. Mol. Bio.*, 90, 414–443.
53. Chan, P. L. S., and Holford, N. H. G. (2001), Drug treatment effects on disease progression, *Ann. Rev. Pharmacol. Toxicol.*, 41, 625–659.
54. Holford, N. (2007), Modelling disease progression; available at: <http://www.page-meeting.org/page/page2007/ModellingDiseaseProgression.pdf>.
55. Post, T. M., Freijer, J. I., DeJongh, J., et al. (2005), Disease system analysis: Basic disease progression models in degenerative disease, *Pharma. Res.*, 22, 1038–1049.
56. de Winter, W., DeJongh, J., Post, T., et al. (2006), A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying type 2 diabetes mellitus, *J. Pharmacokinet. Pharmacodyn.*, 33, 313–343.

57. Bacchieri, A., and Della Cioppa, G. (2007), *Fundamentals of Clinical Research*, Springer, Milan.
58. Tang, D., Geller, N. L., and Pocock, S. J. (1993), On the design and analysis of randomized clinical trials with multiple endpoints, *Biometrics*, 49, 23–30.
59. O'Quigley, J., and Chevret, S. (1991), Methods for dose finding studies in cancer clinical trials: A review and results of a monte carlo study, *Stat. Med.*, 10, 1647–1664.
60. Holford, N. H. G., Kimko, H. C., Monteleone, J. P. R., et al. (2000), Simulation of clinical trials, *Ann. Rev. Pharmacol. Therapeut.*, 40, 209–234.
61. Sheiner, L. B., and Beal, S. L. (1983), Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine pharmacokinetic data, *J. Pharmacokinet. Biopharmaceut.*, 11, 303–319.
62. Sanathanan, L. P., and Peck, C. C. (1991), The randomized concentration-controlled trial: An evaluation of its sample size efficiency, *Controlled Clin. Trials*, 12, 781–794.
63. Hale, M. D. (1997), Using population pharmacokinetics for planning a randomized concentration-controlled trial with a binary response, in Aarons, L., Balant, L. P., Danhof, M., Eds., *European Cooperation in the Field of Scientific and Technical Research*, European Commission, Geneva, pp. 227–235.
64. Hale, M. D., Nicholls, A. J., Bullingham, R. E. S., et al. (1998), The pharmacokinetic-pharmacodynamic relationship for mycophenolate mofetil in renal transplantation, *Clin. Pharmacol. Therapeut.*, 64, 672–683.
65. Kimko, H. C., and Duffull, S. B. (2003), *Simulation for Designing Clinical Trials*, Dekker, New York.
66. Howard, R. A. (1989), The evolution of decision analysis, in Howard, R. A., and Matheson, J. E., Eds., *Readings on the Principles and Applications of Decision Analysis*, Strategic Decisions Group, Menlo Park, CA, pp. 7–16.
67. Clemen, R. T. (1996), *Making Hard Decisions, An Introduction to Decision Analysis*, 2nd ed., Brooks/Cole, Pacific Grove, CA.
68. Matheson, D., and Matheson, J. (1998), *The Smart Organization*, Harvard Business School Press, Cambridge, MA.
69. Spetzler, C., and Stael von Holstein, C. (1989), Probability encoding in decision analysis, in Howard and Matheson, Eds. *Readings on the Principles and Applications of Decision Analysis*, Vol. II, Strategic Decisions Group, Menlo Park, CA, pp. 601–626.
70. Sharpe, P., and Keelin, T. (1998), How Smith Kline and Beecham makes better research allocation decisions, *Harvard Business Rev.*, March–April, 5–10.
71. Pally, A. (2000), A decision analytic approach to determining sample sizes in a Phase III program, *Drug Info. J.*, 34, 365–377.
72. Berry, D. (2004), Bayesian statistics and the efficiency and ethics of clinical trials, *Stat. Sci.*, 19, 175–187.
73. International Conference on Harmonisation (1998), *General Considerations for Clinical Trials*; available at: <http://www.emea.europa.eu/pdfs/human/ich/029195en.pdf>.
74. International Conference on Harmonisation (2000), *Non-Clinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals*; available at: <http://www.emea.europa.eu/pdfs/human/ich/028695en.pdf>.
75. International Conference on Harmonisation (1998), *Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals*; available at: <http://www.emea.europa.eu/pdfs/human/ich/030295en.pdf>.
76. U. S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (2005), *Guidance for Industry: Estimating the Maximum*

- Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers*; available at: <http://www.fda.gov/cder/guidance/5541fnl.pdf>.
77. Reigner, B. G., and Blesch, K. S. (2002), Estimating the starting dose for entry into humans: Principles and practice. *Eur. J. Clin. Pharmacol.*, 57, 835–845.
  78. Lowe, P. J., Hijazi, Y., Luttringer, O., et al. (2007), On the anticipation of the human dose in first-in-man trials from preclinical or prior clinical information in early drug development, *Xenobiotica*, 37, 1331–1354.
  79. Feng, M. R., Lou, X., Brown, R., et al. (2000), Allometric pharmacokinetic scaling: Towards the prediction of human oral pharmacokinetics. *Pharma. Res.*, 17, 410–418.
  80. Kawai, R., Mathew, D., Tanaka, C., et al. (1998), Physiologically based pharmacokinetics of cyclosporine A: Extension to tissue distribution kinetics in rats and scale-up to human, *J. Pharmacol. Exper. Therapeut.*, 287, 457–468.
  81. Dixon, W., and Mood, A. (1948), A method for obtaining and analyzing sensitivity data, *J. Am. Statist. Assoc.*, 43, 109–126.
  82. Storer, B. (1989), Design and analysis of Phase I clinical trials, *Biometrics*, 45, 925–937.
  83. O'Quigley, J., Pepe, M., and Fisher, L. (1990), Continual reassessment method: A practical design for Phase I clinical trials in cancer, *Biometrics*, 46, 33–48.
  84. Neuenschwander, B., Branson, M., and Gsponer, T. (2008), Critical aspects of the Bayesian approach to phase I cancer trials, *Stat. Med.*, 27, 2420–2439.
  85. Garrett-Mayer, E. (2005), *Understanding the Continual Reassessment Method for Dose Finding Studies: An Overview for Non-Statisticians*, Johns Hopkins University, Dept. of Biostatistics Working Papers, Baltimore, MD, Paper 74.
  86. Thall, P. F., and Cook, J. D. (2004), Dose finding based on efficacy toxicity trade offs, *Biometrics*, 60, 684–693.
  87. Whitehead, J., Zhou, Y., Stevens, J., et al. (2004), An evaluation of a Bayesian method of dose escalation based on bivariate binary responses, *J. Biopharma. Stat.*, 14, 969–983.
  88. Pinheiro, J., Bretz, F., and Branson, M. (2006), Analysis of dose response studies: Modeling approaches, in Ting, N., Ed., *Dose Finding in Drug Development*, Springer, New York, pp. 146–171.
  89. Bates, D. M., and Watts, D. G. (1988), *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.
  90. Hochberg, Y., and Tamhane, A. C. (1987), *Multiple Comparisons Procedures*, Wiley, New York.
  91. Hsu, J. C. (1996), *Multiple Comparisons*, Chapman and Hall, New York.
  92. Bretz, F., Pinheiro, J., and Branson, M. (2005), Combining multiple comparisons and modeling techniques in dose response studies, *Biometrics*, 61, 738–748.
  93. Gallo, P., Chuang-Stein, C., Dragalin, V., et al. (2006), Adaptive designs in clinical drug development—an executive summary of the PhRMA working group, *J. Biopharma. Stat.*, 16, 275–283.
  94. Ette, E. I., and Williams, P. J. (2007), *Pharmacometrics: The Science of Quantitative Pharmacology*, Wiley, Hoboken, NJ.



# 20

---

## *Monitoring*

**Nigel Stallard<sup>1</sup> and Susan Todd<sup>2</sup>**

<sup>1</sup>*Warwick Medical School, University of Warwick, Warwick, United Kingdom*

<sup>2</sup>*Applied Statistics, University of Reading, Reading,  
United Kingdom*

### **Contents**

|        |                                                                                 |      |
|--------|---------------------------------------------------------------------------------|------|
| 20.1   | Reasons for Monitoring Clinical Trial Data                                      | 1020 |
| 20.2   | Monitoring of Clinical Trial Data without Unblinding Treatment Allocations      | 1021 |
| 20.2.1 | Administrative and Safety Data Monitoring                                       | 1021 |
| 20.2.2 | Monitoring for Estimation of Nuisance Parameters and Sample-Size Reviews        | 1022 |
| 20.3   | Dangers of Unplanned Clinical Trial Monitoring                                  | 1026 |
| 20.4   | Methodology for Sequential Monitoring of Unblinded Efficacy Data                | 1027 |
| 20.4.1 | Parameterization of Treatment Difference and Test Statistics                    | 1028 |
| 20.4.2 | Repeated Significance Testing                                                   | 1029 |
| 20.4.3 | Spending Function Approach                                                      | 1030 |
| 20.4.4 | Boundaries Approach                                                             | 1031 |
| 20.4.5 | Analysis after a Sequential Trial                                               | 1032 |
| 20.4.6 | Monitoring and Analysis of More Than One Endpoint                               | 1033 |
| 20.5   | Implementation of Sequential Designs                                            | 1034 |
| 20.6   | Example: Lithium Gamolenate in Patients with Advanced Pancreatic Adenocarcinoma | 1035 |
| 20.7   | Response-Driven and Adaptive Designs                                            | 1038 |
|        | References                                                                      | 1039 |



## 20.1 REASONS FOR MONITORING CLINICAL TRIAL DATA

Many traditional methods for the analysis of data assume that the data are all available contemporaneously. This reflects the setting of agricultural experimentation in which much of the methodology was developed, as data on crop yields from different fields are collected simultaneously after harvesting. In the majority of clinical trials, however, patients are recruited over a period of time, so that data accumulate gradually through the course of the trial, with data from some patients available prior to the recruitment of others. This accumulation of data makes possible the statistical analysis of some of the data collected before the trial is completed. Such analysis of data during the course of the trial is termed *monitoring*.

There are a number reasons why monitoring of clinical trial data might be considered desirable. The first is an ethical reason. In a randomized trial, randomization of subjects to receive one treatment or another is generally considered ethically acceptable provided there is *equipoise*, that is, an uncertainty as to which treatment is superior. If clear evidence was available that one treatment was of superior efficacy to the other, or that one treatment was less safe than the other, continued randomization would be considered unethical. In a single-arm trial, enrollment in the trial would be similarly considered unethical if the treatment being assessed clearly led to an unacceptable level of adverse events. Monitoring of efficacy and safety data allows the results of an ongoing trial to be regularly assessed and enables termination of the trial as soon as the results are sufficiently convincing that continued recruitment cannot be ethically defended.

The second reason for monitoring a clinical trial is administrative. As the trial progresses, it is desirable to ensure protocol compliance, check data collection procedures, and to monitor recruitment rates. The estimation of nuisance parameters (such as the unknown variance for normally distributed responses) used in power calculations can also be conducted to ensure that the trial will have a sufficient sample size before the close of recruitment.

The third reason for monitoring a clinical trial is economic. Clinical trials are expensive, and it is not in the interest of sponsors or patients to continue a trial once the results are clear. By monitoring the accumulating data, a clinical trial can be stopped as soon as sufficient data are available to justify a conclusion that an experimental treatment is superior or inferior to the control, or that the data clearly indicate that any difference is so small that continuation of the trial is futile.

In a blinded randomized trial we may consider data monitoring to be divided into three types: first, administrative monitoring of clinical trial conduct and monitoring of safety data with no monitoring of efficacy data; second, monitoring of efficacy data without unblinding of treatment allocation, for example, for the estimation of nuisance parameters; third, monitoring of efficacy data with treatment allocation unblinded to allow the estimation of the difference in efficacy between the treatments being compared. In the latter two instances, administrative and safety monitoring will most likely be conducted in addition. As will be discussed in Section 20.3, it is the third of these types of monitoring that presents the most ethical and statistical challenges. It is therefore on this type of monitoring that the majority of the chapter will be focused. Before moving on, however, the first two types of monitoring will be discussed briefly in the next section.

## 20.2 MONITORING OF CLINICAL TRIAL DATA WITHOUT UNBLINDING TREATMENT ALLOCATIONS

### 20.2.1 Administrative and Safety Data Monitoring

Monitoring of administrative data such as recruitment rates, accuracy of completion of record forms, treatment compliance, and data collection/transfer procedures is common in large clinical trials. The aim of such monitoring is to ensure that any problems in trial conduct or data collection are identified and rectified as soon as possible. Identification of incomplete record forms from a particular investigator could lead, for example, to collection of the missing data prior to the end of the trial, while observation of a lower than anticipated recruitment rate might justify the introduction of new centers in a multicenter trial. An example of a form to assess ongoing treatment compliance as part of a trial is given in Figure 1. Administrative monitoring of this kind is generally conducted without unblinding treatment allocation since treatment information is generally irrelevant. The risks of increased error rates or release of information about the performance of the study drugs that could lead to bias as discussed in Section 20.3 therefore do not arise from this sort of monitoring.

Many large trials also involve some monitoring of safety data. As discussed above, in a trial in which serious adverse events may be observed, this type of monitoring serves an important ethical function. In order to assess whether there is an excess number of adverse events in one treatment group relative to the other, it is necessary to unblind treatment allocation. In some cases, the two treatment groups are presented simply as groups A and B, without indicating which group is receiving the experimental treatment and which the control. However, it is generally easier to interpret adverse events in the light of the treatment administered, so that full unblinding is considered to be a better practice [1]. The unblinding of treatments introduces the risk of bias as discussed in Section 20.3. In order to minimize this risk, particularly in large clinical trials, the safety data may be reviewed by a *data and safety monitoring board* (DSMB) (sometimes alternatively called a *data monitoring committee*), so that only members of this committee need have access to the unblinded data. This committee decides on the basis of the unblinded data reviewed whether or not it is safe for the trial to continue and makes its recommendation to the *trial steering committee* responsible for the overall conduct of the trial. Further details of the role and function of DSMBs are given by Ellenberg et al. [2] and Whitehead [3]. Two example forms for monitoring adverse events are given in Figures 2 and 3. Figure 2 illustrates the categorizations to be used when determining adverse events and their attributability. Figure 3 shows how a case report form for serious adverse events might be structured.

As the purpose of safety monitoring is solely to recommend termination of the clinical trial if an experimental treatment is associated with too many adverse events, this type of monitoring cannot lead to a positive claim of effectiveness for the experimental treatment based on the monitored data. This means that, in contrast to the type of efficacy monitoring discussed in Section 20.3, there is no risk of a false-positive conclusion arising as a result of this monitoring; that is, the type I error rate cannot be increased. The special statistical methodology that is discussed in the rest of this chapter is therefore not required in this case.

|                  |                      |                                                                                     |
|------------------|----------------------|-------------------------------------------------------------------------------------|
| Treatment Period | PATIENT STUDY NUMBER | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> |
|                  | PATIENT INITIALS     | <input type="text"/> <input type="text"/> <input type="text"/>                      |

**PHARMACY DRUG DISPENSING RECORD**

|                                                         |                                                                                     |
|---------------------------------------------------------|-------------------------------------------------------------------------------------|
| <b>ISSUES</b>                                           |                                                                                     |
| Date study medication dispensed                         | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> |
| Number of capsules dispensed                            | <input type="text"/> <input type="text"/> <input type="text"/>                      |
| or    Number of ampoules dispensed                      | <input type="text"/> <input type="text"/> <input type="text"/>                      |
| Signature of dispensing pharmacist _____                |                                                                                     |
| <b>RETURNS</b>                                          |                                                                                     |
| Date study medication returned                          | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> |
| Number of capsules returned                             | <input type="text"/> <input type="text"/> <input type="text"/>                      |
| or    Number of unopened ampoules returned              | <input type="text"/> <input type="text"/> <input type="text"/>                      |
| Signature of receiving pharmacist _____                 |                                                                                     |
| <b>VERIFICATION</b> <i>(to be completed by monitor)</i> |                                                                                     |
| Date of verification                                    | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> |
| Number of capsules checked                              | <input type="text"/> <input type="text"/> <input type="text"/>                      |
| or    Number of unopened ampoules                       | <input type="text"/> <input type="text"/> <input type="text"/>                      |
| Signature of monitor _____                              |                                                                                     |

**FIGURE 1**    Example form for monitoring treatment compliance.

**20.2.2    Monitoring for Estimation of Nuisance Parameters and Sample-Size Reviews**

Sample size calculations at the planning stage of a clinical trial often require the specification of values for unknown parameters other than the treatment effect of primary interest. These are often known as *nuisance parameters*. For example, if data are assumed to be normally distributed, the primary measure of relative treatment

**GUIDELINES FOR RECORDING OF ADVERSE EVENTS**

|                             |   |                                                                                                                                                                                         |
|-----------------------------|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Severity                    | 1 | Mild                                                                                                                                                                                    |
|                             | 2 | Moderate                                                                                                                                                                                |
|                             | 3 | Severe                                                                                                                                                                                  |
|                             | 4 | Serious – notify the company immediately by telephone. Complete and return Serious Adverse Event form, which can be found in the Investigator File                                      |
| Relation to Treatment       | 1 | Probably                                                                                                                                                                                |
|                             | 2 | Possibly                                                                                                                                                                                |
|                             | 3 | Probably not                                                                                                                                                                            |
|                             | 4 | Related to concomitant therapy                                                                                                                                                          |
|                             | 5 | Unknown                                                                                                                                                                                 |
| Action taken due to this AE | 0 | None                                                                                                                                                                                    |
|                             | 1 | Dosage changed                                                                                                                                                                          |
|                             | 2 | Temporary interruption                                                                                                                                                                  |
|                             | 3 | Permanent discontinuation                                                                                                                                                               |
|                             | 4 | Drug therapy changed or added – note changes on medication page                                                                                                                         |
|                             | 5 | Other therapy – give details below                                                                                                                                                      |
|                             | 6 | Hospitalization or prolonged hospitalization – notify the company immediately by telephone. Complete and return Serious Adverse Event form, which can be found in the Investigator File |
| Outcome                     | 1 | Complete recovery                                                                                                                                                                       |
|                             | 2 | Recovery with sequelae                                                                                                                                                                  |
|                             | 3 | Condition improving                                                                                                                                                                     |
|                             | 4 | Condition unchanged                                                                                                                                                                     |
|                             | 5 | Condition worsening                                                                                                                                                                     |
|                             | 6 | Death due to this AE – notify the company immediately by telephone. Complete and return Serious Adverse Event form, which can be found in the Investigator File                         |

**NB A SERIOUS ADVERSE EVENT is defined as the following:**

- Any untoward medical occurrence that at any dose
- Results in death
  - Requires inpatient hospitalization or prolongation of existing hospitalization
  - Results in persistent or significant disability/incapacity
  - Is life threatening (at the time of the event)

**FIGURE 2** Example categorizations for adverse event monitoring.

efficacy will often be the difference in the mean response for the two treatments. Specification of the variance of the response, which is commonly assumed to be the same in the two groups, is, however, needed for calculation of the required sample size. A second example is for binary data, when the sample-size calculation depends on the average probability of success for the two groups. In practice, in such cases, the guess of the nuisance parameter required before the start of the trial may be inaccurate. This might lead to an inappropriate choice of sample size and to an underpowered or overpowered trial.

**1. PATIENT AND REACTION INFORMATION**

|                                                                                                                                         |                           |         |                                                                                                                                        |  |  |  |     |     |    |     |     |                                                                                                                                         |  |  |  |     |     |    |                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------------------------------------------------------------------------------------------------------------------|---------------------------|---------|----------------------------------------------------------------------------------------------------------------------------------------|--|--|--|-----|-----|----|-----|-----|-----------------------------------------------------------------------------------------------------------------------------------------|--|--|--|-----|-----|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Patient Initials                                                                                                                        | Study Code<br>Patient No. | Country | Date of Birth<br><table border="1"> <tr> <td></td> <td></td> <td></td> </tr> <tr> <td>Day</td> <td>Mth</td> <td>Yr</td> </tr> </table> |  |  |  | Day | Mth | Yr | Age | Sex | Reaction onset<br><table border="1"> <tr> <td></td> <td></td> <td></td> </tr> <tr> <td>Day</td> <td>Mth</td> <td>Yr</td> </tr> </table> |  |  |  | Day | Mth | Yr | Tick all appropriate to adverse reaction<br><input type="checkbox"/> Patient died<br><br><input type="checkbox"/> Involved or prolonged inpatient hospitalization<br><br><input type="checkbox"/> Involved significant disability or incapacity<br><br><input type="checkbox"/> Life threatening<br><br><input type="checkbox"/> None of the above |
|                                                                                                                                         |                           |         |                                                                                                                                        |  |  |  |     |     |    |     |     |                                                                                                                                         |  |  |  |     |     |    |                                                                                                                                                                                                                                                                                                                                                    |
| Day                                                                                                                                     | Mth                       | Yr      |                                                                                                                                        |  |  |  |     |     |    |     |     |                                                                                                                                         |  |  |  |     |     |    |                                                                                                                                                                                                                                                                                                                                                    |
|                                                                                                                                         |                           |         |                                                                                                                                        |  |  |  |     |     |    |     |     |                                                                                                                                         |  |  |  |     |     |    |                                                                                                                                                                                                                                                                                                                                                    |
| Day                                                                                                                                     | Mth                       | Yr      |                                                                                                                                        |  |  |  |     |     |    |     |     |                                                                                                                                         |  |  |  |     |     |    |                                                                                                                                                                                                                                                                                                                                                    |
| DESCRIBE REACTION(S)<br>(Give signs or symptoms, diagnoses, course ( <u>underlying the main event</u> )<br>(Include relevant lab. data) |                           |         |                                                                                                                                        |  |  |  |     |     |    |     |     |                                                                                                                                         |  |  |  |     |     |    |                                                                                                                                                                                                                                                                                                                                                    |

**2. SUSPECT DRUG INFORMATION**

|                                                                                                          |                         |                                  |                                                                                                                                      |
|----------------------------------------------------------------------------------------------------------|-------------------------|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| SUSPECT DRUG (include all information available: generic name, form and strength, dosage, batch number). |                         |                                  | DID REACTION ABATE AFTER STOPPING DRUG?<br>Yes <input type="checkbox"/> No <input type="checkbox"/> NA* <input type="checkbox"/>     |
| DAILY DOSE (with unit)                                                                                   | ROUTE OF ADMINISTRATION |                                  | DID REACTION REAPPEAR AFTER REINTRODUCTION?<br>Yes <input type="checkbox"/> No <input type="checkbox"/> NA* <input type="checkbox"/> |
| INDICATION FOR USE                                                                                       |                         |                                  |                                                                                                                                      |
| THERAPY DATES (Day, Mth., Yr)<br>From _____ to _____                                                     | THERAPY DURATION        | TOTAL DOSE RECEIVED (with units) | *NA: Not applicable<br>e.g., only 1 dose or irreversible outcome                                                                     |

**3. COMCOMITANT DRUG(S) AND HISTORY**

|                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------|
| CONCURRENT DRUG(S) AND DATES OF ADMINISTRATION (Trade names, generic names, form, strength of dosage. Exclude those used to treat reaction). |
| OTHER RELEVANT HISTORY (e.g., diagnoses, allergies, pregnancy with LMP, etc.)                                                                |

**4. REPORTER OR INVESTIGATOR INFORMATION**

|                          |                                                                                                   |                                          |                                                                                                                                                                                                                                        |                                                                                                       |  |  |  |
|--------------------------|---------------------------------------------------------------------------------------------------|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|--|--|--|
| NAME & ADDRESS           | REPORT SENT TO NATIONAL AUTHORITY<br><br><input type="checkbox"/> Yes <input type="checkbox"/> No | CSM ADVERSE REACTION REGISTRATION NUMBER | CAUSALITY ASSESSMENT (CONCERNING DRUG)<br><br><input type="checkbox"/> UNLIKELY<br><input type="checkbox"/> NOT RELATED<br><input type="checkbox"/> POSSIBLE<br><input type="checkbox"/> PROBABLE<br><input type="checkbox"/> DEFINITE | DATE OF REORT: (Day, Mth, Yr)<br><table border="1"> <tr> <td></td> <td></td> <td></td> </tr> </table> |  |  |  |
|                          |                                                                                                   |                                          |                                                                                                                                                                                                                                        |                                                                                                       |  |  |  |
| DATE SENT (Day, Mth, Yr) | SIGNED<br>_____                                                                                   |                                          |                                                                                                                                                                                                                                        |                                                                                                       |  |  |  |

**5. ADMINISTRATIVE INFORMATION (To be completed by company personnel)**

|                                                                                                                                                                |                                                                                       |                                                                                                                                                                             |  |  |                                                                            |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|----------------------------------------------------------------------------|
| DATE RECEIVED: (Day, Mth, Yr)                                                                                                                                  | RECEIVED BY:                                                                          | STAFF ADVISED (LIST):<br>ADVERSE REACTION NUMBER                                                                                                                            |  |  |                                                                            |
| DEPARTMENT OF HEALTH ADVISED?<br><input type="checkbox"/> No <input type="checkbox"/> Yes DATE: <table border="1"><tr><td></td><td></td><td></td></tr></table> |                                                                                       |                                                                                                                                                                             |  |  | CSM ADVERSE REACTION REGISTRATION NUMBER:<br><br>(if not registered above) |
|                                                                                                                                                                |                                                                                       |                                                                                                                                                                             |  |  |                                                                            |
| DATE OF THE REPORT (Day, Mth, Yr)                                                                                                                              | REPORT TYPE<br><input type="checkbox"/> INITIAL<br><input type="checkbox"/> FOLLOW-UP | REPORT SOURCE<br><input type="checkbox"/> STUDY<br><input type="checkbox"/> HEALTH PROFESSIONAL<br><input type="checkbox"/> LITERATURE<br><input type="checkbox"/> CONSUMER |  |  |                                                                            |

**FIGURE 3** Example form for monitoring serious adverse events.

Gould and Shi [4] proposed, as a solution to this problem, the use of an analysis of the accumulating data part way through the trial to estimate the unknown nuisance parameter in what is usually termed a *sample-size review*. A review of the method and further advances in the methodology is given by Gould [5]. The required sample size for the trial can then be obtained based on this estimate of the nuisance parameter, and the sample size for the remainder of the trial adjusted accordingly. For example, suppose a trial with a normally distributed response was initially designed with a sample size of 100 patients. An analysis could be conducted when data were available from 50 patients to calculate the unknown common variance. At this point it might be revealed that the estimated variance based on the data was larger than the value anticipated when the trial was planned, so that the trial is likely to be underpowered. Based on the estimate from the data, the required sample size might be 120 patients rather than the initial 100 planned. The second stage of the trial could then recruit a total of 70 rather than 50 patients to give this total sample size. Simulation studies have shown that such an approach is effective in maintaining the power even if the initial guess of the unknown nuisance parameter is inaccurate.

In the case of normally distributed data, as in the illustration above, the nuisance parameter that is estimated at the sample size review is the common variance. The usual pooled variance estimate is given by

$$s^2 = \frac{1}{n_{E1} + n_{C1} - 2} \left[ \sum_{j=1}^{n_{E1}} (x_{Ej} - \bar{x}_{E1})^2 + \sum_{j=1}^{n_{C1}} (x_{Cj} - \bar{x}_{C1})^2 \right] \quad (1)$$

where  $x_{E1}, \dots, x_{En_{E1}}$ , with mean  $\bar{x}_{E1}$ , are the  $n_{E1}$  observations on the experimental arm and  $x_{C1}, \dots, x_{Cn_{C1}}$ , with mean  $\bar{x}_{C1}$ , are the  $n_{C1}$  observations on the placebo arm. The calculation of this estimate requires a knowledge of which patients are in which treatment group, so that the treatment allocation must be unblinded. As explained below, this is undesirable in clinical trial monitoring as it may lead to bias through the differential treatment and interpretation of further patients in the trial.

Expression (1) for  $s^2$  can be rearranged to give

$$s^2 = \frac{1}{n_{E1} + n_{C1} - 2} \left[ (n_{E1} + n_{C1} - 1)s_0^2 - \frac{n_{E1}n_{C1}}{n_{E1} + n_{C1}} (\bar{x}_{E1} - \bar{x}_{C1})^2 \right] \quad (2)$$

where  $s_0^2$  is the overall estimate of variance, which can be calculated without breaking the blind, given by

$$s_0^2 = \frac{1}{n_{E1} + n_{C1} - 1} \left[ \sum_{j=1}^{n_{E1}} (x_{Ej} - \bar{x})^2 + \sum_{j=1}^{n_{C1}} (x_{Cj} - \bar{x})^2 \right] \quad (3)$$

with  $\bar{x}$  the overall mean.

This suggests a method for estimation of the standard deviation from blinded data. The first term in (2), depending on  $s_0^2$ , can be found from the unblinded data. An estimate of the standard deviation can thus be obtained from blinded data together with an estimate of  $(\bar{x}_{E1} - \bar{x}_{C1})$ , the treatment difference. Gould [5] suggests

using the treatment difference under which the power is calculated to give an estimate of the standard deviation on which a new sample size can be calculated. If we denote this difference by  $\delta_R$ , replacing  $(\bar{x}_{E1} - \bar{x}_{C1})$  in (2) by  $\delta_R$  gives an estimate of the variance of

$$\frac{1}{n_{E1} + n_{C1} - 2} \left[ (n_{E1} + n_{C1} - 1) s_0^2 - \frac{n_{E1} n_{C1}}{n_{E1} + n_{C1}} \delta_R^2 \right]. \quad (4)$$

If an estimate of the unknown variance is based on blinded data, the sample size reestimation procedure does not have any impact on the overall type I error rate, in contrast to the more general sequential monitoring that is discussed in the next section. The type I error rate may be inflated, however, if unblinded data are used to provide the variance estimate used for the calculation of the sample size for the second stage [6]. This gives further reason for the use of methods that do not require knowledge of treatment allocations.

### 20.3 DANGERS OF UNPLANNED CLINICAL TRIAL MONITORING

Although, for the reasons just described, monitoring of accumulating data in a clinical trial is very appealing, there are a number of dangers, particularly when considering treatment efficacy, to which careful attention should be given whenever data monitoring is considered.

The main danger is a statistical one: The conduct of multiple analyses can lead to difficulty in the interpretation of the results [7]. In particular, if significance tests comparing experimental and control treatments are performed more than once, with testing stopped as soon as a significant result is obtained, the type I error rate is inflated as explained below.

When a single significance test is conducted, it is performed in such a way that the probability of declaring the treatments to be different if in fact they are identical, that is, of making a type I error, is controlled, usually to be 5%. If the test is repeated at a number of analyses, with each test conducted at the 5% level and the procedure stopped as soon as a significant result is obtained, there is risk of a type I error at each of these analyses. The overall probability of such an error, the type I error rate, will thus be more than 5%. The effect of repeated looks conducted at the 5% level in the simplified setting of comparison of normal data with known variance was assessed by Armitage et al. [8]. They showed that in this setting while a single analysis has a type I error rate of 5% as planned, if two analyses are conducted, the overall type I error rate is increased to 8%, with this increasing to 14% if five analyses are conducted, and increasing further if the number of analyses is more than five.

The problem of type I error rate inflation means that monitoring of clinical trials cannot be conducted without the use of specialist statistical methodology to ensure the type I error rate is controlled at the nominal level. The remainder of this chapter gives descriptions and examples of such statistical methods, explaining how valid tests may be obtained. A clinical trial conducted using such methodology is called a *sequential clinical trial*, and the formal analysis of data part way through the trial is known as an *interim analysis*.



A further statistical problem, related to that of type I error inflation, is that of estimation at the end of a monitored clinical trial. Just as the type I error rate is inflated because the trial stops exactly when extreme results are obtained, so a conventional estimate obtained at the end of a sequential clinical trial will be biased because the trial stops exactly when positive results are seen at an interim analysis. Again, special statistical techniques are required to provide a valid analysis, as described below.

In addition to the statistical problems associated with the conduct of interim analyses, there are some practical problems that may make monitoring of efficacy data difficult or undesirable. In some trials the primary endpoint may be some measure, such as long-term survival, that is observed only after considerable follow-up. In this case, all patients may be treated before the primary endpoint data from even the first patients enrolled in the trial are observed. This means there is no possibility of stopping the trial as a result of interim analyses. Although there may be some advantage in monitoring the efficacy data even in this case, for example, to obtain the results of the trial more quickly, there is no potential saving in terms of trial length or sample size.

If the data observed early in a trial are not considered typical of all of the data that are likely to be observed, the results of an interim analysis may be misleading. Monitoring of efficacy data in this case is thus undesirable and should be avoided. An example might be a comparison of treatments for cancer when the primary efficacy endpoint is survival, and one treatment is expected to lead to excess early mortality, but to better long-term prognosis. If only early death data were available at an interim analysis, the long-term advantages of this treatment could be missed, and the trial halted erroneously.

The final danger associated with many data-monitoring strategies is the need to unblind the treatment allocation. If treated patients continue to be followed-up, there is a danger that knowing which treatment they are receiving may lead to a difference in the level of care for or interpretation of results from patients in the experimental treatment and control groups. This may give rise to bias in the treatment comparison. The release of interim results may also lead to reluctance among investigators to recruit patients, or reluctance among the patients themselves to consent to randomization, if there is some evidence of a treatment difference, so that equipoise is no longer considered to exist.

## **20.4 METHODOLOGY FOR SEQUENTIAL MONITORING OF UNBLINDED EFFICACY DATA**

As described above, some clinical trial monitoring, such as monitoring of adverse events for safety, or the estimation of nuisance parameters from blinded data for sample-size review may be conducted with no risk of type I error rate inflation or effect on the final analysis. The monitoring of unblinded efficacy data, with the chance to stop the trial early if a large difference between the treatments is observed, does, however, lead to an increase in the type I error and bias in the analysis unless specific statistical methods are employed. The remainder of this chapter describes such methods and illustrates how valid monitoring of efficacy data can be conducted. Most of the statistical methodology in this area has been developed for clinical trials

in which two treatment groups, often corresponding to an experimental treatment and a control treatment, are compared. It is on this type of trial that the rest of the chapter will be specifically focused.

Whitehead [9] gives a list of the four key elements that must be considered in the design and conduct of a sequential clinical trial to assess treatment difference. These are a parameter giving some measure of the advantage of the experimental treatment over the control treatment; test statistics that give information on the size of this advantage on the basis of the data observed at an interim analysis and the amount of information about the treatment difference that is given by the observed data; a stopping rule that determines, on the basis of the observed test statistic, values at an interim analysis whether the trial should continue or should stop; and a final analysis method, valid for the stopping rule used, that enables a conclusion of whether or not the experimental treatment is superior to the control and provides a  $p$  value and point estimate and confidence interval for the treatment difference at the end of the trial.

The four elements are now considered in detail. Specification of the first two, the treatment difference parameter and the test statistics, is required whether fixed sample size or sequential methods are employed, but is discussed here for completeness. The specification of the third, the stopping boundary, represents a solution to the problem of control of the type I error rate. A number of solutions to this problem have been proposed, leading to several different approaches to the way in which the stopping boundary is calculated. These are described below. Although the fourth element, the analysis, is very important in the interpretation of the results from the sequential trial, statistical methodology in this area has generally lagged behind that for the construction of stopping rules. In practice, the use of an appropriate final analysis has often been neglected, in part due to the lack of availability of suitable software. Methods for the valid analysis have been developed, however, and the analysis can now be conducted using commercially available software. The methods are described briefly below and the alternative computer programs are discussed in the next section.

#### 20.4.1 Parameterization of Treatment Difference and Test Statistics

As in a fixed sample-size trial, the first stage in the design of a sequential clinical trial is the specification of some primary measure of treatment efficacy. This measure should be chosen on the basis of clinical relevance, ease and accuracy of measurement, and familiarity to clinicians. Following the definition of the primary endpoint, an associated parameter measuring the difference between the experimental and control treatments must be chosen. This will depend on the type of data collected on the primary endpoint, the interpretability of the parameter, for example, whether a difference or a ratio is more familiar, and the precision of the resulting analysis.

For example, if the primary response is a dichotomous variable such as whether a cancer patient is in remission or not after one year, the chance of being in remission in each group can be measured by the odds of remission, and the difference between the treatment groups can be measured by the ratio of the odds in the two groups, or more commonly the logarithm of this ratio, the *log-odds ratio*. If the primary response is a continuous measure such as weight loss, a difference in true unknown mean weight loss for the two groups would be of interest. If the primary

response is the time to some event, such as death, the difference between treatment groups might be measured by the ratio of the hazards in the two groups or the logarithm of this ratio.

At each interim analysis, some test statistic measuring the difference between the experimental and control treatments based on the observed data is calculated together with a test statistic measuring the amount of information given by the observed data at that interim analysis. Whitehead [10] and Jennison and Turnbull [11] suggest using a measure of difference known as the *efficient score* statistic and a measure of information called the *observed Fisher's information*, though other pairs of statistics may also be used, such as an estimate of the treatment effect and its standard error as suggested by Jennison and Turnbull [12] or a  $p$  value for comparison of the treatments and the planned sample size, as is used in *repeated significance testing* described in the next paragraph and in the *adaptive design* approach described in more detail in the last section of this chapter.

#### 20.4.2 Repeated Significance Testing

As discussed above, if a simple significance test comparing the experimental and control treatment groups is conducted repeatedly at the 5% level until a significant result is obtained, the overall type I error rate will be increased above 5%. Armitage et al. [8] showed how the actual error rate could be calculated when the number of repeated tests is specified. Similar calculations to those performed by Armitage et al. can be used to enable repeated significance testing with overall type I error rate controlled at the 5% level. By conducting each individual test at a more stringent level, the overall type I error rate is reduced, and so may be reduced to be 5% as desired. Numerical computation may be used to calculate the required error rate for each of the repeated tests to ensure that the overall level is 5%.

Pocock [13] proposed such a repeated significance testing approach, giving tables of the nominal type I level at which each test should be conducted to give an overall type I error rate of 5 or 1% for different numbers of tests. For example, to give an overall type I error rate of 5%, if two tests are performed each should be carried out at the 2.94% level, whereas if five tests are conducted, the nominal level for each should be 1.58%.

The repeated significance tests described by Pocock [13] are assumed to all be conducted at the same nominal level so that the same critical value is used for a standardized test statistic at each of the interim analyses. In general, there is no requirement that this should be so. If, say, five interim analyses are planned, it is possible to use different nominal levels for the five significance tests and control the overall type I error rate to be 5%. An alternative pattern of significance levels was proposed by O'Brien and Fleming [14]. They constructed the repeated significance tests in such a way that the level of evidence required to reject the null hypothesis that there is no difference between the two treatments at the earlier analyses is greater than that required to reject it at later analyses. In particular, the critical value for the  $i$ th interim analysis is chosen to be  $k_1/\sqrt{i}$  for some constant  $k_1$  chosen to maintain the overall type I error rate at the desired level. This choice of nominal levels, while maintaining the overall type I error rate, makes stopping at early interim analyses almost impossible, negating some of the advantages of sequential analysis. The Pocock [13] design, on the contrary, has been criticized, for example,

by Pocock and White [15], for leading to too high a chance of early stopping. Alternative approaches have been proposed [16], although the repeated significance testing approach has been largely superseded by the more flexible spending function approach described in the next section.

### 20.4.3 Spending Function Approach

A disadvantage of the repeated significance testing approach is that the number of interim analyses and the timing of these in terms of the amount of information available at each analysis need to be specified in advance in order to calculate the critical values required to control the overall error rate. Lan and DeMets [17] introduced a more flexible approach in which neither the timing nor the number of interim analyses need be given at the design stage.

Rather than specifying the critical values at a number of interim analyses, Lan and DeMets [17] proposed that the stopping rule should be given by a *spending function* that describes how the overall type I error rate, say 5%, is spent through the course of the trial. At the design stage, the maximum amount of information that will be attained if the trial does not stop early is calculated. At each interim analysis the level of information currently available is evaluated and the *information time* calculated to be the proportion of the maximum information available from the current data. The spending function gives, as a function of this information time, the amount of type I error that is to be spent, and the hypothesis test is conducted at such a level to ensure that the overall type I error spent up to and including this interim analysis is equal to this amount.

The spending function is thus an increasing function of the information time. It takes value 0 for information time 0, that is, at the start of the trial, and value equal to the overall type I error rate for information time 1, that is, at the end of the trial. It is common to denote the information time by  $t$ , and the spending function by  $\alpha^*(t)$ . Suppose that interim analyses are taken at information times  $t_1, t_2, \dots$ . The critical values used for the hypothesis tests at each of the interim analyses are chosen so that the probability of stopping and rejecting the null hypothesis at any of the interim analyses up to and including the  $i$ th is equal to  $\alpha^*(t_i)$  if the null hypothesis is really true. Equivalently, if the null hypothesis is true, the probability of stopping and rejecting the null hypothesis at the  $i$ th interim analysis given that the trial has not stopped at an earlier interim analysis is equal to  $\alpha^*(t_i) - \alpha^*(t_{i-1})$ . The critical values to give these stopping probabilities can be calculated using a recursive numerical integration technique similar to that used by Armitage et al. [8] to evaluate error rates for repeated significance tests. A similar technique enables calculation of the probability of stopping under a specified alternative hypothesis, and this allows the maximum information level to be chosen to give a test that satisfies some power requirement. Further details of the spending function approach and the computations required to calculate the critical values and sample size are given by Jennison and Turnbull [11].

As in the repeated significance testing approach, there is considerable flexibility regarding how likely the trial is to stop at the different interim analyses within the constraint that the overall type I error rate is controlled at the desired level. In the spending function approach this flexibility is expressed in the choice of the spending function, which may be any increasing function rising from 0 to the overall type I

error rate. A number of forms for spending functions have been proposed. Lan and DeMets [17] give spending functions to give approximations to the repeated significance testing procedures of Pocock [13] and O'Brien and Fleming [14], while Kim and DeMets [18] and Hwang et al. [19] suggest possible families of spending functions.

The flexibility of the sequential stopping boundaries in terms of the form of the spending function was utilized further by Kim and DeMets [18]. They showed how two different spending functions may be used to construct tests that have different properties for positive and negative treatment differences.

The sequential approach can be extended further to give tests that have different power to detect positive and negative treatment differences. A lower power is associated with a smaller sample size. Since the sample size is no longer fixed, but is now a random variable depending on the observed data, it is possible to stop the trial earlier for certain observed treatment differences than others. In particular, the stopping rule may not be symmetric, so that different decisions may be taken for positive and negative observed differences of the same magnitude. This is valuable, for example, in a clinical trial comparing an experimental treatment with a placebo. If the experimental treatment is superior to the placebo, we desire a high probability of rejecting the null hypothesis that there is no difference, and we are willing to randomize the large number of patients required to demonstrate this. If the experimental treatment is actually inferior to the placebo, demonstrating this is of less interest. It is sufficient to show that the experimental treatment is either the same or worse than the placebo and to continue to randomize patients to discover whether the disadvantage associated with the experimental treatment is significant or not is undesirable. Whitehead [10] calls asymmetric tests of this type tests with *power requirement I*, in contrast to the more usual symmetric tests, which he calls tests with *power requirement II*. Sequential procedures with power requirement type I can also be constructed using the spending function approach [20, 21].

#### 20.4.4 Boundaries Approach

The repeated significance testing approach and the spending function approach provide a stopping rule. In each case, for each interim analysis conducted, a pair of critical values is obtained, and the trial is stopped if the test statistic lies outside of the interval defined by these two values. The critical values are said to form a *stopping boundary* and the set of intervals to form a *continuation region*. The observed values of the test statistic measuring the treatment difference at the interim analyses are said to form a *sample path*, which is compared with the stopping boundary to decide whether or not the trial should continue.

An alternative to the methods for construction of the stopping boundary described above is based on the abstract concept of continuous monitoring, in which it is imagined that the value of the test is observed at all times rather than just at the discrete times given by the interim analyses. In this case, a plot of the test statistic measuring the treatment difference against that measuring the amount of information available forms a continuous line. In the *boundaries approach*, this continuous line is compared with continuous stopping boundaries that are expressed as functions of the information level.

The boundaries approach stems from the work of Wald [22] who developed the sequential probability ratio test (SPRT), initially for the testing of armaments during World War II. In the SPRT the likelihood ratio for a simple alternative hypothesis relative to the null hypothesis is calculated at each interim analysis. The test continues so long as this likelihood ratio falls within some fixed range. Wald derived stopping limits so as to give a test with specified type I error rate and power under the assumption of continuous monitoring. The SPRT has the important optimality property that among all tests with the same type I error rate and power, the SPRT minimizes the expected sample size when either the null or alternative hypothesis holds. Although the expected sample size is minimized by the SPRT, however, the SPRT has no maximum sample size, so that it cannot be guaranteed that the sample size will be below any fixed value. This feature makes the SPRT unsuitable for many clinical trials.

Following the work of Wald [22], a number of alternative forms for boundaries that maintain the overall type I error rate have been proposed. The choice of the form of boundary in this approach is analogous to the choice of spending function in the approach described above, and, as above, trials may be designed with asymmetric (power requirement I) or symmetric (power requirement II) boundaries. Whitehead [10] describes a wide range of such tests, including approximations to the SPRT that do place a limit on the maximum sample size required. One family of boundary shapes that is particularly common in sequential clinical trials is based around the triangular test. If the efficient score statistic is plotted against the observed Fisher information, this test has straight boundaries that form a triangular-shaped continuation region. An example of a double triangular test is illustrated in the example in Section 20.6. The triangular test approximately minimizes the maximum expected sample size among all tests with the same error rates. Unlike the SPRT the triangular test has a maximum sample size. It can be shown that for a triangular test there is a high probability of stopping with a sample size below that of the equivalent fixed sample size test, so that on average the sample size will be reduced by the use of this sequential design.

The stopping boundaries obtained using the boundaries approach maintain the overall type I error rate under the assumption of continuous monitoring of the test statistic measuring the treatment difference. In practice, monitoring is necessarily discrete, even if an interim analysis is conducted after observation of every patient. This means that if the critical values from the boundaries approach are used, the type I error rate will be less than the planned level. Whitehead [10] has proposed a correction, known as the *Christmas tree correction*, to modify the continuous boundaries to allow for the discretely monitored sample path. This correction allows interim analyses to be made at any time provided that the timing is not dependent on the observed treatment difference.

#### 20.4.5 Analysis after a Sequential Trial

At the end of a sequential trial a final analysis of the data observed should be conducted. This analysis should provide a  $p$  value for testing the null hypothesis that there is no difference between the experimental treatment and the control, and a point estimate and confidence interval for the difference between the treatments,



usually as measured by the primary parameter of treatment difference discussed above.

The fact that a sequential stopping rule has been used means that many of the standard analysis methods that would be employed after a fixed sample sized trial are no longer valid. Facey and Lewis [23] reported in 1998 that a standard analysis is performed inappropriately in many sequential trials. This is hopefully now changing.

Suppose that a sequential trial stopped at some interim analysis with the upper critical value exceeded by the test statistic measuring observed treatment difference, leading to the conclusion that the experimental treatment was superior to the control treatment. The trial has stopped precisely because of the large observed value of the treatment difference. This means that standard unbiased estimates of treatment difference will, on average, overestimate the true value of the treatment difference. In a similar way, the  $p$  value from a standard analysis will, on average, be too small, that is, it will tend to overstate the evidence against the null hypothesis. Special methods of analysis are therefore required to allow for the interim analyses that have been conducted in a sequential trial.

The meaning and interpretation of the  $p$  values, point estimates, and confidence intervals obtained after a sequential clinical trial are the same as for a fixed sample size trial, but a method must be employed in their calculation that is valid for the sequential stopping rule used. A number of methods of analysis have been proposed. Most are based on an ordering of all possible data sets to reflect which are considered to provide strongest evidence against the null hypothesis being tested. Further details are given by Whitehead [10] and Jennison and Turnbull [11]. The implementation of these methods in commercially available software, as described in more detail in Section 20.5, enables the easy conduct of a valid analysis following a sequential trial.

#### 20.4.6 Monitoring and Analysis of More Than One Endpoint

In a clinical trial it may be the case that other endpoints in addition to the primary endpoint are important in the clinical assessment of a treatment. If these endpoints are only to be analyzed at the end of the study after complete data on all patients are available, then they are usually termed *secondary endpoints*. No formal monitoring of these is conducted during the trial. If, however, one or more of the endpoints can really be regarded as additional primary endpoints that are important in determining whether a study should stop or continue, then further statistical methodology is required. There are two general approaches to monitoring these *co-primary endpoints* depending on the nature of the endpoints involved.

If several endpoints are measuring a similar aspect of treatment performance, then it may be acceptable to demonstrate efficacy in a collective way across the endpoints. Combining the various endpoints together into a single global measure and then using methods already described in this chapter would then be a suitable approach. As an example, consider the therapeutic area of stroke. There is no single endpoint that is universally accepted. Outcomes of interest include the Barthel index, the modified Rankin score, the Glasgow outcome scale, and the National Institutes of Health (NIH) stroke scale. Investigators do not believe that a positive result in terms of a single one of those outcomes is sufficient. Tilley et al. [24]



describe using a global test statistic in precisely this indication and discuss the advantages and disadvantages of such an approach.

In some indications, with some endpoints, it is not clinically meaningful to combine the endpoints into a single measure. Examples include the need to demonstrate efficacy for two outcomes to gain licensing, such as in Alzheimer's disease where it is generally necessary to consider both mental and physical aspects of a patient's progress. The simultaneous monitoring of efficacy and safety responses would be another example. In these instances, the endpoints should be kept separate and the correlation between them accounted for in the design and analysis of any sequential trial. Due to the complexity of computations, most authors have concentrated on developing *bivariate* methodology, that is, for the case of two co-primary endpoints, see, for example, Jennison and Turnbull [25], Cook and Farewell [26], and Todd [27].

## 20.5 IMPLEMENTATION OF SEQUENTIAL DESIGNS

As illustrated in Section 20.4, there are a number of different approaches to the specification of a sequential design and in particular the form of a stopping rule. When considering the implementation of such methodology, the most important issues to focus upon are the reasons why it is desirable to either stop or continue the trial at an interim analysis. Reasons for stopping were outlined above and may include evidence that the experimental treatment is obviously better than the control, evidence that the experimental treatment is already obviously worse, or indication that there is little chance of showing that the experimental treatment is better than the control so that it is considered futile to continue. Reasons for continuing may include belief that a moderate advantage of the experimental treatment is likely and it is desired to estimate this magnitude carefully, or evidence that the event rate is low and more patients are needed to achieve required power. Criteria such as these will first determine the symmetry requirement that is appropriate for the study under consideration, that is, whether an asymmetric (power requirement I) or a symmetric (power requirement II) design should be considered. Once this has been established, attention turns to the specific properties of different tests within a broad category. A choice between the tests is then made based upon consideration of likely sample sizes under various possible treatment effect scenarios. The wide range of stopping rules now available allow sequential designs to be devised for testing *superiority*, *noninferiority*, *equivalence*, and even formal *safety* aspects of clinical trials. As discussed above, some more complex designs even aim to deal with both efficacy *and* safety aspects in a combined stopping rule.

The availability of appropriate designs is just one element of enabling the implementation of sequential methodology. The key requirement for widespread take-up of the techniques is access to software. There are four major commercial software packages currently available. The package PEST [28] is based on the boundaries approach. The package EaSt [29] implements the spending function approach. An addition to the package S-Plus is the S+SeqTrial module [30], which also implements the spending function approach, in this case for a family of designs based on the methodology described by Kittelson and Emerson [31]. The package AddPlan (see [www.addplan.com](http://www.addplan.com)) is based on the adaptive design approach mentioned in Section 20.7. If a study is to be a comparison of two treatments in respect of a single primary endpoint, with the objective of discovering whether one treatment is superior, non-

inferior, or equivalent to the other, then it will be possible to implement a suitable sequential method using one of the available packages.

When planning to include interim analyses in any clinical trial, the implications of introducing a stopping rule need to be thought through carefully in advance of the study. All parties involved in the production of the protocol should be consulted on the choice of clinically relevant difference, specification of an appropriate power requirement, and the selection of a suitable stopping rule. The operation of any sequential procedure should be clearly described in the statistical section of the trial protocol. Once the trial is underway, it is extremely important that the interim results of an ongoing trial are not circulated widely. Such knowledge may have an undesirable effect on the future progress of the study. Whether a treatment looks good or bad as the trial progresses will obviously affect an investigator's attitude toward the trial. For this reason, it is often the case that a DSMB may be established to consider the progress of a sequential trial as well as assess safety data as discussed in Section 20.2.

If a DSMB is appointed, one of its first duties should be to review the protocol and to scrutinize any proposed sequential stopping rule prior to the start of the trial. The procedure for undertaking the interim analyses should also be finalized in advance. The DSMB would then review results of the interim analyses as they are reported. It is usual for the board to be supplied with full, unblinded information. Ideally, the only other individual to have knowledge of the treatment comparison would be the statistician who performs the actual analyses of the accumulating data. After each interim examination of the data, the DSMB should issue a short report detailing the outcome of its deliberations. The most common outcome of a meeting is likely to be that the study should continue without modification. However, the DSMB may recommend continuation with some protocol modifications, stopping the study due to safety concerns, or stopping as a result of the formal stopping rule for efficacy having been met. An example of the form of report issued by a DSMB following a meeting is given in Figure 4.

Decision making as part of a sequential trial (whether by a DSMB or another party involved in the trial) is both important and time sensitive. A decision taken to stop a study affects both the current trial and impacts on future trials in the same therapeutic area. However, continuing a trial too long puts participants at unnecessary risk and delays dissemination of important results. The issue applies not only to monitoring efficacy data but also to reacting in a timely manner to safety problems. Wondering whether data are accurate and up-to-date makes the decision process harder. It is therefore necessary to have both timely and accurate data. Unfortunately, a trade-off exists. It takes time to ensure accuracy. One option is to establish a "fast-track" system, processing efficacy data and key safety data separately from other information. Fewer data means that they can be validated more quickly. This is a particularly important issue that requires careful thought and planning.

## **20.6 EXAMPLE: LITHIUM GAMOLENATE IN PATIENTS WITH ADVANCED PANCREATIC ADENOCARCINOMA**

The prognosis for pancreatic cancer patients is poor, with an average survival time of a matter of months from diagnosis. Chemotherapy for pancreatic cancer sufferers

Dear

The independent review committee met today to hear in confidence the results of the third interim analysis presented by the independent statistics committee, and then held discussions with representatives of the company.

We noted the incidence of severe side effects but acknowledge that this particular form of cancer is classically associated with thrombotic complications!

We were reassured to hear that centers were now treating patients more expeditiously and with greater familiarity and expertise.

Our decision as to whether to continue the trial was *not* an easy one, but we unanimously agree that the trial should continue under close monitoring and we await the next analysis (70 deaths) with interest.

#### **Other Comments**

- Would it be possible to look statistically at the first few versus the later patients treated at main centers
- For the future – the patient information sheet should tell patients that therapy may be associated with hemoglobinurea and thrombotic problems
- Regarding the proposed protocol amendment, we feel that to issue the results of the company analysis prior to the interim independent analysis was improper, and reject it. In any case such major amendments would, we believe, constitute a new trial.
- We congratulate the company on the preparation and distribution of the weekly data tables.

The independent review committee plans to meet again on June 15.

Yours sincerely

Chairman of the DSMB

**FIGURE 4** *Example letter sent from the DSMB to the Steering Committee.*

offers only very small survival benefits and is associated with considerable side effects. It was essentially felt that there was no standard treatment available. It has been found that unsaturated fatty acids (lithium gamolenate) have an antitumour effect in experimental studies, and in phase II trials few side effects of this treatment were observed. This prompted the conduct of a randomized dose-finding phase III study [32]. It was required to reach a reliable conclusion concerning the optimum dose when considering survival time from randomization.

Patients with a diagnosis of inoperable pancreatic adenocarcinoma were treated with one of three doses of the active treatment, either oral dose, or low or high intravenous doses. When the objectives of the trial were considered, it was decided that a sequential approach was appropriate. The methodology followed was the boundaries approach, as discussed in Section 20.4. The comparison employed was

between the high intravenous (IV) dose and the oral dose. The plan was to recruit patients into the study until a sufficiently large difference in the survival rate was seen between these two doses. Analyses involving the low dose IV treatment were considered secondary to this main analysis. The study was designed to have 90% power to detect a difference in median survival of 4 months between the oral dose and the high dose treatments as significant at the 5% level. The reference median survival for oral treatment was taken to be 120 days.

In evaluating the appropriate dose of the treatment, it was desirable to stop the trial as soon as there was sufficient evidence to indicate that either dose was significantly superior to the other. If there was no difference between the doses, then it was also important to distinguish this quickly. In the latter case, further development of oral dose could be justified as this is a less invasive form of the treatment than the intravenous administration. A sequential procedure that satisfies these requirements and has smaller expected sample sizes than the equivalent fixed sample test for any size of treatment difference is the double triangular test [10].

The data concerning survival times following entry into the trial were modeled as survival data, with the difference between the two treatment groups measured by the log hazard ratio. The statistic signifying the difference between the two doses seen so far was taken to be the efficient score for the log hazard ratio. This is the usual log-rank test for survival data, which will be denoted by  $Z$ . The statistic measuring the amount of information upon which the comparison was based will be denoted by  $V$ . This is the variance of  $Z$  and is approximately one quarter the number of deaths. An independent data and safety monitoring board was established to monitor the progress of the trial at the interim analyses. Figure 5 shows the outcome of five interim analyses. On the diagram the inner dotted boundaries show the Christmas tree correction for discrete looks. These form the stopping boundary, so that the trial should be stopped when one of these values is reached.

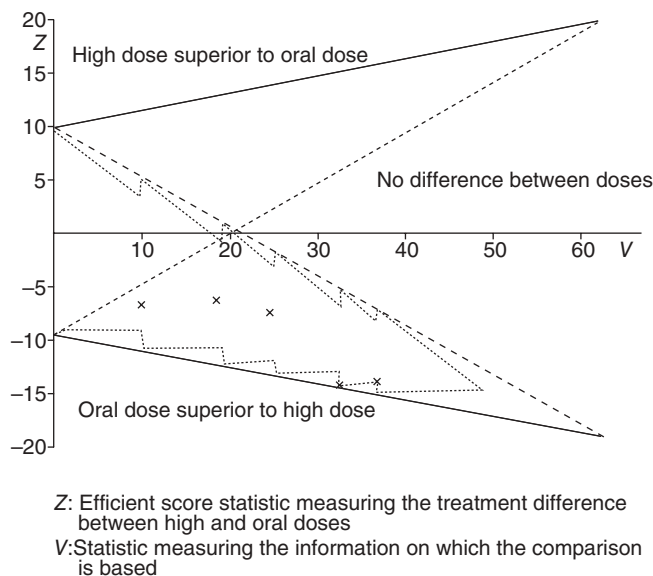


FIGURE 5 Sequential plot for the trial of lithium gamolenate.

A total of 278 patients were recruited to the trial. Of these 93 patients received oral dose, 90 patients received low dose, and 95 patients received high dose. The log-rank statistic  $Z$  crossed the boundary, which indicated that the oral dose was significantly superior to high dose, which was actually contrary to expectations. This occurred at the fourth interim analysis, and recruitment to the trial was stopped at this time. Continuing follow-up of patients already recruited when the trial stopped led to one further *overrunning* analysis, the results of which are also illustrated in Figure 5. A final analysis of the data was conducted using the computer package PEST [28]. The  $p$  value obtained was 0.03, a significant result in favor of the oral dose. Median survival after oral treatment was 129 days, while after high dose treatment it was 94 days.

## 20.7 RESPONSE-DRIVEN AND ADAPTIVE DESIGNS

In the methodology described in detail in Sections 20.4 and 20.5 and illustrated by the example, a single decision is made at each interim analysis: whether the trial should be stopped or should continue to the next interim analysis. Much more general sequential methods can be envisaged in which many aspects of a clinical trial design are reconsidered on the basis of the results from the interim analyses. Decisions could be made after each interim analysis regarding, for example, how many patients should be recruited to the next stage of the trial, whether treatment arms should be dropped from a multiarmed clinical trial, which doses should be used for the treatment of the next group of patients in a dose-finding study, or how the treatment allocation ratio might be changed to favor an apparently superior treatment, in addition to the decision of whether or not to stop the trial. Such general methods are often referred to as *response-driven*, *response-adaptive*, or *data-dependent* designs. They are also sometimes called *sequential designs* or *adaptive designs*, although increasingly these terms are reserved for the methods described in Section 20.4 and methods based on the work of Bauer and Köhne [33], which are described in the next paragraph, respectively.

A statistical approach for clinical trials that compare two treatments that allows more flexibility than the sequential methods described in Section 20.4 is the *adaptive design* approach [33]. In this approach, a  $p$  value is obtained at each interim analysis to test the null hypothesis that there is no difference between the two treatments on the basis of the new data obtained since the previous analysis. The  $p$  values from the different interim analyses are then combined by taking their product, and the value of this product is used to decide whether or not to stop the trial and whether the experimental treatment is superior to the control treatment at the end of the trial. Provided the data from the different stages of the trial are independent, any adaptation of the trial can be conducted without increasing the type I error rate. This method thus provides considerable flexibility in terms of design modification while still enabling valid inference to be drawn. It has been shown, however, that if the flexibility is not utilized, there is a slight reduction in power relative to the sequential methods described in Section 20.4 [34]. Further details of the adaptive design approach are given, together with a review of recent advances, by Posch et al. [35].

Although there has been much interest in the methodology for more general response-driven designs, there has been relative little use of such approaches in

practice. There are two main areas where such designs have been used. The first area is that of small two-arm studies that attempt to modify treatment allocation toward the treatment observed to be more effective. These methods extend the *play-the-winner* designs in which an interim analysis comparing the two treatments is made after observation of each patient, and the next patient is treated with the treatment currently observed to be most effective. A review of statistical methods for the design and analysis of such studies is given by Rosenberger [36].

The second area where response-driven designs have been used is in phase I dose-finding studies. In these studies the aim is to identify a dose that leads to an acceptably low risk of some adverse event in the trial subjects, who may be patients or healthy volunteers. Subjects are often treated in small groups, or *cohorts*, perhaps of three subjects, and the number of subjects to demonstrate the adverse event is recorded. The responses from each cohort are used to decide the dose level or levels to be used for the next cohort. Reviews of methodology for studies of this type are given by Cavalli and Sessa [37] and Whitehead et al. [38].

## REFERENCES

1. Food and Drug Administration (2001), Draft Guidance for Clinical Trial Sponsors on the Establishment and Operation of Clinical Trial Data Monitoring Committees; available at: <http://www.fda.gov/cber/gdlns/clindatmon.pdf>.
2. Ellenberg, S. S., Fleming, T. R., and DeMets, D. L. (2003), *Data Monitoring Committees in Clinical Trials: A Practical Perspective*, Wiley, Chichester.
3. Whitehead, J. (1999), On being the statistician on a data and safety monitoring board, *Stat. Med.*, 18, 3425–3434.
4. Gould, A. L., and Shi, W. J. (1992), Sample size re-estimation without unblinding for normally distributed data with unknown variance, *Commun. Stat. Theory Methods*, 21, 2833–2853.
5. Gould, A. L. (1995), Planning and revising the sample size for a trial, *Stat. Med.*, 14, 1039–1052.
6. Wittes, J., Schabenberger, O., Zucker, D., et al. (1999), Internal pilot studies I: Type I error rate of the naive t-test, *Stat. Med.*, 18, 3481–3491.
7. McPherson, K. (1974), Statistics: The problem of examining accumulating data more than once, *N. Engl. J. Med.*, 290, 501–502.
8. Armitage, P., McPherson, C. K., and Rowe, B. C. (1969), Repeated significance tests on accumulating data, *J. Roy. Stat. Soc. A*, 132, 235–244.
9. Whitehead, J. (1999), A unified theory for sequential clinical trials, *Stat. Med.*, 18, 2271–2286.
10. Whitehead, J. (1997), *The Design and Analysis of Sequential Clinical Trials*, Wiley, Chichester.
11. Jennison, C., and Turnbull, B. W. (2000), *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall/CRC, Boca Raton, FL.
12. Jennison, C., and Turnbull, B. W. (1997), Group sequential analysis incorporating covariate information, *J. Am. Stat. Assoc.*, 92, 1330–1341.
13. Pocock, S. J. (1977), Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 64, 191–199.

14. O'Brien, P. C., and Fleming, T. R. (1979), A multiple testing procedure for clinical trials, *Biometrics*, 35, 549–556.
15. Pocock, S. J., and White, I. (1999), Trials stopped early: Too good to be true? *Lancet*, 353, 943–944.
16. Pocock, S. J. (1982), Interim analyses for randomised clinical trials: The group sequential approach, *Biometrics*, 38, 153–162.
17. Lan, K. K. G., and DeMets, D. L. (1983), Discrete sequential boundaries for clinical trials, *Biometrika*, 70, 659–663.
18. Kim, K., and DeMets, D. L. (1987), Design and analysis of group sequential tests based on the type I error spending rate function, *Biometrika*, 74, 149–154.
19. Hwang, I. K., Shi, W. J., and DeCani, J. S. (1990), Group sequential designs using a family of type I error probability spending functions, *Stat. Med.*, 9, 1439–1445.
20. Stallard, N., and Facey, K. M. (1996), Comparison of the spending function method and the Christmas tree correction for group sequential trials, *J. Biopharma. Stat.*, 6, 361–373.
21. Pampallona, S., Tsiatis, A. A., and Kim, K. (2001), Interim monitoring of group sequential trials using spending functions for the Type I and Type II error probabilities, *Drug Info. J.*, 35, 1113–1121.
22. Wald, A. (1947), *Sequential Analysis*, Wiley, New York.
23. Facey, K. M., and Lewis, J. A. (1998), The management of interim analyses in drug development, *Stat. Med.*, 17, 1801–1809.
24. Tilley, B. C., Marler, J., Geller, N. L., et al., for the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Trial Study Group. (1996), Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA stroke trial, *Stroke*, 27, 2136–2142.
25. Jennison, C., and Turnbull, B. W. (1993), Group sequential tests for bivariate response: Interim analyses of clinical trials with both efficacy and safety endpoints, *Biometrics*, 49, 741–752.
26. Cook, R. J., and Farewell, V. T. (1994), Guidelines for monitoring efficacy and toxicity responses in clinical trials. *Biometrics*, 50, 1146–1152.
27. Todd, S. (1999), Sequential designs for monitoring two endpoints in a clinical trial. *Drug Info. J.*, 33, 417–426.
28. MPS Research Unit (2000), *PEST 4: Operating Manual*, University of Reading, Reading, UK.
29. Cytel Software Corporation (2000), *EaSt 2000: A Software Package for the Design and Interim Monitoring of Group-Sequential Clinical Trials*, Cytel Software Corporation, Cambridge, MA.
30. MathSoft Inc (2000), *S-Plus 2000*, MathSoft Inc., Seattle, WA.
31. Kittelson, J. M., and Emerson, S. S. (1999), A unifying family of group sequential test designs, *Biometrics*, 55, 874–882.
32. Johnson, C. D., Puntis, M., Davidson, N., et al. (2001), Randomized, dose-finding phase III study of lithium gamolenate in patients with advanced pancreatic adenocarcinoma, *Br. J. Surg.*, 88, 662–668.
33. Bauer, P., and Köhne, K. (1994), Evaluation of experiments with adaptive interim analyses, *Biometrics*, 50, 1029–1041.
34. Tsiatis, A. A., and Mehta, C. (2003), On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*, 90, 376–378.
35. Posch, M., Bauer, P., and Brannath, W. (2003), Issues in designing flexible trials, *Stat. Med.*, 22, 953–969.



36. Rosenberger, W. F. (1996), New directions in adaptive designs, *Stat. Sci.*, 11, 137–149.
37. Cavalli, F., and Sessa, C. (1999), Current issues in phase I trials: New study designs and informed consent procedures, *Ann. Oncol.*, 10(Suppl. 6), S147–S148.
38. Whitehead, J., Zhou, Y., Stallard, N., et al. (2001), Learning from previous responses in phase I dose-escalation studies, *Br. J. Clin. Pharmacol.*, 52, 1–7.



## *Inference Following Sequential Clinical Trials*

**Aiyi Liu and Kai F. Yu**

*Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver  
National Institute of Child Health and Human Development,  
National Institutes of Health, Rockville, Maryland*

### **Contents**

|                                                                         |      |
|-------------------------------------------------------------------------|------|
| 21.1 Introduction                                                       | 1043 |
| 21.2 Brownian Motion Paradigm in Sequential Trials                      | 1044 |
| 21.3 Estimating Magnitude of Treatment Difference                       | 1046 |
| 21.4 Calculation of <i>P</i> Values as Evidence Against Null Hypothesis | 1047 |
| 21.5 Construction of Confidence Intervals                               | 1049 |
| 21.6 Inference Concerning Secondary Endpoint                            | 1050 |
| References                                                              | 1051 |

### **21.1 INTRODUCTION**

Mainly in response to ethical and cost concerns, it has become a common practice that clinical trials, especially phase III studies comparing the efficacy of two or more treatment agents, are monitored using the so-called *group sequential testing procedures*. In a group sequential clinical trial a set of stopping boundaries are usually specified for a test statistic at a few time points, either in terms of real calendar time or statistical information. The stopping boundaries are chosen in a way that the type I error (false-positive) and type II error (false-negative) rates are controlled at the desired levels. In the past several decades, extensive research efforts have been devoted to the determination of these stopping boundaries, resulting in a number

of popular methods such as those developed by Pocock [1], O'Brien and Fleming [2], Lan and DeMets [3], Whitehead and Stratton [4], and Jennison [5], to name a few. Jennison and Turnbull [6] and Proschan, and co-workers [7] provide excellent overview of these group sequential methods.

Statistically speaking, the primary aim of a clinical trial is to test the null hypothesis that there is no significant difference between treatment agents under investigation, while maintaining a satisfactory statistical power if certain treatment difference indeed exists. A sequential clinical trial should be terminated as soon as the monitoring test statistic, derived based on the so-termed *primary endpoint*, reaches a stopping boundary, although the Data Safety and Monitoring Board (DSMB) may recommend the trial be terminated earlier before reaching a boundary or be continued after reaching a stopping boundary, based on various reasons, such as safety concerns. Once a boundary is reached and the trial stops, a decision can then be made on whether the null hypothesis should be rejected and one treatment agent is better than the other.

When reporting clinical trial findings, it is not sufficient to simply state *rejection* or *no rejection* of the null hypothesis. Point estimates and confidence intervals, which measure the magnitude of the treatment difference, and  $P$  values, which give the strength of the trial against the null hypothesis, should also be provided as an integrated part of the trial findings. However, from a frequentist's point of view, the usual maximum-likelihood approach, which provides (asymptotically) most efficient fixed sample-size inference, yields biased estimation due to the randomness of the observed sample size and its dependence with the test statistic and is thus no longer appropriate. In fact such randomness and dependence have been long well recognized as the cause of the type I error inflation when the usual fixed-size critical values are used repeatedly at each interim analysis. Unfortunately, when it comes to inferential analysis following a sequential trial, randomness and dependence are often ignored and the usual maximum-likelihood inference is conducted.

In what follows, we summarize several popular approaches developed in the literature to construct point and interval estimation and  $P$  values with respect to the primary endpoint following the termination of a sequential phase III trial. A brief overview will also be given to methods for inference concerning a secondary endpoint. Sequential phase II clinical trials in which interim monitoring is based on a Bernoulli variable (binary outcome) will not be discussed. However, most concepts and principles described here apply analogously to analysis following a sequential phase II trial.

## 21.2 BROWNIAN MOTION PARADIGM IN SEQUENTIAL TRIALS

A phase III clinical trial is usually monitored based on a Brownian motion  $X(t)$  with drift  $\delta$ , provided that the sample size is large enough so that the test statistics can be well approximated by a Brownian motion. A Brownian motion  $X(t)$  with drift  $\delta$  satisfies several nice properties:  $X(0) = 0$ , and  $X(t) \sim N(\delta t, t)$  for  $t > 0$ . Furthermore, for any  $0 < t_1 < t_2 < \dots < t_k < \infty$ ,  $X(t_1)$ ,  $X(t_2) - X(t_1)$ ,  $\dots$ ,  $X(t_k) - X(t_{k-1})$  are mutually independent, a proposition so well termed as *independent increments*, each normally distributed as  $X(t_j) - X(t_i) \sim N(\delta(t_j - t_i), t_j - t_i)$  and  $\text{cov}(X(t_j), X(t_i)) = t_i$ , for  $i < j$ . In practice,  $t$  is usually the Fisher information and  $X(t)$  is usually a test statistic, and

the drift parameter  $\delta$  is a measure of treatment difference with  $\delta = 0$  if there is no treatment difference.

To give the readers a flavor of how such Brownian motion may be constructed, let us consider the case of comparing, with equal allocation, two normal means with known common variance. Let  $X_{1A}, \dots, X_{nA}$  be the outcomes from  $n$  subjects allocated to treatment  $A$ , and  $X_{1B}, \dots, X_{nB}$  be that to treatment  $B$ . Assume that  $X_{iA} \sim N(\mu_A, \sigma^2)$  and  $X_{iB} \sim N(\mu_B, \sigma^2)$ . The null hypothesis is  $H_0: \mu_A - \mu_B = 0$ . The standardized test statistic is  $Z_n = \sum_{i=1}^n (X_{iA} - X_{iB}) / \sqrt{2n\sigma^2}$ , which follows standard normal  $N(0, 1)$ , and a two-sided test with level of significance  $\alpha$  rejects the null hypothesis if  $|Z_n| > \Phi^{-1}(1 - \alpha/2)$ , where  $\Phi$  is the standard normal distribution. To put this into a Brownian motion framework, we redefine (after a little bit pleasant algebraic manipulation) the test statistic as

$$X = \sqrt{\frac{n}{2\sigma^2}} Z_n$$

then  $X$  resembles a Brownian motion  $X(t)$  with drift  $\delta = \mu_A - \mu_B$  and in this example the (discrete) time  $t = n/(2\sigma^2)$ .

This Brownian motion paradigm applies to the two-arm comparison of many clinical endpoints such as proportions, ordinal data, and survival rates. Whitehead [8] presented a nice review of this structure and named it a *unified theory*. See also Lan and Zucker [9] and Proschan and co-workers [7].

With this framework, a group-sequential clinical trial can be formulated as interim monitoring of a Brownian motion  $X(t)$  with drift  $\delta$  at several discrete time points  $t_1, \dots, t_K$ , where  $K$  is the prespecified number of interim and looks to be carried out during the whole course of the trial. For each look  $k$ , let  $C_k = (a_k, b_k)$  be the continuation region and  $S_k = (-\infty, a_k] \cup [b_k, \infty)$  be the stopping region, where  $a_k < 0 < b_k$  define the lower and upper boundary of the sequential test. Then the trial stops at the  $k$ th look if the  $k$ th boundary is reached first, that is, if

$$X(t_i) \in C_i, i \leq k-1 \quad X(t_k) \in S_k$$

Upon stopping at the  $k$ th interim look, we observe two statistics, the stopping time  $T = t_k$  and the position of the Brownian motion  $X(T) = X(t_k)$ . These two statistics together are sufficient for the drift parameter  $\delta$  and hence inference on  $\delta$  can be based solely on  $(T, X(T))$ . The joint distribution  $f_\delta(k, x)$  of  $(T, X(T))$  at  $(T, X(T)) = (t_k, x)$  is  $f_\delta(k, x) = f_0(k, x) \exp\{x\delta - \frac{1}{2}t_k\delta^2\}$ , where the baseline functions  $f_0(k, x)$  can be computed numerically using the following recursive formulas:

$$f_0(1, x) = \frac{1}{\sqrt{t_1}} \phi\left(\frac{x}{\sqrt{t_1}}\right)$$

and

$$f_0(k, x) = \frac{1}{\sqrt{t_k - t_{k-1}}} \int_{C_{k-1}} \phi\left(\frac{x-s}{\sqrt{t_k - t_{k-1}}}\right) f_0(k-1, s) ds \quad k = 2, \dots, K$$

where  $\phi$  is the standard normal density function. Thus, for a statistic  $h(T, X(T))$ , its expectation is given by

$$E(h(T, X(T))) = \sum_{k=1}^K \exp\left\{-\frac{1}{2}t_k\delta^2\right\} \int_{s_k} h(k, x) f_0(k, x) \exp\{x\delta\} dx \quad (1)$$

### 21.3 ESTIMATING MAGNITUDE OF TREATMENT DIFFERENCE

The maximum-likelihood estimate of  $\delta$  is  $\hat{\delta}_{\text{ML}}(T, X(T)) = X(T)/T$ , the same form as that following a fixed-size test ( $K = 1$ ). However, its statistical properties change substantially due to the randomness of  $T$  and further complication of the dependence between  $T$  and  $X(T)$ . As a consequence, many optimal properties that the estimate enjoys under fixed-size sampling vanish under sequential sampling. For instance, the maximum-likelihood estimate is no longer unbiased. When  $K = 2$ , the bias function  $b(\delta)$  can be expressed in a closed form as

$$b(\delta) = \sqrt{t_1} \left( \frac{1}{t_1} - \frac{1}{t_2} \right) \left\{ \phi\left( \frac{b_1 - \delta t_1}{\sqrt{t_1}} \right) - \phi\left( \frac{a_1 - \delta t_1}{\sqrt{t_1}} \right) \right\}$$

See Emerson [10].

A naive approach to reduce the bias is to subtract from  $\hat{\delta}_{\text{ML}}$  the maximum-likelihood estimate of the bias  $b(\delta)$ , yielding an estimate as  $\hat{\delta} = \hat{\delta}_{\text{ML}} - b(\hat{\delta}_{\text{ML}})$ . Since  $\hat{\delta}_{\text{ML}}$  is biased for  $\delta$ ,  $b(\hat{\delta}_{\text{ML}})$  may not be a good estimate for  $b(\delta)$ . Therefore, the resulting estimate  $\hat{\delta}$  may not be a good estimate for  $\delta$ . Whitehead [11] instead suggested replacing  $\delta$  in the bias function by the so-called *bias-adjusted estimate*,  $\hat{\delta}_w$ , to be solved from the equation

$$\hat{\delta}_w + b(\hat{\delta}_w) = \hat{\delta}_{\text{ML}}$$

Whitehead [11] initially proposed his bias-adjusted estimate in the context of *fully sequential tests* with parallel or triangular stopping boundaries. He presented substantial numerical evidence for both boundaries that the bias-adjusted estimate indeed has smaller bias (and mean-squared error as well!) than the maximum-likelihood estimate.

However, the idea is quite general and applicable in reducing the bias of an estimate, so long as the bias function of the estimate can be worked out. The performance of such estimate depends on the shape of the bias functions. If the bias function of the estimate is nondecreasing and can be differentiated at least twice, then the bias-adjusted estimate has uniformly smaller mean-squared error than the initial estimate; see Whitehead [11] and Liu [12].

Emerson and Fleming [13] investigated Whitehead's bias-adjusted estimate in the context of group-sequential tests and compared with an unbiased estimate constructed in their study using conditioning argument:

$$\hat{\delta}_{\text{EF}} = E\left\{ \frac{X(t_1)}{t_1} \middle| (T, X(T)) \right\}$$

This is indeed an unbiased estimate since the expectation of  $X(t_1)/t_1$  is always  $\delta$ , regardless at which stage the test stops. Although  $\hat{\delta}_{\text{EF}}$  nullifies the bias, numerical results from Emerson and Fleming [13] showed that in general it has a larger mean-squared error than Whitehead's bias-adjusted estimate, a sacrifice to be made for enforcing unbiasedness. As expected, the two estimates both have smaller bias and mean-squared error than the maximum-likelihood estimate, as considerable numerical results have shown. However, analytic results are limited, perhaps due to the complicated forms of the density functions, on the statistical properties and comparison of the two estimates.

More can be said about how the sequential sampling [resulting in the randomness of  $T$  and dependence between  $T$  and  $X(T)$ ] changes the statistical property of a parameter estimate. Under fixed-size sampling, the maximum-likelihood estimate  $\hat{\delta}_{\text{ML}}$  is the unique unbiased estimate based on the sufficient statistic  $X(T)$  ( $T$  is a constant now) and has uniformly minimum variance among the unbiased estimates, a fact attributed to the well-known Rao–Blackwell theorem. Under group-sequential sampling, however, as shown in Liu and Hall [14], there are infinitely many sufficient-statistic-based unbiased estimates and none,  $\hat{\delta}_{\text{EF}}$  particularly included, has uniformly minimum variance. Only when restricted to unbiased estimates that are independent of future stopping rules does  $\hat{\delta}_{\text{EF}}$  have uniformly minimum variance.

## 21.4 CALCULATION OF P VALUES AS EVIDENCE AGAINST NULL HYPOTHESIS

The  $P$  values in a clinical trial measure the statistical strength (evidence) against the null hypothesis that there is no difference between treatment agents. It is defined as the probability, computed under the null hypothesis, of obtaining data that yields a statistic of interest being at least as extreme as the observed value of the statistic. Thus, smaller  $P$  values correspond to more extreme data from which the statistic is computed and reflect stronger evidence against the null hypothesis.

In a fixed-size clinical trial ( $K = 1$  and  $t_1$  is a constant), the calculation and interpretation of the  $P$  values is more straightforward and much less problematic. So long as the statistic of interest is a monotone function of  $X(t_1)$ , as is true for most cases, the upper- and lower-tail  $P$  values are given by, respectively,

$$P_{\delta=0}(X(t_1) > \tilde{x}) = 1 - \Phi\left(\frac{\tilde{x}}{\sqrt{t_1}}\right) \quad P_{\delta=0}(X(t_1) < \tilde{x}) = \Phi\left(\frac{\tilde{x}}{\sqrt{t_1}}\right)$$

where  $\tilde{x}$  is the realization (observed value) of  $X(t_1)$ . Doubling the smaller of the upper- and lower-tail  $P$  values yields the two-sided  $P$  value, which, because of symmetry, is given by

$$P \text{ value} = P_{\delta=0}(|X(t_1)| > |\tilde{x}|) = 2 \left\{ 1 - \Phi\left(\frac{\tilde{x}}{\sqrt{t_1}}\right) \right\}$$



Unfortunately, the definition of a  $P$  value in a group-sequential trial is not as clear as it would be in a fixed-size trial and sometime brings considerable ambiguity and confusion to many practitioners. What does it mean by data being *extreme*? What statistic should be used in defining *extreme*? The problem comes from the fact that, instead of summarizing the data by a one-dimensional sufficient statistic  $X(t_1)$  in a fixed-size trial, we now have to summarize the data by two dependent statistics,  $T$  and  $X(T)$ , both of which contain information on the treatment difference  $\delta$ . A sample path (or data) with smaller  $T$  (earlier stopping) and larger  $X(T)$  is without doubt more extreme than a path with larger  $T$  and smaller  $X(T)$ . But what about a path with  $T$  and  $X(T)$  both smaller and a path with  $T$  and  $X(T)$  both larger? In a word, how should the points  $(T, X(T))$  in the sample space in a group-sequential trial,

$$\Omega = \{X(t_1) \in S_1\} \cup \left[ \bigcup_{k=2}^K \{X_{t_i} \in C_i, i < k, X_{t_k} \in S_k\} \right]$$

be ordered so that *being extreme* is well defined? Is there an ordering that is better than the others, statistically and/or practically?

A number of popular orderings have been proposed in the literature, most based on a one-dimensional statistic  $s(T, X(T))$  defined on the sample space  $\Omega$ . A point  $(t_i, X(t_i))$  is more extreme than another point  $(t_j, X(t_j))$  if  $s(t_i, X(t_i)) > s(t_j, X(t_j))$ . With this ordering, suppose the trial stops at stage  $k$  with  $X(t_k) = \tilde{x}$ , then the upper-tail  $P$  value is given by

$$P_{\delta=0} \{s(T, X(T)) > s(t_k, \tilde{x})\}$$

which can be computed numerically using the recurse density functions. The lower-tail  $P$  value is given accordingly, and the two-sided  $P$  value is twice the smaller of the two  $P$  values.

Roser and Tsiatis [15] defined  $s(T, X(T))$  to be the score statistic (the position of the Brownian motion)  $X(T)$ , while Chang [16] suggested using the standardized test statistic  $X(T)/\sqrt{T}$ , and Emerson and Fleming [13] investigated ordering based on the maximum-likelihood estimate  $X(T)/T$ .

The so-called *stagewise ordering* (e.g., Siegmund [17], Fairbanks and Madsen [18], and Tsiatis and co-workers [19]) is somewhat different from the above three orderings and is not based on a one-dimensional statistic. A data point  $(t_i, X(t_i))$  is more extreme than the other  $(t_j, X(t_j))$  if it reaches the upper or lower boundary earlier [ $i < j$  and  $X(t_i) > b_i$  or  $< a_i$ ]. For data points with the same information time, the one with larger value of  $X(t)$  is more extreme.

All four orderings reduce to the same one in the fixed-size setting. However, there are no deterministic results on which ordering is better than the others. Some argue that the stagewise ordering is more preferable since it yields  $P$  values in agreement with the decision from the hypothesis testing; it is less than the significance level  $\alpha$  if and only if the null hypothesis is rejected. Moreover, the stagewise ordering does not depend on the stages beyond the stopping one, making it applicable in some more flexible designs in which the information time points at which interim analysis is conducted are not specified in advance. See Jennison and Turnbull [6].

## 21.5 CONSTRUCTION OF CONFIDENCE INTERVALS

A confidence interval of  $\delta$  with level  $(1 - \alpha)$  is an interval with lower bound  $\delta_L^* = \delta_L^*(T, X(T))$  and upper bound  $\delta_U^* = \delta_U^*(T, X(T))$  such that for any  $\delta$

$$P_\delta \{\delta_L^* < \delta < \delta_U^*\} \geq 1 - \alpha$$

One approach to construct such an interval is to use an ordering of the sample space described previously for calculating  $P$  values. Let the terminal data be  $(t_k, \tilde{x})$ . Then  $\delta_L^*$  is the drift value at which the probability is half  $\alpha$  of observing data being at least as extreme as  $(t_k, \tilde{x})$ , and  $\delta_U^*$  is the drift value at which the probability is  $1 - \alpha/2$  of observing data being at least as extreme as  $(t_k, \tilde{x})$ . That is,  $\lambda(\delta_L^*; t_k, \tilde{x}) = \alpha/2$  and  $\lambda(\delta_U^*; t_k, \tilde{x}) = 1 - \alpha/2$ , where

$$\lambda(\delta; t_k, \tilde{x}) = P_\delta \{(T, X(T)) \text{ is at least as extreme as } (t_k, \tilde{x})\}$$

In order for the two equations to yield an interval, the function  $\lambda(\delta; t_k, \tilde{x})$  needs to be increasing in  $\delta$  for any given data point  $(t_k, \tilde{x})$ . With such monotonicity each equation has a unique solution and  $\delta_L^* < \delta_U^*$  holds. If the  $\lambda$  function is nondecreasing, then  $\delta_L^*$  can be set to be  $\{\delta: \lambda(\delta; t_k, \tilde{x}) = \alpha/2\}$  and  $\delta_U^*$  to be  $\{\delta: \lambda(\delta; t_k, \tilde{x}) = 1 - \alpha/2\}$ . It can be shown analytically that Emerson and Fleming's [13] ordering and the stagewise ordering both yield increasing  $\lambda$  functions; see Emerson [10] or Hall and Liu [20] for the former and Jennison and Turnbull [6, Chapter 8] for the latter. There is no analytic results on this matter for the other two orderings though limited numerical results do not seem to support the monotonicity property.

Jennison and Turnbull [6, Chapter 8] discussed several criteria for defining a good confidence interval such as agreement with the hypothesis testing results and inclusion of the maximum-likelihood estimate. They come to a conclusion that the stage-wise ordering gives a more preferable confidence interval, which, in addition, is the only feasible method to use when future information time for interim analysis is not specified.

Woodroof [21] proposed a different approach to construct a confidence interval by modifying the fixed-size procedure to take the bias into consideration. Write  $Z(\delta) = (X(T) - T\delta)/\sqrt{T}$ . In fixed-size sampling,  $Z$  is normally distributed with zero mean and unit variance and serves as a pivot to construct the usual  $(1 - \alpha)$ -level confidence interval with limits  $X(T)/T \mp \Phi^{-1}(1 - \alpha/2)/\sqrt{T}$ . In a group-sequential sampling, this pivot no longer has a standard normal distribution, so a certain adjustment is needed. A new pivot is defined as  $Z(\delta) = (Z(\delta) - \mu(\delta))/s(\delta)$  where  $\mu(\delta)$  and  $s(\delta)$  are the mean and standard deviation function of  $Z(\delta)$ , respectively. If we assume that this new pivot approximately follows a standard normal distribution, then an approximate  $(1 - \alpha)$ -level confidence interval for  $\delta$  has confidence limits being

$$\left\{ \frac{X(T)}{T} - \frac{\hat{\mu}}{\sqrt{T}} \mp \frac{\hat{s}}{\sqrt{T}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\}$$

where  $\hat{\mu}$  and  $\hat{s}$  are proper estimates of  $\mu(\delta)$  and  $s(\delta)$ , respectively, which are often obtained with  $\delta$  replaced by its maximum-likelihood estimate.

Unlike the confidence intervals based on a sample space ordering, the above pivotal method does not always guarantee the exact coverage probability of  $1 - \alpha$ . However, it works pretty well in most cases, and the coverage probability of the resulting confidence intervals are quite close to the nominal level of  $1 - \alpha$ .

We have seen applications of the orderings of the sample space to constructing confidence intervals and calculating  $P$  values. Another usage of a proper ordering of the sample space is to compute a *median-unbiased* estimate  $\hat{\delta}_M = \hat{\delta}_M(T, X(T))$  of  $\delta$  such that

$$P_{\delta}\{\hat{\delta}_M(T, X(T)) < \delta\} = P_{\delta}\{\hat{\delta}_M(T, X(T)) > \delta\} = \frac{1}{2}$$

That is, a median-unbiased estimate has equal chances to over- or underestimate the parameter. (In contrast, the usual unbiased estimates such as  $\hat{\delta}_{EF}$  are referred to as *mean-unbiased* estimates, if such distinction needs to be made.) For an ordering of sample space with monotonicity property, the corresponding median-unbiased estimate at  $(t_k, \tilde{x})$  can be found by solving for  $\delta$  the equation:

$$\frac{1}{2} = \lambda(\delta; t_k, \tilde{x}) = P_{\delta}\{(T, X(T)) \text{ is at least as extreme as } (t_k, \tilde{x})\}$$

Emerson and Fleming [13] investigated the bias and mean-squared error of the median-unbiased estimates based on certain orderings. Compared with the bias-adjusted estimate  $\hat{\delta}_w$  and the unbiased estimate  $\hat{\delta}_{EF}$ , the median-unbiased estimates in general tend to have larger bias, often coupled with larger mean-squared error as well.

## 21.6 INFERENCE CONCERNING SECONDARY ENDPOINT

Termination of a sequential clinical trial is based on the *primary endpoint*, such as patients' survival in a cancer trial. In contrast, a *secondary endpoint*, such as patients' disease-free survival or treatment  $\times$  strata interaction, refers to an endpoint whose values are also observed from each patient along with the primary endpoint but do not contribute in any way to the termination of the trial. The secondary endpoints are usually analyzed only after the trial stops, though they may be evaluated at certain interim analysis, and such evaluation plays no role in sequential monitoring of the trial.

A secondary endpoint is usually correlated with the primary endpoint since they both are observed from the same patients. Due to the correlation, the usual likelihood analysis should also be adjusted for the random stopping of the trial so that a valid inference on secondary endpoints can be made. For most trials the primary and a secondary endpoint can be put into a random-sampling framework and be (asymptotically) modeled as two correlated Brownian motions with constant (often assumed to be known) correlation and proportional information time. The first Brownian motion, as given in the previous sections, governs the stopping of the trial, and the second one summarizes the observations of the secondary endpoint with

its drift parameter usually measuring the difference in the endpoint between treatment arms. For this drift parameter, Whitehead [22] constructed a bias-adjusted estimate to reduce the bias of the maximum-likelihood estimate, and Liu and Hall [23] investigated its unbiased estimates. Certain confidence intervals are developed by Whitehead and co-workers [24] using Woodroffe's pivotal method; these confidence intervals can also be used for testing hypotheses concerning the secondary endpoints.

Not all secondary endpoints can be modeled (asymptotically) with the primary endpoint as two Brownian motion with constant correlation coefficients. Yakir and Hall [25] and Hall and Yakir [26] gave examples that do not satisfy this model. One example is that a sequential trial with staggered entry is carried out to compare the survival rate of two treatment arms based on the log-rank statistic. Upon termination of the trial, one wants to know whether treatment  $\times$  gender interaction exists. In these examples, the primary and secondary endpoints are modeled as two Gaussian processes with correlation being a function of the information time. Hall and Yakir [26] derived distributional theory and certain optimal properties to construct point estimates and confidence intervals of the parameters associated with the secondary endpoint.

The inference procedures discussed so far are parametric in nature. Chuang and Lai [27, 28] developed nonparametric methods based on a resampling technique for confidence intervals concerning both the primary and secondary endpoints.

## REFERENCES

1. Pocock, S. J. (1977), Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 64, 191–199.
2. O'Brien, P. C., and Fleming, T. R. (1979), A multiple testing procedure for clinical trials, *Biometrics*, 35, 549–556.
3. Lan, K. K. G., and DeMets, D. L. (1983), Discrete sequential boundaries for clinical trials, *Biometrika*, 70, 659–663.
4. Whitehead, J., and Stratton, I. (1983), Group sequential clinical trials with triangular continuation regions, *Biometrics*, 39, 227–236.
5. Jennison, C. (1987), Efficient group sequential tests with unpredictable group sizes, *Biometrika*, 74, 155–165.
6. Jennison, C., and Turnbull, B. W. (2000), *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall/CRC, New York.
7. Proschan, M. A., Lan, K. K. G., and Wittes, J. T. (2006), *Statistical Monitoring of Clinical Trials: A Unified Approach*, Springer, New York.
8. Whitehead, J. (1999), A unified theory for sequential clinical trials, *Stat. Med.*, 18, 2271–2286.
9. Lan, K. K. G., and Zucker, D. (1993), Sequential monitoring of clinical trials: The role of information and Brownian motion, *Stat. Med.*, 12, 753–765.
10. Emerson, S. S. (1988), Parameter estimation following group sequential hypothesis testing [dissertation], University of Washington, Seattle.
11. Whitehead, J. (1986), On the bias of maximum likelihood estimation following a sequential test, *Biometrika*, 73, 573–558.

12. Liu, A. (2003), A simple low-bias estimate following a sequential test with linear boundaries, in Kolassa, J., and Oakes, D., Ed., *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, Institute of Mathematical Statistics Lecture Notes Monograph Series, Beachwood, OH, Vol. 43, pp. 47–58.
13. Emerson, S. S., and Fleming, T. R. (1990), Parameter estimation following sequential hypothesis testing, *Biometrika*, 77, 875–892.
14. Liu, A., and Hall, W. J. (1999), Unbiased estimation following a group sequential test, *Biometrika*, 86, 71–78.
15. Rosner, G. L., and Tsiatis, A. A. (1988), Exact confidence intervals following a group sequential trial: A comparison of methods, *Biometrika*, 75, 723–729.
16. Chang, M. N. (1989), Confidence intervals for a normal mean following a group sequential test, *Biometrics*, 45, 247–254.
17. Siegmund, D. (1978), Estimation following sequential tests, *Biometrika*, 65, 295–297.
18. Fairbanks, K., and Madsen, R. (1982),  $P$  values for tests using a repeated significance design, *Biometrika*, 69, 69–74.
19. Tsiatis, A. A., Rosner, G. L., and Metha, C. R. (1984), Exact confidence intervals following a group sequential test, *Biometrics*, 40, 797–803.
20. Hall, W. J., and Liu, A. (2002), Sequential tests and estimates after overrunning based on maximum-likelihood ordering, *Biometrika*, 89, 699–707.
21. Woodroffe, M. (1992), Estimation after sequential testing: A simple approach for truncated sequential probability ratio test, *Biometrika*, 79, 347–353.
22. Whitehead, J. (1986), Supplementary analysis at the conclusion of a sequential clinical trial, *Biometrics*, 42, 461–471.
23. Liu, A., and Hall, W. J. (2001), Unbiased estimation of secondary parameters following a sequential test, *Biometrika*, 88, 895–900.
24. Whitehead, J., Todd, S., and Hall, W. J. (2000), Confidence interval for secondary parameters following a sequential test, *J. Roy. Statist. Soc., B*, 62, 731–745.
25. Yakir, B., and Hall, W. J. (2003), Testing for a treatment-by-stratum interaction in a sequential clinical trial, in Kolassa, J., and Oakes, D., Ed., *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, Institute of Mathematical Statistics Lecture Notes Monograph Series, Beachwood, OH, Vol. 43, pp. 1–12.
26. Hall, W. J., and Yakir, B. (2003), Inference about a secondary process after a sequential trial, *Biometrika*, 90, 597–611.
27. Chuang, C. S., and Lai, T. L. (2000), Hybrid resampling methods for confidence intervals. With discussion and rejoinder by the authors, *Statistica Sinica*, 10, 1–50.
28. Chuang, C. S., and Lai, T. L. (1998), Resampling methods for confidence intervals in group sequential trials, *Biometrika*, 85, 317–332.

## *Statistical Methods for Analysis of Clinical Trials*

**Duolao Wang,<sup>1</sup> Ameet Bakhai,<sup>2</sup> and Nicola Maffulli<sup>3</sup>**

<sup>1</sup>*Medical Statistics Unit, London School of Hygiene and Tropical Medicine,  
London, United Kingdom*

<sup>2</sup>*Barnet General & Royal Free Hospitals, London, United Kingdom*

<sup>3</sup>*Department of Trauma and Orthopaedic Surgery, Keele University School of Medicine,  
Keele, Staffordshire, United Kingdom*

### **Contents**

|        |                                                                   |      |
|--------|-------------------------------------------------------------------|------|
| 22.1   | Introduction                                                      | 1054 |
| 22.1.1 | Bias and Systematic Errors                                        | 1054 |
| 22.1.2 | Confounding                                                       | 1055 |
| 22.1.3 | Random Error                                                      | 1055 |
| 22.2   | Types of Data, Summary, and Data Presentation                     | 1056 |
| 22.2.1 | Types of Data                                                     | 1056 |
| 22.2.2 | Data Description and Presentation                                 | 1056 |
| 22.2.3 | Summarizing Quantitative Variables                                | 1057 |
| 22.3   | Normal Distribution: Symmetric Frequency Distribution             | 1058 |
| 22.3.1 | What Is a Normal Distribution?                                    | 1058 |
| 22.3.2 | Properties of Normal Distribution                                 | 1059 |
| 22.4   | Principles of Statistical Inference                               | 1059 |
| 22.4.1 | Hypothesis Testing                                                | 1059 |
| 22.4.2 | Alpha (Type I) and Beta (Type II) Errors                          | 1061 |
| 22.4.3 | Confidence Intervals                                              | 1062 |
| 22.4.4 | Relationship between Significant Testing and Confidence Intervals | 1063 |
| 22.4.5 | Examples                                                          | 1064 |
| 22.5   | Comparison of Two Means                                           | 1065 |

|        |                                                                          |      |
|--------|--------------------------------------------------------------------------|------|
| 22.6   | Comparison of Two Proportions                                            | 1067 |
| 22.6.1 | Assessing Size of Treatment Effects in Two-Arm Trial                     | 1068 |
| 22.7   | Survival Analysis                                                        | 1069 |
| 22.7.1 | Example: Pancreatic Cancer Trial                                         | 1070 |
| 22.7.2 | Basic Concepts in Survival Analysis                                      | 1070 |
| 22.7.3 | Assessing Size of Treatment Effect in Two-Arm Trial for<br>Survival Data | 1073 |
| 22.8   | Commonly Used Terms in Clinical Trials                                   | 1074 |
| 22.9   | Concluding Remarks                                                       | 1078 |
|        | References                                                               | 1078 |

**22.1 INTRODUCTION**

Evidence-based medicine is a cornerstone of current medical practice. Randomized controlled trials are regarded as the most robust form of evidence-based medicine. An appreciation of statistical methods is fundamental to understanding randomized trial methods and results.

A randomized controlled trial aims to provide unbiased treatment effect regarding the efficacy and safety of a medicinal product or a therapeutic procedure. The observed treatment effect, however, may represent the “true” difference between the new drug and the comparative treatment or it may not. Therefore, if the trial were repeated with all the available patients in the world, then the outcome would either be the same as the trial (a true result) or different (making the trial result a chance event, or an erroneous false result). Understanding the possible sources of erroneous results is critical in the appreciation of clinical trials. The reasons for erroneous results fall into three main categories.

- First, the trial may have been biased in some predictable fashion.
- Second, it could have been contaminated (confounded) by an unpredictable factor.
- Third, the result may simply have occurred by chance.

**22.1.1 Bias and Systematic Errors**

Bias can influence a trial by the occurrence of systematic errors that are associated with the design, conduct, analysis, and reporting of the results of a clinical trial. Bias can also make the trial-derived estimate of a treatment effect deviate from its true value [1, 2]. The most common types of bias in clinical trials are those related to subject selection and outcome measurement. For example, if the investigators are aware of which treatment a patient is receiving, it could affect the way they collect information on the outcome during the trial, or they might recruit patients in a way that could favor the new treatment, resulting in a selection bias.

In addition, exclusion of subjects from statistical analysis because of noncompliance or missing data could bias an estimate of the true benefit of a treatment, particularly if more patients were removed from analysis in one group than the other [3, 4]. Much of the advanced design strategies seek to reduce these systematic errors.



### 22.1.2 Confounding

Confounding represents the distortion of the true relationship between treatment and outcome by another factor, for example, the severity of disease [5]. Confounding occurs when an extra factor is associated with both the outcome of interest and treatment group assignment. Confounding can both obscure an existing treatment difference and create an apparent difference that does not exist.

If we divided patients into treatment groups based on inherent differences (such as mean age) at the start of a trial, then we would be very likely to find the benefit of the new treatment to be influenced by those preexisting differences. For example, if we assign only smokers to get treatment A, only nonsmokers to get treatment B, and then assess which treatment protects better against cardiovascular disease, we might find that the benefit seen with treatment B is due to the lack of smoking in this group. The effect of treatment B on cardiovascular disease development would therefore be confounded by smoking.

Randomization in conjunction with a large sample size is the most effective way to restrict such confounding, by evenly distributing both known and unknown confounding factors between treatment groups. If, before the study begins, we know which factors may confound the trial, then we can use randomization techniques that force a balance of these factors (stratified randomization). In the analysis stage of a trial, we might be able to restrict confounding using appropriate statistical techniques such as stratified analysis and regression analysis [6].

### 22.1.3 Random Error

Even if a trial has an ideal design and is conducted to minimize bias and confounding, the observed treatment effect could still be due to random error or chance [1, 2, 7]. The random error can result from sampling, biologic, or measurement variation in outcome variables. Since (given specific selection criteria) the patients in a clinical trial are only a sample of all possible available patients, the sample might yet show a chance false result compared to the overall population. This is known as a sampling error. Sampling errors can be reduced by choosing a very large group of patients. Other causes of random error are described elsewhere [2].

Statistical analyses deal with random error by providing an estimate of how likely the measured treatment effect reflects the true effect [7–9]. Statistical testing or inference involves an assessment of the probability of obtaining an observed treatment difference or more extreme difference for an outcome, assuming that there is no difference between treatments. This probability is often called the *P* value or false-positive rate. If the *P* value is less than a specified critical value (e.g., 5%), the observed difference is considered to be statistically significant. The smaller the *P* value, the stronger the evidence is for a true difference between treatments. On the other hand, if the *P* value is greater than the specified critical value, then the observed difference is regarded as not statistically significant and is considered to be potentially due to random error or chance. The traditional statistical threshold is a *P* value of 0.05 (or 5%), which means that we only accept a result when the likelihood of the conclusion being wrong is less than 1 in 20. In other words we conclude that only 1 out of a hypothetical 20 trials will show a treatment difference when in truth there is none.

Statistical estimates summarize the treatment differences for an outcome in the forms of point estimates (e.g., means or proportions) and measures of precision [e.g., confidence intervals (CIs)] [7]. A 95% CI for a treatment difference means that the range presented for the treatment effect contains (when calculated in 95 out of 100 hypothetical trials assessing the same treatment effect) the true value of treatment difference, that is, the value we would obtain if we were to use the entire available patient population is 95% likely to be contained in the 95% CI.

Finally, testing several different hypotheses with the same trial (e.g., comparing treatments with respect to different outcomes or for several smaller subpopulations within the trial population) will increase the chance of observing a statistically significant difference purely due to chance [10]. Even examining the difference between treatments at many time points (interim analyses) throughout the length of the trial could lead to a spurious result due to multiple testing [10, 11]. Therefore, the aim should be to plan a trial in such a way that the occurrence of any such errors is minimal.

## 22.2 TYPES OF DATA, SUMMARY, AND DATA PRESENTATION

### 22.2.1 Types of Data

Clinical studies compare effects of medical treatment or interventions on two or more groups of subjects. Data are collected, however, at the individual subject or patient level. Data are composed of baseline characteristics such as age, gender, height, weight, and so forth, or of disease factors such as presence of arthritis, coronary disease, or severity of a fracture, or of treatment response variables such as reduction in pain, improvement of disease, or prolongation of life. These data come from variables of different types—either continuous (or quantitative), such as age or hemoglobin levels, or they may be categorical (or qualitative), such as race or gender.

Data can be classified further into four main groups:

- Binary (categorical): for example, sex (male and female)
- Unordered (categorical): for example, race (white, black, other)
- Ordered (categorical): for example, severity (mild, moderate, severe)
- Numerical (continuous): for example, age (in years)

### 22.2.2 Data Description and Presentation

In clinical studies, data are summarized for presentation by treatment groups using frequency distributions. These present the distribution of both qualitative and quantitative data, summarizing how often each value of a data point is repeated. With quantitative data, we mostly present a grouped frequency distribution. From a frequency table we can appreciate:

- The frequency (number of cases) occurring for each category or interval—for example, number of 70-year-old patients

- The relative frequency (percentage) of the total sample in each category or interval—for example, 70-year-old patients were 10% of the overall sample
- The highest and lowest or the range of possible values from our patient groups—for example, in that group the oldest patient was 96 and the youngest patient was 25

Although a frequency table provides a detailed summary of the distribution of the data, presentation of the distribution in a graph/chart makes the message from the data more informative. The type of graph depends on the type of data. Generally, if the data are categorical, we use a bar graph or a pie chart. If the data are continuous, a histogram or frequency polygon is more appropriate. As well as for the entire study population, the frequency distributions can then be presented for each treatment group. This is a fast way of ascertaining whether there are broad similarities or differences between treatment groups. Later we can use statistical tests to ascertain whether any differences between the groups are significant.

### 22.2.3 Summarizing Quantitative Variables

Categorical variables may be expressed as percentages and compared between different groups. There is little more that can be done to describe such variables. However, for a quantitative variable we can do more than this. We have other measures with which to summarize the data (summary measures). From the frequency distribution we can calculate the location (or central tendency) that summarizes where the center of the distribution lies and we can also summarize the spread/range (or variation) of the distribution, which describes how widely the values are spread above and below the central value.

There are two measures commonly used to describe the location for the central tendency, depending on whether the distribution is even (such as age) or more biased in one direction (such as physical fitness—many more people are unfit than are Olympic athletes).

- *Mean:* The mean can be calculated by summing all the quantitative observations and divided the sum by the number of observations.
- *Median:* The median is the value that divides the distribution equally. The median is more appropriate for distributions that are skewed, such as physical fitness. When the distribution is symmetrical, the median equals the mean.

There are three measures commonly used to summarize the spread of a variable:

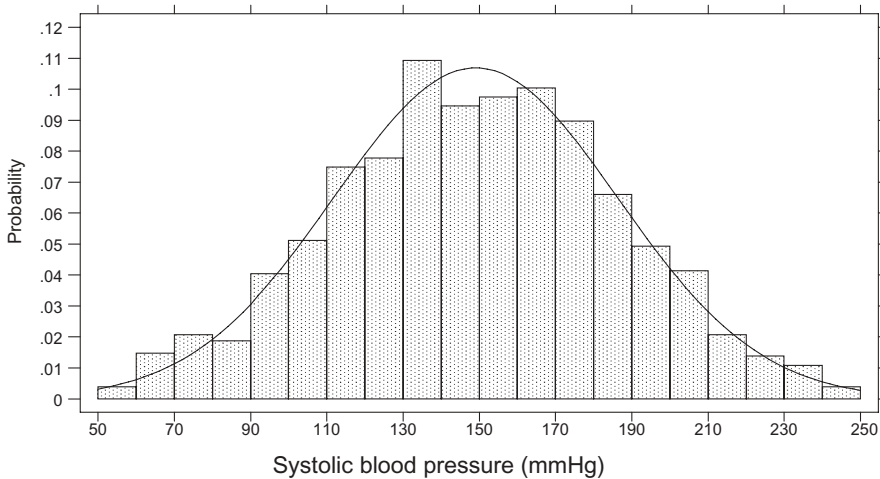
- *Standard deviation:* This gives the average distance of all observations from the mean. The standard deviation has an important role in statistical analysis.
- *Range:* Any data between the lowest and highest values is known as the full range of values.
- *Percentile:* This is the value below which a given percentage of the data observations occur. The most commonly used percentiles are 5 and 95%. Using these overcomes the problem of extreme data values away from the mean or median.

## 22.3 NORMAL DISTRIBUTION: SYMMETRIC FREQUENCY DISTRIBUTION

### 22.3.1 What Is a Normal Distribution?

Quantitative variables can range from very small to very large values, and, in some situations, from negative to positive values. Consider the systolic blood pressure (BP) measurements of 1000 subjects participating in a lifestyle survey. The frequency distribution of these systolic BPs is shown in Figure 1. The height of each vertical bar in this graph shows the proportion (or probability) of subjects whose systolic blood level was between the values at the base of the bar. So, if we summed the heights of all the bars in the histogram, we should find that the total of all the proportions is 1. Proportion is the same as the percentage expressed as a fraction over 100%.

The distribution of these blood pressures conforms approximately to a bell-shaped curve or normal distribution typical for common continuous biological measurements. The curve has particular mathematical properties expressed by the



Normal distribution function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \sigma > 0, -\infty < \mu < \infty, -\infty < x < \infty$$

Properties of normal distribution (curve):

The curve has a single peak at the center; this peak occurs at the mean ( $\mu$ ).

The curve is symmetrical about the mean.

The curve never touches the horizontal axis.

The total area under the curve is equal to 1.

The width or shape of the curve is described by the variance ( $\sigma^2$ ), the squared root of which is the standard deviation ( $\sigma$ ).

**FIGURE 1** Histogram and fitted normal distribution curve for systolic blood pressures from 1000 subjects.

equation shown with the curve. This distribution is one of the most important distributions in statistics and is known as the Gaussian distribution [12].

If we have enough values with smaller bars, the heights of these bars would form the bell-shaped curve shown superimposed on the chart. This curve is defined by the mean, or central value ( $\mu$ ), of 149.044 mmHg, and by its spread, or standard deviation ( $\sigma$ ), of 37.317 mmHg.

### 22.3.2 Properties of Normal Distribution

The properties of a normal distribution are illustrated in the legend of Figure 1. The standard deviation helps to describe the spread of the observations, with about 68% of all observations being captured within one standard deviation at either side of the mean and about 95% of all observations captured within two standard deviations at either side of the mean.

## 22.4 PRINCIPLES OF STATISTICAL INFERENCE

Let us suppose that it is necessary to measure the average systolic blood pressure (SBP) level of all males aged  $\geq 16$  years in the United Kingdom in 2005. For practical and financial reasons, it is not possible to directly measure the SBP of every adult male in the United Kingdom. Instead, we can conduct a survey among a subset (or “sample”) of 500 males within this population. Through statistical inference, we can measure the properties of the sample (such as the mean and standard deviation) and use these values to infer the properties of the entire UK adult male population [11–13].

Population properties are usually determined by population parameters (numerical characteristics of a population) that are fixed and usually unknown quantities, such as the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) in a normal distribution  $N(\mu, \sigma^2)$  [14, 15]. The statistical properties of the sample, such as the mean ( $\bar{X}$ ) and standard deviation ( $S$ ), can be used to provide estimates of the corresponding population parameters. Conventionally, Greek letters are used to refer to population parameters, while the Roman alphabet is used to refer to sample estimates.

Two strategies that are often used to make statistical inference are [7, 13, 14]:

1. Hypothesis testing
2. Confidence intervals (CIs)

### 22.4.1 Hypothesis Testing

Statistical inference can be made by performing a hypothesis (or significance) test, which involves a series of statistical calculations [7, 12–14]. In the sample of 500 adult males, the mean SBP ( $\bar{X}$ ) was 130 mmHg, with a standard deviation ( $S$ ) of 10 mmHg. The empirical estimate for the mean SBP of this population from previous medical literature is reported as 129 mmHg (denoted by  $\mu_0$ ). So, we want to know whether there is any evidence that the mean SBP value for all adult males in the United Kingdom in 2005 ( $\mu$ ) is different from 129 mmHg ( $\mu_0$ ).

We start by stating a hypothesis that the population mean SBP for all adult men in 2005 is 129 mmHg, or  $\mu = \mu_0$  (i.e., no different to that reported in the medical literature). This is referred to as the null hypothesis and is usually written as  $H_0$ , representing a theory that has been put forward as a basis for argument [7, 12–14]. The hypothesis test is a means to assess the strength of evidence against this null hypothesis of no difference.

The alternative hypothesis, usually written as  $H_a$ , is that the mean SBP for the study population is not equal to the specified value, that is,  $\mu \neq \mu_0$ . Note that under the alternative hypothesis, the 2005 population mean could be higher or lower than the reference mean. The statistical test for the above hypotheses is usually referred to as a two-sided test.

Once the null hypothesis has been chosen, we need to calculate the probability that, if the null hypothesis is true, the observed data (or data that were more extreme) could have been obtained [7, 12–14]. To reach this probability, we need to calculate a test statistic from the sample data (e.g.,  $\bar{X}$ ,  $S$ , and  $n$  for quantitative outcomes) using an appropriate statistical method. This test statistic is then compared to the distribution (e.g., the normal distribution) implied by the null hypothesis to obtain the probability of observing our data or more extreme data.

For the SBP data, given the relatively large sample size, we can use the  $Z$  test to calculate the value of the test statistic  $Z$ . The  $Z$  test is expressed by the following formula [7, 12–14]:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

This statistic follows a normal standard distribution under the null hypothesis [7, 12–14]. For the SBP data:

- $\bar{X} = 130$  mmHg
- $S = 10$  mmHg
- $n = 500$
- $\mu_0 = 129$  mmHg

Replacing the values in the formula generates  $Z = 2.24$ . A variety of statistical methods can be used to address different study questions (e.g., comparing treatment difference in means and proportions). The choice of statistical test will depend on the types of data and hypotheses under question [7, 12–14].

Having obtained the appropriate test statistic (in our example, the  $Z$  value), the next step is to specify a significance level. This is a fixed probability of wrongly rejecting the null hypothesis,  $H_0$ , if it is in fact true. This probability is always chosen by the investigators taking into account the consequences of such an error. That is, the significance level is kept low to reduce the chance of inadvertently making a false claim. The significance level, denoted by  $\alpha$ , is usually chosen to be 0.05 (5%). The corresponding  $Z_{\alpha/2}$  is called the critical value of the  $Z$  test. The critical value for a hypothesis test is a threshold with which the value of the test statistic calculated from a sample is compared in order to determine the  $P$  value to be introduced next. For example, if  $\alpha = 0.05$ , we have  $Z_{0.05/2} = 1.96$ ; if  $\alpha = 0.01$ , we have  $Z_{0.01/2} = 2.58$ .

A  $P$  value is the probability of our result ( $Z = 2.24$  for the SBP data) or a more extreme result ( $Z < -2.24$  or  $Z > 2.24$ ) being observed, assuming that the null hypothesis is true. The exact  $P$  value in the  $Z$  test is the probability of  $Z \leq -Z_{\alpha/2}$  or  $Z \geq Z_{\alpha/2}$ , which can always be determined by calculating the area under the curve in two-sided symmetric tails from a statistical table specifically of a normal distribution [7, 14]. For the SBP data, the exact  $P$  value = 0.025.

In a practical application, we often need to determine whether the  $P$  value is smaller than a specified significance level,  $\alpha$ . This is performed by comparing the value of the test statistic with the critical value. Statistically  $P \leq \alpha$  if, and only if,  $Z = -Z_{\alpha/2}$  or  $Z = Z_{\alpha/2}$ . For the SBP data, since  $Z = 2.24 > Z_{0.05/2} = 1.96$ , we can conclude that  $P < 0.05$ . A smaller  $P$  value indicates that  $Z$  is further away from the center (i.e., the null value  $\mu - \mu_0 = 0$ ) and consequently provides stronger evidence to support the alternative hypothesis of a difference.

Although the  $P$  value measures the strength of evidence for a difference, which is largely dependent on the sample size, it does not provide the size and direction of that difference. Therefore, in a statistical report,  $P$  values should be provided together with CIs (described in detail later) for the main outcomes [7, 14].

We are now in a position to interpret the  $P$  value in relation to our data and decide whether there is sufficient evidence to reject the null hypothesis. Essentially, if  $P = \alpha$ , the prespecified significance level, then there is evidence against the null hypothesis, and we accept the alternative hypothesis stating that there is a statistically significant difference. The smaller the  $P$  value, the lower the chance of obtaining a difference as big as the one observed if the null hypothesis were true, and, therefore, the stronger the evidence against the null hypothesis. Otherwise, if  $P > \alpha$ , there is not sufficient evidence to reject the null hypothesis or there is no statistically significant difference.

For our SBP data, since  $P < 0.05$ , we can state that there is evidence to reject the null hypothesis of no difference at the 5% significance level, and, therefore, that the mean SBP for the adult male population is statistically significantly different from 129 mmHg. Furthermore, the actual  $P$  value equals 0.025, which suggests that the probability of falsely rejecting the null hypothesis is 1 in 40 if the null hypothesis is indeed true. On the other hand,  $Z_{0.005} = 2.58 > Z = 2.24$ , and  $P$  is  $> 0.01$ . Now we can state that there is no evidence to reject the null hypothesis of no difference if the significance level  $\alpha$  is chosen as 0.01.

The implementation of the above procedures for hypothesis testing with the SBP data is summarized in Table 1.

#### 22.4.2 Alpha (Type I) and Beta (Type II) Errors

When testing a hypothesis, two types of errors can occur. To explain these two types of errors, we will use the example of a randomized, double-blind, placebo-controlled clinical trial on a cholesterol-lowering drug A in middle-aged men and women considered to be at high risk for a heart attack. The primary endpoint is the reduction in the total cholesterol level at 6 months from randomization.

The null hypothesis is that there is no difference in mean cholesterol reduction level at 6 months postdose between patients receiving drug A ( $\mu_1$ ) and patients receiving a placebo ( $\mu_2$ ) ( $H_0: \mu_1 = \mu_2$ ); the alternative hypothesis is that there is a difference ( $H_a: \mu_1 \neq \mu_2$ ). If the null hypothesis is rejected when it is in fact true, then



TABLE 1 Practical Procedures for Hypothesis Testing

| Step | Procedure                                                                                                                     | Illustration with SBP <sup>a</sup> Data                                                                                                                                                                                                                                                                                               |
|------|-------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | Set up a null hypothesis and alternative hypothesis that is of particular interest to study.                                  | $H_0: \mu = \mu_0 (= 129)$ , i.e., population mean SBP is equal to 129 mmHg<br>$H_a: \mu \neq \mu_0$ , i.e., population mean SBP is different from 129 mmHg                                                                                                                                                                           |
| 2    | Choose a statistical method according to data type and distribution and calculate its test statistic from the data collected. | $Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = 2.24$<br>$\bar{X} = 130$ mmHg<br>$S = 10$ mmHg<br>$n = 500$                                                                                                                                                                                                                                 |
| 3    | Define a significance level $\alpha$ and its corresponding critical value.                                                    | $\alpha = 0.05$ and $Z_{\alpha/2} = 1.96$<br>$\alpha = 0.01$ and $Z_{\alpha/2} = 2.58$                                                                                                                                                                                                                                                |
| 4    | Determine the $P$ value by comparing the test statistic and the critical value, or calculate the exact $P$ value.             | Since $Z = 2.24 > 1.96$ , $P < 0.05$<br>Since $Z = 2.24 < 2.58$ , $P > 0.01$<br>Exact $P$ value = 0.025                                                                                                                                                                                                                               |
| 5    | Make your conclusion according to the $P$ value.                                                                              | As $0.01 < P < 0.05$ , there is evidence to reject the null hypothesis of no difference at the 5% level of significance, but there is no evidence to reject the null hypothesis at the 1% level. The $P$ value of 0.025 means that the probability of falsely rejecting the null hypothesis is 1 in 40 if the null hypothesis is true |

<sup>a</sup>SBP, systolic blood pressure.

a type I error (or false-positive result) occurs. For example, a type I error is made if the trial result suggests that drug A reduced cholesterol levels when in fact there is no difference between drug A and placebo. The chosen probability of committing a type I error is known as the significance level [7, 12–14]. As discussed above, the level of significance is denoted by  $\alpha$ . In practice,  $\alpha$  represents the consumer’s risk [2], which is often chosen to be 5% (1 in 20).

On the other hand, if the null hypothesis is not rejected when it is actually false, then a type II error (or false-negative result) occurs [7, 12–14]. For example, a type II error is made if the trial result suggests that there is no difference between drug A and placebo in lowering the cholesterol level when in fact drug A does reduce the total cholesterol. The probability of committing a type II error, denoted by  $\beta$ , is sometimes referred to as the manufacturer’s risk [2]. The power of the test is given by  $1 - \beta$ , representing the probability of correctly rejecting the null hypothesis when it is in fact false. It relates to detecting a prespecified difference.

22.4.3 Confidence Intervals

The second strategy for making statistical inference is through the use of CIs. In making inference about a population, we might want to know the likely value of the unknown population parameter [e.g., mean ( $\mu$ ), proportion]. This is estimated from the sample statistics. For example, mean ( $\bar{X}$ ), and we call  $\bar{X}$  a point estimate of  $\mu$ .

In addition, we might want to provide some measure of our uncertainty as to how close the sample mean is to the true mean. This is done by calculating a CI (or interval estimate)—a range of values that has a specified probability of containing the true population parameter being estimated. For example, a 95% CI for the mean is usually interpreted as a range of values containing the true population mean with a probability of 0.95 [2]. The formula for the  $(1 - \alpha)\%$  CI around the sample mean ( $\bar{X}$ ) corresponding to the  $Z$  test, is given by

$$\bar{X} \pm Z_{\alpha/2} \text{SE}(\bar{X})$$

where  $\text{SE}(\bar{X})$  is the standard error of  $\bar{X}$ , calculated by  $S/\sqrt{n}$ . This is a measure of the uncertainty of a single sample mean ( $\bar{X}$ ) as an estimate of the population mean [7]. This uncertainty decreases as the sample size increases. The larger the sample size, the smaller the standard error. Therefore, the narrower the interval, the more precise the point estimate.

For our SBP example, the 95% CI for the population mean ( $\mu$ ) can be calculated with the following formula:

$$\bar{X} \pm 1.96S/\sqrt{n} = 129.1\text{--}130.9\text{ mmHg}$$

This means that the interval between 129.1 and 130.9 mmHg has a 0.95 probability of containing the population mean  $\mu$ . In other words, we are 95% confident that the true population mean is between 129.1 and 130.9 mmHg, with the best estimate being 130 mmHg.

Confidence intervals can be calculated not just for a mean but also for any estimated parameter depending on the data types and statistical methods used [12, 14]. For example, one could estimate the proportion of people who smoke in a population, or the difference between the mean SBP in subjects taking an antihypertensive drug and those taking a placebo.

#### 22.4.4 Relationship between Significant Testing and Confidence Intervals

When comparing, for example, two treatments, the purpose of significance testing is to assess the evidence for a difference in some outcome between the two groups, while the CI provides a range of values around the estimated treatment effect within which the unknown population parameter is expected to be with a given level of confidence.

There is a close relationship between the results of significance testing and CIs. This can be illustrated using the previously described  $Z$  test for the SBP data analysis. If  $H_0: \mu = \mu_0$  is rejected at the  $\alpha\%$  significance level, the corresponding  $(1 - \alpha)\%$  CI will not include  $\mu_0$ . On the other hand, if  $H_0: \mu = \mu_0$  is not rejected at the  $\alpha\%$  significance level, then  $(1 - \alpha)\%$  CI will include  $\mu_0$ .

For the SBP data of adult males, the significance test shows that  $\mu$  is significantly different from  $\mu_0 (= 129 \text{ mmHg})$  at the 5% level, and the 95% CI ( $= 129.1\text{--}130.9 \text{ mmHg}$ ) did not include 129 mmHg. On the other hand, the difference between  $\mu$  and  $\mu_0$  is not significant at the 1% level; the 99% CI [ $129 \pm (2.58 \times 10)/\sqrt{500} = 128.8\text{--}131.2 \text{ mmHg}$ ] for  $\mu$  does indeed contain  $\mu_0$ . Further information about the proper use of the above two statistical methods can be found in [7, 16].

### 22.4.5 Examples

Let us assume that four randomized, double-blind, placebo-controlled trials are conducted to establish the efficacy of two weight loss drugs (A and B) against placebo, with all subjects, whether on a drug or placebo, receiving similar instructions as to diet, exercise, behavior modification, and other lifestyle changes. The primary endpoint is the weight change (in kilograms) at 2 months from baseline. The difference in the mean weight change between an active drug and placebo groups can be considered as weight reduction for the active drug against placebo. Table 2 presents the results of hypothesis tests and CIs for the four hypothetical trials. The null hypothesis for each trial is that there is no difference between the active drug treatment and placebo in mean weight change. In trial 1 of drug A, the reduction of drug A over placebo was 6 kg with only 40 subjects in each group. The *P* value of 0.074 suggests that there is no evidence against the null hypothesis of no effect of drug A at the 5% significance level. The 95% CI shows that the results of the trial are consistent with a difference ranging from a large reduction of 12.6 kg in favor of drug A to a reduction of 0.6 kg in favor of placebo.

The results for trial 2 among 400 patients, again for drug A, suggest that mean weight was again reduced by 6 kg. This trial was much larger, and the *P* value ( $P < 0.001$ ) shows strong evidence against the null hypothesis of no drug effect. The 95% CI suggests that the effect of drug A is a greater reduction in mean weight over placebo of between 3.9 and 8.1 kg. Because this trial was large, the 95% CI was narrow and the treatment effect was therefore measured more precisely. In trial 3, for drug B, the reduction in weight was 4 kg. Since the *P* value was 0.233, there was no evidence against the null hypothesis that drug B has no statistically significant benefit effect over placebo. Again this was a small trial with a wide 95% CI, ranging from a reduction of 10.6 kg to an increase of 2.6 kg for the drug B against the placebo.

The fourth trial on drug B was a large trial in which a relatively small, 2-kg reduction in mean weight was observed in the active treatment group compared with the placebo group. The *P* value (0.008) suggests that there is strong evidence against the null hypothesis of no drug effect. However, the 95% CI shows that the reduction is as little as 0.5 kg and as high as 3.5 kg. Even though this is convincing statistically, any recommendation for its use should consider the small reduction achieved alongside other benefits, disadvantages, and cost of this treatment.

Key points from the four trials are summarized in Table 3.

**TABLE 2 Point Estimate and 95% Confidence Interval (CI) for Difference in Mean Weight Change from Baseline between the Active Drug and Placebo Groups in Four Hypothetical Trials of Two Weight Reduction Drugs**

| Trial | Drug | No. of Patients per Group | Difference in Mean Weight Change from Baseline (kg) between the Active Drug and Placebo Groups | Standard Deviation of Difference | Standard Error of Difference | 95% CI for Difference | <i>P</i> value |
|-------|------|---------------------------|------------------------------------------------------------------------------------------------|----------------------------------|------------------------------|-----------------------|----------------|
| 1     | A    | 40                        | −6                                                                                             | 15                               | 3.4                          | −12.6, 0.6            | 0.074          |
| 2     | A    | 400                       | −6                                                                                             | 15                               | 1.1                          | −8.1, −3.9            | <0.001         |
| 3     | B    | 40                        | −4                                                                                             | 15                               | 3.4                          | −10.6, 2.6            | 0.233          |
| 4     | B    | 800                       | −2                                                                                             | 15                               | 0.8                          | −3.5, −0.5            | 0.008          |

**TABLE 3 Summary of Key Points from Results Described in Table 2**

| Key Points from Significance Test and CI                                                                                                                                                                         | Examples       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| In a small study, a large $P$ value does not mean that the null hypothesis is true—absence of evidence is not evidence of absence                                                                                | Trials 1 and 3 |
| A large study has a better chance of detecting a given treatment effect than a small study and is therefore more powerful                                                                                        | Trials 2 and 4 |
| A small study usually produces a CI for the treatment effect that is too wide to allow any useful conclusion                                                                                                     | Trials 1 and 3 |
| A large study usually produces a narrow CI, and therefore a precise estimate of treatment effect                                                                                                                 | Trials 2 and 4 |
| The smaller the $P$ value, the lower the chance of falsely rejecting the null hypothesis, and the stronger the evidence for rejecting the null hypothesis                                                        | Trials 2 and 4 |
| Even if the $P$ value shows a statistically significant result, it does not mean that the treatment effect is clinically significant. The clinical importance of the estimated effects should always be assessed | Trial 4        |

## 22.5 COMPARISON OF TWO MEANS

The statistical methods used to compare two means of different treatment groups depend on how the two means were obtained. Data may be obtained from paired or unpaired samples. Paired samples occur when the source of the observations in the first sample are from the same source as the observations of a second or repeat sample. For quantitative data this means that repeated observations are from the same person using a before and after treatment method. Unpaired samples occur when individual observations in one group are independent of individual observations in the other. That is to say that the groups of subjects are different.

The most commonly used statistical test for the comparison of two means is the  $t$  test. The  $t$  test compares the means of two sets of continuous variables and expresses the probability that any differences are due to the play of chance (i.e., the null hypothesis is supported), or that the differences may be “real” (i.e., the alternative hypothesis is supported). This probability is often determined by a  $t$  value. The  $t$  tests can be used to compare two sample means for both paired and unpaired data, but formulas for calculating  $t$  values are different [12, 14]. An example of a  $t$  test for unpaired data is shown in Table 4.

From Table 4, we see that the  $t$  value of 2.055 corresponds to an observed  $P$  value of 0.040. The  $P$  value is the probability of observing the difference (3.440) or even larger (e.g., absolute difference  $>3.440$ ) from the samples if the null hypothesis were true. The null hypothesis is that the difference in means of the two populations is zero. If there were no difference in systolic blood pressure between two treatment groups, there would be a small chance ( $P = 0.040$ ) that we would observe the difference we did. We can turn this around and say that it is more likely that the two treatment groups differ. We say: “The difference in means between two treatment groups is statistically significant” since the observed  $P$  value is lower than our significance threshold value of  $P = 0.05$ .

The  $t$  test can be performed equivalently by calculating a confidence interval for the difference in means. A confidence interval is a range of values within which the “true” population parameter (such as difference in two means) is likely to lie. Usually 95% confidence limits are quoted, which implies that there is 95%

**TABLE 4** Example of Comparison of Two Means from Independent Samples—*t* Test

An example calculation in a study to evaluate the effects of a treatment to reduce blood pressure are given. A randomized placebo-controlled trial comparing the effect of an antihypertensive therapy against placebo was conducted. The primary endpoint is systolic blood pressure at 6 months after randomization. The results of the two groups of patients were as follows:

Results of Blood Pressures

- In the control group of 548 patients the mean systolic blood pressure was 150.22 with a standard deviation of 27.12 mmHg
- In the treatment group of 550 patients the mean systolic blood pressure was 146.78 with a standard deviation of 28.32 mmHg

Using these data the *t* test can be calculated in the following manner:

Suppose two populations of sizes  $n_1$  and  $n_2$  have blood pressure results forming normal distributions with means  $\bar{X}_1$  and  $\bar{X}_2$  and standard deviations  $S_1$  and  $S_2$ , respectively. Then the efficacy of the treatment, as compared to placebo, can be examined by testing the following hypotheses:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

The statistic (*t*) for the above test is given by

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]}{(n_1 + n_2 - 2)} \sqrt{1/n_1 + 1/n_2}}}$$

Therefore the null hypothesis can be rejected ( $H_0: \mu_1 = \mu_2$ ) at the  $\alpha$  level of significance if  $|t| \geq t(\alpha/2, n_1 + n_2 - 2)$ , where  $t(\alpha/2, n_1 + n_2 - 2)$  is the upper  $(\alpha/2)$ th percentile of the *t* distribution with  $n_1 + n_2 - 2$  degrees of freedom.

In this case, we have  $n_1 = 548$ ,  $n_2 = 550$ ,  $\bar{X}_1 = 150.22$ ,  $\bar{X}_2 = 146.78$ ,  $S_1 = 27.12$ ,  $S_2 = 28.32$ . Substituting the above sample statistics into the *t*-test formula, we obtain  $t = 2.055$ , which equates to a probability of 0.040 that the two populations are the same. The estimated difference in means together with its 95% confidence interval is 3.440 [0.156, 6.724].

As  $P = 0.040 < 0.05$  (or 95% confidence interval does not contain 0), we can say that there is a statistically significant difference in the systolic blood pressures between the two treatment groups. In other words, the drug appears to reduce systolic blood pressure compared to placebo, and the result has a low likelihood of arising by chance.

confidence in the statement that the “true” population parameter will lie somewhere between the upper and lower limits. For the Table 4 data, the estimated 95% confidence interval for the difference in means is [0.156, 6.724]. If the 95% confidence interval does not contain zero, we can say that there is a statistically significant difference in means between two populations. For Table 4 data, as the lower limit of the 95% confidence interval is greater than zero, we can say that two means are statistically significantly different.

Four key assumptions are required in the two-sample *t* test. Firstly, we assume that the two treatment group populations from which the samples are drawn are distributed normally [8, 12, 14]. Second, we assume that the variances (or standard deviations) of the two populations are equal [8, 12, 14], that is,  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ . The equality of variances assumption can be formally verified with an *F* test [8, 12, 14]. We can also perform an informal check by examining the relative magnitude of the two-sample variances  $S_1^2$  and  $S_2^2$ . For example, if  $S_1^2 / S_2^2$  is considerably different

from 1, then the assumption that  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  will be in doubt. In cases where  $\sigma_1^2 \neq \sigma_2^2$ , we need to use a modified  $t$  test or nonparametric method [8, 12, 14, 17]. Third, we assume that the observations in the two treatment groups are independent of each other, that is, no observation in one group is influenced by another observation in the second group [8, 12, 14, 17]. In other words, the value in one treatment group is not affected by that in another group. Finally, we assume that the two populations are homogeneous in terms of the observed and unobserved characteristics of patients (i.e., free from confounding) [7]. These characteristics might be demographics (such as age), prognosis (such as clinical history or disease severity), or baseline measurements of outcome variables.

Although we might never know the unobservable heterogeneity (differences) between two populations, we can assess whether the two populations are comparable by examining the observed summary statistics, such as means or proportions at baseline by treatment. This is why a table that summarizes the baseline information in a clinical trial by treatment group is always provided in a clinical report. If the two treatment groups are not balanced with regard to some of the predictors of outcome, covariate adjustment by means of stratification or regression modeling can be employed [1, 14, 17].

## 22.6 COMPARISON OF TWO PROPORTIONS

Quite often, we have a binary variable result or proportion to analyze from both groups. For example, in a clinical trial comparing a new treatment to reduce mortality after myocardial infarction, the primary endpoint is a binary outcome of death or survival. The numbers of subjects who die or survive in each of the two treatment groups form a  $2 \times 2$  contingency table, from which we can calculate the rate of death or proportion dead for each treatment group [9, 12, 14].

The most common approaches for comparing two proportions are the chi-squared ( $\chi^2$ ) test and the Fisher exact test. The  $\chi^2$  test involves determining the expected number of deaths in the new and standard treatment arms and then comparing these to those observed numbers. The test statistic used to assess these differences in deaths can be expressed as  $\chi^2 = \Sigma(O - E)^2/E$ , where  $O$  represents the observed frequencies and  $E$  the expected frequencies in each cell of the  $2 \times 2$  table. Under the null hypothesis,  $\chi^2$  should follow a chi-squared distribution with one degree of freedom. The Fisher exact test is a little more complex and consists of evaluating the sum of probabilities associated with the observed frequency table and all possible two-by-two tables that have the same row and column totals as the observed data [9, 12, 14]. The  $\chi^2$  test can also be used to compare more than two proportions [9, 12, 14].

When the total study size is large (say over 200), a test for the difference between two proportions also uses a normally distributed test statistic  $Z$  (known as  $Z$  test), which can be easily calculated by hand (Table 5).

The null hypothesis in Table 5 is that there is no difference between the two treatment groups in the death rate among patients after myocardial infarction. The  $P$  value ( $<0.001$ ) gives us the probability of observing the difference seen in the death rates ( $-0.028$ ) if the null hypothesis were true. As the  $P$  value is far less than 0.001, we can reject the null hypothesis and say that the difference in proportions is statistically significant.

**TABLE 5 Example of Comparison of Two Proportions—Z test**

A randomized placebo-controlled clinical trial was conducted to evaluate a cardiovascular drug on reducing mortality on patients with myocardial infarction (MI). The primary endpoint was the death after MI at 30 days after randomization. Following data were obtained from the trial:

- Group 1 (treatment) 110 deaths out of a total of 2045 patients
- Group 2 (placebo) 165 deaths out of a total of 2022 patients

Statistical inference of binary data such as death from parallel two-group trial involves testing the hypotheses regarding the difference in proportions with a yes response between the treatment and the placebo group. This can be expressed as

$$H_0: \pi_1 = \pi_2$$

$$H_a: \pi_1 \neq \pi_2$$

The statistic for the above test is given by

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1 - \hat{p}_c)(1/n_1 + 1/n_2)}}$$

where  $n_1$  and  $n_2$  are the sample sizes (2045 and 2022),  $X_1$  and  $X_2$  the number of events in the two groups (110 and 165), respectively, and  $\hat{p}_1 = X_1/n_1$  ( $= 110/2045 = 0.054$ ),  $\hat{p}_2 = X_2/n_2$  ( $= 165/2022 = 0.082$ ), and  $\hat{p}_c = (X_1 + X_2)/(n_1 + n_2)$  ( $= [(110 + 165)/(2045 + 2022)] = 0.068$ ).

Therefore, the null hypothesis can be rejected ( $H_0: \pi_1 = \pi_2$ ) at the  $\alpha$  level of significance if  $|Z| \geq Z_{\alpha/2}$ , where  $Z_{\alpha/2}$  is the upper  $(\alpha/2)$ th percentile of the standard normal distribution.

In this case the  $Z = -3.548$ , which equates to a probability of  $P < 0.001$ . The estimated difference is  $-0.028$  (95% confidence interval:  $[-0.043, -0.012]$ ).

The above results suggest that patients in the treatment group have a statistically significant lower death rate than patients in the placebo group.

Alternatively, we can calculate the confidence interval for the difference between the two rates. For the Table 5 data, the observed difference is  $-0.028$  ( $-2.8\%$ ) with a 95% confidence interval  $[-4.3\%, -1.2\%]$ . From these results, we can say that there is a 95% probability that the difference that would be seen, if the entire population of all myocardial infarction patients were tested with either the new or standard treatment, would be between  $-4.3$  and  $-1.2\%$ . As the 95% confidence interval does not include zero, we can say that rate of death in the treatment group would be lower than that for standard treatment.

There is one common important assumption for the  $\chi^2$   $t$  test, the Fisher exact test, and the  $Z$  test: The two treatment groups are homogeneous in terms of the patients' characteristics (e.g., demographic information, disease-related risk factors, medical histories, and concurrent medical treatments). We can check whether the two groups are comparable by examining the observed baseline summary statistics. If the two treatment groups are not balanced with regard to some predictor(s) of outcome, logistic regression modeling could be used to adjust for these potential confounding factors [7, 9].

### 22.6.1 Assessing Size of Treatment Effect in Two-Arm Trial

In the previous sections, we have introduced different methods for assessing whether there is any evidence against the null hypothesis of no difference. In this section, we describe three commonly used measurements for assessing the size of any treatment effect for a binary outcome.



**Risk Difference** The difference in the proportion of the outcomes between two groups,  $p_1 - p_2$ , is called the risk difference. In the case of the myocardial infarction (MI) trial data, the risk difference is the risk of death between the active drug and placebo, that is,  $5.4\% - 8.2\% = -2.8\%$ . This means that the estimated absolute risk of death is 2.8% lower (about 3 in 100 MI patients) in the active drug group compared to the placebo group. Statistical inferences about the risk difference, such as point estimate and CIs, can be made by means of a  $Z$  test, as described in the last section.

**Risk Ratio** The *risk ratio* is the ratio of the risks in the active drug treatment group compared to the placebo group. The risk ratio is often abbreviated to RR, and is also sometimes called the relative risk:

$$RR = \frac{p_1}{p_2} = \frac{X_1/n_1}{X_2/n_2}$$

For the MI trial, the  $RR = 0.66$  ( $5.4\%/8.2\%$ ), meaning that the risk of death for the patients in the active drug treatment group is 66% of the risk in the placebo group. Equivalently, we could say that the drug treatment is associated with a 34% ( $100\% - 66\%$ ) reduction in mortality at 30 days.

**Odds Ratio** A third measure of treatment effect is the *odds ratio* (OR). The odds of an outcome event are calculated as the number of events divided by the number of nonevents. For example, in the active treatment arm in the MI trial, the number of deaths is 110 and the number of survivals is 1935, so the odds of death are  $110/1935 = 0.057$ . If the odds of an event are  $>1$ , the event is more likely to happen than not. In particular, the odds of an event that is certain to happen are infinite, and the odds of an impossible event are zero. The OR is calculated by dividing the odds in the active treatment group by the odds in the placebo group.

For the MI trial data, OR is calculated as  $(110 \times 1857)/(1936 \times 165) = 0.64$ , meaning that the odds of deaths after MI in the drug group is 64% of the odds in the placebo group. Clinical trials typically study treatments that reduce the proportion of patients with an event or equivalently have an  $OR < 1$ . In these cases, a percentage reduction in OR is often quoted instead of the OR itself. For the preceding OR, we can say that there is a 36% ( $100\% - 64\%$ ) reduction in the odds of deaths in the active treatment group.

## 22.7 SURVIVAL ANALYSIS

In many clinical trials, the primary outcome is not just whether an event occurs but also the time it takes for the event to occur. For example, in a cancer study comparing the relative merits of surgery and chemotherapy treatments, the outcome measured could be the time from the start of therapy to the death of the subject. In this case the event of interest is death, but in other situations it might be the end of a period spent in remission from cancer spread, relief of symptoms, or a further admission to hospital. These types of data are generally referred to as time-to-event data or survival data even when the endpoint or the event being studied is something

other than the death of a subject. The term “survival analysis” encompasses the methods and models for analyzing such data representing time free of events of interest.

22.7.1 Example: Pancreatic Cancer Trial

The death rate from pancreatic cancer is among the highest of all cancers. A randomized controlled clinical trial was conducted on 36 patients diagnosed with pancreatic cancer. The aim of this trial was to assess whether the use of a new treatment (A) could increase the survival of patients compared to the standard treatment (B). Patients were followed-up for 48 months and the primary endpoint was the time, in months, from randomization to death. Table 6 displays the survival data for the 36 patients. We will use this example to illustrate some fundamental survival analysis methods and their applications.

22.7.2 Basic Concepts in Survival Analysis

*Censoring* In survival analysis, not all subjects are involved in the study for the same length of time due to *censoring*. This term denotes when information on the outcome status of a subject stops being available. This can be because the patient is lost to follow-up (e.g., they have moved away) or stops participating in the study, or because the end of study observation period is reached without the subject having an event. Censoring is a nearly universal feature of survival data. Table 7 summarizes

TABLE 6 Survival Data for 36 Patients with Pancreatic Cancer in Trial of New Treatment versus Standard Treatment

| New Treatment             |                                             | Standard Treatment        |                                             |
|---------------------------|---------------------------------------------|---------------------------|---------------------------------------------|
| Survival Time<br>(months) | Survival Status<br>(0 = survival, 1 = dead) | Survival Time<br>(months) | Survival Status<br>(0 = survival, 1 = dead) |
| 2                         | 0                                           | 3                         | 0                                           |
| 5                         | 0                                           | 5                         | 1                                           |
| 10                        | 1                                           | 6                         | 0                                           |
| 12                        | 1                                           | 7                         | 1                                           |
| 15                        | 0                                           | 8                         | 1                                           |
| 27                        | 1                                           | 10                        | 1                                           |
| 36                        | 0                                           | 11                        | 0                                           |
| 36                        | 0                                           | 12                        | 1                                           |
| 37                        | 1                                           | 13                        | 0                                           |
| 38                        | 0                                           | 15                        | 1                                           |
| 39                        | 0                                           | 16                        | 0                                           |
| 41                        | 0                                           | 23                        | 1                                           |
| 42                        | 0                                           | 30                        | 1                                           |
| 44                        | 0                                           | 39                        | 1                                           |
| 45                        | 1                                           | 40                        | 1                                           |
| 46                        | 0                                           | 45                        | 1                                           |
| 48                        | 0                                           | 48                        | 0                                           |
| 48                        | 0                                           | 48                        | 0                                           |

**TABLE 7** Reasons for Censoring Observations in Clinical Trials

| Reason                                                                                        | Example                                                                         |
|-----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Lost to follow-up                                                                             | Patient moved away or did not wish to continue participation.                   |
| Patient withdrawn                                                                             | Patient withdraws from the study due to side effects.                           |
| Patient has an outcome that prevents the possibility of the primary endpoint (competing risk) | Death from cancer where death from cardiac causes is the primary endpoint.      |
| Study termination                                                                             | All patients who have not died are considered censored at the end of the study. |

the main reasons for censoring that can occur in a clinical trial. Survival analysis takes into account censored data, and, therefore, uses the information available from a clinical trial more fully.

**Survival Function and Hazard Function** In survival analysis, two functions are of central interest, namely *survival function* and *hazard function* [18, 19]. The survival function,  $S(t)$ , is the probability that the survival time of an individual is greater than or equal to time  $t$ . Since  $S(t)$  is the probability of surviving (or remaining event-free) to time  $t$ ,  $1 - S(t)$  is the probability of experiencing an event by time  $t$ . Plotting a graph of probability against time produces a *survival curve*, which is a useful component in the analysis of such data.

The hazard function,  $h(t)$ , represents the instantaneous event rate at time  $t$  for an individual surviving to time  $t$  and, in the case of the pancreatic cancer trial, it represents the instantaneous death rate. With regard to numerical magnitude, the hazard is a quantity that has the form of “number of events per time unit” (or “per person-time unit” in an epidemiological study). For this reason, the hazard is sometimes interpreted as an incidence rate. To interpret the value of the hazard, we must know the unit in which time is measured.

For the pancreatic cancer trial, suppose that the hazard of death for a patient is 0.02, with time measured in months. This means that if the hazard remains constant over one month, then the death rate will be 0.02 deaths per month (or per person-months). In reality, the 36 patients contributed a total of 950 person-months and 16 deaths. Assuming that the hazard is constant over the 48-month period and across all patients, an estimate of the overall hazard is  $16/950 = 0.017$  deaths per person-months.

**Kaplan–Meier Method** The Kaplan–Meier (KM) approach estimated the proportion of individuals surviving (i.e., who have not died or had an event) by any given time in the study [18, 19]. When there is no censoring in the survival data, the KM estimator is simple and intuitive.  $S(t)$  is the probability that an event time is greater than  $t$ . Therefore, when no censoring occurs, the KM estimator,  $\hat{S}(t)$ , is the proportion of observations in the sample with event times greater than  $t$ . For example, if 50% of observations have times  $>10$ , we have  $\hat{S}(10) = 0.50$ .

The KM estimates of the survival curves by the two treatment groups for the pancreatic cancer trial data are displayed in Figure 2. The survival curve is shown in a step function: The curve is horizontal at all times at which there is no event, with a vertical drop corresponding to the change in the survival function at each time,  $t_j$ , when an event occurs.

In reports, KM curves are usually displayed in one of two ways. The curves can decrease with time from 1 (or 100%), denoting how many people survive (or remain event free). However, in general it is recommended that the increase in event rates is shown starting from 0 (or 0%) subjects, with an increasing curve ( $1 - \hat{S}[t]$ ) unless the event rate is high [20]. Placing the curves for different treatment groups on the same graph allows us to graphically review any treatment differences.

**Log-Rank Test** For the two KM curves by treatment group shown in Figure 2, the obvious question to ask is: Did the new treatment make a difference in the survival experience of the two groups? A natural approach to answering this question is to test the null hypothesis that the survival function is the same in the two groups: that is,  $H_0: S_1(t) = S_2(t)$  for all  $t$ , where 1 and 2 represent the new treatment and the standard treatment, respectively.

The above hypothesis can be assessed by performing a log-rank test equivalent to a  $\chi^2$  test [18, 19, 21]. The main purpose of this test is to calculate the number of events expected in each treatment group, and to compare this expected number of events with the observed number of events in each treatment group if the null hypothesis is true.

For the pancreatic cancer trial, the resulting  $\chi^2$  value is 5.424 [22], converted to a  $P$  value of 0.020. As  $P < 0.05$ , the log-rank test has shown a significant survival difference between the new treatment A and standard treatment B. This test readily generalizes to three or more groups, with the null hypothesis that all groups have the same survival function. If the null hypothesis is true, the test statistic has a chi-

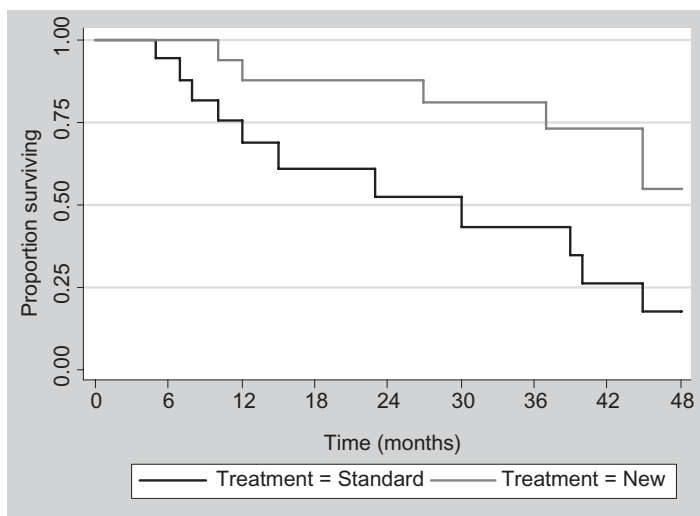


**FIGURE 2** Kaplan–Meier survival functions by treatment group for the pancreatic cancer trial data.

squared distribution with the degrees of freedom equal to the number of groups minus 1.

**Cox Proportional-Hazards Model** The proportional-hazards model relates the hazard function to a number of covariates (such as patient's characteristics at randomization and the treatment received in a clinical trial) as follows [18, 19, 22]:

$$h_i(t) = h_0(t) \exp(b_1 x_{1i} + b_2 x_{2i} + K + b_p x_{pi}) \quad (1)$$

where  $x_{ki}$  is the value of the covariate  $x_k$  ( $k = 1, 2, \dots, p$ ) for an individual  $i$  ( $i = 1, 2, \dots, n$ ).

The equation says that the hazard for individual  $i$  at time  $t$  is the product of two factors:

- A baseline hazard function  $h_0(t)$  that is left unspecified, except that it cannot be negative.
- A linear function of a set of  $p$  fixed covariates, which is then exponentiated.

The baseline hazard function can be regarded as the hazard function for an individual whose covariates all have values of 0 and changes according to time  $t$ .

This is called a proportional-hazards model because, while the baseline hazard can constantly change over time, the hazard for any individual is assumed to be proportional to the hazard for any other individual and will depend on the individual values of covariates. To see this, let us assume that the model only has one covariate (treatment,  $x_{1i}$ ,  $x_{1i} = 0$  for standard treatment and 1 for new treatment). We first calculate the hazards for two individuals 1 and 2 according the Equation (1) and then take the ratio of the two hazards:

$$\begin{aligned} h_1(t) &= h_0(t) \exp(b_1 x_{11}) \\ h_2(t) &= h_0(t) \exp(b_1 x_{12}) \\ \frac{h_1(t)}{h_2(t)} &= \exp[b_1 (x_{11} - x_{12})] \end{aligned} \quad (2)$$

What is important about the equation is that  $h_0(t)$  is canceled out of the numerator and denominator. As a result, the ratio of hazards,  $\exp[b_1 (x_{11} - x_{12})]$ , is a constant over time or proportional.

### 22.7.3 Assessing Size of Treatment Effect in Two-Arm Trial for Survival Data

**Incidence Rate Difference and Ratio** The incidence rate is defined as the number of events divided by the number of units of time [14, 22]. By comparing the incidence rates between treatment groups, we can derive the incidence rate difference and ratios following the procedures described by Kirkwood and Sterne [14, 22]. For the pancreatic cancer trial data, the incidence rates are calculated as 0.9 and 2.9 deaths

per 100 person-months for the new treatment group and the standard treatment group, respectively. The estimates of incidence rate difference and rate ratio together with their 95% CI and  $P$  value are as follows:

- Incidence rate difference:  $-2.0$ , 95% CI  $(-3.9, -0.1)$ ,  $P = 0.034$
- Incidence rate ratio:  $0.30$ , 95% CI  $(0.08, 0.94)$ ,  $P = 0.026$

The above results suggest that the new treatment reduces deaths by 2 per 100 person-months, with a 95% CI of  $0.1$ – $3.9$  per 100 person-months, and that the incidence rate for patients in the new treatment group is only about 30% of the incidence rate for those in the standard treatment group.

Although the incidence rate uses the information on censored observations, it is based on the assumption that the hazard of an event is constant during the study period or has an exponential distribution. In the case of the pancreatic cancer trial, it means that the hazard of death is constant over the 48-month period. However, the risk of an event can change with time. To overcome this problem, the Cox model, which does not require such assumptions, can be used to derive a better measurement for the treatment effect.

**Hazard Ratio** The treatment can be simply measured as a binary covariate (1 for new treatment A and 0 for standard treatment B in the pancreatic cancer trial) and introduced into a Cox proportional-hazards model. In the pancreatic cancer trial, the estimated hazard ratio of death for patients who received treatment A to those who received standard treatment B is  $0.31$ , with 95% CI  $(0.11, 0.89)$ ,  $P = 0.030$ . This means that the new treatment is estimated to reduce the hazard of death by 69%, with 95% CI (11%, 89%) and the reduction in hazard is statistically significant at the 5% significance level. As there is only one covariate (treatment) in the Cox model, the estimated hazard ratio is called a crude or unadjusted treatment effect. The adjusted hazard ratio for the treatment will be generated if other baseline patient characteristics are introduced in the model.

## 22.8 COMMONLY USED TERMS IN CLINICAL TRIALS

**ANOVA (*Analysis of Variance*)** A statistical method for comparing several means by comparing variances. It concerns a normally distributed outcome (response) variable and a single categorical (predictor) variable representing treatments or groups. ANOVA is a special case of a linear regression model by which group means can be easily compared.

**Bias** Systematic errors associated with the inadequacies in the design, conduct, or analysis of a trial on the part of any of the participants of that trial (patients, medical personnel, trial coordinators, or researchers), or in publication of its the results, that make the estimate of a treatment effect deviate from its true value. Systematic errors are difficult to detect and cannot be analyzed statistically but can be reduced by using randomization, treatment concealment, blinding, and standardized study procedures.

**Confidence Intervals** A range of values within which the “true” population parameter (e.g., mean, proportion, treatment effect) is likely to lie. Usually, 95% confidence limits are quoted, implying that there is 95% confidence in the statement that the “true” population parameter will lie somewhere between the lower and upper limits.

**Confounding** A situation in which a variable (or factor) is related to both the study variable and the outcome so that the effect of the study variable on the outcome is distorted. For example, if a study found that coffee consumption (study variable) is associated with the risk of lung cancer (outcome), the confounding factor here would be cigarette smoking since coffee is often drunk while smoking a cigarette, which is the true risk factor for lung cancer. Thus, we can say that the apparent association of coffee drinking with lung cancer is due to confounding by cigarette smoking (confounding factor). In clinical trials, confounding occurs when a baseline characteristic (or variable) of patients is associated with the outcome, but unevenly distributed between treatment groups. As a result, the observed treatment difference from the unadjusted (univariate) analysis can be explained by the imbalanced distribution of this variable.

**Correlation Coefficient ( $r$ )** A measure of the linear association between two continuous variables. The correlation coefficient varies between  $-1.0$  and  $+1.0$ . The closer it is to 0, the weaker the association. When both variables go in the same direction (e.g., height and weight),  $r$  has a positive value between 0 and 1.0 depending on the strength of the relationship. When the variables go in opposite directions (e.g., left ventricular function and life span),  $r$  has a negative value between 0 and  $-1.0$  depending on the strength of this inverse relationship.

**Covariates** This term is generally used as an alternative to explanatory variables in regression analysis. However, more specifically refer to variables that are not of primary interest in an investigation. Covariates are often measured at baseline in clinical trials because it is believed that they are likely to affect the outcome variable and, consequently, need to be included to estimate the adjusted treatment effect.

**Descriptive/Inferential Statistics** Descriptive statistics are used to summarize and describe data collected in a study. To summarize a quantitative (continuous) variable, measures of central location (i.e., mean, median, mode) and spread (e.g., range and standard deviation) are often used, whereas frequency distributions and percentages (proportions) are usually used to summarize a qualitative variable. Inferential statistics are used to make inferences or judgments about a larger population based on the data collected from a small sample drawn from the population. A key component of inferential statistics is hypothesis testing. Examples of inferential statistical methods are  $t$  test and regression analysis.

**Endpoint** Clearly defined outcome associated with an individual subject in a clinical research. Outcomes may be based on safety, efficacy, or other study objectives (e.g., pharmacokinetic parameters). An endpoint can be quantitative (e.g., systolic blood pressure, cell count), qualitative (e.g., death, severity of disease), or time to event (e.g., time to first hospitalization from randomization).



**Hazard Ratio** In survival analysis, hazard (rate) represents instantaneous event rate (incidence rate) at a certain time for an individual who has not experienced an event at that time. Hazard ratio compares two hazards of having an event between two groups. If the hazard ratio is 2.0, then the hazard of having an event in one group is twice the hazard in the other group. The computation of the hazard ratio assumes that the ratio is consistent over time (proportional hazards assumption).

**Hypothesis Testing or Significance Testing** Statistical procedure for assessing whether an observed treatment difference was due to random error (chance) by calculating a  $P$  value using the observed sample statistics such as mean, standard deviation, and so on. The  $P$  value is the probability that the observed data or more extreme data would have occurred if the null hypothesis (i.e., no true difference) were true. If the calculated  $P$  value is a small value (like  $<0.05$ ), the null hypothesis is then rejected, and we state that there is a statistically significant difference.

**Intention-to-Treat Analysis** A method of data analysis on the basis of the intention to treat a subject (i.e., the treatment regimen a patient was assigned at randomization) rather than the actual treatment regimen he or she received. It has the consequence that subjects allocated to a treatment group should be followed up, assessed, and analyzed as members of that group regardless of their compliance to that therapy or the protocol, irrespective of whether they later crossed over to the other treatment group or not or whether they discontinued treatment.

**Kaplan–Meier Estimate and Survival Curve** A survival curve shows an estimate of the fraction of patients who survive over the follow-up period of the study without an event of interest (e.g., death). The Kaplan–Meier estimate is a simple way of computing the survival curve taking into account patients who were lost to follow-up or any other reasons for incomplete results (known as censored observations). It usually provides a staircase graph of the fraction of patients remaining free of event over time.

**Meta-Analysis** The systematic review and evaluation of the evidence from two or more independent studies asking the same clinical question to yield an overall answer to the question.

**Normal Distribution** A bell-shaped symmetric distribution for a continuous variable having highest frequency at a mean value and less frequency further away from this mean value. A normal distribution can be completely described by two parameters: mean ( $\mu$ ) and variance ( $\sigma^2$ ). In the special case of  $\mu = 0$  and  $\sigma^2 = 1$ , it is called the standard normal distribution.

**Number Needed to Treat (NNT)** This term is often used to describe how many patients would need to be given a treatment to prevent one event. It is determined from the absolute difference between one treatment and another. In a randomized study the group receiving treatment A had a death rate of 12.5%, and the group receiving treatment B had a death rate of 15.0%. Both groups are matched for size and length of follow-up. Comparing the two treatments there was an absolute risk reduction of  $15\% - 12.5\% = 2.5\%$  for treatment A. From this we can derive that the

NNT ( $= 1/0.025$ ) is 40. This means 40 patients need to be given treatment A rather than B to prevent 1 additional death.

**Odds Ratio (OR) and Risk Ratio (RR)** These terms compare the probability of having an event between two groups exposed to a risk factor or treatment. The RR is the ratio of the probability of occurrence of an event between two groups. The OR is the ratio of the ratio of patients with and without an event in each group. If the number of deaths in the treatment and control arms (both of sample size 100) of a randomized study are 50 and 25, respectively, the  $RR = (50/100) / (25/100) = 2$ . The treatment group has a twofold relative risk of dying compared with the control group. The  $OR = (50/50) / (25/75) = 3$  indicates that the odds of death in the treatment arm is threefold of the control arm.

**Per-Protocol Analysis** A method of analysis in which only the subset of subjects who complied sufficiently with the protocol are included. Protocol compliance includes exposure to treatment, availability of measurements, correct eligibility, and absence of any other major protocol violations. This approach contrasts with the more conservative and widely accepted “intention-to-treat” analysis.

**Power** The probability of rejecting the null hypothesis (e.g., no treatment difference) when it is false. It is the basis of procedures for calculating the sample size required to detect an expected treatment effect of a particular magnitude.

**Random Error** An unpredictable deviation of an observed value from a true value resulting from sampling variability. It is a reflection of the fact that the sample is smaller than the population; for larger samples, the random error is smaller, as opposed to systematic errors (bias) that keep adding up because they all go in the same direction.

**Regression Analyses** Methods of explaining or predicting outcome variables using information from explanatory variables. Regression analyses are often used in clinical trials to estimate the adjusted treatment effect taking into account differences in baseline characteristics, and in epidemiological studies to identify prognostic factors while controlling for potential confounders. Commonly used regression models include linear, logistic, and Cox regression methods.

**Risk Factor** A risk factor can be defined as anything in the environment, personal characteristics, or events that make it more or less likely one might develop a given disease, reach an adverse event or experience a change in health status. For example, raised cholesterol is a risk factor for heart attacks.

**Standard Error** A measure of the random variability of a statistic (e.g., mean, proportion, treatment effect) indicating how far the statistic is likely to be from its true value. For example, standard error of the mean (SEM) indicates uncertainty of a single sample mean ( $\bar{X}$ ) as an estimate of the population mean ( $\mu$ ). A smaller SEM implies a more reliable estimate of the population mean. Standard error can be used to calculate confidence interval of an estimated population parameter. The

smaller the standard error, therefore, the narrower the confidence interval, and more precise the point estimate of the population parameter.

**Treatment Effect** An effect attributed to a treatment in a clinical trial, often measured as the difference in a summary measure of an outcome variable between treatment groups. Commonly expressed as difference in means for a continuous outcome, a risk difference, risk ratio, or odds ratio for a binary outcome, and hazard ratio for a time-to-event outcome.

**Univariate/Multivariate Analysis** The term variate refers to the term variable. A univariate analysis examines the association between a single variable and an outcome variable (correctly called a bivariate analysis), for example, age and occurrence of stroke. In a multivariate analysis, associations between many variables are examined simultaneously. In particular, multivariate regression analysis can be used to assess the relative importance and contribution of each predictor variable to the outcome variable. For example, a multivariate logistic regression can be undertaken to identify the most important prognostic factors among several risk factors (e.g., age, sex, systolic blood pressure, and cholesterol level) that predict the occurrence of stroke.

## 22.9 CONCLUDING REMARKS

We have introduced some basic statistical approaches to dealing with different types of data and samples. To apply these methods for data analysis, there are a few basic principles we should adhere to. We first need to make the purpose of the analysis clear. For example, is the analysis to compare groups or to estimate associations? Different purposes will require different methods. Second, we need to identify the nature of the data. Are they data from paired samples or two independent samples? Third, we need to examine the characteristics of the data. Are the variables qualitative or quantitative? Symmetrically distributed or not? These properties of the data will guide to use the correct assumptions for analysis. Important as the appropriate use of statistical methods is, we must also ensure that the data are clean, accurate, meaningful, and complete and were gathered reducing any biases occurring in the design, conduct, and analysis of the clinical trial. The advice of a statistician with experience in medical research should always be sought from the design through to completion and analysis phases of a clinical trial.

## REFERENCES

1. Pocock, S. J. (1983), *Clinical Trials: A Practical Approach*, Wiley, Chichester.
2. Chow, S. C., and Liu, J. P. (1998), *Design and Analysis of Clinical Trials: Concept and Methodologies*, Wiley, Chichester.
3. Everitt, B. S., and Pickles, A. (1999), *Statistical Aspects of the Design and Analysis of Clinical Trials*, Imperial College Press, London.
4. Wang, D., and Bakhai, A. (2005), Intention-to-treat Analysis, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 255–264.

5. Wang, D., Clayton, T., and Bakhai, A. (2005), Confounding, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 295–304.
6. Steele, F., and Wang, D. (2005), Regression analyses, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 273–286.
7. Wang, D., Clayton, T., and Yan, H. (2005), Significance tests and confidence intervals, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 185–196.
8. Wang, D., Clemens, F., and Clayton, T. (2005), Comparison of means, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 197–216.
9. Wang, D., Clayton, T., and Clemens, F. (2005), Comparison of proportions, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 217–234.
10. Nitsch, D., Wang, D., Clayton, T., et al. (2005), Multiplicity, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 329–338.
11. Jennison, C., and Turnbull, B. W. (2000), *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall/CRC, New York.
12. Altman, D. G. (1999), *Practical Statistics for Medical Research*, Chapman and Hall, London.
13. Zelen, M. (1998), Inference, in Armitage, P., and Colton, T., Eds., *Encyclopedia of Biostatistics*, Wiley, New York, pp. 2035–2046.
14. Kirkwood, B., and Sterne, J. (2003), *Essential Medical Statistics*, 2nd ed., Blackwell, Oxford.
15. Wang, D., Bakhai, A., and Gupta, A. (2005), Types of data and normal distribution, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 167–184.
16. Sterne, J. A., and Davey Smith, G. (2001), Sifting the evidence—what’s wrong with significance tests? *BMJ*, 322, 226–231.
17. Lee, A. F. S. (1998), Student’s *t* distribution and student’s *t* statistics, in Armitage, P., and Colton, T., Eds., *Encyclopedia of Biostatistics*, Wiley, New York, pp. 4396–4397.
18. Collett, D. (2003), *Modelling Survival Data in Medical Research*, 2nd ed., Chapman and Hall, London.
19. Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall, London.
20. Pocock, S. J., Clayton, T. C., and Altman, D. G. (2002), Survival plots of time-to-event outcomes in clinical trials: Good practice and pitfalls, *Lancet*, 359, 1686–1689.
21. Bland, J. M., and Altman, D. G. (1998), Survival probabilities (the Kaplan–Meier method), *BMJ*, 317, 1572.
22. Wang, D., Clayton, T., and Bakhai, A. (2005), Analysis of survival data, in Wang, D., and Bakhai, A., Eds., *Clinical Trials: A Practical Guide to Design, Analysis and Reporting*, Remedica, London, pp. 235–254.



# 23

---

## *Explanatory and Pragmatic Clinical Trials*

**Rob Herbert**

*The George Institute for International Health, Sydney, Australia*

### **Contents**

|                                                              |      |
|--------------------------------------------------------------|------|
| 23.1 Pragmatic and Explanatory Perspectives                  | 1082 |
| 23.2 Participants                                            | 1083 |
| 23.2.1 Inclusion and Exclusion Criteria                      | 1083 |
| 23.2.2 Run-In Periods                                        | 1084 |
| 23.3 Sample Size                                             | 1084 |
| 23.3.1 Criteria Used to Calculate Sample Size                | 1084 |
| 23.3.2 An Even More Pragmatic Approach: Value of Information | 1086 |
| 23.3.3 Adjusting Sample Size Calculations for Noncompliance  | 1087 |
| 23.4 Interventions                                           | 1088 |
| 23.4.1 Experimental Intervention                             | 1088 |
| 23.4.2 Comparison Intervention                               | 1089 |
| 23.5 Choice of Outcomes                                      | 1091 |
| 23.6 Analysis                                                | 1092 |
| 23.6.1 Who Should Be Included in Analysis?                   | 1092 |
| 23.6.2 Adjustment for Noncompliance                          | 1093 |
| 23.7 Reporting                                               | 1095 |
| 23.8 Recommendations                                         | 1096 |
| Acknowledgments                                              | 1096 |
| References                                                   | 1096 |

23.1    PRAGMATIC AND EXPLANATORY PERSPECTIVES

In 1967 Schwartz and Lellouch published an article in the *Journal of Chronic Diseases* (the precursor to the *Journal of Clinical Epidemiology*) titled “Explanatory and pragmatic attitudes in therapeutical trials” [1]. The article was significant because it clearly distinguished, for the first time, different philosophical attitudes of clinical trialists. Since that time the article has become a classic in the field of clinical trial methodology (the 2007 edition of *Web of Science* records 421 citations), and the ensuing discussion has stimulated development of several methodological advances.

Schwartz and Lellouch argued that clinical trialists can have explanatory or pragmatic attitudes (or perspectives, or orientations) to clinical trials. These attitudes influence many aspects of trial design. In the contemporary literature, explanatory trials may be called “efficacy trials” [2, 3]. Pragmatic trials may be called “effectiveness trials” [2, 3] or “practical clinical trials” [4] or, nearly synonymously, “large simple trials” [5].

Explanatory trialists conceive of clinical trials as being clinical analogs of classic laboratory experiments. The aim is to learn of the potential effects of the specific intervention. Explanatory trials are designed to establish “proof of concept” of a new intervention [3] (Table 1).

In contrast, pragmatic trialists think of clinical trials as a mechanism for informing clinical decisions. Pragmatic trials tell us about the importance of interventions in the real world of clinical practice. They are designed to inform clinical decisions about whether to implement a particular intervention.

We need both explanatory and pragmatic trials. It would appear sensible to first conduct explanatory trials on a particular intervention and then subsequently subject the same intervention to a pragmatic trial. An exception might be where health interventions become established clinical practice before they are subject to clinical trials. Arguably, the first trials of such interventions should adopt a pragmatic approach.

Trialists rarely make explicit statements, in trial reports, about whether their intentions are explanatory or pragmatic. Many trials incorporate some features of explanatory trials and some features of pragmatic trials. This suggests that at least some trialists do not clearly think through whether their intentions are explanatory or pragmatic.

**TABLE 1    Explanatory and Pragmatic Attitudes to Clinical Trials**

| Explanatory Attitudes                                                                                            | Pragmatic Attitudes                                                                                         |
|------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Aim is to learn about effects of an intervention.                                                                | Aim is to make a decision about whether to administer an intervention.                                      |
| Interest is in any effect.                                                                                       | Interest is in clinically important effects.                                                                |
| Ascertains if an intervention <i>can</i> produce effects (proof of concept).                                     | Ascertains if an intervention <i>does</i> confer effects.                                                   |
| Used to make inferences about specific effects of narrowly defined interventions given under optimal conditions. | Used to make inferences about the real-world effects of complex interventions in routine clinical practice. |



Those trialists who explicitly adopt a particular perspective may find themselves misunderstood by reviewers of grant applications or journals. Sometimes pragmatic trials are criticized for failing to satisfy requirements that are particular to explanatory trials and vice versa. Thus it appears that the issues raised by Schwartz and Lellouch over 40 years ago are still not fully appreciated by clinical trialists.

This chapter discusses some of the consequences of the trialist's perspective (explanatory or pragmatic) for design and analysis of randomized trials. The chapter concludes by making some recommendations for the design, analysis, and reporting of randomized trials.

## 23.2 PARTICIPANTS

### 23.2.1 Inclusion and Exclusion Criteria

One of the first issues that confront trialists, when developing a trial protocol, is which patients to include in the trial.

If the intention is to learn whether the intervention *could* be effective (the explanatory perspective), then a sensible course of action would be to recruit those patients in whom effects of intervention are most likely to be apparent. Usually, this involves sampling from a carefully selected and homogenous population. Careful formulation of inclusion and exclusion criteria can ensure that only participants with characteristics likely to be associated with large treatment effects are sampled, and homogeneous sampling can ensure that the variance of outcomes is minimized. Selective, homogenous sampling maximizes the possibility of detecting effects of intervention, but it may also exclude patients who might, in the course of normal clinical practice, be offered the intervention. People with comorbidities, pregnant women, and people who do not speak the local language are typically excluded from explanatory trials.

Pragmatic trialists seek to make inferences about the effects of an intervention on the population for whom the intervention is indicated. Thus pragmatic trials typically sample broadly. Typically, they have broad inclusion criteria and few exclusion criteria. (In some trials the primary inclusion criterion is simply that the relevant clinician might choose to offer the intervention and the participant might choose to accept it. An example of such a trial is the MRC Spinal Stabilisation Trial in which the primary inclusion criterion was that “patients who were candidates for surgical stabilisation of the spine were eligible if the clinician and patient were uncertain which of the study treatment strategies was best” [6].) Pragmatic trials often sample patients under the care of many clinicians and from multiple sites [4].

Ideally, pragmatic trials would do more than sample broadly. Theoretically, at least, pragmatic trials should sample in a representative way from the population about which inferences are to be made. Random sampling is usually difficult or impossible, so there are very few examples of trials that randomly sample participants. However, a small number of studies have made explicit attempts to sample as representatively as possible. Fransen and colleagues describe how they attempted to obtain a representative sample of general medical practitioners and patients in the DIAMOND study of two protocols for management of dyspepsia in primary care [7].

### 23.2.2 Run-In Periods

In some trials, patients are not randomized at the time they are recruited into the trial. Instead, they are asked to participate in a run-in period, during which their compliance with the intervention (or, more usually, with the placebo intervention) is monitored. Only those patients who demonstrate a sufficiently high level of compliance are subsequently randomized into the trial. Thus the run-in period serves to winnow out potential participants who are unlikely to comply with the intervention specified in the trial protocol.

The use of run-in periods offers several potential benefits to explanatory trialists. First, run-in periods can reduce loss to follow-up. Second, the use of a run-in period potentially increases compliance, enabling better explanatory estimates of effects of intervention (see Section 23.6). Third, simulation studies show that when there is both a high degree of noncompliance and the run-in period gives an accurate indication of who will and will not comply with the intervention protocol, the use of a run-in period can produce modest increases in statistical efficiency [8]. (That is, slightly fewer patients would need to be recruited to a trial with a run-in period to yield the same level of statistical power as a trial without a run-in period.) However, these benefits are not guaranteed: When compliance is high or the ability to discriminate between compliant and noncompliant patients is low, the statistical efficiency of trials with run-in periods may actually be lower than in trials without a run-in period [8]. Trials with run-in periods require a longer period of participation in the trial, which could paradoxically reduce participation rates and increase rates of noncompliance and loss to follow-up. Thus, while the use of run-in periods may seem attractive to explanatory trialists, they can be counterproductive.

The use of run-in periods is much less attractive to pragmatic trialists because pragmatists are much more tolerant of noncompliance. From the pragmatic perspective, noncompliance is as much a characteristic of the intervention as it is of the patient. If patients do not comply with an intervention, then the intervention is unlikely to be of any real practical effect, regardless of its explanatory effects. Insofar as a pragmatic estimate of the effect of an intervention depends on the degree of compliance with the intervention, a pragmatic trial of the intervention should involve patients with similar levels of compliance to the population about whom inferences are to be made. This means that, theoretically at least, pragmatic trialists should seek to recruit participants with representative levels of compliance. As discussed above, pragmatic trialists have rarely made explicit efforts to sample representatively, so it may be often true that pragmatic trials do not sample patients with representative levels of compliance.

## 23.3 SAMPLE SIZE

### 23.3.1 Criteria Used to Calculate Sample Size

The classic hypothesis testing approach to sample size calculation is to calculate the sample size required to provide adequate statistical power to detect an effect of intervention while keeping the probability of falsely claiming an effect to acceptably low levels.

A fairly general form of the equation for estimating sample size in two-armed clinical trials with equally sized groups is

$$n \geq 2(z_{1-\alpha/2} + z_{1-\beta})^2 \left( \frac{SD^2}{\delta^2} \right)$$

where  $n$  is sample size per group,  $z_{1-\alpha/2}$  is the  $z$  score corresponding to the probability,  $\alpha$ , of falsely rejecting the null hypothesis,  $z_{1-\beta}$  is the  $z$  score corresponding to the probability,  $\beta$ , of falsely retaining the null hypothesis,  $SD$  is the within-group standard deviation (usually assumed to be the same for both groups), and  $\delta$  is the smallest effect worth detecting [9].

This approach to sample size calculation is widely used but, because it is built on a hypothesis testing approach to statistical inference, and because hypothesis testing is becoming increasingly regarded with suspicion by methodologists, it has been strongly criticized (see, e.g., [10]). Here discussion of issues of sample size calculation is restricted to comparison of explanatory and pragmatic approaches to sample size calculation.

The explanatory trialist might wish to detect *any* effect of intervention, no matter how small. This is because the explanatory trialist's interest is in proof of concept rather than in whether the effect of intervention is large enough to make the intervention worth implementing in clinical practice. But it is not possible to design trials to reliably detect *any* effect because the sample sizes required to detect very small effects become impossibly large. The hypothesis testing approach requires the trialist to nominate  $\delta$ , the smallest effect that is worth knowing about, and calculate the sample size required to detect effects of at least that size. This is problematic for explanatory trialists because in explanatory trials the choice of  $\delta$  seems arbitrary. Consequently, sample sizes calculated for explanatory trials seem arbitrary too.

Pragmatic trialists find the task of nominating  $\delta$  more straightforward. Pragmatic trials are intended to inform decisions about whether or not to implement an intervention, so they should be designed to reliably detect effects of intervention that are just large enough to warrant the risks, costs, discomforts, and inconveniences of intervention [11, 12]. Recently, Barrett and colleagues have developed the benefit-harm trade-off method to establish patient-focused estimates of the smallest important effect of intervention [13–15]. This method can be used to obtain empirical estimates of  $\delta$ .

Paradoxically, while pragmatic trialists may find it easier to nominate  $\delta$  than explanatory trialists, it has been argued that it is not necessary for pragmatic trialists to nominate  $\delta$  [1, 16]. The argument is as follows. The pragmatic trialist does not care about falsely rejecting the null hypothesis because, from the pragmatic trialist's perspective, if the interventions given in the two arms of the trial are equally effective, it does not matter which of the two interventions is administered in clinical practice. Insofar as pragmatic trialists do not care about falsely rejecting the null hypothesis, they should always consider that the intervention with the larger observed effect is the more effective intervention, regardless of any considerations of sampling error. This strategy results in one of three possible scenarios:

1. The two interventions have identical effects, but the trialist claims one is more effective than the other.

2. One intervention is more effective than the other, and the trialist correctly identifies the more effective intervention.
3. One intervention is more effective than the other, and the trialist incorrectly concludes the less effective intervention is the more effective one.

According to Schwartz and Lellouch [1], only the last type of error is of any concern to the pragmatic trialist. They refer to the probability of this kind of error as  $\gamma$ . As the pragmatic trialist wishes only to minimize  $\gamma$ , such studies requires only a sample size of

$$n \geq 2(z_{1-\gamma})^2 \left( \frac{SD^2}{\delta^2} \right)$$

Even if  $\gamma$  is set at just 1%, the sample size calculated in this way will be nearly one-third less than with conventional calculations that assume  $\alpha = 0.05$  and  $\beta = 0.2$ . This observation has been used to defend the conduct of small pragmatic trials [17].

As MacRae has pointed out, Schwartz and Lellouch's approach to sample size calculation for pragmatic trials should only be used when the trialist is prepared to apply an intervention that is no more effective than alternative interventions [16]. In practice, that is often not the case because competing interventions often have quite different risks, inconveniences, and costs. For example, most surgeons would be disinclined to offer spinal stabilization surgery to patients with chronic low back pain if alternatives such as an intensive rehabilitation program have similar effects because the surgery is associated with significantly greater risks, discomfort, inconvenience, and costs than rehabilitation [6].

From a pragmatic perspective, the hypothesis testing approach to sample size calculation is quite unsatisfactory. Pragmatic trials should be designed to assist clinicians and health policymakers make decisions about whether the effects of particular interventions are big enough to be worthwhile. The most useful trials are those that provide precise estimates of the effects of intervention: The more precise the estimate, the more likely it will be that the decision maker will be able to conclude that the intervention definitely is or is not worth implementing [11, 12]. In practice, the hypothesis testing approach to sample size calculation often yields estimates of treatment effects that are too imprecise to be useful for clinical decision making. (That is, confidence intervals or credible intervals are usually so broad that they include both clinically worthless and clinically worthwhile effects.) This suggests an alternative approach to sample size calculation is needed. A better approach may be to explicitly specify the precision required of estimates of effect, and then estimate the sample size required to achieve that precision [18]. Admittedly, the choice of the acceptable precision will usually be quite arbitrary.

### 23.3.2 An Even More Pragmatic Approach: Value of Information

One response to the issues outlined above is to use decision-theoretic approaches to sample size calculation. Decision-theoretic approaches are explicitly pragmatic because they seek to maximize the difference between the cost of a trial and the expected value of the results.

Willan and Pinto have described a decision theoretic approach to sample size calculation in clinical trials [19]. They take a societal perspective, although the broad approach could be modified to take other perspectives such as the perspective of a corporate entity. They estimate the expected value of a trial (the expected value of sample information), the cost of conducting the trial, and the opportunity cost of treating some patients in the trial with the inferior intervention. All three quantities (the value of sample information and the cost and opportunity costs of the trial) are functions of sample size. The optimal sample size is that which maximizes the value of the trial minus its costs. This approach will often result in radically different estimates of sample size to those obtained with traditional methods. Indeed Willan and Pinto show that in some circumstances optimal trial size is zero, and the intervention should be implemented without first conducting a randomized trial.

Willan and Pinto's [19] approach to sample size calculation is attractive because it is rational and because it does not require the trialist to go through the arbitrary exercise of nominating  $\alpha$ ,  $\beta$ , and  $\delta$ . The disadvantage is that the calculations are complex and require that the trialist nominate other difficult-to-estimate quantities such as the number of patients who would, in the future, be eligible to receive the experimental or control interventions.

### 23.3.3 Adjusting Sample Size Calculations for Noncompliance

As noted above, explanatory trialists try to minimize noncompliance with intervention. From an explanatory trialist's perspective, participants who do not receive the experimental intervention cannot provide information about the effects of that intervention. In most trials, however, some degree of noncompliance is inevitable and explanatory trialists need to decide how to deal with it.

One strategy is to omit noncompliant cases in a "per protocol analysis." If the primary analysis is to be conducted per protocol (though it is not clear that it ever should—see Section 23.6), the sample size calculations should anticipate the loss of participants due to noncompliance and apply an appropriate correction. In the simplest scenario in which similar proportions of participants are randomly excluded from two equally sized groups, the sample size should be increased by a factor of  $1/(1 - \text{anticipated proportion excluded})$ .

A better alternative is to retain all participants in the analysis and apply a correction for attenuation of the explanatory effect by noncompliance (see Section 23.6). If the primary analysis is to involve obtaining estimates of explanatory effects of intervention calculated in this way, then the sample size should be adjusted accordingly. Newcombe [20] describes a method for adjusting conventional sample size calculations when an explanatory effect of intervention is to be estimated from the difference between two means, and Sato [21] provides sample size formulas for trials in which the explanatory effect of intervention is to be estimated as a risk difference. Sommer and Zeger tabulate the relative efficiency of pragmatic and explanatory estimates of risk ratios [22].

As discussed in Section 23.2, pragmatic trialists consider that noncompliance is a nonissue. Pragmatic trialists are uninterested in the hypothetical effect that an intervention could have in the presence of complete compliance. To them the parameter of most interest is the average effect of the intervention in clinical populations. It is understood that clinical populations contain some patients who do not

comply with intervention. For this reason pragmatic trialists generally do not adjust sample size estimates for noncompliance.

## 23.4 INTERVENTIONS

### 23.4.1 Experimental Intervention

Some health interventions are constrained by tight protocols that ensure uniformity of clinical practice. More often clinical practice is highly variable. Many clinicians have quite a high degree of autonomy, and they tend to develop their own formal or informal protocols for administering particular interventions. Individual clinicians may develop idiosyncratic decision rules that are not shared by other clinicians. In some settings a particular intervention may be administered in a technically adept manner using optimal protocols by clinicians who appear competent and demonstrate empathy. In other settings the same intervention may be administered tardily, incompetently, or indifferently.

The explanatory trialist wants to know about the effects of an intervention when it is administered optimally. Consequently, explanatory trialists expend considerable effort ensuring that the intervention is administered as well as is possible. Usually, the intervention is administered by a clinician with recognized expertise, and the intervention is administered according to strict protocols. This confers an incidental advantage: It ensures that it is clear precisely what intervention has been provided in the trial, so there need be little ambiguity about precisely what sorts of intervention the trial can be used to make inferences about.

The pragmatic trialist's interest is, instead, in the effects of the intervention as it is administered in real practice. (Or, when the trial is of a *new* intervention, the pragmatic trialist's interest is in what the effects of the intervention *would be* if it was to be administered in real practice.) Seen from this perspective, there is no necessity to enforce optimal intervention in clinical trials other than with the usual mechanisms that regulate the quality of clinical care. The interventions should be applied by clinicians who are representative of the spectrum of clinicians who apply (or would apply) the intervention in clinical practice, not exclusively by expert clinicians.

Useful pragmatic estimates of the effects of an intervention may be difficult to obtain when the intervention is new and administration of the intervention is difficult. An important example is new surgical techniques. The problem arises because new surgical techniques are pioneered by surgeons who have specific expertise in that type of surgery, but with time those techniques come to be more widely used by surgeons with less specific expertise. Moreover, the techniques themselves usually evolve rapidly, so that over the months and years that follow the introduction of the technique the standard approach becomes refined. Pragmatic interest should be in the effects of the intervention some time after the introduction of the new technique, at a time when the intervention has become, or could have become, part of routine surgical practice. By that time the technique may have been greatly refined, many more surgeons (including those with less expertise in the field) may be using the technique, and good training programs may have been introduced. A trial conducted soon after the introduction of a new surgical technique might neces-

sarily involve sampling unusually gifted surgeons who are relatively unpracticed in the technique, and might involve the administration of relatively unrefined surgical methods, compared to the same trial conducted several years later. Thus, while it may be possible to conduct explanatory trials seeking proof of concept soon after the introduction of a new surgical technique, it may be impossible to obtain realistic and generalizable pragmatic estimates of effect until the surgical technique has become widely established. Only then may it become possible to observe the effects of surgeons with widely representative competence performing refined surgical techniques.

#### 23.4.2 Comparison Intervention

The choice of comparison intervention should reflect the orientation of a trial. Explanatory trialists usually aim to determine the specific effects of an intervention. To them, nonspecific effects associated with the rituals of administration of the intervention and the interactions between the clinician and patient are usually of no interest and should not be allowed to contaminate estimates of the specific effects of the intervention. It is possible to estimate the specific effects of intervention by ensuring that both the experimental and control groups experience exactly the same nonspecific effects. That is, explanatory trialists can ensure that estimates of the effects of intervention are uncontaminated by nonspecific effects by ensuring that experimental and control interventions are “equalized” [1].

Experimental and control interventions can be equalized by providing a placebo (or sham) intervention to the control group. The ideal placebo intervention is one that involves identical nonspecific rituals of administration but lacks the specific effects of the experimental intervention. Thus most phase III drug trials compare the outcomes of a group of patients given pills (or injections, or ...) containing the experimental compound with the outcomes of a group given pills that are identical in shape, size, and color but that do not contain the experimental compound. Then both the experimental and control groups experience, on average, similar nonspecific effects associated with administration of the intervention. As a consequence, estimates of the effects of the intervention (the difference in the outcomes of the two groups) are believed to be uncontaminated by these nonspecific effects.

Similar reasoning is used to justify blinding of the clinician who administers the intervention. Blinding of the clinician ensures that, on average, the clinician provides the same degree of enthusiasm and positive regard when administering both interventions [2].

There is almost universal agreement that phase III drug trials should blind both the participants (patients) and the clinician administering the intervention. In contrast, there is much more ambivalence among researchers about blinding patients and clinicians in trials of other sorts of interventions. Trials of surgical interventions, medical devices, and programs of care such as educational interventions, exercise programs, and psychological or behavioural interventions frequently do not blind patients or clinicians. No doubt this is due, at least in part, to issues of feasibility: It can be difficult to design a sham intervention that is both comparable to the experimental intervention and therapeutically inert, ethical considerations may preclude the provision of a convincing sham, and it may be difficult to blind clinicians to whether they have administered the experimental or sham intervention. But there



are also theoretical objections to the use of sham interventions: Placebo controls are important in explanatory trials but may be unnecessary, or less necessary, in pragmatic trials.

To the pragmatic trialist, nonspecific effects are a potentially important component of the total effect of an intervention. Therefore, from the pragmatic trialist's perspective, there is no need to separately estimate specific and nonspecific effects. The pragmatic trialist needs only to know the total effect (the sum of nonspecific and specific effects) of administering the intervention. It is this effect that can be anticipated when the intervention is administered to patients in clinical practice. From this perspective there is no need to equalize exposure to nonspecific factors in experimental and control groups. Instead it is more important that the control group is offered a clinically relevant alternative intervention. In Schwartz and Lellouch's terminology, the goal of the pragmatic approach is to optimize rather than equalize the control intervention [1].

More generally, pragmatic trialists need to know whether their patients will experience better outcomes if prescribed the experimental intervention than if prescribed the best available alternative intervention. In general, then, pragmatic trials should compare outcomes of participants receiving the experimental intervention with the outcomes of participants receiving the best available alternative intervention. When the best alternative intervention is minimal intervention (such as advice and support), the most useful trial is one that compares the outcome of participants given the experimental intervention with the outcomes of participants that receives minimal intervention. When the best alternative intervention is a particular program of treatment, the most useful trial is one that compares the outcome of participants given the experimental intervention with the outcome of participants that receive that program of treatment. The principle is that pragmatic trialists should conduct trials with clinically reasonable (relevant) comparison groups [4]. For this reason pragmatic trialists should seriously consider *not* conducting placebo-controlled trials [23].

This advice needs to be qualified. First, the nonspecific effects of intervention may be context-specific. For example, it may be that a particular intervention has placebo effects that manifest in the context of a particular trial (where, e.g., there may be a particularly high degree of regard paid to the patient) but not in the context of broader clinical practice. Context-specific nonspecific effects are of little interest and should be controlled for. Also, the imperative to obtain clinically relevant comparisons may be outweighed by the need to prevent bias. In clinical trials, blinding of participants and clinicians serves additional purposes, over and above controlling for nonspecific effects associated with the rituals of administration of the intervention. At least theoretically, blinding could reduce differential loss to follow-up. More importantly, whenever outcomes are measured subjectively (as they often must be in pragmatic trials) blinding of participants prevents any possibility that the effects of treatment are biased by participants anticipating the responses that they believe the researchers want to hear (the "polite-patient effect" [12]). One of the greatest advantages of blinding of assessors in trials is that it means subjective outcomes can be evaluated without fear of assessor bias ("the clinical impression can be included and given full weight in the analysis" [24]). These important considerations may cause even the most pragmatic trialist to implement placebo controls in their trials.

### 23.5 CHOICE OF OUTCOMES

Unlike pragmatic trialists, explanatory trialists are not preoccupied with the issue of whether the effects of intervention justify adoption of the intervention in clinical practice; they simply want proof of the concept that the intervention produces certain specific effects. Thus explanatory trialists often use measures of disease activity or disease severity as the primary trial outcomes or endpoints. Examples include measures of tumor size, blood pressure, biochemical markers, muscle strength, bone density, and so on.

In the clinical trials literature these sorts of outcomes are often called surrogate outcomes [25, 26]. Trials that monitor surrogate outcomes are generally considered to be less useful than trials that monitor clinical outcomes such as mortality or quality of life because, as history has shown, surrogate outcomes can provide a misleading indication of the true clinical effects of intervention on mortality or quality of life [26].

The contemporary suspicion of surrogate outcomes is underpinned by two very pragmatic ideas. First, pragmatic decisions about whether or not to implement an intervention involve weighing up all of the potential good and all of the potential harm caused by the intervention. This is more easily achieved by examining global measures of benefit and harm than by assessing the effects of the intervention on specific measures of disease activity. Second, an intervention is ultimately only of clinical benefit if it increases length of life or quality of life. For this reason, the primary outcomes of pragmatic trials should be measures of mortality (survival) or quality of life.

The assessment of mortality is relatively straightforward. A minor controversy concerns which deaths should be counted. Explanatory trialists may prefer to measure disease-specific mortality because analyses of disease-specific mortality may have more statistical power, and therefore be more able to detect effects of intervention, than analyses of all-cause mortality [27]. Pragmatic trialists are more inclined to count all deaths, regardless of their cause, rather than disease-specific mortality. All-cause mortality provides a more global measure of the effects of an intervention because it allows for the possibility that reductions in disease-specific mortality are offset by increases in other causes of mortality [28].

Measurement of quality of life is a more complex task. Quality of life can be measured with generic scales or with any of numerous disease-specific quality-of-life measures [29]. Some trialists are reluctant to measure quality of life because they believe measures of quality of life are relatively insensitive to change and may obscure real effects of intervention. But such reluctance may not always be justified. The *measure* of quality of life must be sensitive to changes in the underlying construct (the latent variable of true quality of life) but, at least from a pragmatic perspective, there need not be any requirement that the underlying construct of quality of life should be sensitive to change.

Outcome measures can be considered to lie on a spectrum. At the explanatory end are measures of disease activity, like biochemistry results, and at the pragmatic end are mortality and quality of life. Somewhere between the ends of this spectrum are outcome measures that are not formal measures of quality of life but could reasonably be expected to reflect quality of life. Examples include measures of depression, or disability, or incidence of injury, or severity of incontinence. From an

extremely pragmatic perspective these measures are really only of interest insofar as they can be used to predict quality of life or costs. Nonetheless, because it is reasonable to expect that these sorts of outcomes are likely to be important determinants of quality of life, they may provide suitable endpoints for pragmatic trials.

Policy-level decisions about whether to implement a particular intervention should be informed by cost-effectiveness (or cost-utility) analyses. For this reason pragmatic trials often measure costs of intervention. Costs to the consumer and health service providers can usually be quantified with relatively little difficulty, although costs to society may be considerably more difficult to estimate. Cost-effectiveness and cost-utility analyses are naturally pragmatic in their orientation so they should always be based on pragmatic outcomes such as quality of life.

## 23.6 ANALYSIS

### 23.6.1 Who Should Be Included in Analysis?

Even the best conducted trials experience protocol violations. For example, participants in the experimental group may fail to be given, or comply with, the experimental intervention. Participants in the control group may fail to comply with the control intervention. Participants in either group may be given co-interventions other than those specified in the protocol. People who do not satisfy the inclusion criteria may be incorrectly admitted to the trial. Outcome measures may be obtained at times other than (usually later than) specified in the protocol.

Some trialists may be tempted to identify participants who experience serious protocol violations and exclude those participants' data from the analysis (a per protocol analysis). This strategy might be particularly attractive to explanatory trialists because it might appear that excluding noncompliant participants' data from the analysis would provide a better estimate of the effect of intervention in compliant participants. However, per protocol analyses are potentially seriously biased [30]. Insofar as protocol violations are related to allocation, exclusion of randomized participants can produce imbalance in observed or unobserved baseline covariates. The potential for bias is great when the proportion of excluded participants is moderate or large. In addition, because per protocol analyses are conducted on a subsample of data, they may be underpowered. Another problem is that inference in per protocol analyses is to a relatively poorly and retrospectively defined population.

For these reasons the primary analysis of most clinical trials should be conducted on all available data from all randomized patients. More generally, the analysis should be conducted on an "intention-to-treat" basis. Different meanings are ascribed to the term intention to treat [31], but here the term is used to mean that (a) wherever possible outcome data are obtained from all trial participants (i.e., from every person who was randomized, not just those for whom the trial protocol was followed), (b) all available data (not just data from participants for whom the protocol was followed) are included in analysis, and (c) each participant's data are analyzed in the groups to which the participant was assigned (not according to treatment actually received). In short, the intention-to-treat policy is one of "ever randomised, always analysed" [32].

There may be some circumstances under which departures from an intention-to-treat policy may be reasonable. Fergusson and colleagues [33] have argued that it is reasonable to exclude from the analysis those participants who were clearly ineligible to participate in the trial at the time they were randomized (i.e., people whose inclusion in the trial was a result of the researchers' failure to follow the protocol). On the other hand, participants should not be excluded from the analysis if the information required to determine they did not meet eligibility criteria only became apparent sometime after randomization. For example, participants who were thought to have a particular diagnosis, but who were found some time after randomization to have an alternative diagnosis that would have excluded them from the trial had the alternative diagnosis been known at the time of allocation, should not be excluded from the primary analysis. These recommendations have a distinctly pragmatic flavor.

### 23.6.2 Adjustment for Noncompliance

In the preceding section it was argued that per protocol analyses should not be used to obtain explanatory estimates of the effects of intervention. The primary analysis should be by intention to treat.

The intention-to-treat approach presents little difficulty for pragmatic trialists. Pragmatic trialists understand estimates of the effects of interventions as estimates of the effect of the *intention* to provide the intervention, or of a *policy* to provide the intervention, rather than in terms of the effect of the intervention itself. More subtly, pragmatic trialists appreciate that the estimate is of the effect of the intervention when it is given in good faith to the people for whom it is thought to be appropriate, even though the intervention may sometimes, for one reason or another, actually be given to patients for whom it may be inappropriate. Similarly, effects are considered to be those measured at as close as possible to a target follow-up time, understanding that in practice the timing of this effect may be variable in practice. An intention-to-treat analysis provides estimates of effects of intervention that can be understood in these pragmatic terms.

Explanatory trialists may, however, be unsatisfied by estimates of effects of intervention based on simple intention-to-treat analyses. Explanatory trialists wish to know the effects of actually receiving an intervention, as distinct from the effects of intending to provide the intervention. This interest may stem from a desire to understand how the effects of intervention are mediated or, alternatively, from a concern about the generalizability of pragmatic estimates of effects of interventions [34, 35]. In recent years there has been a rapid development of methods that can be used to adjust intention-to-treat-based estimates of the effects of interventions to take account of protocol deviations. The focus has been on obtaining intention-to-treat-based estimates of the effects of intervention in compliant patients.

The simplest method for obtaining compliance-adjusted estimates of effects of intervention from all randomized participants applies to the situation in which participants either complied fully with the intervention or received, from the moment of allocation, the intervention they were not allocated. Newcombe has shown that, in this special case, the effect of intervention is attenuated by an amount that depends only on the frequency with which participants receive the intervention they were not allocated to [20]. The explanatory estimate of the treatment effect is the

conventional pragmatic estimate divided by an attenuation factor, which is simply the proportion of subjects in one group receiving the allocated intervention minus the proportion of subjects in the other group receiving the same intervention. The attenuation factor is always less than or equal to one, so the explanatory effect is always greater than or equal to the pragmatic effect. The validity of this estimate is contingent on the assumption that participants complied perfectly either with the allocated intervention or the alternative intervention from the time of randomization (i.e., it assumes participants did not “cross-over” some time after they had already experienced some of the effect of the allocated intervention or too late to experience some of the effect of the intervention they crossed over to). Insofar as participants cross-over after they have already experienced some of the effect of the allocated intervention, or too late to experience some of the effect of the intervention they crossed over to, this explanatory estimate will be biased upward.

Pocock and Abdallah describe a number of simple analyses they conducted to obtain explanatory estimates of the effects of intervention from a placebo-controlled trial of dexfenfluramine for obesity [36]. As they had data on the degree of compliance, they were able to regress outcome against compliance for participants in the experimental group. The regression equation was used to estimate the outcome of perfect compliers in the experimental group. An explanatory measure of the effect of the intervention could then be obtained by subtracting from this value the mean outcome for the control group. (A better approach, though one that was not possible with the data available in that trial, would be to subtract the estimated outcome of perfect compliers in the control group from the estimated outcome of perfect compliers in the experimental group.) A difficulty with this approach is that compliance may be both a determinant and an effect of outcome [37]. Comparison of outcomes of treated and control groups at the same level of compliance are potentially biased when the treatment and control conditions have different influences on compliance.

Sommer and Zeger circumvented this problem in their analysis of a trial of vitamin A supplementation on child mortality [22]. Compliance was treated as a dichotomous variable, so participants were considered to be either compliant or noncompliant. The approach was to infer virtual compliance levels in the control group by assuming that compliance levels in the control group were the same in the control and experimental groups. Unbiased estimates of the effects of intervention in compliant participants could be obtained by comparing mortality risk in the subgroup of compliant participants in the experimental group with the inferred mortality risk in the virtual subgroup of compliant participants in the control group.

Little and Rubin show that both Newcombe’s approach and Sommer and Zeger’s approach can be derived with causal models [38]. Causal modeling provides a rigorous framework for development of explanatory estimates of effects of intervention. The fundamental idea is that the explanatory effect of an intervention on any individual is the difference between the outcome that individual would have experienced with and without intervention. The obvious statistical problem is that, in a clinical trial, each individual either receives the intervention or does not. Nonetheless, given some often plausible assumptions, it is possible to estimate the average causal effect in compliant people. Causal models have been subject of active development in recent years. Bellamy and colleagues provide an introduction to the use

of causal models for obtaining explanatory estimates of effects of intervention in randomized trials [39].

There are two distinctly different ways of conceiving of explanatory effects of intervention. One possibility is to think of the explanatory effect of intervention in terms of the effect of intervention on the subpopulation of patients who comply (the conditional compliance-related treatment gain [35]). Alternatively, the explanatory effect of intervention could be thought of as the effect intervention would have had if all patients had complied with the intervention (the unconditional compliance-related treatment gain [35]). Estimation of these two quantities involves invoking different assumptions and involves different types of statistical models [39].

It goes without saying that any attempts to estimate the effects of intervention in compliant people can only be conducted if good data are available on compliance. Compliance is a complex construct that includes dimensions of consistency, intensity, and sustainability. Noncompliance may manifest as missed treatments, treatments in which the intervention was administered at an inappropriately high or low dose, or failure to continue with treatment. Unfortunately, it is often difficult to obtain good data on all these dimensions. Measurement of compliance is most often based on pill counts, patient diaries, and participants' self-reports [40], all of which may be biased. Usually, compliance data are reduced to a simple dichotomy, so that participants are classified as compliant or noncompliant.

## 23.7 REPORTING

The primary theme of this chapter has been that the design of a clinical trial should be strongly influenced by whether the intention is explanatory or pragmatic. Explanatory trials and pragmatic trials differ in the way they sample participants, administer the intervention and control interventions, measure outcomes, and analyze data. Authors of clinical trials should state explicitly, both in trial protocols and trial reports, whether their intention was explanatory or pragmatic. This will go some way to reducing misunderstandings that arise from the different intentions of trialists and readers of trial reports.

Pragmatic trialists are more interested than explanatory trialists in generalizing their findings to clinical contexts. (Perhaps a more accurate way to say this is that pragmatic trialists are more concerned that their findings can be used to make inferences about the care of individual patients.) The extent to which this is possible depends on the extent to which the characteristics of the trial participants and trial interventions resemble the participants and interventions about which inferences are to be made. Consequently, it is particularly important that reports of pragmatic trials provide detailed information about the characteristics of trial participants (patients and clinicians) and how the intervention was administered. Specific details should be provided about the strategies used to facilitate compliance with the intervention.

The conclusions of trial reports should be consistent with the design of the trial. Importantly, authors of explanatory trials should restrict their conclusions to claims about proof of concept. In the event that an explanatory trial provides evidence in support of the specific effects that were hypothesized, the conclusions should state just that. Explanatory trials should not make global claims about the effectiveness

of an intervention. Such claims should await demonstration of clinically important effects with pragmatic trials.

### 23.8 RECOMMENDATIONS

1. Clinical trials of interventions that are already in widespread practice should usually be pragmatic in orientation.
2. Run-in periods should not be used in pragmatic trials.
3. Pragmatic trials should sample participants (patients and clinicians) in as representative a way as is possible.
4. If sample size calculations for pragmatic trials are to be based on a hypothesis testing approach, they should incorporate empirical estimates of the smallest worthwhile effect of intervention, such as can be obtained with the benefit-harm trade-off method. Alternatively, sample size calculations for pragmatic trials could be based on considerations of the precision of estimates of effect, or of the value of sample information.
5. If the primary analysis of an explanatory trial is to involve adjustment of estimates for noncompliance, sample size should be adjusted accordingly.
6. Explanatory trials should blind participants and clinicians wherever possible. Pragmatic trialists should consider not blinding participants or clinicians.
7. In pragmatic trials the control group should receive the best available treatment alternative.
8. Pragmatic trials should have as endpoints either survival or measures that reflect quality of life or both.
9. The primary analysis of all trials, whether explanatory or pragmatic, should be conducted on an intention to treat basis.
10. Reports of clinical trials should explicitly indicate if the intention is to answer an explanatory or pragmatic question.
11. Pragmatic trials should provide detailed information about the characteristics of trial participants (patients and clinicians), how the intervention was administered, and the degree of compliance with the intervention.
12. Explanatory trials should not make global claims about the effectiveness of an intervention. Such claims should await demonstration of clinically important effects with pragmatic trials.

### ACKNOWLEDGMENTS

The author acknowledges helpful comments on the manuscript made by Luciana Machado and Lisa Harvey. He is supported by a fellowship from the Australian NHMRC.

### REFERENCES

1. Schwartz, D., and Lellouch, J. (1967), Explanatory and pragmatic attitudes in therapeutic trials, *J. Chronic Dis.*, 20, 637–648.



2. Nathan, P. E., Stuart, S. P., and Dolan, S. L. (2000), Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychol. Bull.*, 126, 964–981.
3. Depp, C., and Lebowitz, B. D. (2007), Clinical trials: Bridging the gap between efficacy and effectiveness, *Int. Rev. Psychiatry*, 19, 531–539.
4. Tunis, S. R., Stryer, D. B., and Clancy, C. M. (2003), Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy, *JAMA*, 290, 1624–1632.
5. Yusuf, S., Collins, R., and Peto, R. (1984), Why do we need some large, simple randomized trials? *Stat. Med.*, 3, 409–422.
6. Fairbank, J., Frost, H., Wilson-MacDonald, J., et al. (2005), Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: The MRC spine stabilisation trial, *BMJ*, 330, 1233.
7. Fransen, G. A., van Marrewijk, C. J., Mujakovic, S., et al. (2007), Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND study, *BMC. Med. Res. Methodol.*, 7, 16.
8. Brittain, E., and Wittes, J. (1990), The run-in period in clinical trials. The effect of misclassification on efficiency, *Control. Clin. Trials*, 11, 327–338.
9. Julious, S. A. (2004), Sample sizes for clinical trials with normal data, *Stat. Med.*, 23, 1921–1986.
10. Rothman, K. J., and Greenland, S. (1998), *Modern Epidemiology*, 2nd ed., Lippincott-Raven, Philadelphia.
11. Herbert, R. D. (2000), How to estimate treatment effects from reports of clinical trials. I: Continuous outcomes, *Australian J. Physiother.*, 46, 229–235.
12. Herbert, R. D., Jamtvedt, G., Mead, J., et al. (2005), *Practical Evidence-Based Physiotherapy*, Elsevier, Oxford.
13. Barrett, B., Brown, R., Mundt, M., et al. (2005), Using benefit harm tradeoffs to estimate sufficiently important difference: The case of the common cold, *Med. Decis. Making*, 25, 47–55.
14. Barrett, B., Brown, D., Mundt, M., et al. (2005), Sufficiently important difference: Expanding the framework of clinical significance, *Med. Decis. Making*, 25, 250–261.
15. Barrett, B., Harahan, B., Brown, D., et al. (2007), Sufficiently important difference for common cold: Severity reduction, *Ann. Family Med.*, 5, 216–223.
16. MacRae, K. D. (1989), Pragmatic versus explanatory trials, *Int. J. Technol. Assess. Health Care*, 5, 333–339.
17. Powell-Tuck, J., MacRae, K. D., Healy, M. J., et al. (1986), A defence of the small clinical trial: Evaluation of three gastroenterological studies, *BMJ*, 292, 599–602.
18. Beal, S. L. (1989), Sample size determination for confidence intervals on the population mean and on the difference between two population means, *Biometrics*, 45, 969–977.
19. Willan, A. R., and Pinto, E. M. (2005), The value of information and optimal clinical trial design, *Stat. Med.*, 24, 1791–1806.
20. Newcombe, R. G. (1988), Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur, *Stat. Med.*, 7, 1179–1186.
21. Sato, T. (2000), Sample size calculations with compliance information, *Stat. Med.*, 19, 2689–2697.
22. Sommer, A., and Zeger, S. L. (1991), On estimating efficacy from clinical trials, *Stat. Med.*, 10, 45–52.
23. Vickers, A. J., and de Craen, A. J. (2000), Why use placebos in clinical trials? A narrative review of the methodological literature, *J. Clin. Epidemiol.*, 53, 157–161.

24. Bradford Hill, A. (1961), *Principles of Medical Statistics*, 7th ed. London: The Lancet.
25. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1998), *ICH harmonised tripartite guideline*. Statistical principles for clinical trials.
26. Biomarkers Definitions Working Group (2001), Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
27. Yusuf, S., and Negassa, A. (2002), Choice of clinical outcomes in randomized trials of heart failure therapies: Disease-specific or overall outcomes? *Am. Heart J.*, 143, 22–28.
28. Black, W. C., Haggstrom, D. A., and Welch, H. G. (2002), All-cause mortality in randomized trials of cancer screening, *J. Natl. Cancer Inst.*, 94, 167–173.
29. Guyatt, G. H., Feeny, D. H., and Patrick, D. L. (1993), Measuring health-related quality of life, *Ann. Intern. Med.*, 118, 622–629.
30. Lee, Y. J., Ellenberg, J. H., Hirtz, D. G., et al. (1991), Analysis of clinical trials by treatment actually received: Is it really an option? *Stat. Med.*, 10, 1595–1605.
31. Hollis, S., and Campbell, F. (1999), What is meant by intention to treat analysis? Survey of published randomised controlled trials, *BMJ*, 319, 670–674.
32. Cox, D. (1998), Discussion, *Stat. Med.*, 17, 387–389.
33. Fergusson, D., Aaron, S. D., Guyatt, G., et al. (2002), Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis, *BMJ*, 325, 652–654.
34. Mark, S. D., and Robins, J. M. (1993), A method for the analysis of randomized trials with compliance information: An application to the Multiple Risk Factor Intervention Trial, *Control. Clin. Trials*, 14, 79–97.
35. Goetghebuer, E. J., and Shapiro, S. H. (1996), Analysing non-compliance in clinical trials: Ethical imperative or mission impossible? *Stat. Med.*, 15, 2813–2826.
36. Pocock, S. J., and Abdalla, M. (1998), The hope and the hazards of using compliance data in randomized controlled trials, *Stat. Med.*, 17, 303–317.
37. Rochon, J. (1995), Supplementing the intent-to-treat analysis: Accounting for covariates observed postrandomization in clinical trials, *J. Am. Stat. Assoc.*, 90, 292–300.
38. Little, R. J., and Rubin, D. B. (2000), Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches, *Ann. Rev. Pub. Health*, 21, 121–145.
39. Bellamy, S. L., Lin, J. Y., and Ten Have, T. R. (2007), An introduction to causal modeling in clinical trials, *Clin. Trials*, 4, 58–73.
40. Gossec, L., Tubach, F., Dougados, M., et al. (2007), Reporting of adherence to medication in recent randomized controlled trials of 6 chronic diseases: A systematic literature review, *Am. J. Med. Sci.*, 334, 248–254.

# 24.1

---

## *Ethics of Clinical Research in Drug Trials*

**Roy G. Beran**

*Strategic Health Evaluators, Chatswood NSW 2067 Australia*

### **Contents**

|                                     |      |
|-------------------------------------|------|
| 24.1.1 Foundations for Trial Design | 1099 |
| 24.1.2 Investigator Selection       | 1101 |
| 24.1.3 External Study Approval      | 1102 |
| 24.1.4 Consent Process              | 1103 |
| 24.1.5 Conduct of Trial             | 1105 |
| 24.1.6 After Study Completion       | 1106 |
| 24.1.7 Conclusion                   | 1107 |
| References                          | 1107 |

### **24.1.1 FOUNDATIONS FOR TRIAL DESIGN**

The ethics of clinical drug trialing start with the concept of the trial that is to be undertaken. The trial must aim to answer a question that has not yet been answered, thereby necessitating the need for such a trial [1]. It should not be a substitute for a marketing strategy in which investigators place patients onto expensive agents for which either the patient or the formulary (namely the list of drugs available within a specific facility) cover the real costs of the trial, while the yield from the study contributes very little to scientific/clinical knowledge and more to treatment commitment [2].

The latter group of studies have a distinct advantage for the pharmaceutical company that is sponsoring them because they represent a pool of patients who would, otherwise,

possibly not receive the medication, were it not for the conduct of the clinical trial [3]. These studies generally lack scientific validity, and investigators need to think very carefully before committing to such questionable trials. Ethics committees also need to review the content of studies to make sure that they are not merely marketing tools masquerading as clinical trials.

The design of a study should include calculations to determine the sample size necessary to allow sufficient mathematical power to achieve a reliable response with adequate scientific rigor [4].

There needs to be consideration of the size of the response expected, the right dosage needed to attract such a response, the correct comparator (be it placebo, which theoretically equates to no comparator, or a proven therapeutic agent already known to be effective in the management of the condition under review), and the potential subject “dropout rate” that needs to be accommodated to achieve the desired final sample size upon which to base the study’s analysis with adequate power [5].

It is often impossible to determine these parameters accurately as they may rely on earlier animal studies for extrapolation, and the animal data may not properly translate into human experience [6]. Alternatively, the concept may require an initial pilot study, in which far fewer subjects are involved but in which either the idea or the methodology under investigation can be tested. Such pilot studies may not truly reflect definitive data, but they will provide invaluable evidence to either support or refute the initial hypothesis [7, 8].

Pilot trials may be the only source of information upon which to base a foundation with which to ascertain indicative information to be used as the basis for calculations to determine a subsequent study that allows adequate power to truly answer the scientific question being posed. Often such pilot studies are difficult to publish as they lack the power to answer the question reliably, but their value should not be underestimated nor should the ethics of publication of such information be dismissed as they may represent the only proof of concept information [9].

The dosage of medication to be used in the trial may prove critical. Studies that have used too large a dosage cause unnecessary adverse events and hence delay or impede a drug’s development or acceptance, as was the case with topiramate, which used a dosage of 200mg in its trials [10, 11]. Conversely, the dosage that was used within the trial may be too low to achieve adequate response rate, thereby suggesting the agent to be less effective than it truly is, as was found with gabapentin [12]. Such errors have direct negative implications for the target patient population and may deny patient access to what amounts to an effective therapy because the study design was formulated on false expectations. It can be seen that innocent, but incorrect, assumptions may carry with them ethical consequences that may ultimately affect patient well-being because they affect the availability of appropriate therapeutic modalities [13].

The cost of conducting a trial has direct implications upon the adopted methodology. A crossover design is far less expensive than is a parallel arm design, and the purpose of the study must be decided. If the purpose of the study is to satisfy a regulatory requirement, then a parallel arms study design may be mandated by those regulatory authorities that are to determine the fate of the product under review [14].

The reason for reduced costs, in a crossover design, is that the subject acts as his or her own control, thereby allowing a significantly reduced sample size, as one patient serves in both arms of the trial to answer the study's question, which may effectively half the necessary sample size. This obviously reduces both the time delay and the effort it takes to recruit double the sample size and hence also reduces the need for additional investigators with inherent monitoring and infrastructure costs. Thus the purpose of the study may determine the study design, beyond any strict scientific or ethical requirements.

Where there is a need to have large numbers in a trial, then it often follows that the trial sponsors will resort to cost-cutting measures that may impact on the scientific rigor adopted for the protocol. An example of this is the use of large open-label studies to answer questions of comparative efficacy where the cost of a double-blinded methodology would be prohibitive due to the need for a "double-dummy approach" (in which both treatment A and treatment B need to have mirrored therapy with matched placebos such that the patient receives either treatment A, plus placebo, or placebo plus treatment B, thereby offering a potentially truly double-blind methodology) [15, 16]. The costs of such an approach might mean that the study would never have been feasible while the open-label design may still provide some very useful information.

The need to have large numbers recruited to a study may also translate into the inclusion of less skilled investigators, with far greater propensity for protocol violation and hence possible pollution of data. It is hoped that such potential problems do not translate into ethical expedience with insufficient external scrutiny of procedure.

It can be seen that the ethics of clinical research, and drug trialing, commence long before the first investigators are approached, the first study site is selected, or the first patient is recruited into the trial. The consideration of ethics starts with the selection of the scientific question to be asked, the capacity to answer that question, and the study design to be adopted for that purpose [17].

### **24.1.2 INVESTIGATOR SELECTION**

For an investigator to accept to be part of a clinical trial requires that the investigator be confident that he or she can fulfill the requirements of the trial. This dictates that the investigator conducts a review of the purpose of the study, the question(s) to be asked in the trial, the methodology to be adopted, the types and numbers of patients sought for inclusion (reviewing inclusion and exclusion criteria), and the feasibility of conducting the study within the constraints of his or her center, while recognizing that all conduct within the trial must adhere to good clinical practice [18].

Many international trial sponsors will conduct a feasibility study of proposed investigators, but, as these are based on personal estimations, their validity is quite suspect [19]. Despite this shortcoming, the concept that each proposed investigator should at least assess his or her capacity to adequately conduct such a trial is very appropriate.

It is inappropriate, and ethically unjustified, to nominate oneself for inclusion as a prospective investigator within a trial if it is highly doubtful that one could accom-

modate the proposed trial's needs [20]. If selected as an investigator, the study sponsor automatically accumulates direct costs: such as investigator meetings, with its transportation and accommodation costs, costs for ethics committee submissions, and infrastructure costs [21].

There are basic considerations applicable to each nominated investigator. These include having adequate knowledge of the subject matter under consideration, having the resources to conduct the trial (be it space, staff, storage facilities, or standard operating and accounting procedures), having access to an acceptable and properly constituted ethics committee, and expecting to be able to recruit adequate numbers of appropriate patients to justify inclusion [22].

Without these ingredients, the proposed investigator should exclude him or herself.

The purpose of this discussion is to demonstrate that there are unequivocal ethical obligations on the investigator even before agreeing to be part of any investigative team. These obligations are often overlooked but to do so may cause unnecessary burden to a prospective sponsor. This may cause a sponsor to think twice about underwriting future studies, if investigators failed to meet their agreed commitments.

### 24.1.3 EXTERNAL STUDY APPROVAL

Once the aims of the study have been determined, the protocol established, and investigator sites selected, the study must then be subjected to external scrutiny.

All those who are involved with the conduct of the study (namely the company initiating the trial, the sponsor, the clinical research organization that will supervise the study, and the investigators) must review all the requirements [23], which may include contractual negotiations, privacy agreements, financial incentives, and infrastructure requirements, all of which should be available to the external review process as required by the relevant committees.

Submission to review boards, be they for scientific validity or for human research ethics committee (HREC) evaluation, are usually prepared in a predetermined fashion. The purpose of the HRECs is to protect the interests of trial subjects. Some, but not all, HRECs ask for a relevant expert scientific evaluation of a proposed protocol, before deliberating on the trial, so as to ensure that it is seeking valid scientific answers within the study design. This process should ensure that the study is not simply a marketing ploy. Such appraisal also seeks to confirm the feasibility of a study, its merits, and any potential risk that may, or may not, have been identified at the time of the submission. Some HRECs rely on their own members, which might include the relevant scientific expertise to provide the necessary scientific input. In either context, the role of the HREC is less to do with the scientific value of the proposed research (for which it can rely on others), and its main function is to focus its principal deliberations on the ethics of patient involvement [24].

The HREC reviews the protocols, the reasons for the study, patient safety, access to necessary information, format, content, and language of both patient information and consent documentation, access to investigators in case of any problems, and clear evidence of lack of coercion when recruiting patients [25].

This is of particular importance when the sample population is recruited from the clinical practice of an investigator, be it in the private or public practice of medicine. Patient protection is the fundamental obligation of the HREC [25], while at the same time it should not obstruct worthwhile research that may benefit the wider community.

Clinical trials can expose research candidates to new, and often unknown, risks when consenting to be subjects for a trial. To protect patients against such risks, the sponsors are expected to take out insurance to cover those risks should they eventuate [26]. More recently, members of HRECs have also sought insurance to indemnify their membership against potential litigation consequent to any adverse events that may occur within a trial. The HREC will often seek confirmation of appropriate insurance being in place before approving any study that involves human participation.

Studies that include live subjects (human or animal) cannot proceed without ethics committee approval, and, in the case of humans, a written consent or signed surrogate consent is needed from either the patient or someone responsible. Surrogate consent is “ideally a substituted judgment made by a person responsible for healthcare decision-making for a particular patient under the relevant legislation” [27]. The situations in which such surrogate consent may be required include evaluation of therapy for certain conditions, such as treatment in stroke patients for whom the capacity to decide is no longer valid. It may also be necessary to seek surrogate consent in underage subjects for whom guardian approval is mandatory. A similar situation exists for patients who lack the intellectual capacity to comprehend what is being asked of them within the trial protocol, as is the case with intellectually handicapped patients. For these patients it is also necessary to seek guardianship board approval, so as to relegate the authority to consent for such a patient to be included within any given trial, to a “responsible other” person who will have the patient’s best interests as his or her primary concern [28].

#### 24.1.4 CONSENT PROCESS

“Consent is widely believed to affirm the autonomous decision-making rights of prospective research participants, and their capacity to protect their own interests” [27]. Once the study has been approved by the external review process, each investigator assumes responsibility for the conduct of the trial within his or her institution, be it in private or public practice. This starts with a site initiation visit in which the sponsor, or surrogate, usually a clinic, also termed contract research organization (CRO), which is specifically contracted for the task, confirms that all those at the site are fully conversant with their obligations and the procedures to be adopted within the trial and that the center has the wherewithal to conduct the study [29].

With the emergence of CROs there has also emerged an understanding that they too have unequivocal ethical obligations. The level of sophistication impacting upon CROs have developed to the point of these being umbrella organizations designed to oversee compliance of its voluntary membership [30].

The investigator(s) then approaches suitable candidates regarding their willingness to participate within the trial. Each candidate must provide informed consent, based on provision of HREC approved study information, which clearly outlines



the purpose of study, the obligations imposed on both subjects and investigators, and the potential known risks and hazards that accrue consequent to participation in the trial. Each subject must have exercised free will, devoid of the imposition of undue influence, and must have the capacity to ask any questions that may have arisen, following the reading of the study information, before signing the informed consent document [22].

This process must be conducted in an open and unbiased fashion and the written consent should be witnessed and signed by a competent adult, recognizing that the written consent may be provided by a surrogate, representing the patient's best interests. Some may allow for the lack of equipoise in the doctor–patient relationship and have the trial coordinator initiate the consent process, following which the doctor will be available to answer any outstanding questions [26, 31].

It is imperative that the investigator must still be available to answer more difficult questions (which were beyond the capacity of the trial coordinator) and that investigator should countersign the consent form and confirm its authenticity [26, 31].

This goes some way to overcome any suggestion that patients, selected from the clinical practice of an investigator, have been coerced into participating in the trial. It is far easier to refuse participation within a trial if the request comes from a research coordinator rather than the doctor, especially if the HREC-approved study information includes a passage that guarantees ongoing clinical care by the doctor, irrespective of whether the prospective subject is willing, or not, to participate in the trial [26, 31].

The consent documents must acknowledge that trials of new therapeutic agents may not have encountered all possible risks and hazards that attach to the medication under experimentation. Potential subjects must be warned of the known risks and must also be advised that no one can guarantee that the trial will provide a direct benefit for them. Where a trial compares an investigative agent to a placebo, trial subjects must understand that they may not receive the medication under review, within the trial, and may receive the equivalent of a “sugar pill.” They also need to be appraised of the risk ratio of this happening [29]. Similar advice needs to be given where differing doses of the study medication are to be given so that consent is truly informed. In short, the study protocol must be adequately described within the patient information and consent documents to clearly demonstrate that there has been no attempt to hide vital information or to give a false risk–benefit ratio [29]. The HREC has the responsibility to ensure that this has occurred and that the language used is of a sufficiently simple nature to avoid excessive use of jargon that may confuse a prospective subject. It must also be satisfied that the concepts, including the consent of information documents, are provided in a manner that is not too complicated for the average man/woman to understand [27].

Many trials of new medications adopt a blinded methodology, be it single or double blinded, and the potential subject must appreciate that he or she will not know what medication he or she will receive in a blinded study [15]. It is even more important for the subject to understand that, in the double-blinded methodology, neither the patient nor the doctor knows which medication the patient is taking during the trial. As mentioned earlier, the trial may use a placebo as the comparator. In such trials subjects will not know if they are receiving an active compound at all, in addition to the standard medications they were receiving before the trial. This

encapsulates the very essence of a randomized, placebo-controlled, double-blinded protocol. Most placebo-controlled studies do make provision for the patient to receive the experimental compound in an open-labeled, long-term safety and efficacy trial that follows the blinded study and that continues until the study medication is freely available. This must also be explained to potential trial subjects [23, 32].

Only subsequent to the free giving of informed consent, based on full disclosure, can the patient be included within the trial procedures [33]. Part of the consent process also includes provision for the withdrawal of consent, at any time throughout the study, without fear of compromising future ongoing clinical care by the doctor [33].

#### 24.1.5 CONDUCT OF TRIAL

There is usually very little latitude to alter the conduct of a trial. All requirements and expectations of a trial are set out in the HREC-approved trial protocol, which stipulates the timing and conduct of the trial visit [22]. Unfortunately, all trials that include human subjects will also include errors made by either the investigating team or the subjects themselves. Some of these errors represent only minor infringements for which a waiver may be sought from the sponsor [29]. Other errors may be so significant as to dictate immediate exclusion and withdrawal from any further trial involvement. Should a waiver be approved, then the patient can continue in the trial and allowance is made for the deviation from trial protocol requirements.

If the patient experiences a serious adverse event (SAE), such as the occurrence of something that causes hospitalization or is life-threatening, then both the HREC and the sponsor need to be advised, as soon as is possible, following the investigator having been made aware of the SAE [34]. The necessary action that is to be taken, regarding the experimental agent, is at the discretion of the investigator, although everyone needs to be advised of the action taken. Depending on the nature of the SAE, and its cause, it may be necessary to advise all patients within a trial of the SAE, and it may also be necessary to amend the patient information and informed consent documents. The HREC, itself, may insist on this action and it may go so far as to terminate the trial if its members believe that continuation of the trial is no longer ethical or in the patients' best interests. Investigators must make the same considerations, and advise the HREC accordingly, thereby demonstrating that the investigator also holds the well-being of the trial subjects as his or her primary concern for continuation of any given trial [34].

Others who are trialing the same medication—either for a similar indication or an alternatively indication—and who encounter an SAE—must also notify their appropriate authorities, such as their HREC, sponsor, or CRO. They, meaning the sponsor or CRO, must notify all other investigators who are/were trialing the same agent. Once notified of these SAEs encountered by others, it is imperative that the investigator also bring these to the attention of their HREC in the same fashion as they would have done had they personally encountered the SAE [29]. This further ensures maintenance of informed consent, should any of the SAEs have a bearing on the decision to continue a trial, be it from the perspective of the investigating team or subject.

Based on the SAE reporting, the HREC may seek clarification to determine if: the trial should continue with its currently recruited subjects; if continuation of recruitment should be allowed, if the entire process should be aborted, or if there needs to be an amendment to the current informed consent and information documents; and if the informed consent process needs to be repeated with each subject who is already within the trial to expedite their withdrawal should they choose not to face further previously unforeseen risk [29].

Depending on the nature of an adverse event, there may be need to unblind the blinded randomized data to determine whether the patient was exposed to the experimental compound and, if so exposed, the dosage of the exposure. Again this is at the discretion of the investigator, but it remains imperative that the patient's well-being must assume primal consideration [34]. This takes priority over any secondary ethical obligation that the investigator may have to the sponsor of the study to adhere to trial protocol as it relates to blinding of treatment arms.

Upon completion of the trial, the subject may either elect to continue on the study agent, if an open-label study is to follow, or choose to stop whatever it was that they were taking during the trial [23]. It must be remembered that trial subjects may not be advised of what they were taking during a double-blinded study, even after they complete that part of the trial, if not all subjects have completed the study. This is considered important to maintain mathematical integrity when analyzing the study data, but it does raise important ethical considerations as the decision to continue within the open-label arm is often based on inadequate information, in the absence of knowing what was taken during the trial. Some patients may think that they had a real benefit from the study medication when in fact they were on placebo, while others who were on the study medication may have thought they were on placebo [35]. Thus far there has not been an adequate resolution to this mathematical and ethical dilemma, which still needs to be resolved on a consensus basis.

The question of conducting an open-label study, following completion of a blinded protocol, is also of ethical importance [32]. Access to the study medication, following involvement in a blinded study, is often considered as an inherent ethical consideration, even at the time of the initial HREC approval for the blinded study to be undertaken. It is considered important to ensure that patients, who consider that they have benefited from the study medication, be allowed to continue on it until the medication is available on the market at a reasonable price, especially where a formulary exists, as is the case with the government-sponsored Pharmaceutical Benefits Scheme (PBS) in Australia [23, 36, 37].

#### **24.1.6 AFTER STUDY COMPLETION**

Responsibility for the trial consequences does not automatically stop once the last subject has completed his or her trial commitments. Insurance against the delayed expression of adverse effects is expected to be maintained long after the study has been completed and at least for one year. Ongoing support of patients, access to effective treatment, and notification of subsequent SAEs remains obligatory to ensure that patient safety is maintained [29, 34].

What is often less clear-cut is the ethical requirement to publish the findings of the study once the results become available. Current opinion supports an ethical

obligation to make the findings of any study publicly accessible, to ensure the widest dissemination of all results, be they positive or negative [38]. To satisfy this obligation, it is mandatory for journals to provide equal space for both negative and positive studies, something that is often claimed but not necessarily satisfied or realized with the current perception being that positive studies are far easier to publish than our negative studies [39, 40].

### 24.1.7 CONCLUSION

The ethics of clinical drug trials start at the time of establishing what question(s) should be asked, and answered, as a consequence of the trial. The response to these issues influences every step along the way, up to and including posttrial follow-up.

Issues that have not been canvassed within this treatise include those ethical obligations that represent the responsibility of trial subjects to adhere to their commitments to follow trial dictates. It behoves all those concerned with the conduct of the trial to recognize that trials represent a very expensive and demanding exercise that must adhere to the requirements as stipulated within the relevant protocol. Other issues, not addressed in the above analysis, include the ethical obligations that should be addressed by the company that sponsors the trial. These obligations should include the maintenance of rigorous postmarketing surveillance concerning the product being trialed, to protect against unexpected SAEs, as occurred with such medications as vigabatrin, which caused specific visual field defects that were only recognized years after the drug had come to market [41]. It may also be relevant to conduct postmarketing studies that better reflect the true use of products to see if clinical practice truly reflects what was demonstrated as a consequence of regulatory determined trials [42–44].

## REFERENCES

1. Ablett, S. (1997), Clinical trials: Towards good practice, *Arch. Dis. Child*, 55, 283–286.
2. Altman, D. I. (2002), Poor-quality medical research: What can journals do? *JAMA*, 287, 2765–2767.
3. Editorial (2001), Sponsorship, authorship and accountability. Revision of statement on publication ethics by the International Committee of Medical Journal Editors, *MJA*, 175, 294–296.
4. Chan, Y. H. (2003), Randomised controlled trials (RCTs)—Sample size: The magic number? *SMJ*, 44(4), 172–174.
5. Kirby, A., GebSKI, V., and Keech, A. C. (2002), Determining the sample size in a clinical trial, *MJA*, 177(5), 256–257.
6. Roberts, I., Kwan, I., Evans, P., et al. (2002), Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation, *BMJ*, 324, 474–476.
7. Lancaster, G. A., Dodd, S., and Williamson, P. R. (2004), Design and analysis of pilot studies: Recommendations for good practice, *J. Eval. Clin. Pract.*, 10(2), 307–312.
8. Wittes, J., and Brittain, E. (1990), The role of internal pilot studies in increasing the efficiency of clinical trials, *Stat. Med.*, 9(1–2), 65–71.

9. Workman, P. (2004), Inhibiting the phosphoinositide 3-kinase pathway for cancer treatment, *Biochem. Soc. Trans.*, 32, 393–396.
10. Stephen, L. J., Sills, G. J., and Brodie, M. J. (2000), Topiramate in refractory epilepsy: A prospective observational study, *Epilepsia*, 41, 977–980.
11. Salinsky, M. C., Storzback, D., Spencer, D. C., Oken, B. S., Landry, T., and Dodrill, C. B. (2005), Effects of topiramate and gabapentin on cognitive abilities in healthy volunteers, *Neurology*, 64, 792–798.
12. Chadwick, D., Leiderman, D. B., Sauermann, W., et al. (1996), Gabapentin in generalised seizures, *Epilepsy Res.*, 25(3), 191–197.
13. Kane, J. M. (2002), Issues in clinical trials designs, in Davis, K. L., Charney, D., Coyle, J. T., Nemeroff, C., Eds. *Neuropsychopharmacology: The Fifth Generation of Progress*, Lippincott Williams & Wilkins, Philadelphia, pp. 537–546.
14. FDA (1999), Guidance for industry: Clinical development programs for drugs, devices and biological products for the treatment of rheumatoid drugs, devices and biological products for the treatment of rheumatoid arthritis (RA), U.S. Food and Drug Administration, Rockville, MD.
15. Grimes, D. A., and Schulz, K. F. (2002), Clinical research in obstetrics and gynecology: A Baedeker for busy clinicians, *Obstet. Gynecol. Surv.*, 57(9), S35–S53.
16. Talley, N. J., Moore, M. G., Sprogis, A., et al. (2002), Randomised controlled trial of pantoprazole versus ranitidine for the treatment of uninvestigated heartburn in primary care, *MJA*, 177(8), 423–427.
17. The Therapeutic Goods Administration (1991), *Guidelines for Good Clinical Research Practice (GCRP) in Australia*, Cth Dept of Health Housing and Community Services (DEB 3), Aust. Gov. Printers, Canberra, Commonwealth Dept. of Health, Housing & Community Services.
18. World Health Organisation (2002), *Handbook for Good Clinical Research Practice (GCP): Guidance for Implementation*, WHO, Geneva.
19. Committee for Proprietary Medical Products, *Guidelines for Good Clinical Practice (GCP): ICH Harmonised Tripartite Guideline* (originally approved 17 July 1996).
20. Lader, E. W., Cannon, C. P., Ohman, E. M., et al. (2004), The clinician as investigator: Participating in clinical trials in the practice setting, *Circulation*, 109, 2672–2679.
21. Kovacevic, M., Odeleye, O. E., Sietsema, W. K., et al. (2001), Financial concepts to conducting and managing clinical trials within budget, *Drug. Info. J.*, 35, 1031–1038.
22. The Therapeutic Goods Administration (2006), *The Australian Clinical Trial Handbook: A Simple Practical Guide to the Conduct of Clinical Trials to International Standards of Good Clinical Practice (GCP) in the Australian Context*, Dept. of Health and Ageing, Aust. Gov., Canberra.
23. World Medical Association Declaration of Health, Helsinki (1989), *Recommendations Guiding Physicians in Biomedical Research Involving Human Subjects*, Adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964, and amended by 29th World Medical Assembly, Tokyo, Japan, Oct. 1975; 35th World Medical Assembly, Venice, Italy, Oct 1983; 41st World Medical Assembly, Hong Kong, Sept 1989; 48th General Assembly Somerset West, Republic of South Africa, Oct 1996; 52nd World Medical Assembly Edinburgh, Scotland, Oct 2000; AMA Code of Ethics, Canberra 1996, 2003, 2004; Fluss S S, *International Guidelines on Bioethics*—Informal listing of selected International Codes, Declarations, Guidelines etc. on medical ethics/bioethics, health care ethics, human rights

- aspects of health, *EFCGP News*, Sept. 1998; rev. ed., Dec 1999; 2nd rev. ed., Autumn 2000.
24. National Health & Medical Research Council—Human Research Ethics Committee (1999), *National Statement on Ethical Conduct in Research Involving Humans*, Commonwealth of Australia, Canberra.
  25. Commission of the European Communities (1997), Proposal for a European Parliament and Council Directive on the Approximation of Provisions Laid down by Law, Regulation or Administrative Action Relating to the Implementation of Good Clinical Practice in the Conduct of Clinical Trials on Medicinal Products for Human Use [COM (97) 369 final], Brussels.
  26. Beran, R. G. (2005), Ethical considerations within clinical research with special focus upon clinical drug trials, *Med. Law*, 25, 411–436.
  27. National Health and Medical Research Council (2001), *Human Research Ethics Handbook*, NHMRC, Canberra.
  28. Devereux, J. A. (2002), Guardianship and Consent, in Beran, R. G., Eds., *Epilepsy: A Question of Ethics*, Yozmot, Tel Aviv.
  29. Anon (2000), *Note for Guidance on Good Clinical Practice* (CPMP/ICH/135/95): Annotated with TGA comments. Canberra: Therapeutic Goods Administration, Department of Health and Aged Care. Available at: <http://www.tga.gov.au/docs/html/ich13595.htm>; accessed April 4, 2006.
  30. Beran, R. G. (2003), Informed consent, a legal requirement in the management of patients with epilepsy, *Med. Law*, 22(1), 155–184.
  31. Chase, D., Gierend, M., Rettig, S., et al. (2005), Quality assurance in clinical trials: results of the follow-up systems of member companies in the BVMA, *Int. J. Pharm. Med.*, 19(5–6), 275–276.
  32. Wainwright, P. (2002), Consent to open label extension studies: Some ethical issues, *J. Med. Ethics*, 28, 373–376.
  33. Donnellan, P., and Smyth, J. (2001), Informed consent and randomised controlled trials, *J. R. Coll. Surg. Edinb.*, 46(2), 100–102.
  34. Keech, A. C., Wonders, S. M., Cook, D. I., et al. (2004), Balancing the outcomes: Reporting adverse events, *MJA*, 181(4), 215–218.
  35. Oh, V. M. S. (1994), The placebo effect: Can we use it better? *BMJ*, 309, 69–70.
  36. Graham, D. (1995), The Australian pharmaceutical benefits scheme, *Aust. Prescr.*, 18, 42–44.
  37. Eadie, M. J. (2004), The Australian pharmaceutical benefits scheme, in Beran, R. G., Eds., *Epilepsy and the Law of Therapeutics*, Yozmot, Tel Aviv, Israel.
  38. Wager, E., Field, E. A., and Grossman, L. (2003), Good publication practice for pharmaceutical companies, *Curr. Med. Res. Op.*, 19, 149–154.
  39. Hensley, S., and Abboud, L. (2004), Medical research has “black hole”: Negative results often fail to get published in journals; some blame drug industry, *Wall. St. J. (East Ed)*, June 4, B3.
  40. Berlin, J. A., and Wacholtz, M. C. (2005), Selective reporting, publication bias and clinical trial registry: An industry perspective, *Int. J. Pharm. Med.*, 19(5–6), 277–284.
  41. Beran, R. G. (2001), The ethics of post-marketing surveillance of therapeutic agents, *Med. Law*, 20(4), 587–594.
  42. Beran, R. G., Berkovic, S. F., Black, A. B., et al. (2005), Efficiency and safety of levetiracetam 1,000 to 3,000 mg/day in patients with refractory partial seizures: A multicentre, open-label single arm study, *Ep. Res.*, 63, 1–9.

43. Beran, R. G., Berkovic, S., Black, A., et al. (2001), Australian Study of Titration to Effect Profile of Safety (AUS-STEPS): High-dose gabapentin (Neurontin®) in partial seizures, *Epilepsia*, 42(10), 1335–1339.
44. Berry, D. J., Beran, R. G., Plunkett, M., et al. (2003), The absorption of gabapentin following high dose escalation, *Seizure*, 12, 28–36.



# 24.2

---

## *Ethical Issues in Clinical Research*

**Kelton Tremellen<sup>1</sup> and David Belford<sup>2</sup>**

<sup>1</sup>*Repromed, Dulwich, South Australia*

<sup>2</sup>*GroPep Limited, Adelaide, South Australia*

### **Contents**

|          |                                                                                                        |      |
|----------|--------------------------------------------------------------------------------------------------------|------|
| 24.2.1   | Introduction                                                                                           | 1112 |
| 24.2.2   | Ethical Dilemmas Central to Conduct of Clinical Research                                               | 1112 |
| 24.2.3   | Philosophical Theory and Clinical Trial Ethics                                                         | 1114 |
| 24.2.3.1 | Deontological Theory (Kantianism)                                                                      | 1114 |
| 24.2.3.2 | Utilitarianism                                                                                         | 1114 |
| 24.2.3.3 | Principlism                                                                                            | 1115 |
| 24.2.4   | Historical Triggers for Development of Codes of Ethical Research Practice                              | 1116 |
| 24.2.5   | Development of Guidelines for Ethical Conduct of Medical Research                                      | 1117 |
| 24.2.5.1 | Hippocratic Oath (Fourth Century BC)                                                                   | 1117 |
| 24.2.5.2 | Nuremberg (1949)                                                                                       | 1118 |
| 24.2.5.3 | Declaration of Helsinki (1964)                                                                         | 1118 |
| 24.2.5.4 | National Research Act (1974) and Belmont Report (1979)                                                 | 1119 |
| 24.2.5.5 | CIOMS (1982)                                                                                           | 1120 |
| 24.2.5.6 | National Guidelines on Ethical Conduct of Human Research                                               | 1121 |
| 24.2.6   | Ethical Issues Design, Conduct, and Reporting of Medical Research                                      | 1121 |
| 24.2.6.1 | Gold Standard: Randomized Clinical Trial                                                               | 1121 |
| 24.2.6.2 | Ethical Frameworks for Conduct of Clinical Trials                                                      | 1123 |
| 24.2.6.3 | Ethical Issues in Clinical Trial Design: Study Power and Control Groups                                | 1125 |
| 24.2.6.4 | Informed Consent                                                                                       | 1126 |
| 24.2.6.5 | Transparent Clinical Trial Culture: Financial Disclosure, Trial Registration, and Reporting of Results | 1129 |

|          |                                                                  |      |
|----------|------------------------------------------------------------------|------|
| 24.2.7   | Ethically Challenging Scenarios in Clinical Research             | 1133 |
| 24.2.7.1 | Children                                                         | 1133 |
| 24.2.7.2 | Mentally Impaired                                                | 1135 |
| 24.2.7.3 | Women and Pregnancy                                              | 1137 |
| 24.2.7.4 | Critically Ill Patient                                           | 1138 |
| 24.2.7.5 | Research in Undeveloped World                                    | 1139 |
| 24.2.8   | Four Golden Rules of Ethical Conduct in Clinical Research        | 1140 |
| 24.2.8.1 | A: Respect for Patient Autonomy                                  | 1141 |
| 24.2.8.2 | I: Maximization of Research Impact on Medical Treatment          | 1141 |
| 24.2.8.3 | M: Minimization of Risk to Research Participants                 | 1142 |
| 24.2.8.4 | S: Scientific Integrity                                          | 1142 |
| 24.2.9   | Conclusion                                                       | 1142 |
|          | Appendix A: Conclusion                                           | 1143 |
|          | Appendix B: World Medical Association Declaration of Helsinki:   |      |
|          | Ethical Principles for Medical Research Involving Human Subjects | 1144 |
|          | References                                                       | 1148 |

**24.2.1 INTRODUCTION**

“It is not enough that you should understand about applied science in order that your work may increase man’s blessings. Concern for man himself and his fate must always form the chief interest of all technical endeavors.”

—*Albert Einstein, Address to the California Institute of Technology, 1931*

“There is the more reliable safeguard provided by the presence of an intelligent, informed, conscientious, compassionate, responsible investigator.”

—*Harry Beecher, New England Journal of Medicine, 1966*

The quest for new treatments to alleviate suffering and prevent disease is indeed a very noble cause. Medical research has the potential to create huge benefits to society and enhance the professional reputation of the researchers who are responsible for these discoveries. However, a researcher must never forget that concern for the participants of human research is paramount, irrespective of how important the research outcomes are to society in general. A researcher must never knowingly seriously compromise the welfare of a research participant—“the ends must never justify the means.” According to Harry Beecher, careful training of researchers in both biomedical and ethical concerns is the key to preventing or minimizing serious mishaps. Therefore, the aim of this chapter is to educate clinical researchers in the common pitfalls of human experimentation and ways to avoid these ethical problems in pharmaceutical trials.

**24.2.2 ETHICAL DILEMMAS CENTRAL TO CONDUCT OF CLINICAL RESEARCH**

Ethics is a branch of philosophy that is concerned with answering the age-old questions about duty, honor, integrity, and justice. Research ethics is a branch of applied ethics that studies the ethical problems, dilemmas, and issues that arise in the conduct of research. Within the field of clinical trials three key ethical dilemmas often arise [1]:

1. *Good of Individual Versus Good of Society* Central to all moral theories is the need to respect and promote the rights of an individual. Natural or universal rights are those rights which are inherent in the nature of people and are not contingent on human actions or beliefs nor do they require legal enforcement by government or society. Philosophers believe that the rights to life and liberty (ability to act according to his or her own will) are the two guiding principles of natural rights. On the other hand, the promotion of social welfare of the community in general is also a central tenet in moral philosophy. Unfortunately these moral principles of respect for an individual's natural rights and the needs of society are often in conflict in the setting of clinical research. It is vitally important to use human subjects in research in order to gain valuable scientific knowledge that can benefit all of society, but this research has the potential to place individuals in harm's way and compromise their natural rights. Therefore the central ethical question in all clinical research is how best to protect the universal natural rights of an individual without compromising the scientific validity and social value of the research.

2. *Net Result of Risks Versus Benefits* In any clinical research setting with the potential to harm the participant physically or mentally, the researcher has an ethical obligation to ask the question whether the benefits of the research to the participant and/or the community in general outweigh the risks to the individual. It is normal to assess risks/benefits in terms of probability and magnitude and in the context of the underlying health status of the trial participants. For example, it would be considered unethical to conduct a research study for a medication with even a low risk of death (1%) in the setting of a non-life-threatening medical condition such as dermatitis or alopecia. However, it may be ethically acceptable to enroll terminally ill patients with grade IV astrocytoma in a trial of a new chemotherapy medication with a 1% chance of medication-induced death if this trial treatment gives them any real chance of cure. Conversely, it may be acceptable to place healthy phase I trial volunteers at significant risk of mild reversible adverse events such as headaches or nausea, but it is totally unacceptable to place them at any material risk of death or irreversible disability.

3. *Just Distribution of Benefits and Harms* All human experimentation raises the important ethical dilemma of equitable distribution of benefits and harms. The potential benefits to an individual of being involved in medical research include access to new potentially more efficacious treatments, supervision of clinical care by highly qualified medical staff with intensive monitoring, identification of secondary pathology during the trial entry examination, financial support, and improved self-esteem. Many people gain considerable benefit from knowing that their participation in research is advancing medical science, thereby helping other people in the future with the same medical condition that they are experiencing. Such altruistic motivation is a common positive psychological benefit of trial participation. Conversely, involvement in a clinical trial may create harm. The trial medication may produce physical or psychological adverse effects ranging from the relatively benign through to death. For example, if the participant is allocated to the placebo arm of a study, their involvement in research may delay or block their access to already established effective treatments. Finally, many trial designs require extensive monitoring which can be intrusive and tiresome to an individual whose quality of life may already be diminished by their underlying disease. If the benefits of participa-

tion in a trial are likely to exceed the potential risks, natural justice would dictate that any individual, regardless of their age, race, gender, religion, or financial resources, should have equal access to the trial.

### **24.2.3 PHILOSOPHICAL THEORY AND CLINICAL TRIAL ETHICS**

Philosophers have developed several different ethical theories/principles which may assist in making the morally correct decision in the field of clinical research [2]. The following discussion will summarize the main components of the three most useful theories/principles in the field of research ethics. Each of these theories/principles has its own strengths and weaknesses, making it important to not analyze a clinical research dilemma from a single philosophical point of view.

#### **24.2.3.1 Deontological Theory (Kantianism)**

Deontological theory is an approach to ethics that focuses on the rightness or wrongness of actions themselves, as opposed to the rightness or wrongness of the consequences of those actions [2]. It can be summarized succinctly as an ethical theory based on “moral duty or obligation.” The German philosopher Immanuel Kant (1724–1804) is generally considered as the father of deontological theory. Kant claimed that an action is morally wrong, irrespective of the final outcome, if it is inconsistent with the status of a person as a free and rational being [3]. The duty to promote an individual’s freedom and rationality is the guiding ethical principle of Kantianism, referred to as the categorical imperative. Stated differently, Kantianism requires an individual to treat both themselves and others as ends in themselves, never only as means to an end. All human beings have intrinsic moral dignity and we should not abuse, manipulate, harm, exploit, or deceive people in order to achieve specific goals, no matter how desirable that goal may be. Put quite simply, deontological theory can be condensed into “the ends never justify the means” and “do unto others as you would have done to yourself.”

#### **24.2.3.2 Utilitarianism**

The basic guiding principle of utilitarianism is that the morally right thing to do is to produce the best overall outcome for the most number of people. The English philosophers Jeremy Bentham (1748–1832) and John Mill (1806–1873) developed the theory of utilitarianism [4]. Mills stated that the morally correct action was the one which created the greatest balance of happiness for the greatest number of people—somewhat akin to a moral “cost–benefit” analysis. Where deontological theory prohibits placing clinical trial participants at significant personal risk for the benefit of science or the community in general, utilitarianism states that this is ethically sound provided the expected outcomes for society are favorable. Utilitarianism is a commonly used theory in medical research as investigators and institutional review boards (IRBs) are always trying to find the appropriate balance between personal risk and benefit for the trial participant versus the benefits to society in general.

### 24.2.3.3 Principlism

The term *principlism* is used to describe four standard principles outlined in the Declaration of Helsinki [5] and the Nuremberg Code [6] that govern ethically responsible research practice. These principles include respect for autonomy, non-maleficence, beneficence, and justice [7]. Principlism is not derived from a single philosophical theory, but rather combines several different philosophical perspectives [1]. For example, the right of individual freedom or autonomy is central to deontological theory while beneficence or the generation of “greater good” is a key feature of utilitarianism. As principlism borrows the best features of the available philosophical theories, it is generally considered the most applicable to resolving ethical dilemmas in medical research:

(a) *Autonomy* Autonomy is the basic human right to respect an individual’s ability to make their own decisions regarding their own health and future. If a researcher is to respect an individual’s right for self-determination, he or she must create a situation where the research subject is able to make an autonomous decision regarding participation in the research. This can only be enabled by providing adequate information about the participant’s involvement in the trial (informed consent) and avoiding undue influence or coercion. To achieve informed consent, subjects must be given adequate information regarding the research procedures that they will undergo, the risks and benefits of being involved, and alternative medical therapies of proven benefit outside the clinical trial. This information must be provided in both oral and written form in simple lay language that can be easily understood by a person with only basic language skills. Furthermore, all study participants must enter voluntarily, free from coercion or undue influence, and understand that they may withdraw at any time. The provision of excessive financial reward or the threat of withdrawal of standard medical care would be considered an unethical violation of the ethical principle of subject autonomy.

(b) *Beneficence* Beneficence refers to actions likely to promote individual well-being and/or benefit others.

(c) *Nonmaleficence* The concept of nonmaleficence is best embodied by the Latin phrase *primum non nocere* (“first do no harm”). To avoid doing harm has been a guiding principle of medical treatment since the advent of the Hippocratic oath. Medical researchers have both a moral and legal obligation to avoid “unjustified” harm to trial participants. Some harm may be justifiable provided that the benefits outweigh the negatives. All medical research must be assessed on the balanced probability of harm versus benefit according to the ethical principles of beneficence and nonmaleficence.

(d) *Justice* Justice requires that there be fair procedures and outcomes in the selection of research subjects. The burdens and benefits of research must be equally distributed to include individuals of both genders, different ages, and various racial and ethnic backgrounds, not just the easily available or vulnerable to facilitate trial recruitment. In the past researchers have targeted middle-aged white men for trials, avoiding recruitment of children and women of reproductive age. This produces an injustice as it creates an inequality in medical understanding of how a particular treatment may work (or not work) in children and women while robbing these

disadvantaged groups of the opportunity to benefit from new clinical trial treatments.

#### **24.2.4 HISTORICAL TRIGGERS FOR DEVELOPMENT OF CODES OF ETHICAL RESEARCH PRACTICE**

Clinical research using the established scientific method of hypothesis generation, observation, and experimentation has been conducted throughout the world for many centuries. However, only relatively recently have such medical experiments been placed under the scrutiny of ethical analysis and the nature of human experimentation limited by published ethical guidelines. Unfortunately many researchers have either been ignorant or intentionally subjugated these established ethical principles, often leading to disastrous consequences. Some of the best known and studied examples of human experimentation gone awry are outlined in the paragraphs below as examples of why clinical research requires enforceable ethical guidelines to protect research participants and society alike.

The Tuskegee experiment is an example of medical research that breaks almost every ethical principle [1]. In 1929 the U.S. Public Health Service commenced a study to examine the prevalence of syphilis among poor uneducated blacks and its possible mechanisms of treatment. The study confirmed that mass treatment of syphilis was feasible and desirable, but by 1932 funding for treatment was suspended due to lack of money created by the Great Depression. However, in the town with the highest prevalence of syphilis (Tuskegee, Alabama) it was decided that researchers would continue to monitor participants for the long-term complications of untreated syphilis, thereby ascertaining the natural history of the disease in African Americans [8, 9]. Participants were administered “therapy” which was known to be ineffective (mercurial ointment and spinal taps) and coerced into being involved in the research by being offered financial incentives such as provision of free medical “care” and payment of funeral expenses! The Tuskegee research participants were not told that they had syphilis but rather “bad blood” and they were not offered alternative medical treatment, despite penicillin becoming an established medical cure for syphilis as early as 1945. It was not until 1972, following considerable public outcry, that the U.S. Department of Health, Education and Welfare (DHHS) closed the study. This unfortunate experiment breaks several established ethical principles, including the right of participants to make an informed unbiased choice regarding their involvement in research (autonomy), the need to maximize individuals’ well-being while minimizing harm (beneficence, nonmaleficence), and justice [9].

Medical experiments conducted by Nazi doctors of the Dachau, Auschwitz, and Buchenwald concentration camps are one of the worst examples in history of a researchers’ subjugation of their ethical duties to research participants. These doctors were proponents of racial hygiene theory, which stated that people of non-Aryan background such as Jews were subhuman and therefore should not be afforded the same basic human rights as the Aryan people. As such, the Nazis believed that they were free to conduct experiments on concentration camp prisoners that can only be described as barbaric. These experiments included subjecting prisoners to extremes of temperature and pressure, ingestion of seawater, limb transplants

without medical need, injection of infectious agents to determine the effectiveness of new antibacterial drugs or vaccines, deliberate inflicting of skin wounds to provide an experimental model to assess new forms of wound therapy and exposure of subjects to high doses of radiation [1, 10]. In all of the experiments the aim was to gain knowledge that was helpful to the German military, not the treatment of underlying prisoner ailments. Furthermore, prisoners were not told what would happen to them, were not given the opportunity to refuse consent, and were not offered established medical treatment when complications arose. These gross violations of basic human rights and ethical principles by the Nazis during World War II were the impetus for the development of the first formal international guideline on the ethical conduct of medical research, the Nuremberg Code (1949).

There are many examples of researchers using vulnerable subjects for research to help facilitate recruitment [1]. In the 1950s New York University conducted experiments in which mentally retarded institutionalized children from Willowbrook State School were deliberately infected with hepatitis virus to enable study of the natural progression of the disease. These children's parents had unfair inducement applied by researchers to give their consent by offering to expedite admission of their children to the school if they agreed to be involved. Between 1960 and 1972 the Defense Atomic Support Agency funded a research program in which low-income, poorly educated principally black patients with cancer were administered massive doses of whole-body irradiation as a "treatment." Despite there being no scientific justification for treating a localized cancer with whole-body irradiation, this study was conducted so that the military could assess human tolerance to radiation exposure and effective methods of treatment. Finally there are multiple examples of prisoners being subjected to dangerous exposures of infectious agents and radiation in the interests of medical research. While many of these prisoners were informed of the nature of the experiments, they were coerced into involvement by incentives such as early release from prison or gifts of money and cigarettes. In the prison system, where an individual's liberties are severely limited, such coercion does not allow unbiased informed consent.

#### **24.2.5 DEVELOPMENT OF GUIDELINES FOR ETHICAL CONDUCT OF MEDICAL RESEARCH**

##### **24.2.5.1 Hippocratic Oath (Fourth Century BC)**

The Hippocratic oath is an oath taken by physicians governing the ethical practice of medicine. It is believed that the oath was written by Hippocrates, the father of medicine, in the fourth century BC. While the contents of the oath do not pertain directly to medical research, they do set down some valid ethical practices to which physicians conducting clinical trials should aspire. Two components of the oath have special relevance to the ethical practice of clinical research. First, Hippocrates outlines the need for nonmaleficence and beneficence in the phrase "I will prescribe regimes for the good of my patients according to my ability and my judgment and never do harm to anyone" and the decree to keep "myself far from all intentional ill-doing." Second, he speaks of the importance of educating the next generation of physicians. In the broadest sense, one can draw parallels between this requirement



for education and the ethical requirement for researchers to share knowledge gained during research with the general community.

#### 24.2.5.2 Nuremberg (1949)

Following the end of World War II trials of war criminals were conducted before the Nuremberg Military Tribunal in order to investigate and punish those individuals who committed crimes against humanity. Specifically, they were in response to the inhumane Nazi human experimentation carried out during the war by individuals such as Dr. Josef Mengele [2, 10]. During the trial several of the accused argued that their experiments differed little from prewar experiments and that there was no law that set the boundaries differentiating legal from illegal human experimentation. Dr. Leo Alexander then submitted to the tribunal a list of six points that defined what should be considered legitimate human medical research. These six points were adopted by the trial verdict and an additional four points were added—together making the 10-point Nuremberg Code governing ethical human experimentation [6]. The contents of the Nuremberg Code are outlined in Appendix A. In summary, the code outlines the need for ensuring participant autonomy (Articles 1 and 9), beneficence (Article 2), and nonmaleficence (Articles 3–8 and 10) during human experimentation.

While an excellent starting point in the development of guidelines to govern ethical human experimentation, the Nuremberg Code was lacking in several key areas. First, the code contained no guidelines concerning the ethical conduct of research on individuals who are unable to give informed consent (e.g., children, mentally impaired, and critically ill). Second, it totally ignored the need for maintaining participant confidentiality and the need for justice regarding individual's opportunity to engage in medical research. These deficiencies were later addressed in the Declaration of Helsinki [5].

#### 24.2.5.3 Declaration of Helsinki (1964)

The Declaration of Helsinki [5] was developed by the World Medical Association (WMA) as a set of ethical principles for physicians and other participants in medical research. It was originally adopted in 1964 in Helsinki, Finland, and has since undergone eight revisions, the latest significant revision occurring in 2000 (Appendix B). The declaration covers all of the major ethical protections outlined in the Nuremberg Code but also contains several important additions:

1. *Article 1* The declaration states that human material and identifiable human data require the same ethical protections given to individuals undergoing direct human experimentation. This means that researchers are not able to conduct experiments using human tissue/blood samples collected for a non-related purpose without the consent of the original research participant. Similarly, researchers are not able to use databases containing confidential medical details of individuals for the purposes of research without first gaining these individuals' informed consent.
2. *Articles 8 and 23–25* Here the declaration discusses restraints required for the ethical conduct of research on “vulnerable persons” such as children, the

mentally incompetent, and subjects in a dependent therapeutic doctor–patient relationship with the researcher.

3. *Article 9* The declaration states that the ethical protections set forward by the declaration cannot be reduced or eliminated by local legal or regulatory requirements but these protections should be considered universal worldwide.
4. *Article 12* Here the declaration states that due care for the environment and welfare of animals must be taken into consideration during the conduct of medical research.
5. *Article 13* There is an absolute requirement that all human research protocols should first be reviewed by an independent committee of relevant experts (ethical review committee) to ensure that the study design is scientifically robust and provides adequate protection of research participants.
6. *Article 21* This stresses the need for maintaining participant privacy and confidentiality.
7. *Article 27* Both authors and publishers of scientific journals have an obligation to honestly disclose the results of both positive and negative outcomes of human experimentation. Furthermore, any potential source of conflict of interest must be declared in all publications.
8. *Articles 29 and 30 and Associated Clarifications* Following public outcry concerning the use of a placebo in a U.S. Center for Disease Control study of the antiviral zidovudine to prevent perinatal transmission of HIV in the developing world, the declaration's provisions on acceptable "alternative care" were revised (2000). This revision held that developed-world standards of care should apply to the conduct of clinical trials in the developing world, irrespective of whether this standard of care is actually available to the public through the existing health system. While the current declaration does not expressly prohibit placebo-controlled studies, it does set out some significant limitations. Article 29 states that "the benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic, or therapeutic method exists." The declaration then goes on to state that it is acceptable to use a placebo in a research study where there is a proven alternative therapy provided that the condition being studied is a relatively minor one and the research participant will not suffer any irreversible harm and all research participants allocated to the placebo arm will later be offered the best available proven therapy at the conclusion of the study.

#### **24.2.5.4 National Research Act (1974) and Belmont Report (1979)**

Although the Nuremberg Code and Declaration of Helsinki provided excellent guidelines on what constituted the ethical practice of medical research, they did not have the power of legal mandated enforcement within individual countries. Starting in the United States, the Surgeon General (1966) and later the U.S. Food and Drug Administration (FDA) (1971) mandated prior peer review of all

human research protocols funded by the National Institutes of Health or regulated by the FDA. This in turn led to the development of the National Research Act (1974), which mandated that any government-funded medical research would first require review and a pass judgment on the acceptability of the proposed research by an independent IRB. Furthermore, the regulations set down rules which would govern the composition of these IRBs, such as a mixture of men and women, scientific and nonscientific members, as well as general community members. The composition of these boards was to ensure that they had the capacity to assess the scientific validity of the research proposal as well as that ethical principles such as distributive justice, participant safety, and autonomy were respected while blocking research with the potential for inciting moral outrage in the general community.

When the National Research Act was signed into law on the July 12, 1974, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was formed. This commission was charged with developing guidelines that should be followed during all human medical experimentation. The results of nearly 4 years of deliberations by the commission were published in 1979 as the landmark Belmont Report [11]. As would be expected, the report focuses on the need to respect participant autonomy, the need to maximize good (beneficence), and the application of justice.

#### **24.2.5.5 CIOMS (1982)**

The Council for International Organizations and Medical Sciences (CIOMS) was formed in 1949 by the World Health Organization (WHO) and the United Nations Scientific and Cultural Organization (UNESCO) with the goal to facilitate and promote international activities in the field of the biomedical sciences. As part of this mandate, CIOMS undertook important work in the field of bioethics during the 1970s, culminating in the publication of the “International Ethical Guidelines for Biomedical Research Involving Human Subjects” in 1982. This work has since been updated in 1993 and 2002 and is available in print [12] or online via the CIOMS website (<http://www.cioms.ch/>; CIOMS 2002 guidelines). The CIOMS document consists of 21 guidelines governing the ethical conduct of human biomedical research. These guidelines are heavily based on those contained within the Declaration of Helsinki but also contain several important practical discussions on how best to implement ethical principles in the design and conduct of a research. For example, guideline 4 stresses that informed consent is a continuing process that should be conducted throughout the life of a study, not just an event at the commencement of a participant’s enrollment in a research study. The CIOMS document provides a very useful 26-point checklist that helps ensure that all the requirements of proper informed consent are met by a study [12]. Guideline 7 outlines in very practical terms what level of financial reimbursement is ethically acceptable for involvement in a trial and what would be considered unacceptable financial inducement. Appendix 1 of the CIOMS guidelines outlines the 48 individual points of discussion that need to be addressed by all research submission to a Human Research Ethics Committee/IRB to enable proper determination of the ethical merits of the proposed research.

#### **24.2.5.6 National Guidelines on Ethical Conduct of Human Research**

Most countries have their own guidelines governing what is ethically acceptable practice in the field of human experimentation. The Office of Human Research Protections, DHHS, publishes a summary of the individual regulations and legislation covering human research in 79 individual countries and two confederations (European Union and the Commonwealth of Independent States). This summary is updated regularly and can be located on the Office for Human Research Protections website [13]. While the legislation enacted in each country may differ, it should be stressed that no national ethical, legal, or regulatory requirement should be allowed to reduce or eliminate any of the protections for human subjects set forth in the Declaration of Helsinki (Article 9, 2000).

### **24.2.6 ETHICAL ISSUES IN DESIGN, CONDUCT, AND REPORTING OF MEDICAL RESEARCH**

The aim of this section is to make the reader aware of aspects of the ethical debate surrounding the design, conduct, and reporting of a clinical trial. Several issues which are discussed in previous chapters are not discussed in detail here. These include stopping a clinical trial (Chapter 20), the role of Data and Safety Monitoring Committees (DSMCs) and ethics committees (Chapter 8), and patient records and privacy (Chapter 15). It is further recognized that as the ethical debate is evolving, guidance documents and ethical codes are revised at regular intervals, and new legislative changes are under consideration. For example, at the time of this writing, the Declaration of Helsinki is being revised [5], an update of the CONSORT statement is being prepared [14], and federal legislation requiring disclosure of clinical trial results is being considered in the United States [15]. Therefore, where possible, we have provided current website addresses that direct the reader to the latest versions of the relevant documents (Table 1).

It is further noted that the ethical conduct of clinical trials extends beyond the study design, conduct, and reporting. Appropriate attention to the quality of manufacture of investigational drugs, validation of methods, and demonstration of stability underpins ethical evaluation of new drug candidates. Such considerations are now mandatory in most regulatory territories. In addition, when conducting non-clinical safety studies to support clinical trial programs, researchers must be aware of the applicable quality systems—Organisation for Economic Co-operation and Development (OECD) principles for good labor practices (GLP) [16]; GLP regulations, Code of Federal Regulations (CFR), Title 21, Part 58] [17]—and the Declaration of Helsinki, Article 12, which states that “the welfare of animals used for research must be respected” [5].

#### **24.2.6.1 Gold Standard: Randomized Clinical Trial**

A necessary, although not sufficient, condition for any experiment on human subjects is that the study is scientifically sound [18]. The principle of beneficence includes the requirement that the hypothesis being tested has scientific merit, that the design of the study adequately addresses the hypothesis, and the outcomes

**TABLE 1 Summary of Ethical Codes for Clinical Research—Web-Based Resources**

| Guidance/Ethical Code                                                                                                                                                                                              | URL or Literature Reference                                                                                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Ethical codes and acts</b>                                                                                                                                                                                      |                                                                                                                                                                          |
| Nuremberg Code (1949)                                                                                                                                                                                              | <a href="http://www.hhs.gov/ohrp/references/nurcode.htm">http://www.hhs.gov/ohrp/references/nurcode.htm</a>                                                              |
| Declaration of Helsinki (2004; first adopted by the 18th WMA General Assembly, Helsinki, Finland, 1964)                                                                                                            | <a href="http://www.wma.net/e/ethicsunit/helsinki.htm">http://www.wma.net/e/ethicsunit/helsinki.htm</a>                                                                  |
| Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research (1979)                                                                                                          | <a href="http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm">http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm</a>                                      |
| CIOMS International Ethical Guidelines for Biomedical Research Involving Human Subjects (2002; first issued 1982).                                                                                                 | <a href="http://www.cioms.ch">http://www.cioms.ch</a>                                                                                                                    |
| U.S. FDA 21 CFR 50, Protection of Human Subjects                                                                                                                                                                   | <a href="http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=50">www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=50</a> |
| U.S. FDA 21 CFR 56, Institutional Review Boards                                                                                                                                                                    | <a href="http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=56">www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=56</a> |
| ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6 (R1)                                                                                                                                  | <a href="http://www.ich.org">www.ich.org</a>                                                                                                                             |
| International Compilation of Human Research Protections, 2008 Edition, compiled by Office for Human Research Protections, U.S. Department of Health and Human Services.                                            | <a href="http://www.hhs.gov/ohrp/international/HSPCompilation.pdf">http://www.hhs.gov/ohrp/international/HSPCompilation.pdf</a>                                          |
| U.S. FDA Best Pharmaceuticals for Children Act                                                                                                                                                                     | <a href="http://www.fda.gov/opacom/laws/pharmkids/contents.html">http://www.fda.gov/opacom/laws/pharmkids/contents.html</a>                                              |
| <b>Clinical trial registration</b>                                                                                                                                                                                 |                                                                                                                                                                          |
| WHO International Clinical Trials Registry Platform (ICTPR)                                                                                                                                                        | <a href="http://www.who.int/ictpr/en/">http://www.who.int/ictpr/en/</a>                                                                                                  |
| Ottawa Statements                                                                                                                                                                                                  | <a href="http://ottawagroup.ohri.ca/">http://ottawagroup.ohri.ca/</a>                                                                                                    |
| Financial conflict of interest: U.S. Department of Health and Human Services, Guidance Document: Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection | <a href="http://www.hhs.gov/ohrp/humansubjects/finreltn/fguid.pdf">http://www.hhs.gov/ohrp/humansubjects/finreltn/fguid.pdf</a>                                          |
| <b>Publication of clinical trials</b>                                                                                                                                                                              |                                                                                                                                                                          |
| International Committee of Medical Journal Editors (ICMJE) Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication                                   | <a href="http://www.icmje.org/index.html">http://www.icmje.org/index.html</a>                                                                                            |
| The CONSORT Statement                                                                                                                                                                                              | <a href="http://www.consort-statement.org/">http://www.consort-statement.org/</a> , <i>JAMA</i> 2001; 285(15), 1987–1991.                                                |
| Good Publication Practice (GPP) for Pharmaceutical Companies                                                                                                                                                       | <a href="http://www.gpp-guidelines.org/">http://www.gpp-guidelines.org/</a> , <i>Curr. Med. Res. Opin.</i> 2003; 19(3), 149–154.                                         |
| Design of clinical trials: ICH Harmonised Tripartite Guideline: Choice of Control Group and Related Issues in Clinical Trials E10                                                                                  | <a href="http://www.ich.org">www.ich.org</a>                                                                                                                             |

Note: All URLs last accessed December 2007.

of the study are valid. Beneficence further requires that investigators are competent to conduct the study and they seek to minimize the risks and maximize the benefits of the research [12]. The “gold standard” for determining the efficacy and safety of a new therapeutic intervention is the randomized controlled trial (RCT).

*Randomization* aims to eliminate systematic differences between treatment groups [19]. It removes any selection bias that may be introduced by those conducting the study and, ideally, balances confounding variables (some of which may be unknown) among the treatment groups, thus preventing “covariate imbalance” [20]. By diminishing the chance that baseline confounding variables differ among the treatment groups, the validity of conclusions relating the study outcomes to the intervention under investigation is enhanced. A more detailed discussion on randomization procedures is provided in Chapters 11 and 12.

*Blinding* involves the effective masking of both the investigational drug and control treatments (e.g., a placebo or comparator drug) in an attempt to ensure that all those involved in the study (i.e., trial participants and trial staff) do not alter their behavior in response to the knowledge of which treatment is being administered. In a “double-blind” trial, treatment allocation is unknown to the participant, investigator, sponsor, and trial staff, including those assessing outcome measures. Blinding thus aims to minimize potential bias in the management, treatment, and assessment of trial subjects that may otherwise occur if the treatments were known [19]. The term “single blind” refers to studies where one party (usually the subject) is blinded to treatment allocation.

The design of the RCT raises many complex ethical issues. For example, how can investigators be asked to suspend the principles of individualized patient care inherent in the Hippocratic oath and, blinded to prospective treatments, allow their patients to be randomized receive an experimental or control therapy? Further, can the use of placebo as a control be justified if currently available therapies are known to effectively treat the condition under investigation? How can a clinician condone trial-based medical procedures (e.g., the taking of biopsies) that have no therapeutic benefit for the patient? Indeed, should the roles of treating physician and clinical investigator be separated and filled by different individuals?

From the trial participant’s perspective, does an RCT—in requiring patients to delay knowledge of their treatment until the randomization code is broken—contravene the principle of autonomy that would allow individual patients to determine which treatment they receive?

Ethical frameworks for the conduct of a clinical trial, with reference to ethical codes, are briefly introduced in the following sections.

#### **24.2.6.2 Ethical Frameworks for Conduct of Clinical Trials**

The concept of uncertainty in outcome, or “equipoise,” has been invoked by many authors to resolve ethical considerations inherent in the conduct of an RCT. According to this paradigm, equipoise exists if the outcome of a trial is sufficiently uncertain such that an individual participating in the trial will not be disadvantaged; under these circumstances, enrollment in the trial is therefore considered “an equal bet in prospect” [21] and participants should not be harmed by random

allocation to a given treatment arm [22]. A trial in equipoise should, therefore, be acceptable to both Kantian and utilitarian philosophies [21], both of which seek to balance the duties and rights of the individual and society (see a brief description of each above). Thus, in the presence of true equipoise, individuals can, in theory, contribute to the wider medical knowledge at no cost to themselves [23]. The application of the concept of equipoise does, however, raise several questions. For example:

1. *From whose point of view should equipoise exist?* Options include the individual investigators, the clinical research community (“clinical equipoise,” after Freedman [24]), or the prospective patient. Many commentators would consider equipoise as essentially a characteristic of the informed, competent patient [21]. Importantly, a trial that may be considered to be in equipoise for one trial participant may not be for another [23].

2. *What degree of equipoise should exist?* Any given outcome in a clinical study is rarely an “even bet.” As a therapeutic candidate moves through the development process, a significant body of knowledge accrues, and it is unlikely sponsors would continue development unless previous nonclinical and clinical studies were favorable. Moreover, most published industry-funded trials report a positive outcome in favor of the drug candidate under investigation [22, 25]. This finding is not explained by publication bias alone and suggests that, in reality, very few industry-sponsored trials are in true equipoise. The question then becomes: At what point of “equipoise imbalance” does a trial become unethical? One study found that 50% of subjects perceived a trial to be unethical when 70% of experts favored one arm of the study; if this figure rose to 80% of experts, less than 3% of subjects thought a trial is justifiable [26].

3. *At what point is a decision regarding equipoise made?* The key decision faced by most trial participants is whether or not to consent to randomization. Equipoise in this context would mean the prospective participant perceives that agreeing to randomization gives them an equal chance of a good outcome as not participating in the trial [27].

The concept of equipoise has many critics. Some consider that the term lacks consistency, reality, and utility [28], while others maintain that it disregards the autonomy of the subject [22] and is fundamentally incoherent [29, 30]. Alternative ethical frameworks for the conduct of clinical trials have been proposed, including the “positive expected outcomes” [22]. Under this model, the trial participant would base a decision to participate in a given study on the pooled average of expected outcomes in all arms of the trial before randomization. In this context, some authors have noted that many trial participants achieve above average clinical outcomes even though the trial may have reported a negative result or the patient received a control treatment [31, 32]. Several reasons for this observation have been cited, including the cyclical nature of the disease being studied [20], the contention that clinicians who recruit to trials tend to be better clinicians [31], and the possibility that the strict protocol-driven procedures (often overseen by expert committees) provide a superior standard of care. It is emphasized, however, that improved medical care should not be offered as an inducement to participate in a clinical trial.



### 24.2.6.3 Ethical Issues in Clinical Trial Design: Study Power and Control Groups

**Study Power** The social value and scientific validity of a study are critical to the ethical conduct of clinical research. Exposure of trial participants to even minor risks cannot be justified if the study is poorly conducted or is of little scientific merit. It is often argued in this context that underpowered trials that yield outcomes that are not clear-cut have the potential to mislead and are therefore unethical. In contrast, others would maintain that, if equipoise truly exists, then there is no ethical objection per se to participating in an underpowered study [21]; the patient does not “lose out in prospect.” While this contention would seem to contravene a central tenet of ethical clinical research that demands validity and scientific rigor, there are examples where an underpowered trial may be better than no trial at all or where an underpowered trial represents a key step in the drug development process. These include studies on rare conditions, where alternative sources of data may be used to support the trial results, and early-phase studies that aim to determine estimates of efficacy and safety [33]. The authors point out that, in both cases, prospective participants should be informed of how the trial data may be used (e.g., inclusion in a subsequent meta-analysis) and that the trial may only indirectly contribute to health care benefits.

**Placebo-Controlled Trials** Significant ethical concern and debate surround the use of placebo treatments as a control in a clinical study. Some authors have invoked the concept of equipoise to argue that, if an effective treatment is available, then placebo-controlled trials are unethical because equipoise cannot exist [21]. In contrast, others have argued that methodologically sound placebo-controlled trials, where participants are not exposed to excessive risks of harm, are ethical and that equipoise in this context does not provide an adequate ethical framework [30, 34].

Article 29 of the Declaration of Helsinki (issued in October 2000) states: “The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists” [5]. This article was initially interpreted by some, in its strictest sense, to mean that the use of placebo controls was ruled out whenever an effective therapy was available. In 2002, the WMA added a note of clarification to Article 29, reaffirming “its position that extreme care must be taken in making use of a placebo-controlled trial and that in general this methodology should only be used in the absence of existing proven therapy. However, a placebo-controlled trial may be ethically acceptable, even if proven therapy is available, under the following circumstances:

- Where for compelling and scientifically sound methodological reasons its use is necessary to determine the efficacy or safety of a prophylactic, diagnostic or therapeutic method; or
- Where a prophylactic, diagnostic or therapeutic method is being investigated for a minor condition and the patients who receive placebo will not be subject to any additional risk of serious or irreversible harm” [5].

The CIOMs guideline 11 [12] elaborates on these points, further stating that a placebo control can be used “when withholding an established effective intervention would expose subjects to, at most, temporary discomfort or delay in relief of symptoms.” Examples cited include a slightly raised blood pressure or serum cholesterol, common headache, or treatments for baldness or common cold.

In relation to the term “compelling and scientifically sound methodological reasons,” it has been argued that the use of “active controls” does pose methodological concerns [35, 36]. Such views would hold that a noninferiority or equivalence study showing no difference between a study drug and active comparator could be interpreted to mean that both were effective, both were ineffective, or the trial (for whatever reason) was simply not able to detect a difference between them. In this regard it has been noted that, in several areas of medicine, not all “known” active treatments will reliably prove to be superior to a placebo in every study. This particularly applies to analgesics, antidepressants, antianxiety drugs, antihypertensives, antiangina drugs, anti-heart failure drugs, antihistamines, and drugs for asthma prophylaxis [35]. Moreover, trials designed to demonstrate that a new drug and an active control possess similar efficacy do not provide sponsors and investigators with the same “built-in incentives for trial excellence” [19] as studies designed to show a treatment difference. It is relatively easy for poorly designed and conducted studies to fail to show a difference between treatment arms. Thus, for some therapeutic areas, it is widely considered that reliable evidence of efficacy can only be obtained by superiority trials or the internal validation of active-comparator equivalence or noninferiority trials using a placebo. In the former case, the requirement to demonstrate superior efficacy over an authorized treatment may be considered too restrictive. Moreover, an active comparator trial may expose patients to risk and fail to deliver a clear-cut answer.

Thus, if invoking this clause to justify inclusion of a placebo control, the possibility of “additional risk of serious and irreversible harm,” including the period of time a placebo treatment may be administered, needs to be carefully considered. All other ethical safeguards should apply to placebo-controlled trials including ethical and scientific review, full disclosure and informed consent, and the right to withdraw at any time. Additional strategies to minimize the risk of participants include the incorporation of predefined “stop” rules in the study protocol and the monitoring of data collected during the study by a Data Safety and Monitoring Board (Chapter 20). Where appropriate, interim analyses should be undertaken to ensure trial participants are not exposed to the placebo for any longer than necessary or to assess the likelihood that the trial will deliver a clear-cut answer. If, in the latter case, this probability is low, the trial should be terminated. Study designs that seek to address the ethical and practical issues associated with the conduct of placebo-controlled trials are described in the International Conference on Harmonisation (ICH) guideline “Choice of Control Group and Related Issues in Clinical Trials” [19], including “early escape” and “randomized withdrawal” designs.

#### **24.2.6.4 Informed Consent**

Informed consent is a central tenet of clinical research. It is fundamental to all codes of ethical clinical research. The first principle of the Nuremberg Code states that “the voluntary consent of the human subject is absolutely essential” [6]. The

Declaration of Helsinki (Articles 20–26) clearly addresses the principles of informed consent, including reference to subjects who may have a dependent relationship with their physician, subjects who are legally incompetent or physically and mentally incapable of giving consent, and subjects with a condition that renders them unable to give informed consent. In obtaining consent, clinical researchers should adhere to the requirements outlined in the ICH “Guideline for Good Clinical Practice” (Section 4.8) [37], the CIOMS guidelines 5 and 6 [12], and any additional codes published by regulatory authorities (e.g., 21 CFR 50) [38].

Consent must be voluntary and based on sufficient information about the proposed study and the implications of participation to allow potential trial participants to make an autonomous decision whether or not to enroll. The Belmont report states that “no undisclosed risks may be more than minimal” [11]. Importantly, subjects must have realistic expectations as to the study outcomes, including any potential benefit to both the individual participant and society. The language used on the informed consent documents should be “as non-technical as practical and should be understandable to the subject or the subject’s legally acceptable representative and the impartial witness, where applicable” [37]. It is important that the informed consent process continue throughout the course of the study [12].

The source of funding, financial relationships, and financial interests of the parties involved should be disclosed in line with the recommendations in the FDA guidance “Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection” [39] or local institutional policies. This FDA guidance further recommends that, if a potential or actual financial conflict of interest exists that could “influence the tone presentation or type of information presented during the consent process,” then a third party who has no conflict of interest should be involved in the consent process.

**Coercion and Inducement** Coercion and undue inducement are ethically unacceptable and can invalidate consent. While the issue of patient dependence is addressed in Article 23 of the Declaration of Helsinki, researchers should be alert to the fact that more subtle forms of coercion (or intimidation) can result from the position of authority held by an investigating clinician. Indeed, in one study of patients consenting to a major surgical procedure, consent was based more on the trust rather than an assessment of information provided to them [40]. All subjects should be assured that a decision whether or not to participate in the proposed clinical trial will not affect their medical care [12].

It is generally accepted that, for studies that aim to recruit patients into later phases (phases II and III trials), financial reimbursement should only compensate for direct costs of participation (such as travel to and from the clinic). Exceptions to this position may be considered in situations where administration or outcome-testing schedules cause significant inconvenience or when little or no benefit is expected to result from trial participation.

In the case of phase I trials, bioequivalence or pharmacokinetic studies, where participants receive no medical treatment that may be of benefit, a payment above out-of-pocket costs as compensation for inconvenience can be offered. However, such a payment should not be so great as to encourage participants to expose themselves to excessive risk. Moreover, payment contingent on completion of the trial

may prevent participants from leaving a study against their better judgment and could be deemed to constitute undue inducement.

Reasons for leaving a trial should be documented. In cases of withdrawal due to adverse reactions to the study drug, trial subjects should be encouraged to fully articulate all reasons for leaving the trial, and full payment should be offered. Subjects who withdraw due to personal reasons should be considered for a prorated payment. Willful noncompliance on the part of a participant, resulting in expulsion from the study, may be grounds for withholding all payment.

Any incentives offered to trial participants must be reviewed by an ethics committee before recruitment activities commence (see the Declaration of Helsinki, Article 13).

One interpretation of inducement deserves further comment. Although the benefits of participation in clinical trials have been discussed by several authors [20, 31, 32], the prospect of improved medical care cannot be offered as inducement to participate in a trial. However, application of this ethical principle in the context of (for example) teaching hospitals that service populations who may not have access to or cannot afford ongoing quality medical care presents several ethical challenges [18]. Does the prospect of access to medical care (including, e.g., transport to and from the clinic) constitute undue inducement? Does the recruitment of such individuals raise issues of distributive justice that would require the burdens and benefits of clinical research to be evenly distributed (see CIOMs guideline 12 [12])? Does participation in a trial by those seeking access to treatments not otherwise available to them constitute exploitation? In considering these issues, clinical researchers should be cognizant of the principles of justice implicit in relevant articles in the Declaration of Helsinki. In particular, Article 19 states: "Medical research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research." It is noted here that the definition of "populations" is open to interpretation, and there are many clinical research situations where the intended benefit is directed at the wider community. Further, Article 22 states: "The subject should be informed of the right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal." Lastly, in relation to posttrial access of beneficial medicines, in its clarification to Article 30 of the Declaration of Helsinki, the WMA "hereby reaffirms its position that is necessary during the study planning process to identify post-trial access by study participants to prophylactic, diagnostic and therapeutic procedures identified as beneficial in the study or access to other appropriate care." This article raises several ethical and practical issues; importantly, posttrial access arrangements should be agreed to prior to commencing the study and such arrangements require prior review by an ethics committee. For a discussion of the wider ethical issues of research in populations or communities with limited resources, the reader is referred to CIOMS guidelines 10 and 11 [12].

**Competence** Consent should be given by competent individuals with the capacity to understand the information provided to them. It is recognized that conditions for informed consent may vary with the nature of the study and the cultural and religious sensitivities of the participants and their communities. The Belmont report states: "it is necessary to adapt the presentation of the information to the subject's capacities" [11]. It is important that the ethical review process (see Chapter 11)

provides a further level of protection to the informed consent process. The ICH “Guideline for Good Clinical Practice” [37] requires that, before initiating a trial, the investigator must have obtained “written approval of the informed consent form and any other written information to be provided to subjects.” Further, consent by a subject alone is not reason enough to justify participation in the trial. The Declaration of Helsinki (Article 15) states: “The responsibility for the human subject must always rest with a medically qualified person and never rest on the subject of the research, even though the subject has given consent.”

Despite the best attempts of sponsors and investigators to clearly present information about a clinical trial to potential participants, several studies have shown that some degree of communication difficulty is inevitable [21]. For example, the concept of randomization and its justification may not be fully grasped by all individuals [29, 41], and patients who have recently received a diagnosis of a life-threatening disease may not be in a psychological state that enables them to absorb all the information provided to them [42]. In this context, provision of information should not be conducted in a manner that may cause distress; “respect for persons” does not only equate with autonomy [11, 43]. The degree of protection provided by investigators and IRB/Institutional Ethics Committees (IECs) should take into account both a patient’s capacity for self-determination and the risk of harm and degree of benefit that participation in a clinical trial may entail.

Finally, all investigators and sponsors should carefully review their patient information sheets and informed consent forms before initiating their study. Are all risks adequately disclosed? Is the language in any way coercive? Are the chances of receiving a placebo treatment or active therapy fully spelled out? Are financial conflicts of interest disclosed? On a more operational note, Good Clinical Practice (GCP) audits commonly reveal shortcomings in the informed consent process, including use of outdated forms or forms that have not been approved by the relevant ethics committee, timing of signing of the informed consent form in relation to commencement of trial procedures, and forms not appropriately signed or dated by the subject or witness.

#### **24.2.6.5 Transparent Clinical Trial Culture: Financial Disclosure, Trial Registration, and Reporting of Results**

Several related initiatives that seek to increase the transparency of the clinical trials process have arisen in response to (1) concerns that financial interests in clinical research may be placing trial participants at risk [44]; (2) the delayed disclosure or nondisclosure of trial results and publication bias [45], including selective publication [46], selective reporting [47], and inadequate reporting [48]; and (3) inappropriate authorship and ghost writing of publications [49]. There is little doubt these concerns have significant ethical implications and momentum is building on several fronts to address them.

**Financial Disclosure** Clinical researchers should be aware of and adhere to the ethical codes and guidance documents relevant to both their regulatory territories and institutional affiliations. The Declaration of Helsinki (Article 13) states: “The researcher should also submit to the (ethical review) committee, for review, information regarding the funding, sponsors, institutional affiliations, other potential conflicts of interest and incentives for subjects.”

The document “Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection” was released by the DHHS in 2004 to help research institutions, IRBs, and investigators identify potential and actual conflicts of interest and suggest means to eliminate or manage such conflicts with the interests with the welfare of the trial subjects in mind. This guidance document, which is based on the ethical principles described in the Belmont report (respect for persons, beneficence, and justice), poses several questions that IRB members, investigators, and researchers should consider when determining whether or not a conflict of interest exists; it then suggests several “actions to consider” in each case. Many research institutions have adopted these principles.

It should be noted that some consider that the requirement for investigators to disclose any actual or potential conflicts of interest on the informed consent documents does not go far enough to adequately protect research subjects. Indeed, there is a lack of data on how potential clinical trial participants use disclosures in their decision-making process [44]. Moreover, disclosure alone, without further safeguards, may even confer a “moral license to exaggerate” on the part of investigators that could potentially increase any bias in the study that exists as a result of a conflict of interest [43]. Some therefore propose that additional mechanisms must be found to protect subjects; in many cases, the only answer would seem to be to encourage investigators conducting a trial to divest any possible conflicts of interest.

**Registration of Clinical Trials** The participation of any subject in a clinical trial that does not contribute to the medical knowledge of the wider research community contravenes the obligations of clinical researchers. Nondisclosure or delayed disclosure of trial results may mean that the welfare and safety of trial participants could be adversely affected by subsequent clinical trial designs and ethical review processes that do not build on prior research knowledge. On a wider scale, failure to consider all available data may misdirect those responsible for health policy decisions and resource allocation. The Declaration of Helsinki includes several statements relevant to these considerations:

*Article 16* “Every medical research project involving human subjects should be preceded by careful assessment of predictable risks and burdens in comparison with foreseeable benefits to the subject or to others.

The design of all studies should be publicly available.”

*Article 11* “Medical research involving human subjects must ... be based on a thorough knowledge of the scientific literature, other relevant sources of information. ...”

*Article 19* “Medical research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research.”

Registration of clinical trials on publicly available databases represents a key mechanism to promote transparency, knowledge sharing, and the timely and accurate disclosure of clinical information. A global movement for registration of clinical trials is gaining momentum, driven by bodies such as the Ottawa Group [50, 51], WHO [52], and the International Committee of Medical Journal Editors (ICJME)



[53, 54] (see below) and the trial registries. Many trial registries currently exist, including the International Standard Randomized Controlled Trial Number (ISRCTN) register and ClinicalTrials.gov, which is administered by the National Library of Medicine. The WHO International Clinical Trials Registry Platform (ICTRP) project aims to set international standards for trial registration, ensure a minimum set of results are reported, and provide global trial identification and search capability. The ICJME policy, requiring trial registration as a prerequisite to publication, has resulted in a marked increase in the number of trials registered [15, 54].

However, registration of trial information does not currently require that trial outcomes will be disclosed and there are no international standards for reporting of clinical trial results [55]. In response, many now advocate systems to facilitate full public disclosure of published and unpublished trial results. Significantly, in the United States, both federal and state legislation are addressing such concerns. In Maine, a law that came into effect in October 2005 requires companies wishing to advertise prescription drugs to disclose trial results, including adverse effects. In addition, federal legislation currently under consideration will require reporting of results in government databases [15]. For example, the Fair Access to Clinical Trial Act of 2005 would see the ClinicalTrials.gov database expanding to include a clinical trials registry and clinical trials results.

Significant challenges to the registration of clinical trials and disclosure of results lie ahead, including agreement between the pharmaceutical industry and other stakeholders as to the minimal data set for trial registration [51], validation of the registry data, prevention of duplicate registrations, and defining and naming of interventions and the format of reporting and validation of results [15]. Nevertheless, the climate is changing. As noted by a recent ICMJE editorial, “three years ago trial registration was the exception; now it is the rule” [54].

**Publication of Clinical Trial Results** “Both authors and publishers have ethical obligations. In publication of the results of research, the investigators are obliged to preserve the accuracy of the results. Negative as well as positive results should be published or otherwise publicly available” (Declaration of Helsinki, Article 29).

In addition to the registration of trial information and initiatives under way to promote the public disclosure of trial results, several important policy and guidance documents have been developed by professional bodies to promote the quality of reporting of clinical trials in biomedical journals.

**International Committee of Medical Journal Editors** Three policy initiatives of the ICMJE, a committee representing 12 member biomedical journals, are discussed below. These include the requirement to register trials as a prerequisite to publication and requirements for authorship and disclosure of any conflicts of interest on published manuscripts. Additional information can be found in the Uniform Requirements for Manuscripts (URM) Submitted to Biomedical Journals, a set of ethical principles and recommendations developed by the ICJME to promote the accurate reporting of biomedical studies [56]. These guidelines have been adopted by many journals.



1. *Requirement to Register Trial: Prerequisite to Publication* In 2005 the ICMJE established a policy that requires information about trial design to be prospectively registered in recognized databases as a prerequisite to publication in member journals as follows [53]:

The ICMJE member journals will require, as a condition of consideration for publication, registration in a public trials registry. Trials must register at or before the onset of patient enrollment. This policy applies to any clinical trial starting enrollment after July 1, 2005. For trials that began enrollment prior to this date, the ICMJE member journals will require registration by September 13, 2005, before considering the trial for publication.

A marked increase in the number of registrations lodged with the five recognized registries was seen around the time of implementation of the policy. The largest existing registry, ClinicalTrials.gov, which registered around 30 trials per week before September 2005, received over 200 registrations per week after September 2005 [15]. The number of registrations continued to grow as many other biomedical journals embraced this policy.

Although early-phase studies (including phase 1 trials and pharmacokinetic studies) were initially excluded from this requirement [53], the ICMJE has recently adopted the WHO definition of clinical trials as “any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes” [54]. Thus, as of July 1, 2008, all early-phase studies will require, as prerequisite to their publication, registration on a recognized database. Importantly, in addition to the five previously recognized registries, the ICMJE will now accept registration on primary registries participating in the WHO ICTRP project [54].

2. *Requirements for Authorship* The URM lists three requirements for authorship as follows [56]:

- Substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data
- Drafting the article or revising it critically for important intellectual content
- Final approval of the version to be published

Authors should meet all these conditions and no authors who fulfill these criteria should be omitted. The URM further states that “acquisition of funding, collection of data, or general supervision of the research group, alone, does not justify authorship.” While these guidelines have been widely adopted, the requirements of individual journals should be observed.

3. *Conflict of Interest* In-line with the Declaration of Helsinki (Article 27), which states, “Sources of funding, institutional affiliations and any possible conflicts of interest should be declared in the publication,” the URM requires: “When authors submit a manuscript, whether an article or a letter, they are responsible for disclosing all financial and personal relationships that might bias their work. To prevent ambiguity, authors must state explicitly whether potential conflicts do or do not exist.” Again, individual journals often have their own conflict of interest policies.

*Good Publication Practice (GPP) for Pharmaceutical Companies* The GPP guidelines set forth principles that aim to increase the transparency of the processes

involved in the publication of industry-sponsored trials, reduce publication bias, and clarify the roles of the sponsor company and academic investigators [57, 58]. The guideline includes topics such as publication standards, premature publication, duplicate publication, authorship, and the role of professional medical writers.

A further comment on the role of professional medical writers is made here. Ghost writing occurs when a professional writer controls the content of the manuscript and discloses neither their involvement nor their funding source. This is ethically unacceptable. However, working with a professional medical writer should pose no ethical concerns provided that (1) the medical writing assistance is acknowledged and (2) their source of funding is disclosed [59]. Throughout the writing of the manuscript, authors should maintain responsibility and control for the content of the publication, including key messages and data to be presented.

*CONSORT (Consolidated Standards of Reporting Trials) Statement* The CONSORT group (which comprises trialists, methodologists, and medical journal editors [60]) has developed several initiatives to improve the reporting of randomized clinical trials, including the CONSORT statement. The CONSORT statement is an evolving guideline that was originally published in 1996 [61] and revised in 2001 [48]. Following a meeting of the CONSORT group in January 2007, a further update of the CONSORT statement is due for release. The latest versions of all CONSORT documents can be found on the CONSORT group website [60].

In its current form, the CONSORT statement includes a 22-item checklist and flow diagram that is intended to improve the quality of the writing and assist in the review and evaluation of RCT reports. In particular, the checklist items encourage transparency in the reporting of the trial methods and results, which together account for 17 of the 22 items. The flow diagram depicts the number of participants progressing through the enrollment, intervention allocation, follow-up, and data analysis stages of a clinical trial, allowing the reader to evaluate the basis of any analysis undertaken. Although the main CONSORT statement is directed at two-group parallel RCTs, extensions to the statement address alternative trial designs, including cluster trials [62] and noninferiority and equivalence randomized trials [63]. A checklist of items that should be included in conference or journal abstracts is due for publication in 2007 [60]. It is highly recommended that the requirements of the CONSORT statement be considered early in the trial design process.

## 24.2.7 ETHICALLY CHALLENGING SCENARIOS IN CLINICAL RESEARCH

### 24.2.7.1 Children

Children, by virtue of their developmental and cognitive abilities, are a vulnerable population in relation to research participation. The Willowbrook Hepatitis Study [2] and many other similar questionable pediatric research studies have shown us that unscrupulous researchers have resorted to the use of children in research trials because they were seen as easy to recruit and manipulate. With the advent of the Nuremberg Code, which required informed consent, research in children declined because it was seen as almost impossible to get fully informed consent from a child of limited cognitive ability. For this reason and the fact that pharmaceutical manu-

facturers saw pediatric markets as too small to warrant expensive pediatric clinical trials, the majority of medications in clinical use today have not been properly tested in children but are being used “off label.” This in itself is a major ethical dilemma as several important differences in disease pathophysiology, pharmacokinetics, pharmacodynamics, and adverse drug reactions occur between children and adults. For example, the effective dose of a single application of an asthma nebulizer will be higher in an adult with effective hand/breathing coordination than a relatively uncoordinated small child. The ability of diazepam to cause sedation in adults but agitation in some children is an example of the wildly varying physiological responses to a drug. The absence of adequate studies in the pediatric population hinders the effective provision of the medical care of children. As such, there has been a recent push by U.S. and European regulators to require drug manufacturers to assess the safety and effectiveness of medications in pediatric patients when pediatric use is anticipated [64, 65]. The U.S. Best Pharmaceuticals for Children Act—2002 creates financial incentives for pharmaceutical companies to engage in pediatric research by allowing a 6-month extension to their exclusive patent protection and funding for research of drugs that no longer have patent protection.

While distributive justice mandates that research should be conducted on children to enhance their health care, there are some instances when it is absolutely inappropriate to conduct research on children [66]. It is ethically appropriate to exclude children from research when:

- (a) The research topic is irrelevant to children (e.g., ischemic heart disease, alzheimer’s disease).
- (b) Existing data in adults are sufficient to judge risks in children.
- (c) Local laws or regulations bar inclusion of children in this type of research.
- (d) A nontherapeutic study which places a child at greater than small increase in minimal risk. Minimal risk is defined as a risk that is “greater than those already encountered in daily life or during the performance of routine physical or psychological examinations or tests.” The FDA, DHHS, and National Institutes of Health (NIH) have produced guidelines on categories of risk in pediatric research and the associated approval requirements [66]. While the degree of risk associated with “routine life” is open to varying interpretations by individual IRBs, it would generally be considered unethical to expose healthy children to any material risk of a serious adverse outcome, such as a phase I study of a new class of drug.

Pediatric research raises two potential dilemmas above and beyond the ethical dilemmas routinely seen in adult research [65–67]:

- (a) *Informed Consent* Informed consent, a process where research participants are given sufficient information about a study to allow them to determine if it the risks and benefits warrant their own involvement, is not possible for a child with limited cognitive powers below the legal age of “consent.” Pediatric research therefore requires a two-part consent process—parental permission and the child’s assent. Central to the requirement of parental permission is the assumption that a loving, protective relationship exists between the

parent/guardian and the child. This assumption has two weaknesses. First, the theologian Paul Ramsey [68] states that a parent does not have the moral authority to permit a child to be put at risk of harm in an experiment from which the child could not benefit medically as a child is not a piece of property but an autonomous individual. Second, in dysfunctional relationships the parents may not always have the child's best welfare at hand. Therefore, if the researcher has any suspicion of a dysfunctional relationship between child and parent, they should defer any decisions concerning that child's participation to an independent IRB. After a parent has given permission for their child to be involved in research, the child's assent must be sought. The child should be given a simple explanation about why the research is being conducted and what procedures and discomfort may occur if they are to be involved. Researchers also need to define risks with the child's perspective in mind, not that of an adult. For example, separation from parents and unfamiliar surroundings may be considered a risk from the perspective of a child but not an adult. Assent should be sought apart from the parents' direct influence and the child must be informed that it is their right to refuse involvement. While childhood assent is always preferable, there are several instances where it is not mandated. For example, assent is not required if the child's capacity to understand the research is limited (intellectual age of less than 7 years, American Academy of Pediatrics) or the potential direct benefit to the health of the child is high and only available through direct participation in the research (e.g., new chemotherapy for a childhood cancer). Apart from these limitations, failure to obtain a child's assent overrides a parent's permission and prohibits that child's involvement in research.

- (b) *Payment for Research Participation* Payment to pediatric research participants becomes particularly problematic because of the child's vulnerability and dependence on their parents as the primary decision makers. Poor families are particularly vulnerable to monetary rewards, and therefore offering large financial incentives may create unnecessary coercion by the parents for their children to participate in research. Only small token gestures of appreciation or reimbursement of direct costs of involvement should be made to ensure payment is not the reason parents are enrolling their children in the research. Furthermore, incentives given to children should not be excessive and, according to American Academy of Paediatrics guidelines [69], should only be revealed after enrollment so that the child's participation is voluntary. This however is contrary to usual informed consent guidelines where incentives should be discussed before enrollment. A reasonable compromise may be to mention in the consent process that a gift will be given to participants without outlining its exact nature.

#### **24.2.7.2 Mentally Impaired**

Over the last 50 years very effective pharmacological treatments for mental disorders such as depression, psychosis, bipolar disorder, and obsessive-compulsive disorder have seen the transition of mental health care from institutions to the general community, resulting in huge benefits to patients and society in general. Without research, none of these life-changing benefits would have materialized.

Guideline 15 of the International Ethical Guidelines for Biomedical Research Involving Human Subjects [12] states that research on mentally impaired individuals is permissible provided that the purpose of the research is to obtain knowledge relevant to the particular health needs of the mentally impaired and that the research cannot be equally well carried out on persons with full mental capacity.

According to Michels [70] the conduct of clinical research in mentally ill patients raises three key ethical questions. Are mentally ill subjects especially vulnerable to exploitation? Are they competent to give informed consent? Are psychiatric research methods particularly dangerous? Following the suicide of a patient with schizophrenia after being asked to cease their antipsychotic medication in the “wash out” period of a clinical trial, regulators have been struggling with the ethical dilemma of whether mentally ill patients should be allowed to participate in research. The U.S. National Bioethics Advisory Commission (NBAC) and DHHS have both published guidelines regarding the ethical conduct of clinical research in mentally ill patients [71, 72]. A summary and discussion of these recommendations are presented below:

- (a) *Impaired Capacity* According to the NBAC, mental capacity to consent to research requires that the subject understand the difference between treatment and research, the nature of the research being conducted, its risks and benefits, available alternatives, and the fact that he or she is making a decision and that this decision can be changed. The subject must not be swayed by a pathological affective state, a false belief, or a dependent relationship that might interfere with the decision or his or her autonomy and must be capable of making a stable, reasoned choice and communicating it. It should be noted that merely having a mental illness does not mean that an individual’s mental capacity to give informed consent is impaired. Different mental illnesses have varying effects on a person’s decision-making capacity. Furthermore, illnesses such as schizophrenia can produce fluctuating levels of capacity as an individual passes through stages of good and bad mental health. It is therefore advisable that researchers carefully consider the timing of recruitment of participants into research studies to avoid periods of heightened vulnerability due to acute mental illness. Furthermore, while not yet mandated, it may be preferable to get an independent expert in mental health to assess a potential research subject’s capacity to give informed consent, especially if this research places a participant at more than minimal risk. Finally, as informed consent is a continual process, a research participant’s capacity to consent should be continually monitored throughout the entire life of the study, preferably by an independent monitor.
- (b) *IRB Provisions* Under U.S. regulations, IRBs that regularly review research involving vulnerable persons should have one or more individuals on that IRB who are expert within the field. In the area of mental health research it has been suggested that an expert in the area of mental health research and an individual from a relevant mental health support group would provide the best assessment of research protocols and therefore protection of research participants.
- (c) *Surrogate Permission* Where permitted by law, individuals with impaired capacity may have a family member or other legally authorized representa-

tive serve as a surrogate for research decisions. Surrogates should be informed of the risks, benefits, and alternatives to the research when they are providing permission. The surrogate should be educated that it is their duty to first act in a manner that reflected the views of the mentally incapacitated individual while they were decisionally capable (substituted judgment). If the values of the individual are unknown, a “best interest” standard should be used. The autonomy of the patient should be respected and therefore their assent to participation in research should be obtained whenever possible. Their right to withdraw from research should be respected unless the research treatment may be in the best interests of the mentally incapacitated patient and not available outside of a research setting.

#### **24.2.7.3 Women and Pregnancy**

Prior to the 1990s women were systematically excluded from participating in research studies primarily because of the fear that unrecognized pregnancy might place the early embryo at risk, increasing liability risk for researchers. Such a policy of exclusion is not just since it leads to an unequal knowledge base regarding the correct use of medications in the male and female population, potentially reducing the quality of clinical care provided to women. Differences in body composition (fat to muscle ratio), hormone fluctuations during the menstrual cycle, and changes in physiology specific to pregnancy make it highly likely that significant differences in pharmacokinetics and pharmacodynamics exist between the two sexes, therefore mandating more balance research protocols [73].

Guideline 16 of CIOMS International Ethical Guidelines [12] states that the potential for becoming pregnant during a study should not in itself be used as a reason for precluding or limiting participation. A paternalistic approach where an investigator, sponsor, or IRB limits a woman’s capacity to be involved in research because of a potential risk to the fetus infringes a woman’s autonomy. McCullough [74] argues that a previsible fetus does not have the moral or legal status of an autonomous individual and as such a woman should have the absolute right to make an autonomous decision concerning her own involvement in potentially harmful research without regard for the well-being of a nonviable fetus. However, in order to sensibly minimize risk to a potential unborn child, ethical guidelines mandate two precautions when conducting research in women of reproductive age [75]. First, research participants should be screened for pregnancy with sensitive and reliable pregnancy tests before enrolment if their participation in the research poses more than a minimal risk to the fetus or if being pregnant may materially increase their own risk during the conduct of the study. Second, researchers should guarantee access to effective contraceptive methods before enrollment in research. Research protocols should not mandate certain types of contraceptive use (e.g., IUCD, contraceptive pill) and respect the women’s autonomy to choose a method, including abstinence, according to her own needs and values.

Clinical research on pregnant women carrying a potentially viable fetus (post 23 weeks gestation) is a larger ethical dilemma as here the fetus is an individual by virtue of its ability to survive *ex utero*. A general prohibition of termination of pregnancy beyond the period of potential viability provides legal support for this moral concept that a fetus’s own health must be considered when conducting research



on women in the later stages of pregnancy. McCullough [74] outlines four criteria for the ethical conduct of phases I–II studies in pregnant women:

- (a) The medication should be reliably predicted to alter the course of pregnant women's condition based on prior animal and human studies.
- (b) Previous animal and/or human studies have documented no deaths or serious or irreversible injury to pregnant women that was caused by the medication.
- (c) Previous animal and/or human studies have documented no deaths or serious or irreversible injury to the fetus that was caused by the medication.
- (d) Previous animal and/or human studies have reported no or only a very low risk of less serious injury to the fetus that was caused by the medication.

Pregnant women should be provided with adequate information regarding the known benefits and risks of research to their own health and that of the fetus to ensure adequate informed consent. Appropriate follow-up of health outcomes of children exposed to an investigational agent during pregnancy should also be arranged to confirm that the investigational agent is not harmful to the fetus.

Consent of a pregnant woman alone is usually sufficient for most research which may have an impact on the fetus. In the United States 45 CFR 46 states that “when research holds out the prospect of direct benefit solely to the fetus then the consent of the pregnant woman and the father should be obtained.” However, international guidelines such as CIOMS [12] and the American College of Obstetricians [73] do not support this need for paternal consent as paternal rights may weaken or infringe maternal autonomy—the right of a woman to take action in decisions that affect her own body and health. At present, the need for maternal, not paternal, informed consent is standard ethical practice in medical research.

#### **24.2.7.4 Critically Ill Patient**

Clinical research involving critically ill patients is necessary to reduce the extreme morbidity and mortality encountered in the intensive care unit (ICU). However, this type of research is ethically challenging for two principal reasons [76]. First, critically ill patients are captive and thus vulnerable as they are dependent on the ICU team for all aspects of care or even life itself. This vulnerability raises concerns about the patient or their surrogate's ability to give informed, autonomous consent to participate in clinical research, even if they can judge the pros and cons of participation. Second, most critically ill patients lack decisional capacity due to their illness (head injury, delirium, and drug-induced coma) or the drugs (sedatives, analgesics) that they are receiving. As a result, their family members are usually asked to provide permission for research participation. While a recent study concerning research in the ICU suggests that surrogates do predict the choices of patients with a reasonably high degree of accuracy [77], there is the potential for a surrogate to make a decision that ultimately contravenes the participant's real wishes.

At the present moment there is no legislation in place governing research in the setting of the critically ill patient. The American Thoracic Society [78] and the American Heart Association [79] have published some very useful guidelines concerning involvement of critically ill patients in clinical research. These guidelines



mirror issues discussed in previous paragraphs related to research in vulnerable persons such as the mentally ill. However, we will outline some potential solutions specific to the ICU setting:

- (a) *Separation of Clinical and Experimental Team* Due to the extreme vulnerability of the critically ill patient and their dependence on the ICU team, it is probably best that researchers separate from the clinical team ask patients or their surrogates for permission to be involved in any study. This ensures that patients/surrogates do not confuse research with therapeutic intent and can feel free to refuse involvement without any perceived pressure that such refusal may harm the standard of routine care being provided to the patient.
- (b) *Consent* As most patients are not in a position to give their own consent, it is permissible for family members to provide surrogate consent provided that the research is of minimal risk or may have some direct benefit to the patient. It is probably unethical to allow a surrogate to make a decision regarding a third person (the patient) that places that individual at more than minimal risk. Alternatives to the use of surrogate consent include retrospective consent, prospective consent, and waiver of consent by the IRB. An example of retrospective consent would be when a small sample of blood is taken as part of routine clinical care and then stored for potential later use in research once the patient has recovered and given consent. In prospective consent a noncritically ill patient is consented for research in the event that he or she may become critically ill later. For example, preoperatively a patient undergoing major surgery may be consented to research on a new therapy to prevent lung infections related to prolonged artificial ventilation in the ICU. Finally, in limited studies on true emergencies where the intervention is minimal, an IRB may waive the need for patient or surrogate consent. For example, a randomized trial between a manual CPR technique and a new but FDA-approved mechanical CPR machine may be undertaken without consent as both are reasonable care and there is no time for consent in a cardiac arrest [29].

#### 24.2.7.5 Research in Undeveloped World

The conduct of clinical research in undeveloped countries raises three principal ethical concerns [80, 81]. First, many individuals from the undeveloped world have only very basic education, making it difficult for them to understand abstract concepts such as placebos, making informed decisions on the benefits and disadvantages of research involvement difficult. Furthermore, health care expenditure in the undeveloped world is paltry compared to western society, with even basic medical care often lacking. If an individual has the chance of receiving some medical treatment in the form of a trial medication, compared to no chance of medical care outside of a research protocol, does this really allow for an autonomous unpressured decision? For these reasons it is potentially more difficult to obtain truly informed consent in the undeveloped world.

The second major ethical issue surrounding research in the developing world is exploitation of communities that are unlikely to benefit themselves from the medical

advances developed by the research. Guideline 10 of the CIOMS ethical guidelines [12] states that “research must be responsive to the needs and the priorities of the population or community in which it is carried out.” Furthermore, the guideline goes on to state that “any intervention or product developed, or knowledge generated, must be made reasonably available for the benefit of that population or community.” It is only just that a community that may potentially be harmed by research should also share in its final benefits (new proven therapy). Unfortunately this has not always been the case. For example, the AIDS Clinical Trials Group protocol 076 was a study in which pregnant women from the undeveloped world were randomized to receive either placebo/routine care or oral antiviral medication throughout pregnancy, intravenous antiviral during labor, and postpartum oral antiviral for the baby in order to determine if such intensive treatment could reduce the vertical transmission of HIV to the child. While this trial did show a very significant reduction in vertical transmission, it was heavily criticized as unethical because the proposed treatment is unlikely to ever be reasonably available to mothers in the third world because of significant cost. The trial protocol cost \$800 per woman and child, 600 times the annual per-capita allocation for health care in Malawi [82]. As such, while proven to be an effective therapy for women in the western world, the results of the trial are unlikely to ever be practically applicable to the community in which the information was gained, making the study exploitative and unethical. Clinical trials can only be ethically performed in the developing world if the pathology in question is clinically relevant to the developing community and the trial intervention, if proven to be effective, is likely to be reasonably available to the community after completion of the research.

The final ethical concern for research in the developed world is an appropriate choice for the “alternative” control treatment in a trial. Some commentators suggest that it is ethical to use a placebo in a clinical trial in the undeveloped world, as no treatment is the default “standard of care” present in these poor countries [83]. These individuals suggest that since half the participants will be exposed to a potentially useful study medication, the “greater good” is still being done, making the study ethical. However, the majority view is to the contrary. The CIOMS guideline [12] states that research participants in the control arm of a trial should receive an established effective intervention (guideline 11). The use of a placebo is only ethically justifiable if there is no established effective intervention or delaying the use of this established intervention would not add any serious risk or cause any irreversible harm. As the CIOMS Guideline 11 states [12], “An economic reason for unavailability of an established effective intervention cannot justify a placebo-controlled study in a country of limited resources when it would be unethical to conduct a study with the same design in a population with general access to the effective treatment outside the study.” The reader is referred to the CIOMS Guideline for a wider discussion of this topic.

#### **24.2.8 FOUR GOLDEN RULES OF ETHICAL CONDUCT IN CLINICAL RESEARCH**

After consideration of the various philosophical theories related to medical ethics and review of the relevant published guidelines it is possible to summarize ethical

research practice into four guiding principles. These principals, or golden rules, are summarized by the acronym AIMS to help facilitate recall, where A refers to respecting patient autonomy, I stands for maximizing the potential impact of the research on society in general, M refers to the mitigation of potential harm to research participants, and S stands for maintenance of scientific integrity in all aspects of clinical research. The following sections outline the key features of each of these golden rules.

#### **24.2.8.1 A: Respect for Patient Autonomy**

Researchers should respect an individual's right to make decisions regarding their future by providing them with information which honestly outlines the risks and benefits of participating in the research in terms that are easily understandable to the average lay person of limited education (informed consent). Researchers should avoid offering excessive financial or other incentives that may impede unbiased decision making on the part of the potential research participant. When a subject lacks the ability to make an informed decision themselves (child, mental incapacity), the researcher must gain informed consent from a legally authorized representative. Research participants must feel free to exit a study at any stage without fear of compromising their standard medical care. Research participants should expect the same degree of respect of their privacy and dignity as would be afforded patients undergoing standard medical care. Maintenance of individual participant's privacy during the conduct of the trial and in the publication of trial results is paramount.

#### **24.2.8.2 I: Maximization of Research Impact on Medical Treatment**

The impact of a piece of research on contemporary medical understanding/treatment can be maximized by considering four issues. First, research protocols should be scientifically well designed so that they have the capacity to answer the question of interest. Poorly designed trials with clinically irrelevant endpoints, inadequate sample sizes, or inherent biases that make firm conclusions impossible are unethical as they consume scarce resources while placing participants at risk of adverse events for no real gain. Proper peer review and IRB assessment should prevent such studies from even commencing. Second, it is important that trials are not designed with very limited inclusion criteria (e.g., white, male, normal weight, 30–40 years age) as the results of these types of trials may not be applicable to the population in general. Justice requires that, if possible, a broad cross section of the community should be approached to be involved in research so that the potential benefits are equitably distributed throughout society. Third, it is a moral imperative that the results of research must be published within the scientific literature, irrespective of the research outcome. Failure to publish negative outcomes could create a situation where an identical study is replicated by an independent research group unaware of the original study outcome, thereby exposing the second group of research participants to unnecessary risk while wasting valuable research resources. Finally, researchers should only conduct experiments that have scientific and social worth and not use human beings in frivolous research.

**24.2.8.3 M: Minimization of Risk to Research Participants**

According to the theory of utilitarianism, it is ethically acceptable to expose an individual to justifiable risk provided the greater good is served. However, there is a moral obligation on the part of the researcher to mitigate risk by designing protocols that minimize risk or provide backup that reduces the consequences of an adverse event if it were to occur. For example, before conducting a phase I study investigating a new drug in healthy volunteers it is imperative that extensive testing in several different animal models has shown this investigational agent to be relatively safe. Furthermore, clinical trials should be supervised by appropriately trained and qualified medical personnel that are able to identify and treat any adverse reaction to a trial medication. The research should be conducted in a facility that has the appropriate equipment available to treat any medical complication of the research. Data-monitoring committees, independent of the research sponsor, should monitor trial data during the conduct of the trial to ensure the welfare of trial participants. These committees significantly enhance the safety of clinical research as they enable early termination of unsafe trials. Finally, it is a legal and moral imperative that all research participants are covered by insurance that will provide free medical care and compensation if they are inadvertently injured as a result of their participation in the research.

**24.2.8.4 S: Scientific Integrity**

Researchers have an ethical obligation to objectively analyze and report data. Improper data analysis or biased reporting of results in the scientific literature is unethical as it could lead to the adoption of medical treatments as standard proven therapy when in fact these treatments may be at best useless or worse even harmful. Careful analysis and recording of data, sharing of data and resources, and respect of research colleagues are all important components of ethical research conduct.

**24.2.9 CONCLUSION**

Our understanding of what constitutes ethical conduct in clinical research has come a long way since the advent of the Nuremberg Code in 1949. Today, the clinical researcher has several excellent guidelines [5, 11, 12] to help with the design and conduct of clinical studies that maximize participant safety and autonomy while still allowing appropriate scientific experimentation. This chapter has outlined the ethical principles used in the development of these clinical research guidelines. It is hoped that with an in-depth knowledge of these principles, researchers will be in a better position to design and conduct scientifically and ethically appropriate clinical trials. We hope that this chapter will facilitate a better understanding of ethical principles and the importance of ethics to clinical research practice.

**APPENDIX A NUREMBERG CODE\***

1. The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonable to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment.

The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity.

2. The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.
3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment.
4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.
5. No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.
6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.
7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death.
8. The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.
9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.

\*Reprinted from *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10*, Vol. 2, pp. 181–182. Washington, DC: U.S. Government Printing Office, 1949.

10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful judgment required of him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

## **APPENDIX B WORLD MEDICAL ASSOCIATION DECLARATION OF HELSINKI: ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS\***

Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964, and amended by the

29th WMA General Assembly, Tokyo, Japan, October 1975

35th WMA General Assembly, Venice, Italy, October 1983

41st WMA General Assembly, Hong Kong, September 1989

48th WMA General Assembly, Somerset West, Republic of South Africa, October 1996 and the 52nd WMA General Assembly, Edinburgh, Scotland, October 2000

Note of Clarification on Paragraph 29 added by the WMA General Assembly, Washington 2002.

Note of Clarification on Paragraph 30 added by the WMA General Assembly, Tokyo 2004.

### **A. Introduction**

1. The World Medical Association has developed the Declaration of Helsinki as a statement of ethical principles to provide guidance to physicians and other participants in medical research involving human subjects. Medical research involving human subjects includes research on identifiable human material or identifiable data.

*\*Clarification of paragraph 29:* The WMA hereby reaffirms its position that extreme care must be taken in making use of a placebo-controlled trial and that in general this methodology should only be used in the absence of existing proven therapy. However, a placebo-controlled trial may be ethically acceptable, even if proven therapy is available, under the following circumstances:

- where for compelling and scientifically sound methodological reasons its use is necessary to determine the efficacy or safety of a prophylactic, diagnostic, or therapeutic method or
- where a prophylactic, diagnostic, or therapeutic method is being investigated for a minor condition and the patients who receive placebo will not be subject to any additional risk of serious or irreversible harm.

All other provisions of the Declaration of Helsinki must be adhered to, especially the need for appropriate ethical and scientific review.

*Clarification of paragraph 30:* The WMA hereby reaffirms its position that it is necessary during the study-planning process to identify posttrial access by study participants to prophylactic, diagnostic, and therapeutic procedures identified as beneficial in the study or access to other appropriate care. Posttrial access arrangements or other care must be described in the study protocol so the ethical review committee may consider such arrangements during its review.

2. It is the duty of the physician to promote and safeguard the health of the people. The physician's knowledge and conscience are dedicated to the fulfillment of this duty.
3. The Declaration of Geneva of the World Medical Association binds the physician with the words, "The health of my patient will be my first consideration," and the International Code of Medical Ethics declares that, "A physician shall act only in the patient's interest when providing medical care which might have the effect of weakening the physical and mental condition of the patient."
4. Medical progress is based on research which ultimately must rest in part on experimentation involving human subjects.
5. In medical research on human subjects, considerations related to the well-being of the human subject should take precedence over the interests of science and society.
6. The primary purpose of medical research involving human subjects is to improve prophylactic, diagnostic and therapeutic procedures and the understanding of the aetiology and pathogenesis of disease. Even the best proven prophylactic, diagnostic, and therapeutic methods must continuously be challenged through research for their effectiveness, efficiency, accessibility and quality.
7. In current medical practice and in medical research, most prophylactic, diagnostic and therapeutic procedures involve risks and burdens.
8. Medical research is subject to ethical standards that promote respect for all human beings and protect their health and rights. Some research populations are vulnerable and need special protection. The particular needs of the economically and medically disadvantaged must be recognized. Special attention is also required for those who cannot give or refuse consent for themselves, for those who may be subject to giving consent under duress, for those who will not benefit personally from the research and for those for whom the research is combined with care.
9. Research Investigators should be aware of the ethical, legal and regulatory requirements for research on human subjects in their own countries as well as applicable international requirements. No national ethical, legal or regulatory requirement should be allowed to reduce or eliminate any of the protections for human subjects set forth in this Declaration.

## **B. Basic Principles for All Medical Research**

10. It is the duty of the physician in medical research to protect the life, health, privacy, and dignity of the human subject.
11. Medical research involving human subjects must conform to generally accepted scientific principles, be based on a thorough knowledge of the scientific literature, other relevant sources of information, and on adequate laboratory and, where appropriate, animal experimentation.
12. Appropriate caution must be exercised in the conduct of research which may affect the environment, and the welfare of animals used for research must be respected.



13. The design and performance of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol. This protocol should be submitted for consideration, comment, guidance, and where appropriate, approval to a specially appointed ethical review committee, which must be independent of the investigator, the sponsor or any other kind of undue influence. This independent committee should be in conformity with the laws and regulations of the country in which the research experiment is performed. The committee has the right to monitor ongoing trials. The researcher has the obligation to provide monitoring information to the committee, especially any serious adverse events. The researcher should also submit to the committee, for review, information regarding funding, sponsors, institutional affiliations, other potential conflicts of interest and incentives for subjects.
14. The research protocol should always contain a statement of the ethical considerations involved and should indicate that there is compliance with the principles enunciated in this Declaration.
15. Medical research involving human subjects should be conducted only by scientifically qualified persons and under the supervision of a clinically competent medical person. The responsibility for the human subject must always rest with a medically qualified person and never rest on the subject of the research, even though the subject has given consent.
16. Every medical research project involving human subjects should be preceded by careful assessment of predictable risk and burdens in comparison with foreseeable benefits to the subject or to others. This does not preclude the participation of healthy volunteers in medical research. The design of all studies should be publicly available.
17. Physicians should abstain from engaging in research projects involving human subjects unless they are confident that the risks involved have been adequately assessed and can be satisfactorily managed. Physicians should cease any investigation if the risks are found to outweigh the potential benefits or if there is conclusive proof of positive and beneficial results.
18. Medical research involving human subjects should only be conducted if the importance of the objective outweighs the inherent risks and burdens to the subject. This is especially important when the human subjects are healthy volunteers.
19. Medical research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research.
20. The subjects must be volunteers and informed participants in the research project.
21. The right of research subjects to safeguard their integrity must always be respected. Every precaution should be taken to respect the privacy of the subject, the confidentiality of the patient's information and to minimize the impact of the study on the subject's physical and mental integrity and on the personality of the subject.
22. In any research on human beings, each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail. The subject should

be informed of the right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal. After ensuring that the subject has understood the information, the physician should then obtain the subject's freely-given informed consent, preferably in writing. If the consent cannot be obtained in writing, the non-written consent must be formally documented and witnessed.

23. When obtaining informed consent for the research project the physician should be particularly cautious if the subject is in a dependent relationship with the physician or may consent under duress. In that case the informed consent should be obtained by a well-informed physician who is not engaged in the investigation and who is completely independent of this relationship.
24. For a research subject who is legally incompetent, physically or mentally incapable of giving consent or is a legally incompetent minor, the investigator must obtain informed consent from the legally authorized representative in accordance with applicable law. These groups should not be included in research unless the research is necessary to promote the health of the population represented and this research cannot instead be performed on legally competent persons.
25. When a subject deemed legally incompetent, such as a minor child, is able to give assent to decisions about participation in research, the investigator must obtain that assent in addition to the consent of the legally authorized representative.
26. Research on individuals from whom it is not possible to obtain consent, including proxy or advance consent, should be done only if the physical/mental condition that prevents obtaining informed consent is a necessary characteristic of the research population. The specific reasons for involving research subjects with a condition that renders them unable to give informed consent should be stated in the experimental protocol for consideration and approval of the review committee. The protocol should state that consent to remain in the research should be obtained as soon as possible from the individuals or a legally authorized surrogate.
27. Both authors and publishers have ethical obligations. In publication of the results of research, the investigators are obliged to preserve the accuracy of the results. Negative as well as positive results should be published or otherwise publicly available. Sources of funding, institutional affiliations and any possible conflicts of interest should be declared in the publication. Reports of experimentation not in accordance with the principles laid down in this Declaration should not be accepted for publication.

### **C. Additional Principles for Medical Research Combined with Medical Care**

28. The physician may combine medical research with medical care, only to the extent that the research is justified by its potential prophylactic, diagnostic or therapeutic value. When medical research is combined with medical care, additional standards apply to protect the patients who are research subjects.
29. The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic

methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists.

30. At the conclusion of the study, every patient entered into the study should be assured of access to the best proven prophylactic, diagnostic and therapeutic methods identified by the study.
31. The physician should fully inform the patient which aspects of the care are related to the research. The refusal of a patient to participate in a study must never interfere with the patient-physician relationship.
32. In the treatment of a patient, where proven prophylactic, diagnostic and therapeutic methods do not exist or have been ineffective, the physician, with informed consent from the patient, must be free to use unproven or new prophylactic, diagnostic and therapeutic measures, if in the physician's judgement it offers hope of saving life, re-establishing health or alleviating suffering. Where possible, these measures should be made the object of research, designed to evaluate their safety and efficacy. In all cases, new information should be recorded and, where appropriate, published. The other relevant guidelines of this Declaration should be followed.

## REFERENCES

1. Shamoo, A., and Resnik, D. (2003), *Responsible Conduct of Research*, Oxford University Press, Oxford.
2. Loue, S. (2000), *Textbook of Research Ethics. Theory and Practice*, Kluwer Academic/Plenum, New York.
3. Kant, I. (1998), *Groundwork of the Metaphysics of Morals*, Cambridge University Press, Cambridge.
4. Shaw, W. (1998), *Contemporary Ethics. Taking Account of Utilitarianism*, Blackwell, London.
5. World Medical Association Declaration of Helsinki (2000), Ethical Principals for Medical Research Involving Human Subjects; available at: <http://www.wma.net/e/policy/b3.htm>.
6. The Nuremberg Code (1949), *International Principals for Human Experimentation, from the Trials of War Criminals before the Nuremberg Military Tribunals under the Control Council Law No. 10*, Vol. 2, Government Printing Office, Washington D.C., pp. 181–182.
7. Beauchamp, T., and Childress, J. (1994), *Principals of Biomedical Ethics*, Oxford University Press, New York.
8. Vonderlehr, R., Clark, T., Wenger, O., et al. (1936), Untreated syphilis in the male Negro: A comparative study of treated and untreated cases, *JAMA*, 107, 856–860.
9. Jones, J. H. (1993), *Bad Blood: The Tuskegee Syphilis Experiment*, Free Press, New York.
10. Roelcke, V. (2004), Nazi medicine and research on human beings, *Lancet*, 364(Suppl.), 6–7.
11. The Belmont Report (1979), *Ethical Principals and Guidelines for the Protection of Human Subjects of Research*, Government Printing Office, Washington, D.C.
12. Council for International Organizations of Medical Sciences (CIOMS) (2002), *International Ethical Guidelines for Biomedical Research Involving Human Subjects*, World Health Organization, Geneva.

13. Office of Human Research Protections, U.S. Department of Health and Human Services; available at: <http://www.hhs.gov/ohrp/international/HSPCompilation.pdf>, accessed December 2007.
14. CONSORT Statement; available at: <http://www.consort-statement.org/?o=1399>, accessed December 2007.
15. Zarin, D. A., Ide, N. C., Tse, T., et al. (2007), Issues in the registration of clinical trials, *JAMA*, 297(19), 2112–2120.
16. OECD Principles of Good Laboratory Practice (GLP), available at: <http://www.oecd.org>, accessed December 2007.
17. Good Laboratory Practice (GLP) regulations; available at: <http://www.cfsan.fda.gov/~dms/opa-pt58.html>; and <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=58>, accessed December 2007.
18. Ashcroft, R. E., Chadwick, D. W., Clark, S. R., et al. (1997), Implications of socio-cultural contexts for the ethics of clinical trials, *Health Technol. Assess.*, 1(9), i–iv, 1–65.
19. ICH (2000), *ICH Harmonised Tripartite Guideline: Choice of Control Group and Related Issues in Clinical Trials E10*, Federal Register, vol. 66, no. 93, 24390–24391. May 14, 2001
20. Simon, S. D. (2001), Is the randomized clinical trial the gold standard of research? *J. Androl.*, 22(6), 938–943.
21. Edwards, S. J., Lilford, R. J., Braunholtz, D. A., et al. (1998), Ethical issues in the design and conduct of randomised controlled trials. *Health Technol. Assess.*, 2(15), i–vi, 1–132.
22. Fries, J. F., and Krishnan, E. (2004), Equipoise, design bias, and randomized controlled trials: The elusive ethics of new drug development, *Arthritis Res. Ther.*, 6(3), R250–255.
23. Lilford, R. J. (2003), Ethics of clinical trials from a bayesian and decision analytic perspective: Whose equipoise is it anyway? *BMJ*, 326(7396), 980–981.
24. Freedman, B. (1987), Equipoise and the ethics of clinical research, *N. Engl. J. Med.*, 317(3), 141–145.
25. Djulbegovic, B., Lacey, M., Cantor, A., et al. (2000), The uncertainty principle and industry-sponsored research, *Lancet*, 356(9230), 635–638.
26. Johnson, N., Lilford, R. J., and Brazier, W. (1991), At what level of collective equipoise does a clinical trial become ethical? *J. Med. Ethics*, 17(1), 30–34.
27. Olson, L. G. (2002), Patient-centred equipoise and the ethics of randomised controlled trials, *Monash Bioeth. Rev.*, 21(2), S55–67.
28. Sackett, D. L. (2000), Equipoise, a term whose time (if it ever came) has surely gone, *CMAJ*, 163(7), 835–836.
29. Ashcroft, R. (1999), Equipoise, knowledge and ethics in clinical research and practice, *Bioethics*, 13(3–4), 314–326.
30. Miller, F. G., and Brody, H. (2007), Clinical equipoise and the incoherence of research ethics, *J. Med. Philos.*, 32(2), 151–165.
31. Braunholtz, D. A., Edwards, S. J., and Lilford, R. J. (2001), Are randomized clinical trials good for us (in the short term)? Evidence for a “trial effect,” *J. Clin. Epidemiol.*, 54(3), 217–224.
32. Walach, H., Sadaghiani, C., Dehm, C., et al. (2005), The therapeutic effect of clinical trials: Understanding placebo response rates in clinical trials—a secondary analysis, *BMC Med. Res. Methodol.*, 5, 26.
33. Halpern, S. D., Karlawish, J. H., and Berlin, J. A. (2002), The continuing unethical conduct of underpowered clinical trials, *JAMA*, 288(3), 358–362.
34. Miller, F. G., and Brody, H. (2002), What makes placebo-controlled trials unethical? *Am. J. Bioeth.*, 2(2), 3–9.

35. FDA (1998), FDA Guidance for Institutional Review Boards and Clinical Investigators; available at: <http://www.fda.gov/oc/ohrt/irbs/drugsbiologics.html#study>.
36. EMEA/CPMP (2001), *Position Statement on the Use of Placebo in Clinical Trials with Regard to the Revised Declaration of Helsinki*, EMEA Workshop on Ethical Considerations in Clinical Trials, London.
37. ICH (1996), *ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6*, Federal Register, Vol. 62, no. 90, 25691–25709, May 9, 1997.
38. 21 CFR part 50, Protection of Human Subjects; available at: [www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRsearch.cfm?CFRPart=50](http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRsearch.cfm?CFRPart=50).
39. Guidance Document: Financial Relationships and Interests in Research Involving Human Subjects: Guidance for Human Subject Protection; available at: <http://www.hhs.gov/ohrp/humansubjects/finreltn/fguid.pdf>, accessed December 2007.
40. McKneally, M. F., and Martin, D. K. (2000), An entrustment model of consent for surgical treatment of life-threatening illness: Perspective of patients requiring esophagectomy, *J. Thorac. Cardiovasc. Surg.*, 120(2), 264–269.
41. Robinson, E. J., Kerr, C. E., Stevens, A. J., et al. (2005), Lay public's understanding of equipoise and randomisation in randomised controlled trials, *Health Technol. Assess.*, 9(8), 1–192, iii–iv.
42. Bernstein, M. (2005), Fully informed consent is impossible in surgical clinical trials, *Can. J. Surg.*, 48(4), 271–272.
43. De Vries, R., and Elliott, C. (2006), Why disclosure? *J. Gen. Intern. Med.*, 21(9), 1003–1004.
44. Weinfurt, K. P., Friedman, J. Y., Dinan, M. A., et al. (2006), Disclosing conflicts of interest in clinical research: Views of institutional review boards, conflict of interest committees, and investigators, *J. Law Med. Ethics*, 34(3), 581–591, 481.
45. Dickersin, K. (1990), The existence of publication bias and risk factors for its occurrence, *JAMA*, 263(10), 1385–1389.
46. Song, F., Eastwood, A. J., Gilbody, S., et al. (2000), Publication and related biases, *Health Technol. Assess.*, 4(10), 1–115.
47. Chan, A. W., Hrobjartsson, A., Haahr, M. T., et al. (2004), Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles, *JAMA*, 291(20), 2457–2465.
48. Moher, D., Schulz, K. F., and Altman, D. G. (2001), The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials, *Ann. Intern. Med.*, 134(8), 657–662.
49. Sismondo, S. (2007), Ghost management: How much of the medical literature is shaped behind the scenes by the pharmaceutical industry? *PLoS Med.*, 4(9), e286.
50. <http://ottawagroup.ohri.ca/>.
51. Krleza-Jeric, K. (2005), Clinical trial registration: The differing views of industry, the WHO, and the Ottawa Group, *PLoS Med.*, 2(11), e378.
52. The WHO International Clinical Trials Registry Platform; available at: <http://www.who.int/ictpr/en/>.
53. Clinical trial registration: A statement from the International Committee of Medical Journal Editors; available at: [http://www.icmje.org/clin\\_trial.pdf](http://www.icmje.org/clin_trial.pdf).
54. Laine, C., Horton, R., DeAngelis, C. D., et al. (2007), Clinical trial registration: Looking back and moving ahead, *JAMA*, 298(1), 93–94.
55. <http://www.who.int/ictpr/results/en/>.

56. International Committee of Medical Journal Editors: Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication; available at: <http://www.icmje.org/>.
57. Good Publication Practice Guidelines for Pharmaceutical Companies; available at: <http://www.gpp-guidelines.org/>.
58. Wager, E., Field, E. A., and Grossman, L. (2003), Good publication practice for pharmaceutical companies, *Curr. Med. Res. Opin.*, 19(3), 149–154.
59. Woolley, K. L. (2006), Goodbye ghostwriters!: How to work ethically and efficiently with professional medical writers, *Chest*, 130(3), 921–923.
60. Consort: Transparent Reporting of Clinical Trials; available at: <http://www.consort-statement.org/>.
61. Begg, C., Cho, M., Eastwood, S., et al. (1996), Improving the quality of reporting of randomized controlled trials. The CONSORT statement, *JAMA*, 276(8), 637–639.
62. Campbell, M. K., Elbourne, D. R., and Altman, D. G. (2004), CONSORT statement: Extension to cluster randomised trials, *BMJ*, 328(7441), 702–708.
63. Piaggio, G., Elbourne, D. R., Altman, D. G., et al. (2006), Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement, *JAMA*, 295(10), 1152–1160.
64. McKinney, R. E. (2003), Congress, the FDA, and the fair development of new medications for children, *Pediatrics*, 112, 669–670.
65. Raymond, A. S., and Brasseur, D. (2005), Development of medicines for children in Europe: Ethical implications, *Paediatric Respir. Rev.*, 6, 45–51.
66. Knox, C. A., and Burkhart, P. V. (2007), Issues related to children participating in clinical research, *J. Pediatric Nurs.*, 22(4), 310–318.
67. Burns, J. P. (2003), Research in children, *Crit. Care Med.*, 31(3), S131–136.
68. Ramsey, P. (1976), The enforcement of morals: Nontherapeutic research on children, *Hastings Center Rep.*, 6, 21–30.
69. American Academy of Pediatrics, Committee on Drugs (1995), Guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations, *Pediatrics*, 95, 286–294.
70. Michels, R. (1999), Are research ethics bad for our mental health? *N. Engl. J. Med.*, 340, 1427–1430.
71. NBAC (1998), *Research Involving Persons with Mental Disorders That May Affect Decision Making Capacity*, National Bioethics Advisory Commission, Rockville, MD.
72. Research involving individuals with questionable capacity to consent: Points to consider, U.S. Departments of Health and Human Services; available at: <http://grants1.nih.gov/grants/policy/questionablecapacity.htm>.
73. American College of Obstetricians and Gynecologists (2007), Research involving women, *Obstet. Gynecol.*, 110, 731–736.
74. McCullough, L. B., Coverdale, J. H., and Chervenak, F. A. (2005), A comprehensive ethical framework for responsibly designing and conducting pharmacologic research that involves pregnant women, *Am. J. Obstet. Gynecol.*, 193, 901–907.
75. McCullough, L. B., Coverdale, J. H., and Chervenak, F. A. (2006), Preventative ethics for including women of childbearing potential in clinical trials, *Am. J. Obstet. Gynecol.*, 194, 1221–1227.
76. Bigatello, L. M., George, E., and Hurford, W. E. (2003), Ethical considerations for research in critically ill patients, *Crit. Care Med.*, 31(Suppl 3), s178–181.
77. Coppolino, M., and Ackerson, L. (2001), So surrogate decision makers provide accurate consent for intensive care research? *Chest*, 119, 603–612.

78. American Thoracic Society (2004), The ethical conduct of clinical research involving critically ill patients in the United States and Canada, *Am. J. Respir. Crit. Care Med.*, 170, 1375–1384.
79. American Heart Association (2006), Recommendations for implementation of community consultation and public disclosure under the FDA “exception from informed consent requirements from informed consent”; available at: <http://www.americanheart.org/emergencyexception>.
80. Lackey, D. P. (2001), Clinical trials in developing countries: A review of the moral issues, *M. Sinai J. Med.*, 68, 4–12.
81. Wendler, D., Emanuel, E., and Lie, R. K. (2004), The standard of care debate: Can research in developing countries be both ethical and responsive to those countries needs? *Am. J. Public Hlth.*, 94, 923–928.
82. Varmus, H., and Satcher, D. (1997), Ethical complexities of conducting research in developing countries, *New Engl. J. Med.*, 337, 1003–1005.
83. Phanuphak, P. (1997), Ethical issues in studies in Thailand of the vertical transmission of HIV, *New Engl. J. Med.*, 338, 834–835.



# 25

---

## *Regulations*

**Ramzi Dagher,<sup>1</sup> Rajeshwari Sridhara,<sup>2</sup> Nallaperumal Chidambaram,<sup>3</sup>  
and Brian P. Booth<sup>4</sup>**

*<sup>1</sup>Pfizer, Inc., New London, Connecticut*

*<sup>2</sup>Office of Translational Science, Office of Biostatistics, Division of Biometrics,  
Food and Drug Administration, Rockville, Maryland*

*<sup>3</sup>Office of New Drug Quality Assessment, Division of Post-Marketing Evaluation,  
Food and Drug Administration, Rockville, Maryland*

*<sup>4</sup>Office of Translational Science, Office of Clinical Pharmacology, Division of Clinical  
Pharmacology, Food and Drug Administration, Rockville, Maryland*

### **Contents**

|                                                                                                |      |
|------------------------------------------------------------------------------------------------|------|
| 25.1 Introduction                                                                              | 1154 |
| 25.2 Regulatory History                                                                        | 1154 |
| 25.3 Clinical Investigations                                                                   | 1155 |
| 25.4 NDA Review Process                                                                        | 1157 |
| 25.5 Pediatric Initiatives                                                                     | 1157 |
| 25.6 Role of Statistics in Regulatory Setting                                                  | 1158 |
| 25.7 Role of Chemistry, Manufacturing, and Controls (CMC) Information in<br>Regulatory Setting | 1160 |
| 25.8 Role of Clinical Pharmacology in Regulatory Setting                                       | 1162 |
| 25.8.1 Biopharmaceutics Guidances                                                              | 1163 |
| 25.8.2 Clinical Pharmacology Guidances                                                         | 1164 |
| 25.8.3 Timings of Studies                                                                      | 1165 |
| 25.9 Drug Regulation in a Global Environment                                                   | 1167 |
| Appendix: Clinical Pharmacology Drug Development Questions                                     | 1168 |
| References                                                                                     | 1169 |

The views expressed are those of the authors and do not reflect official policy of the FDA. No official endorsement by the FDA is intended or should be inferred.

---

*Clinical Trials Handbook*, Edited by Shayne Cox Gad  
Copyright © 2009 John Wiley & Sons, Inc.

25.1 INTRODUCTION

The U.S. Food and Drug Administration (USFDA) is a scientific, regulatory, and public health agency that oversees what is estimated to be 25% of the consumer economy. The agency has jurisdiction over most food products, human and animal drugs, therapeutic biologic agents, medical devices, radiation-emitting products, cosmetics, and animal feed [1]. The FDA has a staff of approximately 10,000 employees including chemists, pharmacologists, toxicologists, physicians, microbiologists, veterinarians, pharmacists, lawyers, statisticians, project managers, and others. In this chapter, the role of the FDA and different scientific disciplines in the marketing approval of drug products in the United States are presented.

25.2 REGULATORY HISTORY

The FDA’s mandate has evolved based on a number of legislative acts that have been introduced over the last 100 years. These are discussed below and outlined in Table 1. Regulatory functions were incorporated into the agency’s mission in 1906 with passage of the Federal Food and Drugs Act. The 1906 law forbade interstate and foreign commerce in adulterated and misbranded foods and drugs. Drugs had to abide by standards of purity and quality set forth in the *U.S. Pharmacopeia* and the *National Formulary*, works prepared by committees of physicians and pharmacists, or meet individual standards chosen by their manufacturers and stated on their labels. The law prohibited the adulteration of foods by the removal of valuable constituents, the substitution of ingredients so as to reduce quality, the addition of deleterious ingredients, or the use of spoiled animal and vegetable products.

Another important milestone in the agency’s history took place in 1938 with passage of the Federal Food, Drug, and Cosmetic (FDC) Act, which, for the first time, required new drugs to be shown safe before marketing. The FDC Act also extended the FDA’s jurisdiction to cosmetics and therapeutic devices, authorized factory inspections, and added the remedy of court injunctions to the previous penalties of seizures and prosecutions.

In the late 1950s and early 1960s, a drug used in Europe for its sedative effects (thalidomide) was found to have caused birth defects in thousands of infants. The drug had been kept off the market in the United States. This event, as well as other factors, aroused public support for stronger drug regulation. In 1962, the

TABLE 1 FDA Regulatory Landmarks

| Year | Legislation                     | Features                                                               |
|------|---------------------------------|------------------------------------------------------------------------|
| 1906 | Food and Drugs Act              | Defined FDA regulatory functions                                       |
| 1938 | Food, Drug and Cosmetics Act    | Requirement for safety data                                            |
| 1962 | Kefauver–Harris Drug Amendments | Required demonstration of safety and effectiveness prior to marketing  |
| 1992 | Subpart H                       | Accelerated approval of drugs for serious or life-threatening diseases |
| 1992 | Prescription Drug User Fee Act  | Fees for product applications                                          |
| 1997 | Food and Drug Modernization Act | Several reforms                                                        |

Kefauver–Harris Drug Amendments passed. In addition to more rigorous requirements for the demonstration of drug safety, this legislation required manufacturers to demonstrate to the FDA the effectiveness of their products prior to marketing.

In the 1980s and early 1990s, there was a growing recognition for the need to encourage the timely development and approval of drugs for serious or life-threatening illnesses such as HIV/AIDS (the human immunodeficiency virus/acquired immunodeficiency syndrome) and cancer. In 1992, subpart H was added to the new drug approval (NDA) regulations, allowing for accelerated approval (AA) of drugs for serious or life-threatening diseases where a drug demonstrates an advantage over available therapy or demonstrates its effects in a setting where no available therapy exists. Approval is based on a surrogate endpoint “reasonably likely to predict clinical benefit.” The sponsor must study the drug further to demonstrate clinical benefit in subsequent or ongoing clinical trials [2, 3].

Also in 1992, the Prescription Drug User Fee Act (PDUFA) was enacted, requiring drug and biologic drug manufacturers to pay fees for product applications and supplements. The act also required the FDA to use these funds to increase the number and breadth of reviewer expertise for product applications. In 1997, the Food and Drug Modernization Act reauthorized PDUFA and mandated several reforms in agency practices [4]. Measures were introduced to accelerate review of devices, to regulate advertising of unapproved uses of approved products, and to regulate health claims for foods.

### 25.3 CLINICAL INVESTIGATIONS

In the United States, investigational drugs must be administered under an investigational new drug (IND) application submitted to the FDA. The regulations describe two parties involved in IND submission, the *sponsor*, who is responsible for reporting to the FDA, and the *investigator*, who performs the trial [5]. The sponsor may be a pharmaceutical company, an academic institution, or an individual. Sponsors are to select only investigators “qualified by training and experience as appropriate experts to investigate the drug.”

When an IND is submitted to the FDA, a team of scientific reviewers evaluate the safety data from animals or other sources, evaluate the proposed study, and determine whether patients would be exposed to an unreasonable and significant risk. These issues are discussed individually in the sections below. All studies of nonapproved drugs must be done under an IND. For approved drugs, however, some studies require an IND and others are exempt from the IND requirement.

To determine that an IND is not needed, the following conditions should be met: The study (1) is not intended to support approval of a new indication or a significant change in the product labeling, (2) is not intended to support a significant change in advertising, (3) does not involve a route of administration or dosage level or use in a patient population or other factor that significantly increases the risks associated with the use of the drug product, (4) is conducted in compliance with institutional review board (IRB) and informed consent regulations, and (5) will not be used to promote unapproved indications [5].

The IND process spans the entire period of drug investigation (Fig. 1). It includes the initial IND application and later amendments to provide safety reports or

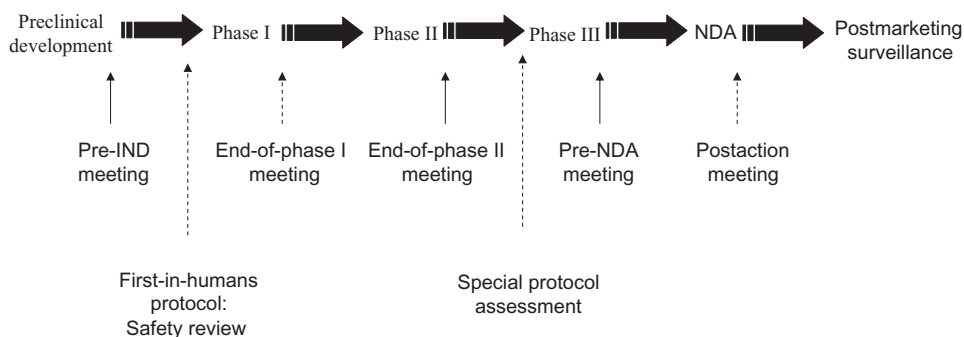


**FIGURE 1** Drug development timeline. The solid arrows indicate standard time points for meeting with drug developers. The dashed arrows indicate additional time points when the FDA sometimes meets with sponsors.

additional protocols. The initial IND application usually consists of a phase I clinical protocol and data to support the safety of the proposal. The latter would include in vitro, animal, and/or human evidence describing drug toxicity and allowing prediction of a safe starting dose. Manufacturing data describing the composition, manufacture, and control of the drug substance and drug product is also included. After the FDA receives the initial IND, sponsors are required to wait 30 days prior to initiating the proposed study unless they request and receive a waiver of the 30-day review period from the FDA. A multidisciplinary team of FDA reviewers, including physicians, pharmacologist/toxicologists, chemists, statisticians, and clinical pharmacologists determine whether the study is safe to proceed. The FDA may put an IND “on hold” if the agency believes subjects would be exposed to unreasonable risk of injury or if there is insufficient information to assess the risks.

The FDA frequently meets with sponsors and investigators in pre-IND meetings to review proposed IND plans and to clarify IND requirements. Animal studies should use the same or a similar schedule of drug administration proposed for the phase I clinical study. The starting dose for investigational drugs used in human studies is usually based on a preclinical animal model, which may differ based on the type of product and the therapeutic area of development. FDA advice on the specific requirements for a given product should be sought from representatives of the responsible FDA review division.

A critical FDA role in drug development is to meet with sponsors to provide advice on the design and conduct of phase III (and sometimes phase II) clinical trials that will support marketing applications. The multidisciplinary FDA team attending these meetings includes physicians, statisticians, pharmacologists/toxicologists, clinical pharmacologists, and often external expert consultants and patient advocates. FDA chemists also meet with sponsors, often separately, to discuss manufacturing and quality control issues. Sponsors can submit protocols subsequent to these meetings and request a Special Protocol Assessment, which provides for a binding agreement on protocol design [6]. After the clinical trials have been conducted and trial results are available, sponsors again meet with the FDA in pre-NDA meetings to discuss whether a new drug application (NDA) may be warranted and, if so, to discuss details of an NDA submission.

After clinical trials have been completed and an NDA has been submitted, the FDA verifies data quality and judges whether trial results demonstrate that the drug is safe and effective for the proposed use. After approval, the FDA continues to evaluate drug safety and regulate drug marketing.

#### 25.4 NDA REVIEW PROCESS

The package insert describes clinical trial results from data that have been reviewed and validated by FDA review teams. Regulations require that NDAs contain all relevant information about manufacturing, preclinical pharmacology and toxicology, human pharmacokinetics and bioavailability, clinical data, and statistical analyses. NDA applicants must submit financial disclosure information about the investigators. The FDA review of the NDA involves a team of chemists, pharmacologists/toxicologists, clinical pharmacologists, physicians, statisticians, microbiologists, site inspectors, and a project manager. The FDA reviewers evaluate the primary data, verify analyses, and where appropriate perform additional analyses. The FDA field inspectors verify that information on case report forms is supported by source data such as hospital charts and that biological drug concentrations are documented. Inspectors also inspect manufacturing facilities for suitability.

Based on the PDUFA, the FDA performs NDA reviews with either a 6-month or a 10-month goal. Applications representing a significant improvement compared to products that are currently available are assigned priority status and a 6-month review period, whereas standard applications have a 10-month review goal.

The FDA routinely seeks external advice on the design, analysis, and interpretation of clinical trials. These consultants are screened to exclude any potential conflicts of interest. Individual consultants advise the FDA during the design of clinical trials and during the NDA review. After an initial NDA review, the FDA presents selected NDAs to the appropriate advisory committee. This group is composed of medical and/or surgical specialists, statisticians, patient advocates, consumer representatives, and a nonvoting industry representative. At the public meetings of an advisory committee, the NDA applicant summarizes the study results, and the FDA presents its review of the findings. Members of the public may comment on the issues being discussed during the public speaker portion of the meeting. The committee finally votes on the specific questions submitted by the FDA. Although the discussion and the votes by the committee are taken into consideration, the FDA is not obligated to adhere to the advice provided.

#### 25.5 PEDIATRIC INITIATIVES

There has been general recognition of the lack of adequate labeling of drug products for children. In order to encourage evaluation of new therapies in pediatric populations and to optimize submission of pediatric data to support labeling of commercially available drugs, the FDA has undertaken two initiatives, which are outlined in Table 2 and summarized as follows.

A voluntary program, commonly referred to as “pediatric exclusivity,” was outlined in the Food and Drug Modernization Act of 1997. Under this program,

**TABLE 2    FDA Pediatric Initiatives**

| PREA                                           | BPCA                            |
|------------------------------------------------|---------------------------------|
| Applies to drugs and biologics                 | Applies only to drugs           |
| Studies mandatory                              | Studies voluntary               |
| Required only for drug/indication under review | Studies on entire active moiety |
| Orphan indications exempted                    | Includes orphan drugs           |

a commercial sponsor may receive a 6-month extension of existing marketing exclusivity by submitting data from pediatric clinical trials in support of an NDA. The general design of studies to be conducted and data to be submitted are outlined in a pediatric written request letter issued by the FDA to the sponsor. This program applies to drug products, including those with orphan designation. However, biological therapies such as monoclonal antibodies and vaccines are excluded. The Best Pharmaceuticals for Children Act (BPCA) subsequently provided some refinements to the program, including a framework for making summaries of medical and pharmacology reviews available to the public [7].

A mandatory program applying to indications where the disease in pediatric patients is similar to the adult disease, previously known as the “pediatric rule,” was subsequently formally legislated in the Pediatric Research Equity Act (PREA) [8]. This mandatory requirement for sponsors to submit data from pediatric studies when an adult indication is granted applies to a specific drug and indication under review. In order to avoid the potential delay in drug development and in providing access to therapies for life-threatening diseases that can arise when the adult data support approval and pediatric studies are still ongoing, a deferral for submission of the pediatric data may be granted. Both drug and biologic products are subject to the PREA. However, products with orphan designation are excluded from this requirement.

**25.6    ROLE OF STATISTICS IN REGULATORY SETTING**

In 1962 the Kefauver–Harris amendment to the Food, Drug and Cosmetic Act (FD&C Act) led to the provision that the manufacturing companies need to provide substantial evidence of effectiveness and that the evidence show adequate and well-controlled investigations, including clinical investigations, by qualified scientific experts, that proves the drug will have the effect claimed by its labeling” [Section 505(d) FD&C Act of 1962 as amended]. Statistical methodology provides tools for objectively planning, conducting, and evaluating clinical trials. When evaluating the efficacy and safety of a drug from a regulatory perspective, two objectives have to be met: First, the claimed effect is true and, second, the effect is clinically meaningful with acceptable toxicity. The role of a statistical expert at the FDA is, using the data submitted by the applicant, to evaluate whether the claimed efficacy is in fact true. The statistical team works in conjunction with the medical team to evaluate whether the claimed effect is clinically meaningful and has an acceptable toxicity profile.

Before an application [an NDA or biologics license application (BLA)] for marketing approval is submitted to the FDA, product development would typically have

gone through several phases: preclinical phase and clinical phases I–III. Although statisticians are consulted at the design stage of each of these phases, the greatest impact of statistical advice is in planning and designing comparative, randomized, well-controlled studies (phase II or III registration studies). A randomized study mitigates potential selection and reporting biases and balances any known and unknown prognostic factors that may influence the outcome of the study. As detailed in Section 25.3, new product manufacturers can consult FDA staff by submitting the planned study protocol as an IND application. These protocols are reviewed by the FDA staff and comments are sent to the applicant. In reviewing these protocols including special protocol assessments (SPAs), the statisticians critique both the design and planned analysis of the study.

The protocol should clearly identify the study objective, the primary hypothesis (the benefit that is to be demonstrated), study population, and the primary outcome (or endpoint) that is being measured based on which efficacy claim is planned. The primary endpoint could be a binary outcome (e.g., proportion of patients cured of an infection), a continuous variable (e.g., lowering of cholesterol), or a time-to-event endpoint (e.g., overall survival, time to progression of disease). An endpoint may also be defined as a composite endpoint (e.g., myocardial infarction or death). In certain disease settings more than one endpoint may be of primary interest (e.g., the rate of acute rejection and survival for an organ transplant drug). The design and analysis of a trial depends on the endpoint that is being measured. In conducting and evaluating clinical studies, it is important to control false-positive claims or the so-called type I error rates ( $\alpha$ ). The study also should have enough power or low type II error rates ( $\beta$ ), so that a true effect, if it exists, can be detected with the number of patients enrolled in the study. Subjective endpoints, such as symptom measurements, are difficult to evaluate in open-label studies, and potential biases can be minimized in double-blinded studies [9].

In confirmatory, phase III comparative studies [10, 11], there are basically three types of hypotheses that are tested: (1) the new product is superior [new product compared to placebo; new product compared to an active control (standard treatment); the new product has a dose–response relationship]; (2) the new product is noninferior (the new product is not worse than the existing treatment—it may be better or it may have a similar effect); or (3) two products are bioequivalent [absence of any difference between the two products (e.g., generic drugs)]. Marketing applications with claims of superiority or noninferiority of new drugs are reviewed at the Center for Drug Evaluation and Research. For consideration of a noninferiority claim, the assumption that the patient population and the conditions of the current study with the new treatment are similar to the historical placebo-controlled trials with the existing treatment should be met, the effect of size of the existing treatment must be well established, and that a certain prespecified percentage of the historical control effect size is maintained. Because of these assumptions, in general, noninferiority claims are difficult to establish and require large studies. In hypothesis 3, bioequivalence studies of generic products are generally reviewed by the Office of Generic Drugs at the FDA.

The sample size required to conduct a clinical trial that will ensure that the study is informative to test the primary hypothesis depends on the primary endpoint of interest, the smallest difference in treatment efficacy, which is important enough to detect, the probability of type I error (significance level), the probability of type II



error (1, power of the test), the number of interim analyses that are planned, the number of endpoints, and the comparisons that are being evaluated. If interim analyses or multiple comparisons are planned, then the type I error rate allocated for testing the final analysis has to be adjusted so that the overall familywise type I error rate is maintained at the specified level [10]. Sometimes, meta-analysis, a technique combining results from different studies, is used to provide benchmarks for new agents and new indications and in the design stage of clinical trials.

The statistical analysis plan should be prespecified. The statistical analysis plan should include definitions of the primary and secondary endpoints, and the hypotheses and the analysis plan for each of the endpoints. Comparative analyses based on secondary endpoints are considered exploratory and hypothesis generating if the familywise type I error rate is not maintained at the prespecified level. A post hoc analysis and claim based on a subgroup of patients, when the efficacy could not be demonstrated in the overall study population, is considered exploratory and hypothesis generating, and are generally not considered suitable as a basis for drug approval.

During an ongoing clinical trial, any changes in the definition of endpoint, changes in the type of hypothesis being tested, change in study population, additional interim analyses, increasing sample size, and other study modifications are problematic if the modifications or adaptations are not based on established statistical methods. Such modifications could compromise on the integrity of the study and introduce bias in the interpretation of the results, particularly in an open-label study. Too many modifications and adaptations of the study design and conduct are considered as exploratory. In general drug product approval is based on statistically persuasive results that are clinically meaningful with adequate efficacy and safety.

## **25.7 ROLE OF CHEMISTRY, MANUFACTURING, AND CONTROLS (CMC) INFORMATION IN REGULATORY SETTING**

As per regulations [12], the sponsor must submit CMC information that will describe the general method of preparation of the drug substance, physical, chemical, and biological characteristics, drug product formulation, manufacturing process, and controls to provide assurance of the identity, strength, quality, purity, and potency of the investigational drug. The CMC information that is expected in an initial IND (phase I) is more to assure patient safety and less on product and process controls. However, for phase II and phase III studies, as more manufacturing experience is attained with the manufacturing process and the final formulation is chosen, more exhaustive details regarding process and product controls should be provided. The amount of CMC information required is also dependent on the type of molecule to be tested (e.g., synthetic, semisynthetic, naturally derived, biotechnological) and whether there is any previous clinical experience with the study drug.

The amount of CMC information that will be required in an IND will also be dependent on the marketing status of the drug product. For a drug that is already marketed in the United States, the investigator should provide the current approved labeling and/or package insert of the drug. If a drug is not marketed in the United States, but studied under a cross-referenced IND, the sponsor should provide a letter of authorization (LOA) from the sponsor of the cross-referenced IND to allow the

agency to access to the referenced IND. The LOA should clearly specify the name, strength, and dosage form of the drug. If a drug is not marketed in the United States or it is not referenced to an IND, the sponsor should provide CMC information on the study drug appropriate to the phase of the study. If the drug is intended to be administered for intravenous administration after dilution, compatibility data with different diluents and in-use stability data for the duration of administration should be submitted. Please refer to the following guidance documents regarding the CMC information that needs to be submitted to support the proposed clinical study in the IND [13, 14].

If a drug is not marketed in the United States, but approved and marketed in a foreign market, or it is a legally marketed dietary supplement, the sponsor should provide full CMC information [13, 14]. For a limited phase I study, this requirement may be fulfilled by providing the following information: components and composition of the drug product, name of manufacturer/supplier of the drug product, and the labeling and/or package insert.

The sponsor should also provide a certificate of analysis for the lot that is intended to be used in the proposed clinical trial. Additional product-specific information may be required for the drugs approved and marketed outside the United States or for a legally marketed dietary supplement in the United States. When a sponsor is faced with a situation that is unique and not covered by the above guidance(s) [13, 14], the FDA recommends that the sponsor obtain clarification from the review division with regard to the CMC information necessary to support the proposed clinical study. For a botanical drug, the CMC information that needs to be submitted may be different from that of conventional small-molecule drugs [15].

For all drugs, the sponsor should provide a statement in the CMC section (as well as under dosage and administration section of the clinical protocol), as to the labeled dosage form and the strength of the drug product and how the drug will be administered. If the sponsor intends to change from the labeled dosage form, strength, or route of administration, the sponsor should provide relevant information such as release and stability data to support the change as well as to assure that the change will not adversely affect the identity, strength, quality, purity, and potency of the investigational drug. For all investigational drugs, the following labeling and environmental assessment information should be provided; the label for the immediate packaging of the drug product should be provided that includes the following cautionary statement: "Caution: New Drug—Limited by Federal (or United States) law to investigational use" [16].

A statement requesting a categorical exclusion from environmental assessment under provisions provided for in 21 CFR § 25.31(e); Part 25, Environmental Impact Considerations; Subpart C, Categorical Exclusions; Section 25.31, Human Drugs and Biologics, Paragraph (e)—Action on an IND [17].

If a device is used in conjunction with the study drug (e.g., a nebulizer for an inhalation drug or a pump for continuous infusion for home use), information on the manufacturer and model of the device to be employed and a general description of relevant conditions of use (e.g., carrier gas, flow rate, temperature) should be provided. A statement as to whether the device is FDA approved or not approved should also be included. When appropriate, submit adequate information regarding safety instructions for use of equipment and disposal of unused portion of the drug.

## 25.8 ROLE OF CLINICAL PHARMACOLOGY IN REGULATORY SETTING

The purpose of clinical pharmacology and biopharmaceutics data is to support the determination that the drug to be marketed is safe and effective for use in the indicated patient population. The clinical pharmacology data in conjunction with the safety and effectiveness data from the registration clinical trial(s) can be used to understand the relationships between drug dose or concentration and effects, namely patient outcome and toxicity (exposure–response relationships, ER). These relationships allow for better choices regarding optimizing the dose for the patient population, adjusting doses in special clinical situations, and providing information in the product labeling. Some of these studies provide direct support for the assessment of ER, such as pharmacokinetic information derived from clinical studies (see Fig. 1). On the other hand, some clinical pharmacology and biopharmaceutic data provide indirectly to understanding the behavior of the drug in the patient. For instance, in vitro cytochrome P-450 (CYP 450) screening often indicates whether the drug is metabolized by or modulates any of these enzymes. These in vitro data may suggest whether or not in vivo drug–drug interactions might occur between the drug being developed and other medications that may be commonly administered to a particular patient population that may lead to toxicity or reduced effectiveness.

The summation of these studies should provide appropriate information on drug behavior and its safe and effective use in the product labeling. With that goal in mind, the clinical pharmacology and biopharmaceutics studies can be used collectively to address some basic questions about the drug and how to use it appropriately. With respect to labeling the drug, the main questions are: Is this dose the optimal dose for this patient population? Are any dose or dose regimen adjustments needed for patients with specific clinical characteristics? The answer to these questions are quite complex. Many different chemical, pharmacological, and clinical aspects of the drug contribute to the answers. In the Appendix, many questions about the clinical pharmacology of the drug are posed, and these specific answers, once determined, contribute to the overall issues regarding the optimal dose, dose adjustments, and the safe labeling of the drug.

To address many of these specific issues, the FDA has publicly posted or is planning to post Guidances for Industry that describe the current thinking of the FDA on a specific topic [18]. Each of these is intended to be a general set of guidelines to deal with a specific issue; they may not address every situation, but they should be the starting point for dealing with the issue of interest. Essentially, these are located on the FDA website under the Biopharmaceutics, Chemistry, Clinical, and Clinical Pharmacology headings [18]. The Biopharmaceutics and Chemistry Guidances address issues that affect the drug formulation and may have an impact on how the drug product behaves after administration to the patient (i.e., how much drug may be absorbed). For instance, changing the ingredients of a tablet may alter how quickly the tablet dissolves after administration to the patient, which may have an impact on the absorption of the drug and therefore its effect(s). The Clinical Pharmacology Guidances describe what happens to the drug, and hence its impact on the patient once it is administered to the patient. For instance, renal impairment may cause the exposure of drug to increase because drug elimination is reduced. This in turn may increase the rate of drug toxicity and may necessitate a dose adjustment in these patients.

### 25.8.1 Biopharmaceutics Guidances

The Bioanalytical Method Validation Guidance [19] is designed to verify that the analytic method used to measure drug concentration is accurate and reliable for the purpose it is intended. The Food Effect Guidance [20] describes the general issues to be considered when assessing the effect of food on a per oral formulation. It addresses the biopharmaceutics issue of how food taken with a medication can alter the dissolution and absorption of a given product. The outcome of this study is often used to determine how the drug should be administered in future trials and, eventually, the intended patient population. During drug development, it is common for a drug developer to alter the drug formulation. These changes may impact the pharmacokinetics and pharmacodynamics of the drug in the patient, and these formulation changes raise questions about how to “bridge” (compare) data from different studies developed with different formulations. The Dissolution Guidance [21] and the Bioavailability and Bioequivalence Guidance [22] are closely related in purpose. Both methodologies are intended to assess the effect of changes in the product formulation on the pharmacokinetics and, by extension, the pharmacodynamics of a drug. The dissolution guidance describes how an *in vitro* method can be developed using well-defined apparatus and criteria. The purpose is to develop a discriminatory *in vitro* method to assess alterations in the formulation of a per oral drug product (similar approaches can be used to develop an *in vitro* drug release methodology for a liposome drug product). For instance, changing an ingredient may have an impact on the rate that the tablet dissolves in the gastrointestinal milieu. One could expect that this could alter the extent to which a drug is absorbed and the timing of this event. This could result in undesired effects in the patient (reduced effectiveness or increased toxicity). However, significant changes in dissolution do not necessarily indicate clinically significant changes in the behavior of the drug, so monitoring changes in dissolution is the first step in assessing the impact of a formulation change. Such an *in vitro* observation is likely to be further assessed by observing the effect of the new formulation *in vivo*. The Bioavailability/Bioequivalence Guidance [22] describes in general terms the elements of a study for comparing the pharmacokinetics and pharmacodynamics of a preexisting drug formulation with a newly developed one. Studies of dissolution and pharmacokinetics conducted early in drug development can lead to the development of *in vitro*–*in vivo* correlations [23] in which a clinically meaningful dissolution method can be used to predict *in vivo* performance of new formulations. These IVIVC relationships are often used in the development of modified-release drug products.

With regard to clinical pharmacology studies, the characterization of the drug often begins with *in vitro* studies. Protein binding, although not the subject of a specific guidance, is usually assessed with *in vitro* incubations for each new drug. (There are examples of *ex vivo* protein binding determinations as well.) Current scientific thinking is that only free or unbound drug can mediate drug action. Therefore, highly protein-bound drugs ( $\geq 90\%$ ) have their pharmacological effects mediated by only a small fraction of drug. Furthermore, small changes in protein binding for these drugs might be expected to mediate a substantial effect. Therefore, the extent of protein binding to human serum proteins (e.g., albumin,  $\alpha$ -acidic glycoprotein) is commonly performed.

The Drug Metabolism Guidance describes both *in vitro* and *in vivo* studies [24, 25]. The *in vitro* studies describe cytochrome P-450 (CYP 450) screening. The CYP 450 superfamily of enzymes accounts for the majority of drug metabolizing capability in humans. *In vitro* screening often identifies the main isozyme(s) responsible for the metabolism of a particular drug and the extent of its contribution. New drugs can also be screened for their ability to either inhibit CYP 450 isozyme(s) activity or induce their activities and expression. The end result may be reduced drug clearance or increased drug clearance, respectively, which may affect effectiveness and/or safety. These *in vitro* findings can then be used as the basis for deciding on the need to conduct *in vivo* drug–drug interaction studies. It is noteworthy that these *in vitro* studies do not account for the activity of all metabolizing enzymes, and the metabolism and potential drug–drug interactions on non-CYP 450 metabolized drugs are often difficult to assess. One non-CYP 450 enzyme system that has been studied further is the P-glycoprotein (PGP) transporter system. A newly published draft Guidance on Drug Metabolism describes both *in vitro* and *in vivo* studies to assess the impact of new drugs on PGP [26].

Related to the *in vitro* metabolism studies in purpose are the *in vivo* “mass balance” studies. These studies are not the subject of a specific guidance but are referred to in some agency publications. The purposes of these types of studies are generally twofold. The first is to determine whether a novel metabolite is generated in humans that has not been previously characterized (at least in part) in preclinical studies. The second purpose is to assess the route of excretion/elimination from the patient. This latter purpose is of importance because it can indicate that patients with organ dysfunction may have difficulty eliminating the drug, which in turn can lead to excessive drug exposure and toxicity. The results of these studies are used to determine whether pharmacokinetic studies in patients with either renal and/or hepatic impairment should be conducted. One common misinterpretation regarding mass balance studies is that the administration of radiolabeled drug is required. This, in fact, is untrue. Any appropriate methodology may be used to assess mass balance. However, use of radiolabeled drug is often employed because of the considerable sensitivity imparted by the methodology.

### 25.8.2 Clinical Pharmacology Guidances

The basic elements of the pharmacokinetic characterization (clearance, CL, volume of distribution,  $V_d$ , and elimination half-life,  $t_{1/2}$ ) are characterized in the Format and Content of Human Pharmacokinetics and Bioavailability Section of an Application Guidance [27]. Essentially, the kinetics should be determined to better understand the behavior of the drug. Typically, single- and multiple-dose pharmacokinetics are assessed, usually in early studies in volunteers or patients. Dose proportionality is often assessed with the data generated from these studies. The exposure–response guidance describes the different measures that can be used to evaluate the relationships between drug amount and its effectiveness and toxicity(ies) [28]. These relationships are of great importance because they are the basis for understanding the behavior of the drug and for making rational decisions about choosing the optimum dose or altering the dose when necessary. The Population Pharmacokinetic Guidance [29] describes alternative approaches to assessing the pharmacokinetics of a drug. In some drug development circumstances, the execution of a dedicated phar-

macokinetic study with dense sampling may not be feasible. In these cases, an alternative approach can be employed to collect pharmacokinetic data in a more limited fashion. These data are often combined with previously collected data, and the effects of different factors (covariates) such as disease, patient demographics, and patient characteristics can be assessed, often by using sparse sampling.

The guidances on the effect of renal [30] or hepatic impairment [31] on the pharmacokinetics of a drug also describe important studies. Most drugs are renally excreted or metabolized (predominantly by the liver) prior to elimination or sometimes both. As drug exposure is a function of both the administered dose, and the rate of elimination, renal or hepatic impairment can lead to significant reductions in drug elimination, which in turn produce excessive drug exposures (e.g., higher than anticipated plasma concentrations). Sometimes these elevated concentrations lead to toxicities. Ideally, by understanding the pharmacokinetics (PK) and the pharmacodynamics (PD) of the drug in both the general patient population and in those patients with organ dysfunction, rational dose modifications can be made.

The agency has also published a guidance on the evaluation of QT prolongation [32], which describes how to assess the potential of a drug to adversely affect cardiac conduction.

The agency has also been developing new guidances that deal with issues of drug PK in pregnant women [33] or lactating women [34], as well as in pediatric patients [35].

Recently, the FDA also implemented the EOP1/2a Guidance [36]. The purpose of this guidance is to encourage drug developers to employ all of the available pre-clinical biomarker, toxicity, and pharmacokinetic and chemistry information along with existing clinical data on effectiveness, biomarkers, and pharmacokinetics in a comprehensive model and to subsequently simulate the outcomes of different drug development trials with different regimens, combinations, and dosages to help choose the most effective trial design for the pivotal trial. In a similar vein, the agency has also issued the Guidance for Industry on Pharmacogenomic Data Submission [37] in a effort to stimulate the development of testing platforms and analysis and utilization of pharmacogenomic data in clinical trials in order to develop better, targeted therapies.

### 25.8.3 Timing of Studies

With respect to when these studies should be conducted, there is no regulatory requirement. It is up to the drug development sponsor to decide on when and how to allocate resources to these studies. However, there is some sense to the order and timing of these studies. For instance, there may not be too much value in conducting a mass balance study in parallel with the phase III safety and effectiveness trial, if the active metabolites have been identified and organ dysfunction studies have been conducted (or planned). Figure 2 depicts one possible scheme of clinical pharmacology and biopharmaceutics study conduct. According to this scheme, development begins with preclinical studies. The extent of protein binding is determined, and the compound is screened *in vitro* to evaluate it as a substrate for, or inhibitor and/or inducer of, CYP 450 enzymes. The compound is also screened as a substrate or inhibitor of P-glycoprotein. Additionally, biomarkers may be selected as potential indicators of activity based on animal and *in vitro* models. Other preclinical studies



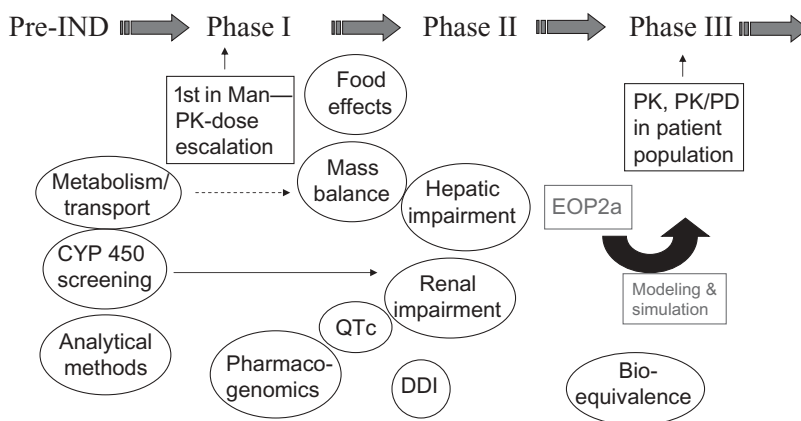


**FIGURE 2** One possible clinical pharmacology drug development scenario is depicted. The FDA does not regulate the timing of studies in drug development.

generate information regarding the toxicity of the compound, including the potential to alter cardiac function (e.g., hERG assay). The analytical method chosen to measure in vivo drug concentrations is often adapted from preclinical to clinical use and validated.

With the first-in-man studies, and ideally throughout the completion of the phase III trial(s), the pharmacokinetics are assessed. Single-dose and multiple-dose pharmacokinetics are usually determined in the dose escalation phase I study(ies), and dose proportionality and time-dependent changes in the pharmacokinetics can be assessed with this data. In some cases, population PK/PD modeling is started with the development of the base model using this data. Collection of samples for pharmacokinetic analyses from phase II may be correlated with markers of effectiveness and toxicity. Drug development according to this scheme would lend itself to an EOP2a analysis, and subsequent simulation of possible phase III trial designs, which may be very helpful in choosing the most efficient regimens, doses, combinations, and/or patient populations. In the late phase I early phase II period, it may be advisable to assess the effect of food on the pharmacokinetics of orally administered drugs in order to decide how to administer the drug in the pivotal trial and the intended patient population. It may also be helpful to consider the effect of two drugs that may be co-administered together. For instance, in oncology, drugs developed for glioblastoma multiforme are often administered with antiepileptic medications, such as phenytoin. The latter drug induces CYP 450 3A4 expression and often reduces the plasma concentrations of the new drug to levels that are too low to be effective. In these cases, the new drug is often developed with concomitant use in all later studies and is labeled for use specifically in combination with each other.

Often in this time frame, thorough QTc studies are performed to assess the potential for the drug to prolong the QT interval, and possibly generate torsades de pointes. Mass balance studies are often conducted some time between the first-in-man studies and phase II. Drug-drug interaction studies and organ dysfunction studies are often conducted in parallel to the phase III, when the decision has been made to pursue the drug marketing. Bioequivalence studies are conducted at various



points during development, from phase I to postdrug approval, depending on the development of the drug formulation and oral preparations with new strengths.

## 25.9 DRUG REGULATION IN GLOBAL ENVIRONMENT

Drug development has become a global effort with clinical trials in several disease areas often involving a multitude of centers across the globe. Even when some individual clinical trials are conducted in one country or region, a package submitted for approval purposes often includes trials conducted in a number of regions around the world. Even when individual trials are conducted in one region, commercial sponsors usually have a global strategy aimed at marketing approval in a number of countries and regions. While international activities take place in virtually every component of the FDA, the Office of International Programs (OIP) is the focal point for the agency's international activities. It provides leadership for FDA's international activities by guiding and supporting agency programs and by developing strategic relationships with other U.S. and foreign governmental agencies and international organizations. While the FDA takes a proactive role in these activities, it is also worthwhile to discuss not only the role of the FDA but to also describe the roles of other regulatory agencies around the world. The following is a brief introduction to drug regulation in Canada, Europe, Australia, and Japan.

The Therapeutic Products Directorate (TPD) is the Canadian federal authority that regulates pharmaceutical drugs and medical devices for human use and is housed within Health Canada (Ministry of Health) [38]. Prior to being given market authorization, a manufacturer must present substantive scientific evidence of a product's safety, efficacy, and quality as required by the Food and Drugs Act. The Biologics and Genetic Therapies Directorate (BGTD) has regulatory oversight for biologics, radiopharmaceuticals, genetic therapies, and the transplantation of tissues, cells, and organs.

With headquarters in London, the European Medicines Agency (EMA) is a decentralized body of the European Union (EU) that coordinates a network of 42 national agencies and has undergone significant changes concurrent with the enlargement of the European Union [39, 40]. Its main responsibility is the protection and promotion of public and animal health through the evaluation and supervision of medicines for human and veterinary use. The EMA began its activities in 1995, when the European system for authorizing medicinal products was introduced, providing for a centralized and a mutual recognition procedure. The EMA has a role in both, but is primarily involved in, the centralized procedure. Where the centralized procedure is used, companies submit one single marketing authorization application to the EMA. A single evaluation is carried out through the Committee for Medicinal Products for Human Use (CHMP) or Committee for Medicinal Products for Veterinary Use (CVMP). If the relevant committee concludes that safety and efficacy is proven, it adopts a positive opinion. This is sent to the Commission to be transformed into a single marketing authorization for the whole European Union. Since 2005, applications for marketing authorization within the European Union for diabetes, AIDS, cancer, neurodegenerative disorders, and for products with orphan designation have required submission exclusively through the centralized procedure [40].

In Australia, the Therapeutics Goods Administration (TGA) has authority in the regulation of prescription and nonprescription medicines, medical devices, and gene technology products. The TGA guidelines are aligned with those of the European Union and the International Conference on Harmonisation (ICH) of drug regulatory activities. Further information on drug regulation in Australia can be found on the official webpage [41].

In Japan, regulatory activities have been consolidated into the Pharmaceuticals and Medical Devices Agency (PMDA) since 2004. The PMDA is authorized by the Ministry of Health, Labor and Welfare (MHLW) and includes a staff of approximately 250 personnel including reviewers. The PMDA is comprised of four offices: Office of Relief Funds, Office of Review, Office of Safety, and Office of Research and Development Promotion. Further information on drug regulation in Japan can be found on the official PMDA webpage [42].

In China, regulation of foods, drugs, cosmetics, and traditional Chinese medicinal preparations is the purview of the State Food and Drug Administration (SFDA). Drugs are regulated by the Center for Drug Evaluation (CDE). Further information on drug regulation in China can be found on the official SFDA webpage [43].

## **APPENDIX CLINICAL PHARMACOLOGY DRUG DEVELOPMENT QUESTIONS**

### **Exposure–Response Questions**

1. What are the concentration/dose–response relationships? (ER, population PK)
  - What are the effectiveness endpoints? In vivo
    - Biomarkers?
    - Clinical outcomes?
  - What are the safety endpoints? In vivo
2. What are the factors that affect the concentration–response relationships? (ER)
  - Do pharmacogenomics affect the ER? In vitro, In vivo (VGDS)
  - What is the extent of protein binding? Does it impact ER? In vitro
  - Are there active metabolites that contribute to ER? In vitro (drug metabolism/drug interaction)
  - What are the characteristics of metabolism and elimination that affect the ER?
  - Is the drug eliminated by metabolism, renal excretion, or both (some other route)? In vitro, in vivo (drug metabolism/drug interaction)
  - Is the drug a CYP 450 isozyme(s) substrate, inhibitor, induce? In vitro
  - Are in vivo drug–drug interaction studies needed to assess impact? In vivo
  - Are labeling instructions necessary?
  - Is the drug a P-glycoprotein substrate or inhibitor? In vitro (drug metabolism/drug interaction)
  - Are in vivo drug–drug interaction studies needed to assess impact? In vivo
  - Are labeling instructions necessary?
  - Are there metabolic polymorphisms that affect metabolism/elimination?
    - In vitro, In vivo (drug metabolism/drug interaction)

- Are in vivo drug–drug interaction studies needed to assess impact?
- Are labeling instructions necessary?
- Are in vivo studies in patients with hepatic impairment needed to assess the need for dose adjustments? In vivo (hepatic impairment)
- Are in vivo studies in patients with renal impairment needed to assess the need for dose adjustments? In vivo (renal impairment)
- Does age have any effect? In vivo (21 CFR 314)
- Does race have any effect? In vivo (21 CFR 314)
- Does gender have any effect? In vivo (21 CFR 314)
- Are the ER the same in pediatric populations? In vivo (Pediatric PK)
- Are specific dosing instructions/regimens required?
- 3. What biopharmaceutics issues affect ER of the drug?
  - Can the active moieties of the drug be accurately measured? In vitro, In vivo (Bioanalytical method validation)
  - What is the extent of absorption? In vitro, in vivo (21 CFR 320, BA/BE, BCS Classification)
  - Are there different formulations of the drug that affect the ER? In vitro, In vivo (BA/BE, dissolution, IVIVC, SUPAC, IR, MR)
  - Are there different oral formulations that absorbed differently?
  - Is there a modified intravenous formulation that affect ER? In vitro, in vivo (Liposome drug products) e.g., liposomal drug, antibody-conjugated drug, etc?
  - How does food affect absorption of an oral formulation? In vivo (Food effect)

## REFERENCES

1. Hilts, P.J. (2003), *Protecting America's Health: The FDA, Business, and One Hundred Years of Regulation*, Random House, New York.
2. 21 Code of Federal Regulations Part 314.530.
3. Dagher, R., Johnson, J., Williams, G., Keegan, P. and Pazdur, R. (2004), Accelerated approval of oncology drug products: A decade of experience, *J. Natl. Cancer Inst.*, 96, 1500–1509.
4. The Food and Drug Modernization Act, 21 USC 355a, Public Law # 105–115, 1997.
5. 21 Code of Federal Regulations Part 312.23(b)(1).
6. Guidance for Industry: Special Protocol Assessment, May 2002; available at: <http://www.fda.gov/cder/guidance/3764.htm>.
7. Best Pharmaceuticals for Children Act (BPCA) Public Law #107–109, January 4, 2002.
8. Pediatric Research Equity Act (PREA) Public Law #108–155, December 3, 2003.
9. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims; available at: <http://www.fda.gov/cder/guidance/index.htm>.
10. FDA (1998), Guidance for Industry: International Conference on Harmonization E9 Statistical principles for Clinical Trials Center for Drug Evaluation and Research.
11. Guidance for Industry: International Conference on Harmonization E10 Choice of Control Group and Related Issues in Clinical Trials, 2001.
12. The Code of Federal Regulations Title 21 part 312.23(a) (7), 2006.

13. Guidance for Industry: Content and Format of Investigational New Drug Applications (INDs) for Phase 1 Studies of Drugs, Including Well Characterized, Therapeutic, Biotechnology-derived Products, 1995.
14. Guidance for Industry: INDs for Phase 2 and Phase 3 Studies Chemistry, Manufacturing, and Controls Information, 2003.
15. Guidance for Industry: Botanical Drug Products, 2004.
16. The Code of Federal Regulations Title 21 CFR § 312.6 (a), 2006.
17. The Code of Federal Regulations 21 CFR § 25.31(e); Part 25, 2006.
18. Available at: <http://www.fda.gov/cder/guidance/index.htm>.
19. Guidance for Industry: Bioanalytical Method Validation. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2001.
20. Guidance for Industry: Food-Effect Bioavailability and Fed Bioequivalence Studies, 2002.
21. Guidance for Industry: Dissolution Testing for Immediate Release Solid Oral Dosage Forms. Center for Drug Evaluation and Research, United States Food and Drug Administration, 1997.
22. Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2003.
23. Guidance for Industry: Extended Release Oral Dosage Forms: Development, Evaluation, and Application of In Vitro/In Vivo Correlations, 1997.
24. Guidance for Industry: Drug Metabolism/Drug Interaction Studies in the Drug Development Process: Studies In Vitro, 1997.
25. Guidance for Industry: In Vivo Drug Metabolism/Drug Interaction Studies—Study Design, Data Analysis, and Recommendations for Dosing and Labeling, 1999.
26. Guidance for Industry: Drug Interaction Studies—Study Design, Data Analysis, and implications for Dosing and Labeling, 2006.
27. Guidance for Industry: Format and Content of the Human Pharmacokinetics and Bioavailability Section of an Application, 1987.
28. Guidance for Industry: Exposure-Response Relationships—Study Design, Data Analysis, and Regulatory Applications. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2003.
29. Guidance for Industry: Population Pharmacokinetics. Center for Drug Evaluation and Research, United States Food and Drug Administration, 1999.
30. Guidance for Industry: Pharmacokinetics in Patients with Impaired Renal Function. Center for Drug Evaluation and Research, United States Food and Drug Administration, 1998.
31. Guidance for Industry: Pharmacokinetics in Patients with impaired Hepatic Function: Study Design, Data Analysis and Impact on Dosing and Labeling. Center for Drug Evaluation and Research, United States Food and Drug Administration, 2003.
32. Guidance for Industry: International Conference on Harmonization E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs, 2005.
33. Guidance for Industry (draft): Pharmacokinetics in Pregnancy—Study Design, Data Analysis, and Impact on Dosing and Labeling, 2004.
34. Guidance for Industry (draft): Clinical Lactation Studies—Study Design, Data Analysis, and Recommendations for Labeling, 2005.

35. Guidance for Industry (draft): General Considerations for Pediatric Pharmacokinetic Studies for Drugs and Biological Products, 1998.
36. Guidance for Industry: End of Phase 1/2a Meetings, 2004.
37. Guidance for Industry: Pharmacogenomic Data Submission, 2005.
38. Farrell, A. T., Papadouli, I., Hori, A., Harczy, M., Harrison, B., Asakura, W., Marty, M., Dagher, R., and Pazdur, R. (2006), The advisory process for anti-cancer drug regulation: A global perspective, *Ann. Oncol.* 17, 889–896.
39. Pignatti, F., Boone, H., and Moulon, I. (2004), Overview of the European regulatory approval system, *J. Ambul. Care Manage.* 27(2), 89–97.
40. Available at: <http://www.EMEA.eu.int>.
41. Available at: <http://www.tga.gov.au>.
42. Available at: <http://www.pmda.go.jp>.
43. Available at: <http://www.sfda.gov.cn>.



# 26

---

## *Future Challenges in Design and Ethics of Clinical Trials*

**Carl-Fredrik Burman<sup>1</sup> and Axel Carlberg<sup>2</sup>**

<sup>1</sup>*Technical & Scientific Development, AstraZeneca, Mölndal, Sweden*

<sup>2</sup>*Department of Cardiothoracic Surgery, Lund University Hospital, Lund, Sweden*

### **Contents**

|        |                                                   |      |
|--------|---------------------------------------------------|------|
| 26.1   | Introduction                                      | 1174 |
| 26.1.1 | Introducing Public Trust                          | 1174 |
| 26.1.2 | Introducing Productivity                          | 1175 |
| 26.2   | Challenge of Public Trust                         | 1177 |
| 26.2.1 | Requirement of Genuine Consent                    | 1178 |
| 26.2.2 | Requirement of Patient Benefit Optimization (PBO) | 1180 |
| 26.2.3 | Requirement of Nonexploitation                    | 1185 |
| 26.3   | Challenge of Clinical Trial Efficiency            | 1187 |
| 26.3.1 | Background                                        | 1188 |
| 26.3.2 | Decision Analysis Approach to Study Design        | 1188 |
| 26.3.3 | Optimizing Information per Patient                | 1190 |
| 26.3.4 | Optimal Amount of Information to Collect in Trial | 1193 |
| 26.3.5 | Adapting to New Information                       | 1195 |
| 26.4   | Discussion and Conclusions                        | 1197 |
|        | References                                        | 1198 |

Disclaimer: The views expressed in this chapter do not necessarily reflect those of our respective affiliating institutions.

---

*Clinical Trials Handbook*, Edited by Shayne Cox Gad  
Copyright © 2009 John Wiley & Sons, Inc.



## 26.1 INTRODUCTION

Clinical development faces two major challenges: public trust and productivity. Costs in drug development have increased substantially over the past decades [1]. Notwithstanding this development, the number of new drugs that are implemented in clinical practice has not increased but rather decreased in recent years [2]. At the same time, recent opinion polls [3] show that the public trust in pharmaceutical companies has fallen sharply and is now almost at a level comparable to that of the oil and tobacco industries. While some may argue that these challenges of public trust and productivity are difficult to align, we consider both to be essential and indeed mutually dependent. If public trust in clinical trials decreases, recruitment of patients for clinical trials will be more difficult, which in turn will result in even poorer productivity.

The ambition of improving public trust, especially in view of the tragic incident at Northwick Park Hospital in March 2006, has much to do with heightening ethical standards. The dual objective of this chapter is therefore to clarify the nature of the ethical challenges that lie ahead and to specify the design features that can increase productivity in individual clinical trials under ethically acceptable standards. Our question could be formulated as follows: Given that clinical trials should meet stringent requirements, how can productivity be maximized? In order to formulate an answer to this question, we will first have to specify the ethical requirements. Once these have been formulated, we will provide a method of analyzing trial efficiency and will give examples of how productivity can be increased. In doing so, we will consider specifically how to optimize the information per patient, approach sample size determination from a business perspective, and choose between adaptive design options.

This reflection is inspired by decision analysis (DA), which is a method that aims at optimizing different kinds of decisions, often by analyzing trade-offs. The essential assumption of DA is that different objectives can be valued on a common scale. For example, if one treatment offers better health benefits while the other provides better safety, it is possible to compare the two and choose the treatment that maximizes the net utility for the patient. Furthermore, even if the respective health benefits and safety features of these treatments are uncertain, DA offers the possibility of comparing the expected utility of the two treatment options. In this study, the first DA will be patient-centered, meaning that the net utility of the *patient* will be focused. After and secondarily to that, the interests of science and the sponsor of the trial will be considered, and a commercial optimization of clinical trials will be outlined. This priority is, in our minds, the only one that is ethically acceptable.

### 26.1.1 Introducing Public Trust

The study will therefore begin by a broad review of the ethical issues facing clinical development and more specifically randomized clinical trials (RCTs). In doing so, we formulate three minimal ethical requirements that should be met before patients can be enrolled in RCTs: genuine consent, patient benefit optimisation (PBO), and nonexploitation. These requirements are addressed principally to physicians who ultimately are responsible for their patients' health and for the enrollment of their

patients in any RCT. Our requirements do not cover properly scientific or regulatory concerns, and it is taken for granted that any clinical trial must also be examined and approved by an ethical committee or an investigational review board (IRB). In addition, we take for granted that regulatory approval is needed to administer new pharmaceuticals outside clinical trials. A more detailed list of ethical requirements covering such scientific and regulatory issues has been formulated by Emanuel et al. [4].

The core issue being treated here is therefore the physician/patient relationship and how it is affected by the collaboration of the former in clinical development. We argue that this collaboration does call into question the physician's fiduciary commitment to the well-being of his patient, as he or she must concurrently serve the interests of science. Such collaboration, which constitutes a threat to patient trust, can only be ethically justified through the consent process in which the patient acquiesces to participate in an RCT. The patient's informed consent in this matter should therefore not only be a formal but genuine and be based on relevant and appropriate information allowing the patient to formulate a well-reasoned judgment about such participation.

While this genuine consent is a necessary condition, it is not a sufficient prerequisite for trial participation. There are instances when the patient might accept to take part in a trial but when such participation would nonetheless be unethical. Such situations might arise if the patient is not sufficiently informed or free to consent or decline trial participation or when the trial design itself is morally skewed. That is why the physician, who is ultimately responsible for the patient's medical well-being, must be certain that he or she can, in good conscience, propose an RCT to his or her patient. Traditionally, this ethical peace of mind has been sought in the state of equipoise, understood as a state of "genuine uncertainty" on the part of the physician as to the merits of the respective arms of the trials. We argue, in a Bayesian perspective, that such genuine uncertainty in reality never exists. The physician has always a hunch or can second guess, on the basis of medical competence and experience, which arm in the RCT will receive the best treatment. An appeal to equipoise alone will therefore not do. We believe that the only way to ethically justify physician collaboration and patient participation in a RCT is the requirement of patient benefit optimization, that is, the judgment that the patient will benefit more from inclusion in the trial than from exclusion [5].

Besides genuine consent and patient benefit optimization, there is a third ethical requirement that we believe is essential: the principle of nonexploitation. It stipulates that no coercion should be used in the recruitment of patients. In other words, a patient or volunteer must be provided with sufficient freedom to willingly accept or decline an invitation to enroll in a RCT. The research subject might not, for instance, feel free to decline an offer to participate in a trial if this would result in, for example, inferior treatment. Such a situation can arise when research is carried out in developing countries or if patients are required to participate in a clinical study in order to qualify for medical coverage.

### 26.1.2 Introducing Productivity

Within the framework of these ethical requirements, the sponsor must design the most efficient trial. What does efficiency mean, however? Just as different

therapeutic treatments have medical pros and cons, trial designs have also relative benefits and drawbacks. Thus, DA is useful, not only in evaluating the ethical issues involved but even more so in the analysis of efficiency. In Section 26.3.2 we propose a detailed 10-step DA approach to clinical development. In this introduction, we limit ourselves to outlining the four most important steps, which are (1) modeling background information, (2) specifying trial objectives and utilities, (3) identifying design options, and (4) optimizing the decision of the three previous steps.

Experimentation is driven by the need to know more about the medical effects and safety features of a certain drug. It is therefore important to evaluate the information that is already available and to quantify the level of uncertainty. Existing information about the drug's characteristics can usually be compiled from a number of sources. Previous clinical trial data are often limited, especially in early clinical development. To fill in the information gap, investigators can refer to preclinical data, literature on the disease and related drugs, and expert input. In a DA approach, a quantitative model must first be constructed based on already existing information and a set of assumptions. These assumptions are necessarily subjective and should therefore be challenged during the analysis process. Then the objective of the trial must be formulated from a wide enough context in order to avoid suboptimization. The overall objective is then converted into quantitative preferences, which can be called utilities. The utility is thus a function of the outcome of the trial and reflects both the value of information as well as the cost of experimentation. In Section 26.3, we will take the commercial net present value (NPV) as an approximation of the utility of the trial sponsor. Third, DA can help to identify the designs and to tailor them in such a way that they are compatible with the objectives of the experiment. Still today, many trials are given standard designs, without taking into account its specific objectives. We therefore argue that standard procedure, at the outset of any trial, should be to generate, through project team brainstorming, a number of designs by considering common options such as cross-over, adaptive, and enrichment designs. Finally, the fourth step consists in predicting the trial results for the different designs. These predictions are made from the model and are typically carried out by computer simulations. When this is done, the predicted trial outcomes can be combined with the utility function. An expected utility can then be calculated and the optimal trial be designed in view of maximizing this value. By questioning the underlying assumptions and making tentative modifications of them, the robustness of this result can be verified.

The DA approach will thus be the dominating framework when discussing different aspects of study design. We will therefore discuss how much information should be obtained from each patient in view of optimizing the design (Section 26.3.3). Multiple measurements and multiple treatments (cross-over study) for each individual will yield more information but might also cost more and cause methodological problems. It is therefore important to analyze the optimal amount of information to purchase, that is, the optimal sample size when information value is compared with the cost of the experiment (Section 26.3.4). Finally, we will consider some sequential and adaptive designs where the trial design decisions depend on accumulating information (Section 26.3.5).

A number of recent trends in clinical study methodology and technology are supporting the DA approach to trial optimization. Model-based drug development,

and especially pharmacokinetic/pharmacodynamic (PK/PD) modeling, are being used more and more by the industry and are being promoted by regulators. The Food and Drug Administration (FDA) has stated that a drug may, in some instances, be approved on the basis of only one single confirmatory trial when supported by modeling results. The increase in computer speed has opened the way for nonlinear mixed-effect models as well as for Bayesian methods, where a probability model is updated with accumulating data. New design options are identified, and there has been an increasing interest in adaptive designs during the last decade. Web-based data capture has greatly facilitated the practical implementation of sequential and adaptive designs. The interest in decision-analytic approaches is on the upswing, and methodological work is being done in related areas, ranging from the psychology of decision making to the integration of commercial values in sample size determination.

## 26.2 CHALLENGE OF PUBLIC TRUST

From the point of view of medical ethics, clinical trials are problematic because they challenge the generally held view that the principal responsibility of physicians is to further the well-being of the individual patient. This responsibility is the cornerstone of the physician–patient relationship. As Ruth Purtilo argues [6], the professional's priority to further the welfare of the patient:

is interpreted to have the stringency of a special moral obligation on the part of the professional to seek a patient's well-being, guided by that person's health-related concerns and needs. Any other worthy goal, such as furthering knowledge about disease and its cure, or earning a just wage, or maintaining the efficiency or financial solvency of the institution, is not an appropriate beacon to guide the professional in this relationship.

This hallow responsibility is enshrined in most codes of ethics regulating medical conduct and has enjoyed almost universal recognition. In the Oath of Hippocrates (Greece, 4th century B.C.E.), the physician swears to apply his art “for the benefit of the sick according to my ability and judgment” adding that he will keep the sick “from harm and injustice.” In the Indian tradition, the Oath of Initiation (Caraka Samhita) from the first century C.E. puts it another way: “Day and night, however you may be engaged, you shall endeavour for the relief of patients with all your heart and soul.” Similar enjoinders can be found in the Advice to a Physician (Persian tradition 10th century C.E.) and in the 17 Rules of Enjuin (Japanese tradition 16th century C.E.). In modern times, the International Code of Medical Ethics adopted by the World Medical Association in 1949 phrases this obligation as such: “A physician shall act only in the patient's interest when providing medical care which might have the effect of weakening the physical and mental condition of the patient” [7].

Insofar as the objective of clinical trials is to advance science through the randomization of patient treatment, the collaboration of physicians and the involvement of their patients in such trials would seem at first glance to be ethically problematic. The question that looms at the outset of this study is whether clinical

trials may be conducted under ethically acceptable standards. To answer that question we must begin by examining the nature of clinical practice and especially the character of the physician–patient relationship. This reflection will help us to determine whether the collaboration of physicians and the participation of patients in such trials can ever be justified, and if so, under what conditions.

### 26.2.1 Requirement of Genuine Consent

The startling breakthroughs of biomedical science during the last decades, and the promise of more revolutionary treatments and drugs, obscure the simple fact that medicine is primarily a relationship between two individuals. As Thomasma [8] has pointed out, “medicine is not a theory about the body or how the body works, so much as a theory about practice, about the ways in which physicians and patients, other health professionals, and institutions interact to bring about healing” (p. 246). This interaction begins when a person in a vulnerable state due to physical or psychological illness or distress requests assistance and is admitted to medical care. In doing so, the patient assumes that the physician, as the person in charge of his or her care, will act in his or her best interests and will prescribe the most effective medication or treatment available. What characterizes this relationship is therefore trust and responsibility. The patient trusts that the physician will act in his or her best interest while the physician accepts to shoulder the responsibility of taking care of the patient to the best of his or her abilities.

In the Hippocratic tradition, which still shapes our understanding of the ethos of medicine, the physician–patient relationship is understood in terms of a certain kind of friendship [9, p. 20]. The friendship of the physician (*philantropía*) sees to the well-being of the patient and is expressed through the love for his art (*philotekhnía*), which is healing. While there were many practitioners of the art of healing in ancient Greece, such as magicians and exorcists, the Hippocratic physician loved his art as a *techné*, that is, as an empirically established technical body of knowledge useful to his craft. The Hippocratic Corpus [10, p. 314] states unequivocally the commitment of the physician to empirical science:

Do not rely on conclusions that result from mere reasoning, but (rely) on evidence. Arguments in the form of plain rhetoric are false and easily defeated. Therefore, you should stick to the facts and scrutinize them, if you are going to acquire faultless capability, which we call Medicine.

Thus, the Hippocratic physician loved his art through his concern for the well-being of his patient and his patient through the practice of his scientifically established craft [9, p. 45]. In other words, the physician loves his patient as a friend because he can come to his aid by applying his craft and knowing that this craft is well founded and likely to produce good results. In this view, science is an essential medium but still plays an ancillary role to the practical object of medical practice, which is the healing of the patient.

In the context of the discussions of the morality of randomized clinical trials, this view of the physician–patient relationship has undergone a radical change. Medical science is no longer seen as primarily an instrument that permits the physician to

exercise his art of healing but as a competing object of his professional responsibility. In other words, it is no longer only a medium but equally an end of the physician's work and thus regulated by an independent set of values. In a very influential article, Lellouch and Schwartz [11] characterize the conflict of interest facing the physician as one between an individual ethics and a collective ethics.

Clayton [12] spells out what these competing sets of values involve. Individual ethics refers, on the one hand, to the duty of the physician "to apply existing knowledge for the best possible treatment of each individual patient." Collective ethics, on the other hand, refers to two duties of the physician: (1) "to acquire new knowledge so that, by that advance, future patients might benefit" and (2) "having acquired that knowledge, to accurately communicate it to other physicians." Gifford [13] notes the existence of the physician's "therapeutic responsibility," which he believes, *prima facie*, to be incompatible with his collaboration in randomized clinical trials. However, mindful of the benefits that clinical trials bring them, patients might allow the physician to collaborate and would willingly submit themselves to randomized treatments. In this contractual scheme inspired by the philosophy of John Rawls and Robert Nozick, randomized clinical trials could be considered ethical to the extent to which the patients acquiesce to participate and permit the physician to override his or her therapeutic responsibility in order to further a more collective good.

Gifford's contractual approach defers the physician's fiduciary responsibility and places the decision of RCT participation on the shoulders of patients themselves. This approach corresponds well with the insistence upon the informed consent process and the principle of autonomy, which permeates today's ethical climate. This principle of autonomy or permission is widely considered today as the lowest common ethical denominator and the foundation of modern ethics. The reasoning behind one of the major proponent of this view, H. T. Engelhardt [14], is the following. There is a plurality of moral goods, which human beings can pursue and enjoy such as freedom, work, health, pleasure, religious faith, and many others. How these values are ranked in the pursuit of a good life depends very much upon one's political, religious, and philosophical convictions. Furthermore, these different values cannot be arbitrated by appealing to an objective standard, to which everybody can agree. In light of this plurality of moral opinions, it is up to the individual to determine what is good for him or her. A man suffering from prostate cancer might, for instance, want to choose between radical prostatectomy and more conservative methods depending on what outcome he values. If he is younger and still wants to father a child, he might value the preservation of fertility very highly. However, if he has no such thoughts, he might opt for the treatment that offers the best possible outcome in terms of life preservation. It all depends on what values he espouses and what plans he has for the future. Free and informed consent is, in this light, the gold standard of medical ethics.

The principle of autonomy or permission, with its stipulation of free and informed consent, is, in the view of most ethicists, a necessary condition for patient participation in any RCTs, with the possible exception of trials carried out in emergency situations where patients or their next of kin cannot give immediate consent. There is also general agreement that participants should be informed specifically about the relevant details of the randomization process. It is not the quantity of information



given to the research subject that determines whether the consent is free and informed but his or her ability to understand relevant facts in order to make an informed decision. That is why some ethicists talk about “genuine” rather than informed consent [15]. This wording emphasizes the fact that true consent requires health care providers to do their best to communicate accurately and in a way that is understood by patients, volunteers, or relatives. It presupposes that these can understand procedures and risks and react to the limits of their understanding. Moreover, it requires care in detecting and eliminating lack of consent.

While there is general consensus about the necessity of informing patients on randomization procedures as part of the consent process, and prior to trial participation opinions differ on the necessity of informing research subjects about interim data. Those who argue against such disclosure [16] fear that interim data can be unreliable and misunderstood by research participants and the media, which in turn could compromise the completion of RCTs. In a contractual understanding of RCT participation, it is possible to conceive that patients would waive the right to such data, and would accept an “incomplete disclosure” [17] as part of the consent process. Lilford et al. [18] disagree and argue that the praxis of withholding interim data has no public endorsement, is contrary to the spirit of regulations, and is the reflection of a “culture of secrecy and scientific imperialism” that is self-fulfilling. Making interim data available, they argue, would not necessarily reduce recruitment to clinical trials but would most likely increase it because it would enhance public trust in the study. As a result, it would also widen the base of the participants, thus improving the trial’s generalizability. The authors provide the example of the second international study of infarct survival (ISIS 2) in which disclosure of improved survival rates of heart attack patients taking clot busting drugs increased recruitment as sceptical physicians became more interested in the interim results. This example shows that more stringent ethical standards do not necessarily decrease productivity. In some instances they could even increase it.

### **26.2.2 Requirement of Patient Benefit Optimization (PBO)**

The Hippocratic model of the physician–patient relationship, where the physician bears the main burden of responsibility and decision making, is considered by many as too paternalistic. Today, patients want to be full-fledged partners in the clinical process, together with the physician. This is all the more important as the physician’s loyalty to the patient competes with his or her scientific ambitions. The patient must therefore also fend for himself. Informed consent procedures are one step on the way to establishing more equality between patient and physician. Furthermore, a more democratic approach might give added credence to clinical trials and does not necessarily result in less efficiency. On the contrary, more public trust through greater transparency could result in better recruitment and more generalizability. But there are instances where the requirements of informed consent might not be sufficient to make a clinical trial ethically acceptable.

In addition, the principle of autonomy with its requirement of informed or genuine consent builds upon a model of the patient as a rational agent. This model does not describe accurately how patients make clinical decisions concerning their own welfare. Individuals often behave and make decisions irrationally. To compare treatments is often a complex task, involving a multitude of effect, safety, and con-



venience dimensions. When asked to choose between two treatments, we may therefore—irrationally—prefer treatment A to treatment B, B to C, but C to A, thus violating the transitivity axiom in decision theory. Furthermore, many people have difficulties in understanding small risks. Lastly, the patient is not always able to give informed or genuine consent as, for example, in an emergency or situation of distress. Such a situation may arise when the patient is informed that he suffers from a chronic or fatal disease. He may feel that the physician's time is limited and he may interpret a question about trial participation as an advice or wish from his physician. If the physician clearly communicates that he does not know which treatment is best, and has no advice about study participation, the patient may give his formal consent by signing a piece of paper without really understanding the content and relevance of the information received and thus without acting in a truly free and informed manner.

Because patients are not always capable or even willing to act rationally, and bearing in mind that genuine consent is sometimes difficult if not impossible to retrieve, the attending physician ultimately bears the responsibility of furthering the well-being of his patient and assuring that the trial itself is ethically sound. How can this be assured?

Over the past 30 years, the notion of equipoise has been used to ethically justify physician collaboration in RCTs. It was first used in this context by Charles Fried [19] in 1974 when speaking about the “balance of opinion” among physicians and investigators, when no treatment option in a multiarm RCT is perceived to be inferior to the other or others. From this balance of opinion the term has taken a more psychological turn to connote the “state of genuine uncertainty on the part of the clinical investigator regarding the comparative therapeutic merits of each arm in a trial” [20]. According to this standard, RCTs would be morally justified to the extent to which the clinical investigator can honestly claim a “state of genuine uncertainty” between the treatment options incorporated in the study design.

Most ethicists and clinicians agree that true equipoise rarely exists. Clinical trials are financed and conducted because there are reasons to believe that a new drug or treatment option will improve care, which is also why clinicians choose to implement them in their practice. Notwithstanding this *a priori*, there might be, however, scholarly discussions in the scientific community about the relative efficacy of a novel therapy. So while absolute or “theoretical equipoise” might not exist, there might still be a nagging uncertainty among investigators and health care providers about the benefits of the different treatment arms. This practical uncertainty or “clinical equipoise” would be the requirement that scientifically justifies such trials. Another term that is used in this context is the ethical theory of “null hypothesis” pertaining to genuine medical uncertainty. The hypothesis entails that clinical trials are justifiable when no consensus exists about which is the better treatment [12, p. 472].

The null hypothesis theory is an easy and questionable justification of RCTs. Consensus among clinicians and researchers about the benefits of a certain treatment versus an alternative is rare, even regarding established treatments. Given that consensus rarely exists, most if not all RCTs could be considered, according to the equipoise theory, as morally justifiable. Moreover, as Lilford [21] has pointed out, even the justification of “uncertainty” regarding treatment outcomes is indeed a

questionable criterion. “Unknown” or “uncertain” is not equivalent to saying that all possible effects are equally likely. Clinicians and investigators always possess some evidence about the merits of the novel treatment or drug before the trial begins. Such information, which can be the basis for drug modeling, is obtained through animal or in vitro experiments, the same treatment in other diseases or, in some cases, even randomized trials. Saying therefore that the effects of the novel treatment are uncertain is certainly ambiguous. The patient might interpret such a statement as meaning that the recruiter has no idea at all about the benefits of the new treatment, which is not the case. Certainty and uncertainty are not absolute terms, and the physician is therefore most likely to have some idea about relative advantages and disadvantages. As Lilford states, “the blanket term of unknown sidesteps any indication of the magnitude of possible effects, reducing the chance that potential participants will be able to appreciate what is really at stake” [21]. In addition, the whole theory of equipoise and ethical null hypothesis is paradoxical because the more an investigator knows about a compound or a treatment the more he would draw away from equipoise. Oddly, it would encourage ignorance.

The notion of equipoise, as used in the debate over RCTs, is therefore problematic because it places too much emphasis on the physician’s state of mind. Pleading uncertainty, as Lilford points out, says very little about what is known and what should be told to a patient concerning the merits or drawbacks of a certain treatment option. What is really ethically at stake in RCTs is not whether the physician considers himself or herself in a state of equipoise but whether the patient will benefit more from inclusion in than from exclusion from the trial. This is precisely what the requirement of PBO entails. If the patient stands to benefit personally by inclusion in the trial, the RCT would meet the requirement. If he or she does not, the requirement would not be met.

Studies in which patients in the control arm receive the standard treatment while patients in the experimental arm are given a novel drug expected to be superior would fulfill the requirement of patient benefit optimization. With all patients receiving standard treatment as background medication, it is often acceptable to compare the addition of a new drug with the addition of placebo [5]. But what about placebo-controlled trials where patients in the control arm would be given no pharmaceutical treatment at all? Most ethicists agree that there are conditions when placebo-controlled trials are acceptable and others when they are clearly not. If the ailment under investigation is not serious, and when there is minimal risk and minimal burden for research participants, the use of placebo is generally considered as ethical. A trial studying allergic rhinitis is a good example since this ailment typically does not impair health or causes severe discomfort. Since the possibility of harm is minimal, the benefit derived for the patient in the control arm could be the satisfaction of helping others. Such a trial, despite the use of placebo, could therefore meet the requirement of patient benefit optimization. There are, however, other situations where the use of placebo is unethical. If effective, life-saving, or life-prolonging treatment is available, and if research subjects who receive placebo are more likely to suffer serious harm, a placebo-controlled trial is unethical. The disagreements among ethicists, investigators, and clinicians center on whether placebo controls can ever be used in trials where a treatment known to be effective is available and when there is some potential harm to participants being randomized [22].

The Declaration of Helsinki [23], the most authoritative text on matters of research ethics, is not totally clear on this matter. Paragraph 32 states: “The benefits, risks, burdens and effectiveness of a new intervention must be tested against those of the best current proven intervention, except in the following circumstances: the use of placebo, or no treatment, is acceptable in studies where no current proven intervention exists; or where for compelling and scientifically sound methodological reasons the use of placebo is necessary to determine the efficacy or safety of an intervention and the patients who receive placebo or no treatment will not be subject to any risk of serious or irreversible harm. Extreme care must be taken to avoid abuse of this option.” Commentators argue that this wording in the latest revision of the declaration signals a move to a more permissive attitude toward placebo use. This interpretation is motivated by the fact that the 1996 version included the provision that “in any medical study, every patient—including those of a control group, if any—must be assured the best proven diagnostic and therapeutic method” (Paragraph II.3). This sentence, once again, has been eliminated. Despite this omission, the overall tenet of these both versions seems to rule out the use of placebo wherever proven treatment exists. Paragraph 32 in the latest version adopted in 2008 in Seoul, lists two situations where placebo is acceptable: where there are “compelling and scientifically sound methodological reasons,” *or* where the condition under study will not lead “to any additional risk of serious and irreversible harm” [23]. What has led commentators to interpret more leniency regarding placebos is the option of the connector “or” instead of “and” between the two provisions. It would seem therefore that the Declaration of Helsinki, in its present version, could be interpreted as saying that the use of placebos for “compelling and scientifically sound methodological reasons” is permissible even when there is an increased risk of serious harm.

The spirit of the text does not go in that direction and the wording of Paragraph 6 ultimately invalidates such an interpretation. This paragraph states that “in medical research on human subjects, considerations related to the well-being of the human subject should take preference over all other interests.” However, much of the confusion surrounding Paragraph 32 has to do with the terms “best current proven” or “proven” intervention. Does the text mean “proven” in existence, in a local context, according to the judgment and knowledge of the attending physician, or according to the best interests of the patient? The Nuffield Council on Bioethics [25] suggests that “best proven” should mean “the minimum standard of care that should be offered [in the control arm] is the best intervention available as part of the national public health system.”

The Declaration of Helsinki encourages ethically optimized designs that exclude the use of placebos. Consider, for example, a situation where it is scientifically interesting to compare two different drugs, X and Z, which both are available on the market. By giving each patient the drug that is expected to be best for him or her, and assuming that this choice will be different for different patients, it is possible to learn more about the relative merits of the two drugs. Say that X is perceived to be better in patients with mild disease, expressed as a baseline value of a certain covariate below a cut-off  $c$ , while Y is thought to be better for patients with covariate values above  $c$ . If the responses are collected, the results from both groups can be modeled as functions of the covariate and extrapolated so that a comparison of the expected responses of the two treatments can be done. As more and more observations are collected, the estimated cut-off will change and treatment practice will

improve over time. This type of trial would be in conformity with the Declaration of Helsinki ethically optimal but, unfortunately, it is not very efficient or viable compared to standard RCTs.

More often than not, it is desirable for methodological and regulatory reasons that the experimental treatment Y is tested against placebo without using X. Can any such trial withstand the PBO requirement? Imagine a trial where participation is beneficial on average but unfavorable for the patients randomized to one of the treatment arms. Say that mortality is estimated to be 10% for placebo, 9% for the “best proven” therapeutic method C, and 8% for a new nonapproved drug candidate X. In order to get regulatory approval for X, the trial would need to prove its superiority versus placebo, which is much easier and cheaper than to show superiority, or even noninferiority, versus C [26]. Therefore, a blinded trial is proposed where  $\frac{3}{4}$  of the patients are randomized to X and  $\frac{1}{4}$  to placebo. The expected death risk for a patient taking part in this trial is  $\frac{3}{4}(8\%) + \frac{1}{4}(10\%) = 8.5\%$ , which is lower than the 9% risk experienced by the patients that do not take part in the trial and receive C. This trial would meet the requirement of PBO because, statistically, the patient would benefit more from trial inclusion than from exclusion. It is still, however, very debatable whether the trial would be ethically acceptable because it inevitably exposes certain patients to less than “best proven” treatment. This example shows that while the PBO is a necessary ethical requirement, it is not alone a sufficient requirement for ethically validating a trial.

The example above also indicates that the PBO should not be interpreted as an average utility for all participating patients in the trial but should also be relevant and applicable to every individual participant. Retsas [27] has formulated a thought experiment, which illustrates this point through what we will call the VIP principle:

Imagine that you are an acclaimed expert on a particular disease and you are invited to offer your services to your head of state. Will the head of state be enticed to participate in a randomised trial leaving his or her treatment to chance, or will treatment be provided on the best available knowledge of the day? It can be observed safely that there are hardly any prime ministers, secretaries of health, doctors, professors, or similar persons who enlist in the ranks of randomised cohorts that serve the progress of medicine!

Retsas' thought experiment is very general and suggests that all RCTs are unethical. While some ethicists have argued that such is the case [28], most believe that it is not. Retsas is also wrong in stating that “doctors, professors or similar persons” do not participate in RCTs. They often do, and the history of medicine is full of examples of clinical researchers risking their lives with self experiments in the zealous pursuit of scientific knowledge. But the point is well taken that physicians seldom recruit VIPs, be they heads of state or their own children, to RCTs. When human life is truly at stake, leaving treatment option to chance, even when there are reasons to believe that the experimental group stands to benefit, is indeed counterintuitive. The fact that a high percentage of drugs tested in phase III clinical trials never reach the market would lead us to exercise extreme caution. The poignant question that Retsas addresses is this: If VIPs are seldom, if ever, randomized by physicians, why should anybody else be? While the VIP principle might be intuitively sound, reason can still ethically motivate RCTs if every individual patient

can be assured that he or she will receive the best possible treatment or its equivalent.

### 26.2.3 Requirement of Nonexploitation

It has been argued that genuine consent should be a necessary condition for enrollment in a clinical trial. Given that RCTs distort the fiduciary relationship between patient and physician through the latter's participation in what is essentially research and not primarily therapy, the patient's genuine consent is essential in order to reestablish the contractual balance between the two parties. Such a requirement, it has also been argued, is not a sufficient condition for patient enrollment in RCTs. Besides genuine consent, the physician must also ensure that the design of the trial, especially its expected utility, can be aligned with the patient's best overall interests. From an expected utility perspective, the patient should stand to benefit more from inclusion in the trial than from exclusion. There are, furthermore, other cases involving research subjects where the ethical issues involved in the physician–patient relationship are not covered by the two aforementioned requirements, hence the necessity of formulating the requirement of nonexploitation.

A good example of where this requirement is at stake is first-time-in-man (FTIM) trials planned in healthy volunteers. Usually, trial participants are offered economic compensation for the time and discomfort associated with the experiment. Is it unethical to pay these participants for taking medical risks, and would such payment constitute a form of coercion? Assume that there are no discomforts in such a trial, that trial participation is short, that there are no health gains for the volunteers but a slight risk of experiencing an unexpected serious adverse effect. With the possible exception of altruism, there would then be no other reason to participate in the trial other than economic compensation. Assume, furthermore, that a firefighter in the United States is paid to take a small risk in a clinical trial in much the same way that he is compensated for work-related risks. Although it is hard to assess trial risks objectively, it is certainly possible to give some ball-park estimates. Severe adverse events are extremely rare but, as the incident in Northwick Park Hospital in March 2006 shows, they do occur and could potentially lead to death. As a very approximate risk calculation, assume an expected 0.01–0.1 deaths per year globally in FTIM trials, 100 new approved drugs per year, 10 times more drugs tested in humans, and 10 individuals per trial. Then the average death risk for participants in FTIM trials would be about 1 in 100,000 or 1 in 1,000,000. A more careful analysis, using historical data and expert input, may give much better estimates. The risks will also depend on the different kinds of drugs. Hence, there are nonnegligible risks, but these are small and comparable to those taken by some groups as part of their professions. As a comparison, the fatality rate per year in the United States is almost 1 in 1000 for logging workers and fishers and 1 in 4000 for truck and taxi drivers [29]. Healthy volunteers in FTIM studies are paid for their time and for the discomforts they might suffer as a result of the study. From a DA perspective, however, risks and economic gains can be factored in the decision about whether or not to take part in the trial.

The requirement of nonexploitation stipulates that research subjects must be sufficiently free to be able to consent genuinely or, by the same token, to decline trial participation. Generous financial benefits would cast a doubt over certain

persons' capacity to decline an offer of participation, especially those who are economically vulnerable. In this respect, the additional protocol to the Bioethics Convention on biomedical research, adopted by the Council of Europe in January 2005, states very clearly that "no undue influence, including that of a financial nature, will be exerted on persons to participate in research. In this respect, particular attention must be given to vulnerable or dependent persons" [30, Article 12]. The explanatory report of the protocol does not strictly define the term "undue influence" but suggests that it is related to coercion and "may be exerted in particular on a person in a weak or feeble condition, so that very little pressure will overbear the person's will, and make the individual feel that he or she must agree" Furthermore the text specifies that "compensation should not be provided at a level that might encourage participants to take risks that they would not otherwise find acceptable" (Paragraph 64). From an expected utility perspective those who stand to gain the most from trial participation as volunteers, especially the economically weak, would paradoxically be those who are the most discriminated by the requirement of nonexploitation. This short-term discrimination is necessary, however, to protect their long-term interests.

Biomedical research in developing countries is yet another context where the requirement of nonexploitation is especially relevant. Given the accrued risk of exploitation of vulnerable individuals and the potentially huge economic and cultural impact of RCTs in the communities where they take place, the requirement of individual informed consent is not enough to protect the interests of research subjects in this particular context. The European Group on Ethics [31, p. 14] has thus pointed out that "it may be appropriate to seek agreement on the implementation of a research project from persons representative of or invested with a certain authority within the community, or the family." In developing countries, where the capacity of individuals able to consent has been weakened by earlier exploitation or where the principle of autonomy is not widely acknowledged, the consent process must therefore be strengthened by the implication of the larger community. The UK Nuffield Council on Bioethics has also written on the subject and flagged the requirement of nonexploitation as one of three others that should guide biomedical research in developing countries: (1) the duty to alleviate suffering, (2) respect for persons, and (3) sensitivity to cultural differences [25, p. 52].

In most cases, patients in developing countries would receive better medical care while taking part in clinical trials and would therefore generally stand to benefit more from inclusion than exclusion. Seen from an overall benefit perspective, the requirement of nonexploitation could at firsthand discriminate them by placing more obstacles in the way of trial sponsors wishing to conduct such RCTs in developing countries. Such short-term losses would, however, be compensated by creating a better long-term perspectives for individuals as well as developing communities.

The situations detailed above deal with cases where persons able to give consent, and who might wish to participate in a trial, need nonetheless additional protection against possible exploitation at the hands of trial sponsors and collaborating physicians. But what about cases where research might be a condition for treatment? In the United States, the CMS (Centers for Medicare and Medicaid Services) have recently stipulated, for example, that coverage for new users of approved anticancer drugs depends on the willingness of patients to enroll in a clinical trial sponsored by the National Cancer Institute. The CMS policy affects research on tests or treatments



whose efficacy has not been established for the patients being studied. Orentlicher [32, p. 22] argues that such a condition should also extend to research aimed at determining which treatment is most effective when a patient can be offered one of multiple established treatments. Orentlicher's position is motivated by his concern that important studies are being delayed—and medical progress impeded—by difficulties in securing the participation of individuals. In one example he cites, more than half of the patients invited to participate in the study declined with the total recruitment taking twice as long as expected (16 months instead of 8 months).

The Declaration of Helsinki stipulates that participation in any trial is voluntary, that refusal to participate will involve no loss of benefits, and that such participation may be discontinued at any time. Orentlicher's hypothetical protocol would also require voluntary participation and would not deprive the patient of any benefits to which he or she would otherwise be entitled. What he argues for is the individual physician's right to deny care to any patient unwilling to participate in the study. Such a patient would be referred to another physician. The physician–patient relationship is a contractual one and does not oblige the specialized physician to admit any patient who requests medical attention if such medical attention can be received elsewhere and does not represent a danger to the patient in question. Orentlicher argues that trial participation in view of the common good is a valid and ethically justifiable condition for patient care in case the patient can receive equivalent care elsewhere.

Orentlicher's unorthodox position is certainly worth thinking about because it emphasizes that there are societal considerations in any RCT that a strict interpretation of the physician–patient relationship does not cover. Moreover, in the hypothetical protocol that he proposes, the medical interests of the patients are secured given that the methods being compared are well established and that the patients would be treated by another physician if they declined the offer of participation. However, as we stressed in the beginning, the nature of the physician–patient relationship should be inspired both by the sentiments of friendship and care of the doctor toward his patient and the love for his healing art. It is this *philanthropía* that transcends the terms of different protocols or the provisions of trial designs and that ultimately ethically motivates the physician's *philotekhnía* and hence the possibility of carrying out clinical trials.

### 26.3 CHALLENGE OF CLINICAL TRIAL EFFICIENCY

Most experts in the field consider that efficiency in clinical development is far too poor. One reason that explains this is the rather simple design imperfections that can still be found in some trial publications. In addition to weeding out these imperfections, trials can be further enhanced by implementing new design methodology and combining them with novel supporting technologies, such as Web-based data capture. The FDA's Critical Path has identified a large number of possibilities that could increase the productivity in different stages of drug development [33, 34]. We will focus here principally on some of the possibilities that relate specifically to the improvement of clinical trial design. After an overview of the current situation (Section 26.3.1), we will present a DA approach (Section 26.3.2) to a number of design questions. This paradigm is then applied to the different key aspects of a trial



design. Primarily, we will discuss how each patient can contribute more information (Section 26.3.3), analyze the optimal size of a trial (Section 26.3.4), and outline the possibilities that exist to be more adaptive to incoming information (Section 26.3.5).

### 26.3.1 Background

The great advances in medicine have contributed significantly to the prolongation of life expectancy during the last century. As part of this, the role of clinical trials during the second half of the century can hardly be overestimated. New pharmaceutical treatments have almost eliminated many diseases and reduced the risks for and consequences of many others.

Trends in productivity are, however, discouraging for drug development. Despite rapidly increasing development costs, the number of approved new drugs has been declining recently. One important reason for the productivity fall is the increased regulatory demands. A further reason is that drug developers try to solve increasingly difficult medical problems. In areas where there are already at least partly effective treatments, a new drug can often not be compared directly to placebo for ethical reasons. To show an effect versus placebo must thus be done indirectly, by testing the new drug and an active comparator in clinical trials, which requires considerably larger sample sizes [26]. Moreover, diseases for which no effective drugs exist are usually much more complicated and difficult to treat.

As many as half of the drugs entering confirmatory studies (phase III) fail [2]. Some drugs, such as the best-selling Vioxx, have also been withdrawn from the market due to safety findings postapproval. Such failures undermine the public trust in medicines and put increased pressure on regulatory agencies. Regulators may react by increasing the demands for safety documentation before drug approval, which will lead to accelerating development costs. A trend can be seen toward a more frequent need for large outcome trials in order to generate enough data to convince regulators of a positive benefit–risk relation. A net result is that, despite huge investments, too few new pharmaceuticals reach the patients in need.

There are a number of hopeful signs, however. First of all, the FDA has clearly recognized the productivity problems and has launched its Critical Path Initiative [33] as a response. This initiative aims at finding innovative ways to improve drug development, in partnership with academia and industry. The Critical Path Opportunity List [34] from 2006 formulates a large number of suggestions, including the use of different biomarkers, innovative/adaptive study designs, and modeling and simulation. It also flags the need to develop methods for indirect comparisons with placebo, in cases where placebo-controlled trials are unethical [26]. Below, we will focus on issues dealing with clinical study design and will not highlight changes in drug discovery technologies and IT per se. It should be noted, however, that the development of information technologies, such as Web-based data capture, is important for the possibility of conducting sequential and adaptive trials in practice.

### 26.3.2 Decision Analysis Approach to Study Design

As we have noted, the great need for increased efficiency in clinical development is accompanied by new opportunities to improve clinical trial design. A crucial issue

is how to best utilize these possibilities. We will address this challenge by proposing the following systematic, integrated, and quantitative approach:

1. Gather relevant information.
2. Integrate the information into a quantitative model.
3. Define the objectives of the trial.
4. Translate the objectives into a utility function, depending on the trial outcome.
5. Generate a variety of design alternatives.
6. Apply the ethical sift.
7. Simulate the trial outcomes for the interesting designs.
8. Calculate expected utility.
9. Choose the design with maximal expected utility.
10. Check the robustness of the solution to changes in the model assumptions.

The items on this list do not necessarily come in a strict chronological order but are often strongly interrelated. Given that the overall goal is to find a good enough design within the ethical framework, the modeling and DA process will depend on what is needed to accurately answer the design question. Consequently, the proposed approach does not necessarily have to be complicated or time consuming. In many cases, relatively limited work is enough to improve the study design.

This chapter will not describe all the items on the list in depth, as many of them are accurately covered in the existing literature. Much of the literature on modeling is topic specific. A classic text on model building, based in statistics, is Box and co-workers [35]. Pharmacokinetic/pharmacodynamic modeling is described in [36, 37]. Hoppensteadt and Peskin [38] present mathematical modeling for medicine, emphasizing physical processes in physiologic systems but ignoring pharmaceutical interventions. Burman et al. [39] focus on clinical development and provide a shorter, but very similar, version of the list above. Parmigiani [40] treats modeling in medical decision making, taking the perspective of the patient and his treating doctor. This book applies DA and Bayesian methods and describes computer simulations in a relatively nontechnical way. Hunink et al. [41] is an easy read on DA, taking a similar perspective as Parmigiani. It includes a chapter on information gathering (Chapter 8) and notes that subjective components are often needed in the analysis (p. 215). DA can also be viewed from a management perspective [42] and applied to drug development [43]. A common component of decision analysis and decision theory is Bayesian models. Both Spiegelhalter et al. [44] and Berry [45] provide excellent accounts of Bayesian methods in connection to clinical trials and include sections on ethics. More general texts on DA are Raiffa [46] and the broad overview in Howard and Matheson [47].

Simulations, expected utility calculation, and optimization are quite technical matters but specialized programmers and mathematicians do the job. Literature on random simulations includes [48–50]. Mathematical and numerical analysis can sometimes substitute, and often complement, simulations. Many textbooks cover such areas, including optimization. Although items 7–9 on the list above can be carried out by a sufficiently skilled programmer, the robustness checks should be a

concern to the whole scientific team, in interaction with the programmer. The team, and preferably also experts who have not taken part in the modeling, should change the key assumptions and look at how these changes affect the conclusions. If the solution is very sensitive to the value of a parameter, there is a strong indication that more knowledge should be sought regarding its true value. Renewed information gathering or focused experimentation may be indicated.

Naturally, the utilities may be quite different depending on what perspective you take. From a patient perspective, and in order to comply with the spirit and the letter of the Declaration of Helsinki, placebo should preferably not be used. FDA and European regulations stipulate, however, that a positive effect has to be demonstrated versus placebo, by direct or indirect comparison. Finally, a trial sponsor has to consider the return on investments. Although there are complementing company perspectives, we will take the net present value (NPV) as a proxy for the company utility. Profit maximization from the company's side must, however, be made under the constraints outlined above. Among the possible ethically compatible designs, the company will try to maximize the expected NPV for the project or, more generally, for the company. The value of the project will reflect also the regulatory hurdles and the possibilities of convincing payers about the cost-effectiveness of the drug.

Let us now study the key elements of clinical trial design. The following three subsections will address different aspects of this question.

### 26.3.3 Optimizing Information per Patient

The general aim of all experiments is to reduce uncertainty. The obvious way to decrease the uncertainty of trial results is to increase its sample size. In Section 26.3.4, we will apply DA to the problem of finding the optimal sample size. However, the most cost-effective way of retrieving trial information is often to reduce the variance for each patient. There are a number of ways to do this such as devising more than one treatment and several measurements per patient, choosing a more informative response variable, and choosing a patient population with a larger response.

**Covariates and Multiple Measurements** For many chronic diseases, there is a response variable [e.g., cholesterol level, forced expiratory volume (FEV<sub>1</sub>), subjective symptoms, or quality of life (QoL)] that fluctuates over time. Technically, we will assume that the variable follows a linear normal model, implying a linear dependence on covariates as well as normally distributed residual terms. The variability in the response variable during treatment has several different causes. It is standard procedure to include the baseline value of the response variable as a covariate in the linear model, together with other potentially predictive covariates. This will reduce the variability due to explainable differences between patients and may lead to considerably higher efficiency of the trial. These covariate adjustments are part of good statistical practice and concern more the analysis than the study design. Some of the other ways of reducing variability require larger modifications of the study design.

A straightforward example of how DA can be used concerns the decision of how many measurements should be taken per patient. One cause of variability that is not handled by covariate adjustments is the day-to-day fluctuations within each

patient. This variability can be reduced simply by taking measurements on different days and using the average of the observations as the response variable. Many studies collect the necessary information without using it in the analysis. Say, for example, that the objective (item 3 on the list in Section 26.3.2) is to estimate the primary variable in the most cost-effective way as possible. We then proceed by setting up the utility function (item 4). In this case, the cost per information is a function of the number of measurements  $M$  per patient. Whether more than one measurement should be taken is essentially a question of how much of the total (unexplained) variability is attributed to day-to-day variability and how much it costs to reduce it. In a straightforward model (item 2), the variance of the average for a patient is proportional to  $1 + V/M$ , where  $V/(1 + V)$  is the fraction of the observation variance that is attributed to day-to-day variability. Assume (item 2 again) that the cost per patient is proportional to  $1 + CM$ . As information is the inverse of the variance, the cost per unit of information is proportional to  $(1 + CM)(1 + V/M)$ . The number of measurements,  $M$ , can easily be chosen (item 9) to minimize this cost/information ratio, given that the value of the parameters can be estimated (modeled) from existing information (item 1).

Say, for example, that the cost per patient is \$11,000 if only one measurement is taken and that the additional cost per extra measurement is \$1000. Assume as well that 50% of the total response variance of 200 units is due to day-to-day variability. The cost and variance per patient are then  $10,000(1 + 0.1M)$  and  $100(1 + 1/M)$ , respectively. The cost/information ratio,  $1,000,000(1 + 0.1M)/(1 + 1/M)$ , is minimized if  $M = 3$  measurements are made. The cost per patient increases by 18% if  $M = 3$  compared to  $M = 1$ , but the variance per patient decreases by as much as one third. Consequently, the sample size can be decreased by one third, from 200 to 133 per arm if a standard error of 1.0 is to be achieved. The cost per unit of information is reduced by 22%; in a two-arm study the total cost decreases from \$4.4 million to \$3.52 million. A robustness check (item 10) indicates that  $M = 2$  or  $M = 4$  give most of the benefit that is achieved with  $M = 3$ , and that  $M = 2$  is a viable solution for a range of values for the parameters, for example, if the cost per measurement is twice as large as originally assumed.

This example does not indicate that the number of measurements should generally be increased. In other situations the analysis would indicate that costs should be cut by decreasing the number of visits. The point is that the efficiency of a design is often a trade-off between estimation precision and costs. By clearly defining the trial objective, integrating information about variability and economical costs, and quantifying a utility function, the decision problem is clarified. This kind of reasoning is considerably more far-reaching than just deciding on a hunch the number of measurements. In the example above, there could be complications given that the true underlying effect is changing over time, and that there is a nontrivial correlation structure between observations taken on different days. Typically, the effect is increasing toward a steady-state value and the correlation between two observations decrease with the time between the observations. The solution to both these problems is essentially the same: Model the time-effect process and the variability structure and then apply DA to the problem of finding the optimal sample scheme and test variable.

A common way to waste information is to dichotomize data, that is, to replace a continuous response with a zero-one variable in order, for example, to classify

patients as “responders” and “nonresponders” and ignore the rest of the information regarding the response. Occasionally, there are good reasons to classify patients as responders or nonresponders, based on whether a continuous response variable is above or below a specified cut-off value, and base the primary analysis of the constructed zero–one responder variable. However, in most situations, a direct analysis of the original continuous variable gives much more information [51, Section 8.2.4]. If the treatment effect is small, for example, the efficiency gained by using original data is about 57% when the patient population is divided in two approximately equally sized groups based on a normally distributed variable. The efficiency is more than doubled and almost threefold when the dichotomous population split is 80%–20% and 90%–10%, respectively. It is usually possible to model the distribution of the response (item 2), and then the effects of dichotomization are easily investigated by simulating (item 7) the power for the alternative response variables.

**Cross-Over Designs** The between-patient variability is completely eliminated if the patient can serve as his or her own control. This is the idea behind cross-over trials, where two or more treatments are given to each patient during different periods. Such designs are well studied [52, 53] and have been utilized systematically in some areas for many years, especially in phase I trials of pharmacokinetics and also for patient trials in some specific medical indications. However, there are possibilities of expanding the scope of these designs considerably in the future. When a cross-over design can be applied, the efficiency of the trial can normally be reduced drastically. The number of patients in the trial can often be reduced by considerably more than 50%.

There are some clear limitations to the use of cross-over designs, however. Cross-over trials are only possible for chronic diseases and obviously not suitable to investigate, for example, mortality. Their potential is largest in proof of principle and dose finding and for discriminating against similar competitors. They are best suited when the treatment response is relatively fast (asthma, lipids, some lab values, etc.). Carry-over effects have been seen as an important problem. However, there are in many cases possibilities to handle such potential effects by wash-out periods, statistical corrections, or pharmacological modeling.

When a cross-over design is considered as an alternative to a parallel group design, the usual tools of modeling, simulation, and DA can be of help. Using available information (item 1) in terms of earlier trials with the same and similar drugs, between-patient and within-patient variances can often be estimated or at least guessed. The time–effect curve and potential carry-over effects can be modeled in a similar way (item 2). Based on some models, it is possible to predict (item 7), often via simulations, the likely effects in the planned trial. There will often be trade-offs (items 4 and 9), for example, between the treatment duration and factors such as inclusion time, potential bias, and variability. A cross-over design will probably have a shorter inclusion period, as the sample size is lower, but a longer total treatment time. It is sometimes reasonable to shorten the treatment time per period, although this may lead to the steady-state effect not being fully reached.

Cross-over trials make it possible to assess whether there is a between-patient variability in the relative treatment effect. It may even be that treatment A is better in some patients while other patients are better off on treatment B. Parallel trials cannot assess such individual differences in effect between patients with similar

prognostic variables. One extreme version of the cross-over design is worth mentioning in this context. A trial may be conducted in one single patient, exposed to treatments A and B in a series of treatment periods. This design appears to be quite natural. It resembles what a patient suffering from migraine attacks would do to find out which “treatments” are causing the attacks. She occasionally eats chocolate, drinks wine, goes out in heavy sunshine, and notices when the migraine problems occur. Gradually, she will understand her personal risk factors and avoid them. The design, called *N-of-1*, can be much more rigorous, involving blinding and randomization to two or more treatment groups, comparing two or more pharmaceuticals [52, Section 7.4].

*N-of-1* trials may be attractive from an ethical perspective when the aim of the experiment is to find the best treatment for the individual patient, and DA is used to find the optimal criterion for when the testing should stop and the patient receive the better treatment henceforth. However, the signal-to-noise ratio is often too poor to allow firm conclusions on the best treatment for an individual. A combination of data from *N-of-1* trials in several patients can be performed in order to estimate treatment differences more precisely and also estimate how a large percentage of patients would benefit from each treatment.

### 26.3.4 Optimal Amount of Information to Collect in Trial

Having discussed how to optimize the information gathering for each patient, we will now turn to the problem of choosing the optimal amount of information to purchase, that is, to determine the sample size. Clinical trials are commonly dimensioned by specifying a least clinically relevant effect, taking this as the alternative hypothesis, and then calculating the sample size that gives, say, 90% power of detecting any effect. This approach can be severely misleading, we think, as it ignores all relevant information except the least clinically relevant effect. We will sketch how the knowledge and uncertainty regarding effect and safety, regulatory demands, price, market size, and study costs could all help to find the best sample size. However, let us first analyze the notion of “the least clinically relevant effect.” If you have to choose between two drugs with known different effects but identical safety profile, price, dosing, tablet size, and the like, which would you choose? If your 1-year risk of dying is 1% with drug A and 1.1% with drug B, you would clearly choose A. Will the choice be different if the risk is instead 1.001% with drug B? Our conclusion is that all differences are “clinically relevant.” The question about which drug to prefer arises only if both drugs have relative advantages in different aspects. For the patient, there can be a trade-off between effect and safety, for the payor between effect (and safety) and cost.

For the drug developer that tries to maximize the expected net present value (ENPV), study sizing is a trade-off between generating enough information regarding effect and safety on one hand and the cost of the trial on the other. A related and even more fundamental problem is whether any investment at all should be made in developing the drug or if the project should be terminated immediately. It is very complex to optimize a complete clinical development program, as the optimal design of a later phase depends on the data from the earlier phase, and the optimal design of the earlier phase depends on how the results from this phase will be used for designing the later phase. The principle solution to the problem is backward



induction [54], where one starts by analyzing the last phase, given different possible outcomes of earlier phases. After finding how phase III should be designed as a function of the information from phase IIB, phase IIB can be optimized. This process can in principle be carried back to even earlier phases, although computational difficulties and specification of the complex problem may be obstacles.

Consider the problem of dimensioning a phase III trial, assuming that supportive evidence exist and that one successful phase III trial will be enough to get the drug approved. Generally, the utility, as a function of the sample size  $N$  of the trial, is the gain by running the trial minus the cost for doing so. The cost can often be put rather directly in monetary terms. However, the personnel resources may be limited and hard to expand rapidly. Consequently, one may have to consider that if the phase III trial is run, other projects may be delayed or stopped and the net consequences of this should be factored into the cost analysis. As the sample size affects the possible time to market introduction, the time delay may be viewed as a cost. Time effects are, however, related to the trial outcome and are easier to analyze as part of the gain function.

The gain can be complex to analyze, although some market value calculations are routinely done in current practice. The simplest model is to say that the drug is approved if and only if a statistically significant effect versus placebo is shown in phase III, and if it is approved the commercial value is a certain known amount. This model can easily be improved by modeling how the market value will decrease with a delay in study ending time as a result of an increased sample size. There are, however, other factors, such as the market value's dependence on the trial results, that are harder to model. It is often too simplistic, although common, to assume a fixed market value if the drug is approved without considering how good the trial data are. One way of modeling the sales of a drug is to start by analyzing the decision analyses that the patient is doing together with his or her physicians. What is required to choose the new drug in terms of estimated effect, standard error for the effect estimate, safety, price, and the like? Pharmaceutical companies often perform limited market research about such questions by asking physicians to compare different hypothetical drugs. Individual treatment decisions could then be aggregated to form an effect-sales model. It is likely that historical data on, for example, how fast a market is penetrated have to be added in order to find a suitable model.

The commercial and cost models together give a utility function. The design alternatives are all possible sample sizes, including the possibility of stopping the project and not run the trial. Regulators may have requirements for a minimal sample size in order to get enough safety information. With a model for the effect of the drug and for how large the uncertainty is, the trial results can be predicted, by numerical methods or simulations. It is often important to consider the uncertainty in the effect estimate and run Bayesian clinical trial simulations. Combining predicted trial results with the utility function, the expected utility can be calculated for each sample size, and the optimal sample size can be found. The expected utility is relatively flat around the optimum, indicating that the exact sample size is not crucial. The ENPV can be very sensitive to such changes in assumptions concerning costs and gains. However, given that it is optimal to run the trial, the sample size is typically relatively robust against such changes, meaning that the same sample size often gives an ENPV close to the optimal ENPV.



The time to market is essential. If there is an indication that the sample size should be increased, one might therefore consider including a larger number of centers. In analyzing this option, both direct costs and the risk for reduced quality, leading, for example, to a larger unexplained variability, should be considered. The optimal size of an experiment is discussed by Lindley [55] and a review on applications to clinical trials is provided by Pezeshk [56].

**Commercial Limitations** Medical research will likely look for smaller and smaller improvements in many therapeutic areas. Smaller effect differences than what most historic studies have set out to detect may still be of great importance. The vast majority of mortality trials have been dimensioned to detect risk reductions of 20% or more. However, a treatment with 10% additional reduction in, for example, cardiovascular mortality, would of course be of great value. The problem is that smaller effects call for larger trials. The sample size is, at least approximately, proportional to the square of the relative treatment effect. Thus, looking for a risk reduction of 10% instead of 20% will mean that four times more patients are needed in the trial. While the cost for a trial is approximately quadratic in the effect size, the benefit, in the number of spared lives, say, is linear. Thus, small enough relative effect sizes will not be found by clinical trials, although they may translate into a large number of deaths.

Assuming rational payors, the commercial gain of introducing a new pharmaceutical will always be lower than the societal gain, as the price will always be lower, at least for some payors, than the willingness to pay. There are, consequently, research projects that would give a good expected return-on-investment for society, although they would never be run by commercial interest. Noncommercial research therefore has a clear value.

### 26.3.5 Adapting to New Information

For both economic and ethical reasons, it is clearly desirable to adapt the design of a clinical trial in response to accumulating data. Interim data may show, for example, that there is a clear difference between the treatments. The trial can then be stopped, permitting the patients in the control arms of the study to stop receiving the inferior treatment. Alternatively, interim data may indicate that a proposed treatment has no clear benefit over the placebo. The trial can in such a case also be stopped for reasons of futility.

**Group-Sequential Designs** The most well-known example of adaptive designs is the classic group-sequential design [57, 58], which permits a number of interim analyses to be conducted. It allows for the trial to be stopped if interim results are convincing enough according to preset criteria. As a statistically significant result can be obtained both at interim analyses and at the preplanned full sample size, the critical limits for when to claim significance must be adjusted. The result is that a group-sequential trial has to have a somewhat larger maximal sample size than a fixed-size trial if the same power should be obtained. This is the price to pay for the possibility that the trial sometimes can be stopped much earlier than in a trial without interim analyses. With a reasonable group-sequential design, the expected sample size is decreased. Consequently, the cost related to the number of patients

is on average decreased. Among the drawbacks of a sequential design are the cost for the interim analysis per se. Given these costs, it is straightforward to simulate trials and calculate the expected total cost to find which design is best. One can investigate whether it is optimal to include interim analyses, when they should come and what the critical limits for significance should be. There are some additional drawbacks, such as problems with overrunning, that is, that additional trial data may be obtained after an interim analysis that suggested a trial termination. Such data may potentially change the conclusion in the trial. Still, it is possible to simulate this problem and factor it into the DA.

**Flexible Designs** From a DA perspective, the optimal design of a trial may change during a trial if external information is received that changes the utility function or the effect/safety model. Examples of such information is the approval of a competitor drug or the finding of unexpected safety problems in substances with similar chemical properties as the one being developed. It is therefore desirable to be able to change the design during the conduct of a trial. An article by Bauer and Köhne [59] in 1994 gave the impetus to research in flexible designs that allow several types of design modifications in response to interim data or external information. According to Müller and Schäfer [60], “design changes may include increase or reduction of the sample size and of the number and time points of interim analyses, and even changes of the type of statistical test, of the outcome variable, or of the null-hypothesis.” Their stand is that such changes do not have to be preplanned. However, if such flexible designs are getting acceptance and penetrate clinical trial practice, it will constitute a fundamental change compared to the hitherto dominating methodology. Preplanning has always been one of the cornerstones of controlled clinical trials and a basis for the classical statistical analysis of data. Other authors, including Bauer and Köhne [59] in their original work, have been more moderate than Müller and Schäfer [60] regarding when and how flexible designs could be used. The EU reflection study [61] on adaptive designs is quite restrictive to flexible design changes, especially in confirmatory studies.

Sample size reestimation (SSRE) has attracted the greatest interest among the flexible designs. The sample size may, for example, be increased in order to achieve more statistical power if the interim estimate of the treatment effect is lower than anticipated. The fundamentally new technical idea behind flexible designs is to tentatively consider different stages of a trial as different trials altogether. One  $p$  value is calculated for each stage in isolation. These  $p$  values are then combined into a  $p$  value for the complete trial, by weighting together the stagewise  $p$  values. The type I error can be guaranteed, for essentially all kinds of design modifications, as long as the weight for each stage is determined independently of the data from that stage.

The weighted inference is controversial. Jennison and Turnbull [58] have suggested that an adequately developed group sequential design is preferable to an adaptive SSRE that only depends on interim data. The flexible design has, however, the advantage of being able to be modified according to information external to the trial. Another problem with the methodology proposed by Bauer and Köhne [59] is that equally informative observations may receive unequal weight in the analysis, depending on which stage of the trial they are collected. This feature has been criticized for violating the sufficiency principle, leading not only to inefficiency [58] but also in extreme cases to erroneous conclusions [62].

Our tentative conclusions are that careful preplanning of trials is crucial, that there can occasionally still be a genuine need for flexibility, that the weighted test is questionable but that a modification of this test could be of value. The works on flexible designs mentioned above have been based on classical, frequentist statistics. Bayesian statistics can more naturally incorporate sample size modifications and other design changes. Such modifications can be useful in earlier phases but may not be accepted in the confirmatory phase.

## 26.4 DISCUSSION AND CONCLUSIONS

Since World War II, the number of clinical trials has increased rapidly. The importance of these trials can hardly be underestimated. The medical needs continue to be manifold. Despite the clinical needs and the steady increase in the number of clinical trials carried out, there are factors that could threaten their future. In this chapter, we have highlighted the problems concerning ethics and efficiency. Although RCTs represent the gold standard in clinical development, there are alternatives. Observational studies and historic controls can give valuable information, even if their interpretation is not as straightforward as blinded clinical trials. Preclinical experiments and modeling can also substitute some trials.

Clinical trials should be performed in an ethical manner. Failure to do so is not only unethical, per definition, but it also jeopardizes public trust in clinical development. A patient is unlikely to give informed consent unless he or she is convinced that the physician will put the patient's interests first. The concept of informed consent is a fundamental requirement that we have highlighted. In the rare situations where informed consent cannot be given by the patients themselves (studies on unconscious patients or children), the informed consent should instead be given by a guardian *on behalf of and in the interest of* the patient as soon as possible. Likewise, the physician should see principally to the best of his or her patient first and only secondarily to the interests of science. Thus, when the physician also acts as an investigator, he or she is still personally responsible for ensuring that the patient's best interests are safeguarded during trial participation. In addition to genuine consent and patient benefit optimization, an ethical requirement of nonexploitation is needed. Additional requirements may be indicated to cover issues about, for example, payment and trials that are beneficial on average but include nonoptimal treatment arms. While patient enrollment is the responsibility of the physician, the ethical standard of clinical trial is the common responsibility of the investigator, the ethical committee, the sponsor, and the personnel responsible for the planning and the execution of the trial. Personal responsibilities should be stressed to avoid the risk that all actors put the responsibility on someone else, leading to unethical and risky trials.

Ethical requirements constrain trial design, in particular the treatments that can be tested. Depending on the situation and background information, these constraints will sometimes complicate or even obstruct clinical development. Within these limits there are, however, considerable possibilities of optimizing designs that can also be efficient. Several features of a trial design should be considered in the process of trial optimization, such as sample size, patient population, and treatment duration. There are also great possibilities to apply alternatives to the standard

parallel-group nonadaptive design and there is a continuing methodological development. How to choose between all the available options is perhaps the major challenge. Consequently, we anticipate that the most significant changes in trial design will take place in decision process rather than design methodology per se. Model-based drug development is already receiving considerable attention. Still, modeling and DA are clearly underestimated.

## REFERENCES

1. DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003), The price of innovation: New estimates of drug development costs, *J. Health Econom.*, 22, 151–185.
2. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rates? *Nature Rev. Drug Dis.*, 3, 711–716.
3. Santoro, M. A., and Gorrie, T. M., eds. (2005), *Ethics and the Pharmaceutical Industry*, Cambridge University Press. Cambridge, UK, pp. 3–4.
4. Emanuel, E. J., Wendler, D., and Grady, C. (2000), What makes clinical research ethical? *JAMA*, 283, 2701–2711.
5. Senn, S. (2002), Ethical considerations concerning treatment allocation in drug development trials, *Stat. Methods Med. Res.*, 11, 403–411.
6. Purtilo, R. (1995), Professional-patient relationship: Ethical issues, in *Encyclopedia of Bioethics*, vol. 4, Macmillan, New York, pp. 2094–2101.
7. Mason Spicer, C. (1995), Codes, oaths, and directives related to bioethics. Section II. Ethical directives for the practice of medicine, in *Encyclopedia of Bioethics*, vol. 4, Macmillan, New York, pp. 2630–2649.
8. Thomasma, D. (1990), Establishing the moral basis of medicine: Edmund D. Pellegrino's philosophy of medicine, *J. Phil. Med.*, 15, 245–267.
9. Laín Entralgo, P. (1969), *El Médico y el Enfermo*, Ediciones Guadarrama, Madrid.
10. Hippocrates (1984), *Parangeliai (Recommendations, Precepts) 2 Hippocrates 1*, Loeb Classical Library, Harvard University Press, Cambridge, MA.
11. Lellouch, J., and Schwartz, D. (1971), L'essai thérapeutique: Éthique individuelle ou éthique collective? *Rev. Inst. Int. Stat.*, 39, 127–136.
12. Clayton, D. G. (1982), Ethically optimised designs, *Br. J. Clin. Pharma.*, 13, 469–480.
13. Gifford, F. (1986), The conflict between randomized clinical trials and the therapeutic obligation, *J. Med. Phil.*, 11, 347–366.
14. Engelhardt, H. T. (1996), *The Foundations of Bioethics*, 2nd ed., Oxford University Press, New York.
15. Nuffield Council on Bioethics (1995), *Human Tissue: Ethical and Legal Issues*, Nuffield Council, London.
16. Schaffner, K. F. (1986), Ethical problems in clinical trials, *J. Med. Phil.*, 11, 297–315.
17. Kopelman, L. (1986), Consent and randomized clinical trials. Are there moral or design problems? *J. Med. Phil.*, 11, 317–346.
18. Lilford, R. J., Brauholtz, D., Edwards, S., et al. (2001), Monitoring clinical trials—interim data should be publicly available, *BMJ*, 323, 441–442.
19. Fried, C. (1974), *Medical Experimentation: Personal Integrity and Social Policy*, American Elsevier, New York.
20. Freedman, B. (1987), Equipoise and the ethics of clinical research, *N. Engl. J. Med.*, 317, 141–145.

21. Lilford, R. J. (2003), Ethics of clinical trials from a bayesian and decision analytic perspective: Whose equipoise is it anyway? *BMJ*, 326, 980–981.
22. Emanuel, E. J., and Miller, F. G. (2002), The ethics of placebo-controlled trials—a middle ground, *Am. J. Ophthalmol.*, 133, 174–175.
23. World Medical Association (2008), Declaration of Helsinki: Ethical principles for medical research involving human subjects, available at: <http://www.wma.net/e/policy/b3.htm>; accessed February 2009.
24. Carlson, R. V., Boyd, K. M., and Webb, D. J. (2004), The revision of the Declaration of Helsinki: Past, present and future, *Br. J. Clin. Pharmacol.*, 57, 695–713.
25. Nuffield Council on Bioethics (2002), *The Ethics of Research Related to Healthcare in Developing Countries*, Nuffield Council, London.
26. D'Agostino, R. B., Massaro, J. M., and Sullivan, L. M. (2003), Non-inferiority trials: Design concepts and issues of the encounters of academic—consultants in statistics, *Stat. Med.*, 22, 169–186.
27. Retsas, R. (2004), Treatment at random: The ultimate science or the betrayal of Hippocrates? *J. Clin. Oncol.*, 22, 5005–5008.
28. Marquis, D. (1986), An argument that all prerandomised clinical trials are unethical, *J. Med. Phil.*, 11, 367–383.
29. Bureau of Labor Statistics (2005), News: National census of fatal occupational injuries in 2004. Bureau of Labor Statistics, U.S. Department of Labor, Washington D.C., August 25; available at: <http://www.bls.gov/news.release/pdf/cfoi.pdf>; accessed May 2006.
30. Council of Europe (2005), Additional protocol to the convention on human rights and biomedicine, concerning biomedical research. Council of Europe Treaty Series, No 195; available at: <http://conventions.coe.int/>; accessed May 2006.
31. European Group on Ethics (2001), Ethical Aspects of Clinical Research in Developing Countries; available at: [http://ec.europa.eu/european\\_group\\_ethics/docs/avis17\\_en.pdf](http://ec.europa.eu/european_group_ethics/docs/avis17_en.pdf); accessed May 2006.
32. Orentlicher, D. (2005), Making research a requirement of treatment. Why we should sometimes let doctors pressure patients to participate in research, *Hastings Center Rep.*, 35/5, 20–28.
33. FDA (2004), Challenge and opportunity on the critical path to new medical products; available at: <http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html>; accessed May 2006.
34. FDA (2006), Critical Path Opportunities List; available at: [http://www.fda.gov/oc/initiatives/criticalpath/reports/opp\\_list.pdf](http://www.fda.gov/oc/initiatives/criticalpath/reports/opp_list.pdf); accessed May 2006.
35. Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*, Wiley, New York.
36. Gabrielsson, J., and Weiner, D. (1997), *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*, 3rd ed., Swedish Pharmaceutical Press, Stockholm.
37. Sheiner, L. B., and Steimer, J.-L. (2000), Pharmacokinetic/pharmacodynamic modeling in drug development, *Ann. Rev. Pharma. Toxicol.*, 40, 67–95.
38. Hoppensteadt, F. C., and Peskin, C. S. (2000), *Modeling and Simulation in Medicine and Health Sciences*, 2nd ed., Springer, New York.
39. Burman, C.-F., Hamrén, B., and Olsson, P. (2005), Modelling and simulation to improve decision-making in clinical development, *Pharma. Stat.*, 4, 47–58.
40. Parmigiani, G. (2002), *Modeling in Medical Decision Making: A Bayesian Approach*, Wiley, Chichester.
41. Hunink, M. G. M., Glasziou, P. P., Siegel, J. E., et al. (2001), *Decision Making in Health and Medicine: Integrating Evidence and Values*, Cambridge University Press, Cambridge, UK.

42. Goodwin, P., and Wright, G. (1998), *Decision Analysis for Management Judgment*, 2nd ed., Wiley, Chichester.
43. Burman, C.-F., Grieve, A., and Senn, S. (2006), Decision analysis in drug development, in Dmitrienko, A., Chuang-Stein, C., D'Agostino, R., Eds., *Pharmaceutical Statistics*, SAS Institute, Cary, NC.
44. Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluations*, Wiley, Chichester.
45. Berry, D. A. (1993), A case for Bayesianism in clinical trials (with discussion), *Stat. Med.*, 12, 1377–1404.
46. Raiffa, H. (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, McGraw-Hill, New York.
47. Howard, R. A., and Matheson, J. E., eds. (1983), *Readings on the Principles and Applications of Decision Analysis. Volume I–II*, Strategic Decision Group, Menlo Park, California.
48. Ross, S. M. (1997), *Simulation*, 2nd ed., Academic, San Diego.
49. Robert, C. P., and Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer, New York.
50. Thompson, J. R. (2000), *Simulation: A Modeler's Approach*, Wiley, New York.
51. Senn, S. (1997), *Statistical Issues in Drug Development*, Wiley, Chichester.
52. Senn, S. (1993), *Cross-Over Trials in Clinical Research*, Wiley, Chichester.
53. Jones, B., and Kenward, M. G. (2003), *Design and Analysis of Cross-Over Trials*, 2nd ed., Chapman & Hall, Boca Raton, FL.
54. Bather, J. (2000), *Decision Theory: An Introduction to Dynamic Programming and Sequential Decisions*, Wiley, Chichester.
55. Lindley, D. V. (1997), The choice of sample size, *Statistician*, 46, 129–138.
56. Pezeshk, H. (2003), Bayesian techniques for sample size determination in clinical trials: A short review, *Stat. Methods Med. Res.*, 12, 489–504.
57. Whitehead, J. (1997), *The Design and Analysis of Sequential Clinical Trials*, 2nd ed., Wiley, Chichester.
58. Jennison, C., and Turnbull, B. W. (2000) *Group Sequential Methods with Applications to Clinical Trials*, Chapman and Hall, London.
59. Bauer, P., and Köhne, K. (1994), Evaluation of experiments with adaptive interim analyses, *Biometrics*, 50, 1029–1041.
60. Müller, H.-H., and Schäfer, H. (2004), A general statistical principle for changing a design any time during the course of a trial, *Stat. Med.*, 23, 2497–2508.
61. EMEA/CHMP (2006), Reflection paper on methodological issues in confirmatory clinical trials with flexible design and analysis plan; available at: <http://www.emea.eu.int/pdfs/human/ewp/245902en.pdf>; accessed May 2006.
62. Burman, C.-F., and Sonesson, C. (2006), Are flexible designs sound? *Biometrics*, 62, 662–669.

## *Proof-of-Principle/ Proof-of-Concept Trials in Drug Development*

**Ayman Al-Shurbaji**

*Experimental Medicine, International PharmaScience Center, Ferring Pharmaceuticals A/S,  
Copenhagen S, Denmark*

### **Contents**

|        |                                                                                   |      |
|--------|-----------------------------------------------------------------------------------|------|
| 27.1   | Background                                                                        | 1202 |
| 27.1.1 | Why Is Proof-of-Principle/Proof-of-Concept (PoP/PoC) Crucial in Drug Development? | 1202 |
| 27.1.2 | Definitions                                                                       | 1202 |
| 27.1.3 | When Should PoP/PoC Be Established?                                               | 1203 |
| 27.2   | Constantly Improving Toolbox for PoP/PoC                                          | 1204 |
| 27.3   | Biomarkers in PoP/PoC Trials                                                      | 1204 |
| 27.4   | Pharmacogenomics, Proteomics, and Metabonomics                                    | 1205 |
| 27.5   | Imaging                                                                           | 1207 |
| 27.6   | Pharmacological and Nonpharmacological Models for PoP/PoC Trials                  | 1207 |
| 27.6.1 | Pharmacological Models in Healthy Volunteers                                      | 1207 |
| 27.6.2 | Nonpharmacological Models                                                         | 1208 |
| 27.6.3 | Provocation Models in Patients                                                    | 1208 |
| 27.7   | Different Types of PoP/PoC Studies                                                | 1209 |
| 27.7.1 | Exploratory Investigational New Drug PoC Trials                                   | 1209 |
| 27.7.2 | Clinical Trials Using Biomarkers or Models in Healthy Subjects                    | 1210 |
| 27.7.3 | Clinical Trials Using Biomarkers in Patients                                      | 1211 |
| 27.7.4 | Trials Using Clinical Endpoints for PoC                                           | 1212 |



|                                                                                                   |      |
|---------------------------------------------------------------------------------------------------|------|
| 27.8 Knowledge of Exposure–Response Relationship of Drug Is Essential for Successful PoP/PoC      | 1212 |
| 27.9 PoP/PoC Outcomes Can Be “False Dawns”                                                        | 1212 |
| 27.9.1 Summary of Selected Human Models Quoted in This Chapter That Can Be Used in PoP/PoC Trials | 1213 |
| References                                                                                        | 1214 |

## 27.1 BACKGROUND

### 27.1.1 Why Is Proof-of-Principle/Proof-of-Concept (PoP/PoC) Crucial in Drug Development?

Drug development is complex, costly, and time consuming. After decades of fruitful endeavors leading to novel therapies that radically changed the way we treat human disease (e.g.,  $\beta$ -blockers, proton pump inhibitors, SSRIs, HMG-CoA reductase inhibitors, and many others), the pharmaceutical industry began to face increasing and significant challenges. Everyone who is involved in drug development is well aware of issues such as the “innovation gap,” “productivity decline,” and “late-stage attrition.” Indeed, these issues have been on the agenda for some years now and have triggered waves of new paradigms in drug development. Essentially, the pharmaceutical industry has been investing more and more resources in developing new drugs, but this has not been translated into more successful medicines, that is, medicines that are approved by the regulators and well-received by the market. In other words, the output from the R&D machinery has not kept pace with spending. This productivity decline is illustrated in Figure 1 showing escalating drug development costs with no parallel increase in the number of regulatory approvals. The average cost of bringing a new compound to the market is in the neighborhood of \$1.3 billion [1] and the overall clinical success rate is 11% [2]. A nearly 50% attrition rate in phase III is still plaguing pharmaceutical companies [2]; and there is no clear evidence that it is on the decline. Late-stage failures are severe blows that hurt large pharmaceutical companies and might ruin smaller ones, and it is in this context that the importance of early PoP/PoC becomes apparent. Drug developers need to ascertain that only compounds with a high chance of success are selected and progressed through the development process. Although most pharmaceutical companies are trying to adopt “kill early” approaches, this is often hampered by the lack of reliable “killer experiments” and the fear of erroneously discarding valuable assets. For smaller companies such as many in the biotechnology sector, which often lack the resources to take compounds through full clinical development, a positive PoP/PoC significantly improves the chance of finding a partner or successfully licensing out the compound. Regardless of company size, a robust early PoP/PoC has become a key component in contemporary drug development.

### 27.1.2 Definitions

A human PoP/PoC trial is simplistically a trial that provides indications that the compound might work in the intended therapeutic area. Some companies distinguish between a PoP trial in which the pharmacological principle has been

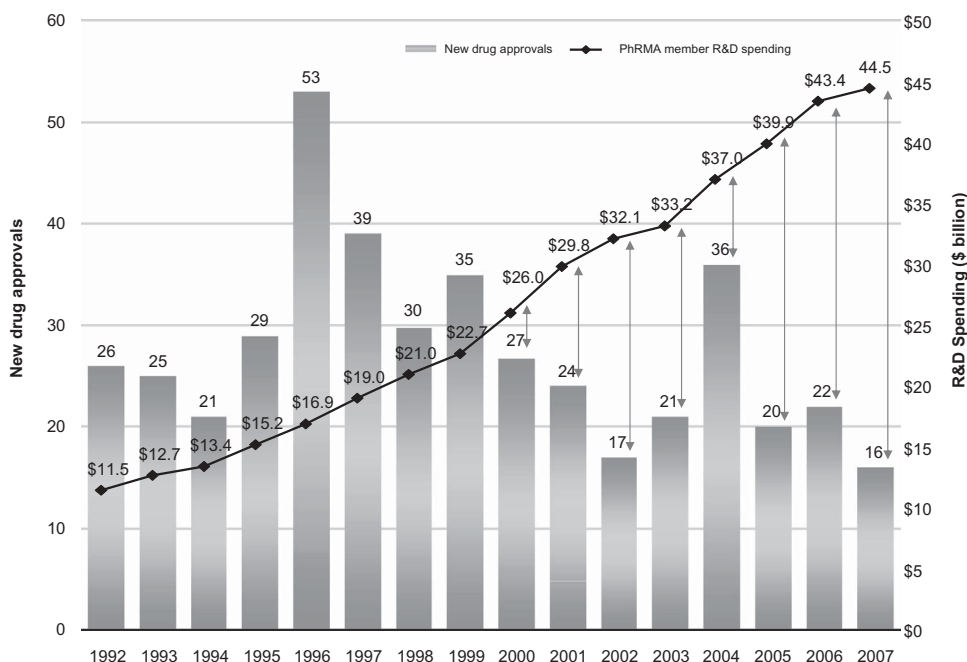


**FIGURE 1** Escalating R&D spending with no increase in output (FDA drug approvals): 2004–2007 approvals include new biologics. (Source: FDA and Pharmaceutical Research and Manufacturers of America, *Pharmaceutical Industry Profile 2008*, PhRMA, Washington, D.C. March 2008).

demonstrated in humans (healthy subjects or patients), usually with the use of a biomarker and a clinical PoC trial where early signs of efficacy are detected, based on solid clinical endpoints or validated surrogates. However, PoP and PoC are often used interchangeably. A broad definition has been proposed by fifth EUFEPS Conference: “A human trial that provides scientifically sound evidence supporting the postulated effects of a new therapeutic drug product, where effects may be relevant pharmacological action or a change in disease biomarkers, established surrogate endpoints, or clinical outcomes, and be beneficial and/or toxic in nature” [3].

### 27.1.3 When Should PoP/PoC Be Established?

As indicated above, spiralling development costs and high attrition rates are forcing drug developers to search carefully for signs of efficacy and potential safety concerns as early in the development process as possible. It is common practice today to conduct a PoP/PoC trial as soon as phase I trials have provided evidence that the compound appears to be reasonably safe in healthy subjects and possesses pharmacokinetic characteristics that are in line with projected clinical use. As shown in Figure 2, a PoP/PoC trial is conducted as a part of phase II.

The PoP/PoC is located at a most crucial “go–no go” decision point in most pharmaceutical R&D organizations. Once PoP/PoC has been established, drug developers (and investors) feel more confident to invest in the subsequent, more costly clinical trials.

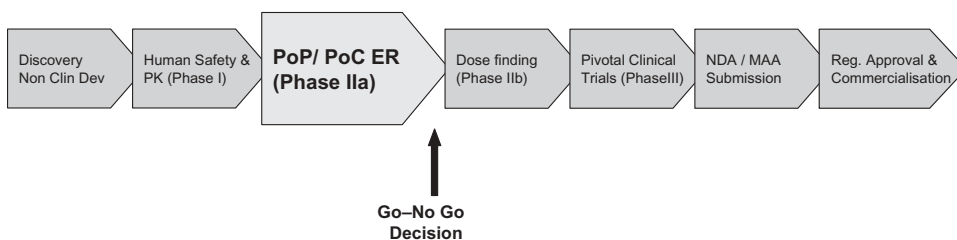


**FIGURE 2** Location of PoP/PoC in the drug development process. ER, Exposure–response; NDA, new drug application; MAA, marketing authorization application.

Although PoP/PoC is often linked to a single trial that provides a dichotomous “yes or no” answer, it is more prudent to think of PoP/PoC as a package of key data that have emerged during the early exploratory phase of development. Thus, a positron emission tomography (PET) study in healthy volunteers may provide evidence that a certain compound crosses the brain–blood barrier and binds to a specific receptor in the brain. A human model (Table 3) may show that the compound is effective in a disease-like setting. Pharmacokinetic/pharmacodynamic (PK/PD) modeling of preclinical and human data is used to characterize the exposure–response (ER) relationship and ensure that safe doses are potentially efficacious. Finally, the integrated data package from all of the above provides a reasonable level of confidence to pursue development, that is, a PoP/PoC-based “go decision.”

## 27.2 CONSTANTLY IMPROVING TOOLBOX FOR POP/POC

Despite the serious issues facing the pharmaceutical industry, the scientific advances in biosciences (chemical and imaging biomarkers, pharmacogenetics and pharmacogenomics, proteomics, toxicogenomics, metabonomics) are providing drug developers with useful tools to explore the potential clinical utility of new targets and novel therapeutic agents. These areas of research have become independent disciplines and there is a wealth of literature on each of them. Since biomarkers constitute a key component in PoP/PoC paradigms, they are discussed in some detail next.

## 27.3 BIOMARKERS IN POP/POC TRIALS

The National Institutes of Health (NIH) Biomarkers Definitions Working Group has defined a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [4]. Although biomarkers have been used in drug development for decades, the new potential of using biomarkers for regulatory purposes (label claims) [5] has intensified the efforts to develop biomarkers that are accepted for ordinary and accelerated approvals. Regulatory initiatives, most notably the Food and Drug Administration’s (FDA’s) Critical Path Initiative have opened new avenues for collaboration between academia,

**TABLE 1** Examples of Biomarkers Accepted by FDA as Surrogate Endpoints

| Condition              | Surrogate                                    |
|------------------------|----------------------------------------------|
| Ischemic heart disease | Plasma cholesterol                           |
| Stroke                 | Blood pressure                               |
| Osteoporosis           | Bone density                                 |
| Diabetes mellitus      | HbA1c                                        |
| Prostate cancer        | Serum testosterone                           |
| HIV/AIDS               | HIV-RNA load and CD4 <sup>+</sup> cell count |

pharmaceutical companies, and regulators aimed at facilitating innovative approaches in the development of novel medicines [6]. A key area in the Critical Path Initiative is biomarkers with a special consortium focusing on the qualification of biomarkers for regulatory decision making. Qualification of biomarkers refers to the process of linking a biomarker with biological processes and clinical endpoints [7, 8]. The highest level of qualification of a biomarker is surrogacy, that is, when the biomarker can substitute for a clinical endpoint [7, 9]. Table 1 shows examples of biomarkers that have been accepted by the FDA as surrogate endpoints.

It should be underscored that while surrogate endpoints and other “valid” biomarkers can and should be used in PoP/PoC trials when available, early clinical development is an exploratory phase, and less well-qualified biomarkers can provide valuable information for internal company decision making. For example, “distal biomarkers”—relative to the clinical endpoint—can be very useful in confirming the pharmacological activity of a given compound in humans [10]. This is especially valuable in two situations: the selection of follow-up or backup compound when the “pathfinder” has demonstrated clinical utility; and the validation of a novel mechanism of action with an uncertain role in the pathogenesis of the disease. Lack of clinical efficacy despite adequate pharmacological effect would disqualify the mechanism of action and stop the development of similar compounds. Also, biomarkers can be very useful in establishing the ER relationship preclinically and translating that relationship into the clinical phases of development (the concept of translational science), facilitating dose selection in PoC studies and beyond [11]. Good examples of the use of biomarkers in early clinical development have been described by Kuhlmann [12]. These included the use of *ex vivo* inhibition of cholesteryl ester transfer protein (CETP) in healthy subjects to establish the pharmacological activity of a CETP inhibitor, the use of leukotriene D<sub>4</sub>-induced bronchoconstriction to establish PoP and dose–response in healthy subjects for a leukotriene receptor antagonist for asthma, and the use of pupillography as a marker for effects of 5-hydroxytryptamine 1A (5-HT<sub>1A</sub>) compounds.

## 27.4 PHARMACOGENOMICS, PROTEOMICS, AND METABONOMICS

The concept of identifying genomic [single nucleotide polymorphisms (SNP), haplotypes, ribonucleic acid (RNA) expression signatures], proteomic [protein profiles in plasma, cerebrospinal fluid (CSF), synovial fluid, etc.], and metabonomic (metabolite patterns in body fluids) markers that identify patients who are likely to respond to treatment and/or to develop serious adverse drug reactions is very exciting albeit

**TABLE 2   Examples of Valid Genetic/Genomic Markers Predicting Clinical Response or Specific ADRs**

| Marker                                       | Drug          | Clinical Relevance                                                                                               |
|----------------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------|
| Her2/neu overexpression                      | Trastuzumab   | Only patients with tumors overexpressing Her2 are eligible for treatment.                                        |
| Philadelphia chromosome (Ph1)                | Busulfan      | Poor response in patients with AML who lack Ph1 chromosome.                                                      |
| PML/PAR alpha gene expression                | Tretinoin     | Clinical response to tretinoin in patients with APL have not been observed in the absence of the genetic marker. |
| Thiopurine methyltransferase (TPMT) variants | Azathioprine  | Increased risk of myelotoxicity in patients with low or absent enzyme activity.                                  |
| HLA-B*1502 allele                            | Carbamazepine | Risk for serious dermatologic reactions in patients positive for the allele.                                     |
| G6PD deficiency                              | Rasburicase   | Drug is contraindicated in patients with G6PD deficiency due to the risk of severe hemolysis.                    |

AML, acute myelogenous leukemia; APL, acute promyelocytic leukemia.

extremely challenging. There are few examples of valid genomic markers that predict clinical outcomes [responder vs. nonresponders; adverse drug reaction (ADR) prone genotypes] and that have been accepted by the FDA as part of the label (Table 2). A comprehensive list of valid genomic biomarkers in the context of FDA-approved drugs is available [13].

However, in the context of early clinical development of novel compounds acting on nonvalidated targets, the identification of such markers remains a major hurdle to overcome. Roses and colleagues described some examples of the use of early pipeline genetics to predict efficacy and side effects of new compounds [14]. These included the identification of gene variant candidates related to efficacy in a phase II obesity drug trial, and the early discovery of the genetic basis (CYP 2C9 polymorphisms) of some of the side effects of the kinase inhibitor lapatinib (Tykerb). The authors rightly underscored that such early findings can only be used to generate hypotheses and that the validation of these markers should await larger clinical trials [14]. The limitations of pharmacogenomic predictor discovery in early clinical trials are discussed in Pusztai [15].

Needless to say, this field is still in its infancy, and the ultimate impact of the “omics revolution” on expediting drug development and reducing attrition rates remains to be seen.

In many pharmaceutical companies, default banking of samples from clinical trials for future genetic/genomic analyses is becoming common practice, although this is sometimes challenged ethically and legally. Since genetically determined variability in drug metabolism, most notably involving CYP 450 enzyme system, impacts significantly on the pharmacokinetics and consequently pharmacodynamics of many drugs, it is prudent to obtain relevant genotypes in healthy volunteers and patients in early clinical trials of compounds known to be mainly metabolized by CYP 450 isoenzymes, primarily CYP 2D6, CYP2C9, and CYP2C19. This would be useful if unexpected PK/PD profiles are observed and potentially reduce the risk of selecting inappropriate doses in PoP/PoC trials resulting in erroneous go–no go decisions.

## 27.5 IMAGING

Advances in medical imaging present a great opportunity in drug development. A number of different imaging technologies are available. These include computed tomography (CT), magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS), positron emission tomography (PET), and single photon emission computed tomography (SPECT). If adequately qualified, imaging biomarkers can be very helpful in the early phases of clinical development. Early evidence of biological/pharmacological activity of a new compound as well as useful pharmacokinetic profiles for future dose selection can be obtained. For example, using fluorodeoxyglucose PET (FDG-PET), it could be shown that glucose metabolism was reduced in gastrointestinal stromal tumors in patients treated with imatinib mesylate (Gleevec), indicating an early response that preceded tumor size reduction (on CT) by weeks [16]. The development of a new antidepressant was terminated based on the finding that the compound lacked specific receptor binding, as shown by cerebral PK imaging [17]. Other good examples of the use of imaging in early clinical trials in different therapeutics are described by Pien et al. [18]. The potential advantages of imaging techniques in the different phases of drug development have recently been reviewed by Strack [19].

## 27.6 PHARMACOLOGICAL AND NONPHARMACOLOGICAL MODELS FOR POP/POC TRIALS

### 27.6.1 Pharmacological Models in Healthy Volunteers

In these models, a pharmacological agent, for example, scopolamine is given to healthy subjects to induce symptoms typical of the targeted disease, in this case dementia (cognitive impairment with deterioration of episodic memory and attention) [20, 21]. Compounds with cognition-enhancing characteristics would reduce these symptoms [22], providing early evidence of potential clinical efficacy. Pharmacological models are often used in psychopharmacological research, and, if adequately adapted, some of them can potentially be useful in early clinical development of psychotropic drugs (see [23] for a review). Asthma is another therapeutic area where pharmacological provocation tests in healthy volunteers; using, for example, methacholine [24, 25] or leukotriene D<sub>4</sub> [26] are utilized to demonstrate a pharmacological activity and capture early signs of potential clinical efficacy. It should be pointed out that unless the model used involves a mechanism that is relevant for the pathophysiology of the disease, a positive effect of a new compound in the model merely demonstrates a pharmacological antagonistic activity and does not provide early evidence of efficacy in the targeted disease. For leukotriene receptor antagonists, a class of drugs that target a mechanism known to be involved in the pathophysiology of asthma, a reduction of bronchoconstriction could be demonstrated in the leukotriene challenge as well as in other provocation models using, for example, allergens, methacholine, and ultrasound-nebulized distilled water [27, 28].

If models are to be reliable, a number of criteria should be fulfilled. These include face validity, test–retest consistency, responsiveness to reference drugs [29], and a reasonably high response rate in healthy subjects [23]. Since the clinical

**TABLE 3 Pharmacological and Nonpharmacological Models Used in PoP/PoC Trials**

| Model                                                                                  | Therapeutic Area                         |
|----------------------------------------------------------------------------------------|------------------------------------------|
| Carbon dioxide inhalation                                                              | General anxiety disorder, panic disorder |
| Conditioned aversive anxiety                                                           | Anxiety                                  |
| Scopolamine test                                                                       | Dementia/cognitive disorders             |
| Pain models (mechanical, chemical, thermal or electrical stimulation)                  | Pain                                     |
| Bronchial challenges (allergens, histamine, methacholine, leukotriene, cold air, UNDW) | Asthma                                   |
| Pentagastrin-induced gastric acid secretion                                            | Peptic ulcer, GERD                       |
| Human endotoxemia model                                                                | Sepsis                                   |
| Phenylephrine challenge                                                                | Benign prostatic hyperplasia             |

UNDW, Ultrasound-nebulized distilled water, GERD, gastroesophageal reflux disease.

manifestations of many diseases are multifaceted and not possible to reproduce in a model involving the administration of a single pharmacological agent, it has been suggested that the model can be useful if it provokes the cardinal symptoms of the disease [23]. Table 3 shows examples of pharmacological models that can be used in PoP/PoC trials.

### 27.6.2 Nonpharmacological Models

These use the same general principles as pharmacological models. However, they require rigorous standardization for reliability [23]. Examples of nonpharmacological models used in PoP/PoC paradigms include models of anxiety and panic such as simulated public speaking [30] and conditioned aversive anxiety [30], and nonpharmacological provocation tests of asthma and human experimental pain models (Table 3). The latter have evolved to provide a useful translational tool linking animal findings with data obtained in healthy subjects and patients. Moreover, by adopting a multimodal (mechanical, chemical, thermal, electrical stimuli), multitissue (skin, muscles, viscera) approach, it is now possible to explore different pain pathways and mechanisms and assess responses induced by novel putative analgesics in PoP/PoC trials. This approach is described in a recent review by Arendt-Nielsen and colleagues [31].

### 27.6.3 Provocation Models in Patients

PoP/PoC can also be explored using provocation models in patients. Both pharmacological and nonpharmacological models are potentially useful to progress a new compound through early clinical development. Asthma and allergic diseases are therapeutic areas where provocation models in patients are most frequently used, and the practical setup of the trials does not deviate from that in healthy volunteers. However, since the trials are conducted in the target patient population [27, 28, 32, 33], a positive outcome is more indicative of a successful PoC compared with a positive result from a provocation experiment in healthy subjects. Provocation tests for PoP/PoC in patients have also been used in other therapeutic areas, such as epilepsy. Among the models tested, the technically challenging intermittent photic stimulation in patients with photosensitive epilepsy has been found reliable in



predicting the antiseizure activity of a number of efficacious drugs that were eventually approved [34]. PoP was recently obtained for two novel antiepileptic drugs, brivaracetam, an SV2A ligand, and carisbamate, a new carbamate using this model [35, 36]. Both drugs suppressed generalized photoparoxysmal electroencephalogram (EEG) responses in treated patients. PoP studies in epilepsy using this and other models are discussed in Schmidt [37].

## 27.7 DIFFERENT TYPES OF POP/POC STUDIES

Considering the broad definition of PoP/PoC (see above), PoP/PoC trials are quite variable in terms of objectives, overall design, trial population, endpoints, as well as regulatory requirements.

### 27.7.1 Exploratory Investigational New Drug PoC Trials

These are early clinical studies conducted under the auspices of FDA's Exploratory Investigational New Drug Guidance [38], which covers microdose studies of pharmacokinetics or imaging as well as phase 0 clinical trials to study pharmacologically relevant doses. A microdose is defined as 1/100th of the dose calculated to yield a pharmacologic effect with a maximum dose of 100  $\mu$ g (or 30 nmol for proteins). Imaging studies using microdoses can be very useful in establishing PoP, for example, by demonstrating significant binding of the investigational compound to the target receptor. Since these studies imply single exposures to minute amounts of the test compound, they can be conducted without the standard safety pharmacology and genotoxicity studies.

Phase 0 trials that are designed to study the pharmacologic effects of the test compounds provide a new avenue to early PoC. The FDA accepts the use of doses at which a PK/PD response or target modulation is observed in such trials. However, prerequisites are the lack of an adverse clinical response and that the dose is less than one fourth of the 2-week rodent no adverse effect level (NOAEL), or that the total exposure to drug expressed as the area under the concentration curve (AUC) does not exceed one half of the AUC at NOAEL in the most sensitive species. Due to the lower doses and consequently lower toxicity risks implied, the preclinical pharmacology and toxicology package is less extensive than that required for traditional phase I investigational new drugs (INDs), although safety pharmacology and genotoxicity studies are required. Moreover, if a "go decision" is made, based on the outcome of the study, that is, target modulation is proven, a traditional IND will have to be submitted in order to progress the compound further in clinical development.

One key therapeutic area in which phase 0 exploratory studies are performed is oncology, where there is a great need for new safe and efficacious agents and where early evidence of target modulation (molecular PoC) would potentially expedite the clinical evaluation and facilitate the development of the most promising agents. In phase 0 cancer trials, low doses are administered to a limited number of patients (10–15 patients) [39].

The most critical factor for the success of a phase 0 PoC study in oncology—and other therapeutic areas—is the availability of a sensitive, robust, and qualified/validated pharmacodynamic assay. Kinders and colleagues proposed the following

set of performance criteria for PD assay validation: accuracy, dynamic range, precision, reproducibility and sensitivity [39]. The same group reported the results of an elegant phase 0 trial [40] in which 14 patients with advanced malignancies were treated with low doses of ABT-888, a poly(ADP-ribose) polymerase inhibitor. Using a PD assay that measured the amount of the enzyme polymer in tumor tissue and blood cells as a surrogate for enzyme inhibition, the authors were able to establish the PK/PD relationship and select a dose that had a potential for efficacy with minimized risk of toxicity.

Although the exploratory IND initiative allows for a more flexible mode of developing drugs in oncology and other therapeutic areas, it remains to be seen whether phase 0 PoC studies will result in a more speedy development of safe and efficacious novel drugs.

Apart from the technical challenges that phase 0 PoC trials pose (e.g., PD assay validation), the use of patients in a trial where there is no therapeutic intent (due to subtherapeutic dosing) introduces new ethical standards in clinical research where “benefit to others” is accepted as the sole benefit. This is discussed in a thoughtful editorial [41], where phase 0 trials are described as “ethically challenged and ethically challenging.”

### 27.7.2 Clinical Trials Using Biomarkers or Models in Healthy Subjects

These trials exploit validated or reasonably qualified “fit-for-purpose” biomarkers or models to elicit a clinically relevant pharmacological response in healthy subjects. For example, an early PoP was obtained for degarelix, a GnRH receptor antagonist currently in clinical development for prostate cancer by Ferring Pharmaceuticals, by demonstrating a rapid, profound and sustained reduction in serum levels of testosterone and LH in healthy volunteers (Figs. 3 and 4). The findings were later confirmed in large trials in the target patient population (unpublished data). The availability of a well-validated biomarker/surrogate (testosterone deprivation in prostate cancer) was crucial for the successful PoP/PoC achieved.

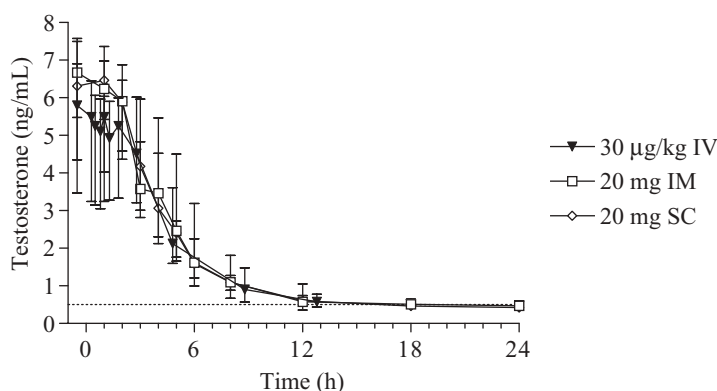


**FIGURE 3** Serum testosterone levels in healthy volunteers after the administration of degarelix intravenously, intramuscularly, and subcutaneously. The horizontal dotted line indicates the level of castration (serum testosterone of 0.5 ng/mL) required for efficacy in prostate cancer trials. Median values and quartiles are shown.

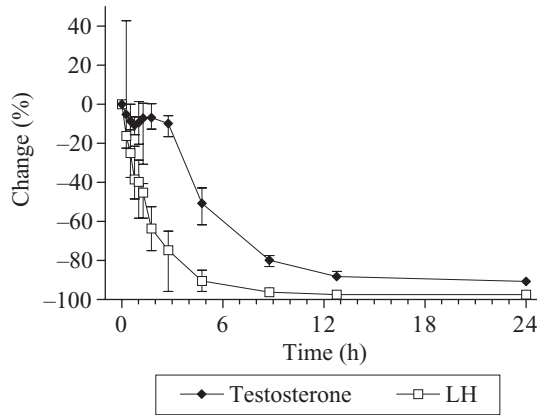


**FIGURE 4** Changes (median and quartiles) in serum levels of testosterone and luteinising hormone (LH) after intravenous administration of degarelix (30 µg/kg) to healthy subjects. The response confirms the mechanism of action of the compound (GnRH antagonism leading to LH decline with resultant chemical castration).

### 27.7.3 Clinical Trials Using Biomarkers in Patients

In these trials, biomarkers related to the pathophysiology of the disease are analyzed and drug-induced modulation of these markers is evaluated. For truly innovative, first-in-class therapeutic agents, a biomarker-based PoP/PoC poses significant challenges. The pathophysiology of the disease is often not completely understood. Consequently, the link between the different biomarkers and the disease biology might not be evident. Therefore, a positive change in a biomarker in a patient trial may constitute a positive PoP but does not guarantee a positive clinical outcome. This is unavoidable in all cases where the biomarkers have not yet been clinically validated. However, as discussed above, the use of biomarkers in PoP/PoC trials is a way of increasing the confidence in a particular therapeutic agent or class of agents to make better internal go–no go decisions. The alternative is to rely primarily on huge clinical trials that might fail late and at a much higher cost.

One therapeutic area in which the use of biomarkers during early drug development is of particular interest is Alzheimer's disease (AD). Despite extensive R&D efforts by many large and small pharmaceutical companies, disease-modifying therapies have been elusive. The identification of presymptomatic neurodegenerative changes that are mirrored by changes in specific markers is crucial and remains to be in focus for AD research. A wide array of biochemical markers such as  $\beta$ -amyloid and tau proteins in CSF [42] and imaging-based biomarkers [43, 44] are available, and evidence seems to suggest that a multiple biomarker approach will be necessary [45]. The logistics of large clinical trials in AD and the fact that available symptomatic treatment of AD (cholinesterase inhibitors, memantine) influences the primary clinical endpoint (cognition), thus blurring the potential effects of a truly disease-modifying agent, make biomarker-based PoP/PoC the most attractive approach to develop such important therapies in a huge area of unmet medical need. Other proposed strategies to enhance drug development in this area include enriched population

PoC studies (patients with mild cognitive impairment who are at higher risk of developing AD), implementing alternative, more sensitive clinical outcome measurements and adaptive dose-finding trials (seamless phase IIb–III) [46].

#### **27.7.4 Trials Using Clinical Endpoints for PoC**

In the absence of qualified biomarkers or acceptable surrogates, PoC can only be established by a successful clinical trial conducted in the target patient population using solid clinical endpoints for efficacy. As discussed earlier, biomarkers that are not fully validated can still be useful in demonstrating a pharmacological effect and selecting doses for the “ultimate” PoC trial. PoC trials using clinical outcomes are usually large and can be of long duration, depending on the natural course of the disease, the drug’s mode of action, and the sensitivity of the clinical measurements to therapeutic intervention.

### **27.8 KNOWLEDGE OF EXPOSURE–RESPONSE RELATIONSHIP OF DRUG IS ESSENTIAL FOR SUCCESSFUL POP/POC**

Adequate characterization of the ER relationship of the drug is vital for robust PoP/PoC conclusions and consequently solid go–no go decisions. This is easily understood if one realizes that a primary objective of PoP/PoC trials is to demonstrate some signs of efficacy, be it direct or via a biomarker or other surrogates at a dose that does not elicit unacceptable adverse effects. Failure to explore the ER relationship properly early on carries a risk of selecting inappropriate dose(s) in the PoP/PoC trial and, consequently, making the wrong go–no go decision. Some drug developers still advocate the use of a high, tolerated dose close to the maximal tolerated dose (MTD) in a PoP/PoC trial; “if this doesn’t work, why bother?” While this approach might work when developing compounds with wide therapeutic windows, it ignores the fundamental principles of clinical pharmacology and carries a significant risk of adverse drug reactions arising later in larger trials or even registering too high a dose of the drug. Finally, with increasing safety awareness, regulatory requirements often include the demonstration of the minimum dose that is maximally effective [11].

It should be underscored that a thoughtful integrated learning approach to PoP/PoC is a prerequisite for well-founded decisions. Notwithstanding the cost constraints that many pharmaceuticals companies experience, PoP/PoC strategies should not be “quick and dirty” or “lean and mean” development approaches. A rational scientifically solid and well-engineered PoP/PoC strategy indeed saves time and money by ensuring the progress of good drug candidates and the termination of compounds with poor chances of making it to the marketplace.

### **27.9 PROOF-OF-PRINCIPLE/PROOF-OF-CONCEPT OUTCOMES CAN BE “FALSE DAWNS”**

Despite the importance of PoP/PoC trials in drug development, as underscored in the literature and reiterated in this chapter, positive trial outcomes are by no means

a guarantee of clinical success. As discussed above, PoP/PoC trials usually rely on biomarkers/surrogates/models that might not adequately predict the clinical outcome and therefore fail the ultimate test of efficacy and safety, that is, the confirmatory phase III pivotal trials. Fleming and DeMets reviewed experiences with biomarkers/surrogates that failed to predict clinical outcomes in different therapeutic areas and discussed reasons for failure [47]. These are: (a) The biomarker/surrogate is not in the causal pathway of the disease. (b) The disease involves several pathways and the drug affects only the one mediated by the biomarker. (c) The biomarker is not in the pathway affected by the drug. (d) The drug affects mechanisms that are independent of the disease process [47]. One of the most classic examples of biomarkers/surrogates failing badly in predicting serious clinical outcomes is the use of reduction of ventricular extra systoles as a surrogate for decreased cardiovascular mortality after myocardial infarction. Flecainide, encainide, and moricizine were found to increase mortality [48, 49] despite the fact that they effectively suppressed ventricular arrhythmias. All three drugs were later withdrawn from the market. This and many other sobering examples reflect the fact that conclusive answers on the efficacy and safety of drugs can only be answered by large clinical trials using solid clinical endpoints. Moreover, long-term safety evaluation often requires extensive clinical use after drug approval. However, this does not decrease the value of biomarker-based PoP/PoC as a means of demonstrating relevant pharmacological activity of new compounds and capturing early signs of potential clinical utility. After all, PoP/PoC is primarily an indication that a certain compound is reasonably good enough to develop further.

### 27.9.1 Summary of Selected Human Models Quoted in This Chapter that Can be Used in PoP/PoC Trials

- *Human Endotoxemia Model* This is an experimental model of sepsis using intravenously administered, purified lipopolysaccharide (LPS) from *Escherichia coli* to healthy volunteers. The endotoxin induces changes that mimic early inflammatory and cardiovascular responses in septic shock [50–54]. The model has been widely used to study the mechanisms underlying the pathophysiology of sepsis and can potentially be useful to establish early PoP/PoC for drugs targeting these mechanisms. One limitation of the model is that it may not capture the increased vascular permeability, an important feature of severe sepsis/septic shock [55, 56].
- *Carbon Dioxide Inhalation Model* Inhalation of 7.5% carbon dioxide induces anxiety manifestations in healthy subjects with a sense of fear and tension together with increases in blood pressure and heart rate [57]. This model has been shown to be sensitive to the effects of lorazepam and to some extent paroxetine [58], and might therefore be useful to establish early PoP for new compounds intended for the treatment of generalized anxiety disorder (GAD) and panic disorder.
- *Scopolamine-Induced Cognitive Impairment Model* Cognitive impairment is induced in healthy subjects using subcutaneously or intravenously administered scopolamine; and the effect of cognition enhancers can be monitored, for example, using computerized neuropsychological test battery (CNTB) and quantitative electroencephalography (qEEG) [20–22].

- *Conditioned Aversive Anxiety Model* In this model, skin conductance responses to auditory stimuli are measured. Ten neutral tones are initially presented (habituation phase), and tone 11 is immediately followed by an aversive white noise (unconditioned stimulus). The response to a second presentation of the tones (extinction phase) is recorded. Amplitude as well as spontaneous fluctuations (variability) of the skin conductance response is registered. A positive control such as diazepam is used, and compounds with putative anxiolytic effects are compared with the control [30, 59].
- *Pentagastrin-Induced Gastric Acid Secretion* In this model, maximal/submaximal gastric acid secretion is induced in healthy subjects by intravenous infusion of pentagastrin. The effect of antiulcer/GERD drugs such as proton pump inhibitors on gastric acid output and intragastric pH is measured [60, 61].
- *Phenylephrine Challenge* This model is used to investigate the effect of alpha blockers on urethral tone and blood pressure in order to estimate the selectivity for the effect on urethral smooth muscles relative to the hypotensive response [62]. Phenylephrine, which raises urethral and blood pressure, is administered by dose escalation infusions at the estimated  $T_{\max}$  for the compound, urethral pressure is monitored using a sensor-equipped self-retaining urethral catheter, and blood pressure is recorded simultaneously. New, more selective compounds can be tested in this model establishing PoP/PoC [62].

## REFERENCES

1. Pharmaceutical Research and Manufacturers of America (2008), PhRMA, Pharmaceutical Industry Profile 2008, PhRMA, Washington, D.C.; available at: <http://www.phrma.org/files/2008%20Profile.pdf>.
2. Kola, I., and Landis, J. (2004), Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug. Discov.*, 3(8), 711–715.
3. Lesko, L. J., Rowland, M., Peck, C. C., et al. (2000), Optimizing the science of drug development: Opportunities for better candidate selection and accelerated evaluation in humans, *Pharmaceutical Res.*, 17(11), 1335–1344.
4. Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., et al. (2001), Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework, *Clin. Pharmacol. Ther.*, 69, 89–95.
5. Lesko, L. J., and Atkinson, A. J. (2001), Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: Criteria, validation, strategies, *Annu. Rev. Pharmacol. Toxicol.*, 41, 347–366.
6. Food and Drug Administration. The “Critical Path Initiative”—Innovation–Stagnation: Challenge and opportunity on the critical path to new medical products; available at: <http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf>.
7. Wagner, J. A., Williams, S. A., and Webster, C. J. (2007), Biomarkers and surrogate endpoints for fit-for-purpose development and regulatory evaluation of new drugs. *Clin. Pharmacol. Ther.*, 81(1), 104–107.
8. Williams, S. A., Slavin, D. E., Wagner, J. A., et al. (2006), A cost-effectiveness approach to the qualification and acceptance of biomarkers, *Nat. Rev. Drug. Discov.*, 5(11), 897–902.

9. Berns, B., Démolis, P., and Scheulen, M. E. (2007), How can biomarkers become surrogate endpoints? *EJC Suppl.*, 5(9), 37–40.
10. Rolan, P. (1997), The contribution of clinical pharmacology surrogates and models to drug development—a critical appraisal, *Br. J. Clin. Pharmacol.*, 44(3), 219–225.
11. Food and Drug Administration (2003), Guidance for industry: Exposure-response relationships—study design, data analysis, and regulatory applications.
12. Kuhlmann, J. (2007), The applications of biomarkers in early clinical drug development to improve decision-making processes, *Ernst Schering Res. Found. Workshop, Rev.*, (59), 29–45.
13. US Food and Drug Administration. Table of valid genomic biomarkers in the context of approved drug labels; available at: [http://www.fda.gov/cder/genomics/genomic\\_biomarkers\\_table.htm](http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm).
14. Roses, A. D., Saunders, A. M., Huang, Y., et al. (2007), Complex disease-associated pharmacogenetics: Drug efficacy, drug safety, and confirmation of a pathogenetic hypothesis (Alzheimer's disease), *Pharmacogenomics J. Rev.*, 7(1), 10–28.
15. Pustai, L. (2007), Limitations of pharmacogenomic predictor discovery in Phase II clinical trials, *Pharmacogenomics*, 8(10), 1443–1448.
16. Stroobants, S., Goeminne, J., Seegers, M., et al. (2003), 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec), *Eur. J. Cancer*, 39(14), 2012–2020.
17. Christian, B. T., Livni, E., Babich, J. W., et al. (1996), Evaluation of cerebral pharmacokinetics of the novel antidepressant drug, BMS-181101, by positron emission tomography, *J. Pharmacol. Exp. Ther.*, 279(1), 325–331.
18. Pien, H. H., Fishman, A. J., Thrall, J. H., (2005), Sorensen, A. G. Using imaging biomarkers to accelerate drug development and clinical trials, *Drug Disc. Today*, 10(4), 259–266.
19. Strack, T. (2007), Imaging as a tool in drug development, *Drugs Today*, 43(10), 725–736.
20. Christensen, H., Maltby, N., Jorm, A. F., et al. (1992), Cholinergic “blockade” as a model of the cognitive deficits in Alzheimer's disease, *Brain*, 115(Pt 6), 1681–1699.
21. Brooks, P., Preston, G. C., Traub, M., et al. (1988), Modelling dementia: Effects of scopolamine on memory and attention, *Neuropsychologia*, 26(5), 685–700.
22. Brass, E. P., Polinsky, R., Sramek, J. J., et al. (1995), Effects of the cholinomimetic SDZ ENS-163 on scopolamine-induced cognitive impairment in humans, *J. Clin. Psychopharmacol.*, 15(1), 58–62.
23. Gilles, C., and Luthringer, R. (2007), Pharmacological models in healthy volunteers: their use in the clinical development of psychotropic drugs, *J. Psychopharmacol. Rev.*, May; 21(3), 272–282.
24. Adamus, W.S., Jansen, U., Schilling, C., et al. (1988), Bronchospasm induced by methacholine inhalation as a model for testing of bronchospasmolytics in healthy volunteers, *Methods Find. Exp. Clin. Pharmacol.*, 10(2), 135–139.
25. Foster, R. W., Jubber, A. S., Hassan, N. A., et al. (1993), Trials of the bronchodilator activity of the xanthine analogue SDZ MKS 492 in healthy volunteers during a methacholine challenge test, *Eur. J. Clin. Pharmacol.*, 45(3), 227–234.
26. O'Shaughnessy, T.C., Georgiou, P., Howland, K., et al. (1997), Effect of pranlukast, an oral leukotriene receptor antagonist, on leukotriene D4 (LTD4) challenge in normal volunteers, *Thorax*, 52(6), 519–522.
27. Carratù, P., Morelli, N., Freire, A. X., et al. (2003), Effect of zafirlukast on methacholine and ultrasonically nebulized distilled water challenge in patients with mild asthma, *Respiration*, 70(3), 249–253.



28. Diamant, Z., Grootendorst, D. C., Veselic-Charvat, M., et al. (1999), The effect of montelukast (MK-0476), a cysteinyl leukotriene receptor antagonist, on allergen-induced airway responses and sputum cell counts in asthma, *Clin. Exp. Allergy*, 29(1), 42–51.
29. Guttmacher, L. B., Murphy, D. L., and Insel, T. R. (1983), Pharmacologic models of anxiety, *Compr. Psychiatry*, 24, 312–326.
30. Deakin, J. F. W., Guimaraes, F. S., and Graeff, F. G. (1994), Testing 5HT theories of anxiety in normal volunteers, in Palomo, T., and Archer, T., Eds., *Strategies for Studying Brain Disorders, Vol. 1, Depressive, Anxiety and Drug Abuse Disorders*, Farrand, London, pp. 211–238.
31. Arendt-Nielsen, L., Curatolo, M., and Drewes, A. (2007), Human experimental pain models in drug development: Translational pain research, *Curr. Opin. Investig. Drugs. Rev.*, 8(1), 41–53.
32. Allocco, F. T., Votypka, V., deTineo, M., et al. (2002), Effects of fexofenadine on the early response to nasal allergen challenge, *Ann. Allergy Asthma Immunol.*, 89(6), 578–584.
33. Greiff, L., Persson, C. G., Svensson, C., et al. (1995), Loratadine reduces allergen-induced mucosal output of alpha 2-macroglobulin and tryptase in allergic rhinitis, *J. Allergy Clin. Immunol.*, 96(1), 97–103.
34. Kasteleijn-Nolst Trenité, D. G. A., Marescaux, C., Stodieck, S., et al. (1996), Photosensitive epilepsy: A model to study the effects of antiepileptic drugs. Evaluation of the piracetam analogue, levetiracetam, *Epilepsy Res.*, 25(3), 225–230.
35. Kasteleijn-Nolst Trenité, D. G. A., Genton, P., Parain, D., et al. (2007), Evaluation of brivaracetam, a novel SV2A ligand, in the photosensitivity model, *Neurology*, 69(10), 1027–1034.
36. Trenité, D. G., French, J. A., Hirsch, E., et al. (2007), Evaluation of carisbamate, a novel antiepileptic drug, in photosensitive patients: An exploratory, placebo-controlled study, *Epilepsy Res.*, 74(2–3), 193–200.
37. Schmidt, B. (2006), Proof of principle studies, *Epilepsy Res.*, 68, 49–52.
38. Food and Drug Administration (2006), Guidance for Industry, Investigators and Reviewers, Exploratory IND Studies.
39. Kinders, R., Parchment, R. E., Ji, J., et al. (2007), Phase 0 clinical trials in cancer drug development: From FDA guidance to clinical practice, *Mol. Interv.*, 7(6), 325–334.
40. Kummar, S., Kinders, R., Gutierrez, M., et al. (2007), Phase 0 Working Group: Inhibition of poly (ADP-ribose) polymerase (PARP) by ABT-888 in patients with advanced malignancies: Results of a phase 0 trial, *J. Clin. Oncol.*, ASCO Annual Meeting Proceedings Part I, 25(18S), 3518.
41. Hill, T. P. (2007), Phase 0 trials: Are they ethically challenged? *Clin. Cancer Res.*, 13(3), 783–784.
42. Sunderland, T., Linker, G., Mirza, N., et al. (2003), Decreased beta-amyloid1-42 and increased tau levels in cerebrospinal fluid of patients with Alzheimer disease, *JAMA*, 289(16), 2094–2103.
43. Frank, R. A., Galasko, D., Hampel, H., et al. (2003), National Institute on Aging Biological Markers Working Group. Biological markers for therapeutic trials in Alzheimer's disease. Proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease, *Neurobiol. Aging Rev.*, 24(4), 521–536.
44. Dickerson, B. C., and Sperling, R. A. (2005), Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease, *NeuroRx. Rev.* 2(2), 348–360.
45. Frank, R., and Hargreaves, R. (2003), Clinical biomarkers in drug discovery and development, *Nat. Rev. Drug Discov.*, 2(7), 566–580.

46. Cummings, J. L. (2008), Optimizing phase II of drug development for disease-modifying compounds, *Alzheimer dementia*, 4, S15–S20.
47. Fleming, T. R., and DeMets, D. L. (1996), Surrogate endpoints in clinical trials: Are we being misled? *Ann. Intern. Med.*, 125(7), 605–613.
48. Echt, D. S., Liebson, P. R., Mitchell, L. B., et al. (1991), Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial, *N. Engl. J. Med.*, 324(12), 781–788.
49. Epstein, A. E., Hallstrom, A. P., Rogers, W. J., et al. (1993), Mortality following ventricular arrhythmia suppression by encainide, flecainide, and moricizine after myocardial infarction. The original design concept of the Cardiac Arrhythmia Suppression Trial (CAST), *JAMA*, 270(20), 2451–2455.
50. Martich, G. D., Boujoukos, A. J., and Suffredini, A. F. (1993), Response of man to endotoxin, *Immunobiology*, 87, 403–416.
51. Suffredini, A. F., Fromm, R. E., Parker, M. M., et al. (1989), The cardiovascular response of normal humans to the administration of endotoxin, *N. Engl. J. Med.*, 321, 280–287.
52. Suffredini, A. F., Shelhamer, J. H., Neumann, R. D., et al. (1992), Pulmonary and oxygen transport effects of intravenously administered endotoxin in normal humans, *Am. Rev. Resp. Dis.*, 45, 1398–1403.
53. Suffredini, A. F., Harpel, P. C., and Parrillo, J. E. (1989), Promotion and subsequent inhibition of plasminogen activation after administration of intravenous endotoxin to normal subjects, *N. Engl. J. Med.*, 320, 1165–1172.
54. Kumar, A., Bunnell, E., Lynn, M., et al. (2004), Experimental human endotoxemia is associated with depression of load-independent contractility indices: Prevention by the lipid analogue E5531, *Chest*, 126, 860–867.
55. van Eijk, L. T. G. J., Pickkers, P., Smits, P., et al. (2005), Microvascular permeability during experimental human endotoxemia: An open intervention study, *Crit. Care*, 9, R157–R164.
56. Anel, R., and Kumar, A. (2005), Human endotoxemia and human sepsis: Limits to the model, *Crit. Care*, 9, 151–152.
57. Poma, S. Z., Milleri, S., Squassante, L., et al. (2005), Characterization of a 7% carbon dioxide (CO<sub>2</sub>) inhalation paradigm to evoke anxiety symptoms in healthy subjects, *J. Psychopharmacol.*, 19, 494–503.
58. Bailey, J. E., Kendrick, A., Diaper, A., et al. (2007), A validation of the 7.5% CO<sub>2</sub> model of GAD using paroxetine and lorazepam in healthy volunteers, *J. Psychopharmacol.*, 21(1), 42–49.
59. Hellewell, J. S. E., Guimaraes, F. S., Wang, M., et al. (1999), Comparison of buspirone with diazepam and fluvoxamine on aversive classical conditioning in humans, *J. Psychopharmacol.*, 13, 122–127.
60. Lind, T., Cederberg, C., Ekenved, G., et al. (1983), Effect of omeprazole—a gastric proton pump inhibitor on pentagastrin stimulated acid secretion in man, *Gut*, 24, 270–276.
61. Pratha, V. S., Hogan, D. L., Lane, J. R., et al. (2006), Inhibition of pentagastrin-stimulated gastric acid secretion by pantoprazole and omeprazole in healthy adults, *Digestive Diseases Sci.*, 51(1), 123–131.
62. Sultana, S. R., Marshall, S., Davis, J., et al. (2007), Experiences with dose finding in patients in early drug development: The use of biomarkers in early decision making, *Ernst Schering Res. Found. Workshop*, 59, 65–79.



# *Index*

- Accelerated approval, 240
- Accuracy, 964
- ACE (angiotensin-converting enzyme), 5
- Acne grading systems, 477
- Adaptive designs, 1044
- Adaptive dose finding, 142
- Adaptive trials, 1018
- Adjusting sample size calculations for
  - noncompliance, 1093
- Adjuvant therapy, 597
- Advanced cancer, 599
- Advanced pancreatic adenocarcinoma, 1041
- Advantages and limitations of web-based clinical trials, 222
- Adverse drug reaction (ADR), 88, 97, 329
  - classification in humans, 89
- Adverse event (AE), 88
  - common terminology criteria (CTCAE), 289
  - guidelines for recording, 1029
  - temporal relationships of, 390
  - unexpected, 83
- Age, 763
- Age-related macular degeneration (AMD), 614
- AIDS, 27
- ALARA (as low as reasonably achievable), 594
- Alzheimer disease, 81
- Alzheimer's disease (AD), 81, 149, 668–670, 871, 887, 1140
  - diagnosis of probable, 670
- Animal efficacy rule, 26
- Annual reports, 362
- Anthrax, 26
- Antimetabolites, 3
- Area under the curve (AUC), 973
- Area under the effect curve (AUEC), 986
- Area under the first moment curve (AUMC), 974
- Atrial fibrillation, 491
- Attrition bias, 515
- Bare-metal stents, 406
- Bayesian analysis, 840
- Bayesian approaches, 146, 267, 268, 1016
- Bayesian methods, 998
- Belmont report, 1125
- Berlin Code of 1900, 903
- Best Pharmaceuticals for Children Act (BPCA), 1164
- Bias, 1060, 1080
- Binary or categorical data, 845
- Bioavailability trials, 960

- Bioequivalence (BE), 248, 960
- Biomarkers, 859, 865, 984
  - life cycle, 861
  - neuroimaging, 881
  - predictive, 889
  - to select patient subgroups, 888
  - soluble, 886
- Biopharmaceutics guidances, 1169
- Bioresearch monitoring (BIMO), 58
- BLA, 29, 234
- Blinding
  - indices for assessment of, 949
  - trials other than oral drug comparisons, 946
  - single, double, and triple blind, 528
  - waiving blindness, 947
- Boundaries approach, 1037
- Breaking code
  - in case of emergency, 947
  - at end of trial, 950
- Brownian motion paradigm, 1050
- Carbon dioxide inhalation model, 1219
- Carcinosarcoma of female genital, 257
- Cardiovascular system, 93
- Case report form (CRF), 151, 176, 177
- Case-control studies, 315
- Caucasian, 762
- Causal chain, 1006
- Censored endpoints, 847
- cGMP quality systems requirements, 38, 40, 42
- Chemistry, manufacturing, and controls (CMC), 166
- Choice of endpoint, 257
- Christmas tree correction, 1038
- CIOMS, 1132
- Classification of adverse drug reactions (ADRs) in humans, 89
- Clinical hold: issues and resolution, 361
- Clinical pharmacology guidances, 1170
- Clinical risk assessment, 58
- Clinical studies necessary for development of new vaccines, 781
- Clinical trials
  - initiating, 907
  - length of, 684
  - oral health, statistical issues, 445
  - registration, 478
  - stimulation, 1011
  - using biomarkers, 1217
- Clinically relevant outcomes, 984
- Cluster randomization, 807
- Coefficients of variation, 991
- Coercion, 1133
- Cognitive tools, 720
- CTCAE (common terminology criteria for adverse events), 289
- Comparison of two means, 1071
- Comparison of two proportions, 1073
- Compartmental models, 977
- Competence, 1134
- Complex response (CR), 287
- Concomitant medications, 164
  - effect of, 391
- Conditioned adverse anxiety model, 1220
- Confidence intervals, 1055, 1068, 1081
  - relationship between significant testing and, 1069
- Confounding, 1061, 1081
- CONSORT (consolidated standards of reporting trials), 940, 1139
- Contamination
  - of early batches of IPV by simian virus 40 (SV<sub>40</sub>), 778
  - of yellow fever vaccine by hepatitis B, 777
- Coronary artery bypass graft, 414
- Coronary stent design, 403
- Correlation coefficient (r), 1081
- Correlative studies, 295
- Council for International Organizations and Medical Sciences (CIOMS), 1126
- Covariate-adaptive randomization, 799
- Covariates, 1081
- Cox proportional-hazards method, 1079
- C-reactive protein (CRP), 859
- Creatinine clearance, 546
- CRF, 195
- Critical path initiatives, 35
- Cross-over designs, 1198
- Cross-sectional observational studies, 313
- Cutter incident, 778
- Cyclo-oxygenase-2 inhibitor, 623
- CYP 1A1, 1A2, and 1B1, 752
- CYP 2C19, 753
- CYP 2C9, 753
- CYP 2D6, 753
- CYP 2E1, 753
- CYP 3A4 and 3A5, 753
- CYP1A2, 967
- CYP2D6, 967
- CYP2E1, 961
- CYP3A4/5, 967

- Data
  - capture, 153
  - cleaning, 208
  - electronic capture (EDC), 211
  - entry methods, 208
  - high-dimensional aspect of, 390
  - management study plan, 186
  - mining, 343, 1002
  - missing, 935
  - protection and security issues, 220
  - queries, 177
  - types, 1062
- Data-Monitoring Committee, 293
- Data-monitoring safety boards, 904
- Data safety and monitoring board (DSMB), 248, 1027, 1050
- Declaration of Helsinki, 127, 903, 1121, 1124, 1189
- Definition of “Japanese,” 761
- Dementia, 712, 716, 719
- Demographic characteristics trials, 960
- Dental implant clinical trials, 455
- Deontological theory, 1120
- Diagnostic trials, 280
- Dietary factors, 755
- Difference in means, 924
- Difference in proportions, 926
- Differences in means adjusted for baseline data, 926
- Disease progression modeling, 1010
- Disease-modifying trials, 693
- Dose-exposure-response modeling, 1003
- Dose-limiting toxicity (DLT), 1015
- Dose-range trials, 960
- Drug formulation trials, 960
- Drug interaction trials, 960
- Drug master file (DMF), 239
- Drug safety, 865
- Drug-eluting stents, 407
- Drusen ablation, 616
- DSMB, 1041, 1042
- DSM-IV criteria for diagnosis of Alzheimer’s disease, 669
- E5 guidelines, 747
- EDC (electronic data capture), 211
- Efficacy, 867, 987
- Electrocardiography (EEG), 82
- Elixir sulfanilamide, 123
- End of phase II (EOP-II), 31, 32
- Endocrine system, 100
- Endpoint, 1081
- Entities involved in clinical trials, 10
- Environmental assessment, 1167
- Epirubicin in nonresponsive breast cancer, 262
- Equivalence trials, 934
- Estimating magnitude of treatment difference, 1052
- Ethical issues in human pharmacology trials, 956
- Ethics committee (HREC), 1108, 904
- Ethics, 8
  - four golden rules of ethical conduct, 1146
- Ethnic factors, 740, 748
- Ethnic pharmaceutical safety and efficacy issues, pharmacodynamics, 757
- Evidence-based medicine (EBM), 116, 568
- Exempted review, 908
- Expanded access, 240
- Expected effect size, 525
- Expedited review, 910
- Exploratory modeling, 1002
- Exponential models, 977
- Extrinsic factors, 749
- Fast-track approval, 366
- FDA
  - conducting and documenting formal meetings, 357
  - consultation with, 356
- FDA-1571, 75
- Federal Food, Drug and Cosmetic Act (FDC), 228, 1160
- FIAU (fialuridine), 249, 99
- Fibonacci sequence, 3
- Financial disclosure, 1135
- First-in-man (FIM) studies, 247, 1013
- First-time-in-man (FTIM), 1191
- 505(b)(2), 239
- Flexible designs, 1202
- Fluoxetine bioequivalence study, 970
- FOB, 72
- Food effect trials, 960
- Functional imaging, 883
- Gastrointestinal system, 96
- Gemcitabine in nasopharyngeal carcinoma, 258, 263
- Generic drugs, 238
- Genital system and teratology, 103

- Genotoxicity, 102
- Geometric averages, 991
- Geriatric population, 541
- GLP regulations in nonclinical investigations, 36
- Good tissue practice compliance, 54
- Grapefruit, 967
- Group sequential designs, 1201
- Group sequential testing procedures, 1049
- Guillain-Barré Syndrome (GBS) and swine influenza vaccine (1976–1977), 779
- Health Insurance Portability and Accountability Act (HIPAA), 300
- Healthy normal versus patient population, 763
- Hemopoietic system, 100
- Hepatic impairment, 550
- Hepatic system, 99
- hERG, 105, 91
- Hierarchy of evidence, 179
- HIPAA, 297
- Hippocratic Oath, 1123
- Historical controls, 817
  - problems with, 818
- Human endotoxemia model, 1219
- Human tissue-based products, (HCT/Ps), 54
- Human vaccine manufacture, 776
- Hypothesis testing, 1065
- Hysteresis PKPD models, 988
- ICH E8, 1014
- ICH E9, 248, 250
- ICH
  - history, 746
  - implications for non-ICH jurisdictions, 750
- ICH M3, 1014
- Imaging, 1213
- Immunological system, 101
- Inactivated influenza vaccine, and Bell's palsy (2000), 779
- Inclusion and exclusion criteria, 283
- IND
  - emergency use of, 27
  - labeling requirements, 55
  - review and approval process, 33
- Induction therapy, 596
- Informed consent forms, 297
- Informed consent, 489, 788, 905, 907, 1140
- Institutional ethics committee (IEC), 902
- Institutional review board (IRB), 26, 28, 46, 47, 49, 79, 231, 482, 644, 902, 904, 1142
  - registration, 905
  - review, basic items, 908
- Integrated dose-exposure-response modeling, 1008
- Intention to treat, 581
- Interim analyses, 292
- Interventional cardiology, 402
- Intrinsic factors, 749
- Investigational new drug applications, 358
- Investigator
  - brochure, 76
  - IND initiated by, 27
  - responsibilities, 912
  - selection, 1107
  - trials initiated by, 27
- IRB. *See* institutional review board
- ISO 9001 requirements, 42
- James Lind, 118
- Japan, traditional practice of medicine in, 742
- Joseph Lister, 120
- Kaplan-Meier method, 1077
- Kefauver-Harris drug amendments, 1161
- Kinetica, 978, 982
- Linearity, 965
- Lithium gamolenate, 1041
- Logistic regression model, 987
- Log-odds ratio, 1034
- Longitudinal models, 1001
- Mainland-Gart test, 845
- Male and female participants, 763
- Masking, 944
- Maximally tolerated dose (MTD), 1015
- Maximization of research impact on medical treatment, 1147
- Maximum concentration ( $C_{\max}$ ), 972
- Maximum effect observed ( $E_{\max}$ ), 986
- Maximum recommended starting dose (MRSD), 1014
- MCP-Mod: unified strategy for dose-finding studies, 1017
- Mean, 1063
- Measures of neuropsychiatric and behavioral changes, 725



- Median approval times, 236
- Median, 1063
- Medical coding, 210
- Medical Dictionary for Regulatory Activities (MedDRA), 210
- Medical registers, 317
- Meta-analysis, 531
- Methotrexate, 3
- Microdialysis, 963
- Mild cognitive impairment (MCI), 668
- Minimization of risk to research participants, 1148
- Misrepresentation of estimate of treatment effect and variance, 929
- Missing information, 385
- Misspecification of rates, 928
- Model-based analysis, 836
- Monitoring visits, 175
- MS
  - diagnosing in clinical practice, 880
  - specific features of, 887
- Multiple mechanisms, 756
- Myocardial infarction, 492
  
- National Formulary, 1160
- National Research Act, 903, 1125
- NDA review process, 1163
- Nervous system, 94
- Net present value (NPV), 1182
- Neuroimaging biomarkers, 881
- Neuropsychiatric symptoms, 676
- New drug application (ANDA), 239
- Nocebo, 250
- Noncompartmental PK parameters, 974
- Non-CYP metabolic enzymes, 754
- Non-Hodgkin's lymphoma, 272
- Noninferiority trials, 934
- NonMEM, 982
- Nonparametric analyses, 844
- NSAIDs, 97
- Nuisance parameters, 1028
- Number needed to treat (NNT), 1082
- Nuremberg Code, 1149, 126
- Nuremberg Military Tribunal, 903
- Nuremberg, 1124
  
- Obesity, 559
- Office of Human Research Protections (OHRP), 902, 904
- Office of Regulatory Affairs (ORA), 229
- Oncology studies, 3, 1015
- Oncology, 15
  
- Oral rotavirus vaccines (1999), 779
- Organ insufficiency trials, 960
- Orphan drug, 44, 240
- Orphan "same drug" category, 45
- Orphan Drug Act, 124
- Outcome measures of function, 721
- Overrunning analysis, 1044
- Over-the-counter (OTC) drugs, 238, 241
  
- Paclitaxel for unresectable hepatocellular carcinoma, 263
- Palliative treatment, 600
- Pancreatic cancer trial, 1076
- Partial response (PR), 287
- Pathways modeling, 1008
- Patient
  - autonomy of, respect for, 1147
  - registration, 786
- PDUFA, 124, 235
- Pediatric drugs, 241
- Pediatric initiatives, 1163
- Pediatric investigation plans, 651
- Pediatric Research Equity Act (PREA), 1164
- Pegaptanib sodium, 622
- Pentagastrin-induced gastric acid secretion, 1220
- Percentile, 1063
- Peripheral stents, 423
- P-glycoprotein, 1171
- Pharmaceutical development in Japan, 743
- Pharmacogenetic trials, 960
- Pharmacokinetic modeling, 977
- Pharmacokinetic-Pharmacodynamic models, 987, 1005
- Phase I trial, 233, 247, 249
- Phase II trial, 233, 257
  - efficacy and toxicity in, 265
  - end of phase II (EOP-II), 31, 32
- Phase III trial, 5, 233, 285, 1049
- Phase IV trial, 6
- Phenylephrine challenge, 1220
- Photodynamic therapy, 618
- Phototherapy and laser trials, 469
- Physiological modeling, 1009
- Placebo effect, 288
- Placebo
  - controlled trials, 1131
  - features, 942
  - responses, 687
- PopKinetics, 982

- Population PK
  - data, 981
  - parameters, 981
- Population trials, 961
- Postmarketing surveillance, 328, 333
- Potency, 987
- Power requirement I, 1037
- P-Pharm, 982
- Precision, 964
- Preexisting disease, effect of, 391
- Pregnancy trials, 960
- Pre-IND meeting, 32
- Prescription Drug User Fee Act (PDUFA), 1161
- Presymptomatic biomarkers, 881
- Prevention trials, 280
- Primary endpoint, 1050
- Principal investigator, 912
- Principlism, 1121
- Progression-free survival (PFS), 282
- Proof-of-concept studies, 1016
- Proof-of-principal/proof-of-concept, 1208
- Protect human subjects, 46
- Pure Food and Drugs Act, 122
  
- Qualified outcome measures for
  - musculoskeletal disorders, 579
- Quality of life (QOL), 9, 570
  - trials, 280
- Quantitative variables, summarizing, 1063
- Quinine, 967
  
- Radiation therapy, 620
- Radiochemotherapy, 597
- Random error, 1061
- Randomized clinical trial, 1127
- Randomized control trials, 509
- Range, 1063
- Ranibizumab, 622
- Rate constant of the terminal phase ( $\lambda_z$ ), 973
- Regional requirements, 748
- Regression models, 1000
- Regulatory requirements and issues, 252
- Renal and urinary system, 100
- Renal impairment, 543
- Repeated significance testing, 1035
- Respiratory system, 96
- Response-driven designs, 1044
- Review boards, 706
- Rheopheresis, 616
- Rheumatoid arthritis (RA), 12
- Rising dose escalation studies, 140
- Robustness, 965
- Run-in periods, 1090
  
- Safety surveillance of antiretroviral drug, 341
- Safety, 250
- Sample size, 172, 580, 848, 1090
  - calculations, 513
  - re-estimation, 932
- Sample-rich trials, 961, 965
- SAS, 982
- Scientific integrity, 1148
- Scopolamine-induced cognitive
  - impairment model, 1219
- Screening trials, 280
- Seamless designs, 147
- Secondary endpoint, 1056
- Selectivity, 965
- Sensitivity, 931, 965
  - to ethnic factors, 557, 749
- Sepsis, 496
- Sequential design, 1040
- Sequential probability ratio test (SPRT), 1038
- Serious adverse event (SAE), 1111, 1029
- Single-stage designs, 260
- Single-threshold design (STD), 268
- Sire performance, 151
- Skin, 101
- Special population studies, 536
  - design features, 538
- Special protocol assessments (SPAs), 1165
- Special senses, 95
- Specificity, 965
- Spending function approach, 1036
- Stable disease (SD), 287
- Standard deviation, 1063
- Starting dose, 285
- Stenting, 417. *See also names of specific stents*
- Stopping rules, 1040
- Stroke, 495, 713
- Study close-out activities, 182
- Study medication supplies, 164
- Subjects, variations among, 390
- Surgical device trials, 707
- Surrogate endpoint, 878
- Surrogate markers, 984
- Survival analysis, 1075
- Systematic errors, 1060

- Terminal half-life, 975
- TGN1412, 1014
- Therapeutic products directorate (TPD), 1173
- Thermal laser photocoagulation, 617
- Threshold of toxicological concern (TTC), 103
- Time of maximum effect ( $t_{\text{Emax}}$ ), 986
- Time to event, 927
- Time to progression, 604
- Tolerance phenomenon, 985
- Torsades de pointes, 91
- Total drug clearance, 975
- Transpupillary thermotherapy, 618
- Trapezoidal rule, 973, 986
- Trauma, 713
- Treatment IND, 27
- Treatment trials, 280
- Trial. *See also* clinical trials
  - design, 284
  - size, methods for determining, 923
  - size, realistic assessment of, 928
  - surgical device, 707
  - treatment, 280
- Troglitazone, 99
- Tuskegee experiment, 1122
- $2 \times 2$  crossover design, 248
- Two-stage analysis, 837
- Two-stage designs, 261
- Type I error (significance level), 1165, 1067
- Type II error (1, power of the test), 1165–1166, 1067
- Types of adaptive techniques, 138
- U.S. Pharmacopeia, 1160
- Unequal allocation, 935
- Unequal randomization, 806
- Unexpected adverse drug experiences, 83
- United Kingdom Prescription Event-Monitoring (PEM) program, 337
- Utilitarianism, 1120
- Vascular dementia (VaD), 668
- Waiving blindness, 947
- Weight, 764
- Weight loss drugs, 1070
- WinBUGS, 982
- WinNonLin, 978
- WinNonMix, 982
- Women
  - in clinical trials, 554
  - and pregnancy, 1143