

# CONTENTS

|                                                                                        |               |
|----------------------------------------------------------------------------------------|---------------|
| <b>Preface to the Third Edition</b>                                                    | <b>xxv</b>    |
| <b>About the Companion Website</b>                                                     | <b>xxviii</b> |
| <b>1 Preliminaries</b>                                                                 | <b>1</b>      |
| 1.1 Introduction, 1                                                                    |               |
| 1.2 Audiences, 2                                                                       |               |
| 1.3 Scope, 3                                                                           |               |
| 1.4 Other Sources of Knowledge, 5                                                      |               |
| 1.5 Notation and Terminology, 6                                                        |               |
| 1.5.1 Clinical Trial Terminology, 7                                                    |               |
| 1.5.2 Drug Development Traditionally Recognizes Four Trial Design Types, 7             |               |
| 1.5.3 Descriptive Terminology Is Better, 8                                             |               |
| 1.6 Examples, Data, and Programs, 9                                                    |               |
| 1.7 Summary, 9                                                                         |               |
| <b>2 Clinical Trials as Research</b>                                                   | <b>10</b>     |
| 2.1 Introduction, 10                                                                   |               |
| 2.2 Research, 13                                                                       |               |
| 2.2.1 What Is Research?, 13                                                            |               |
| 2.2.2 Clinical Reasoning Is Based on the Case History, 14                              |               |
| 2.2.3 Statistical Reasoning Emphasizes Inference Based on Designed Data Production, 16 |               |
| 2.2.4 Clinical and Statistical Reasoning Converge in Research, 17                      |               |

|          |                                                                               |           |
|----------|-------------------------------------------------------------------------------|-----------|
| 2.3      | Defining Clinical Trials, 19                                                  |           |
| 2.3.1    | Mixing of Clinical and Statistical Reasoning Is Recent, 19                    |           |
| 2.3.2    | Clinical Trials Are Rigorously Defined, 21                                    |           |
| 2.3.3    | Theory and Data, 22                                                           |           |
| 2.3.4    | Experiments Can Be Misunderstood, 23                                          |           |
| 2.3.5    | Clinical Trials and the Frankenstein Myth, 25                                 |           |
| 2.3.6    | <i>Cavia porcellus</i> , 26                                                   |           |
| 2.3.7    | Clinical Trials as Science, 26                                                |           |
| 2.3.8    | Trials and Statistical Methods Fit within a Spectrum of Clinical Research, 28 |           |
| 2.4      | Practicalities of Usage, 29                                                   |           |
| 2.4.1    | Predicates for a Trial, 29                                                    |           |
| 2.4.2    | Trials Can Provide Confirmatory Evidence, 29                                  |           |
| 2.4.3    | Clinical Trials Are Reliable Albeit Unwieldy and Messy, 30                    |           |
| 2.4.4    | Trials Are Difficult to Apply in Some Circumstances, 31                       |           |
| 2.4.5    | Randomized Studies Can Be Initiated Early, 32                                 |           |
| 2.4.6    | What Can I learn from $n = 20$ ?, 33                                          |           |
| 2.5      | Nonexperimental Designs, 35                                                   |           |
| 2.5.1    | Other Methods Are Valid for Making Some Clinical Inferences, 35               |           |
| 2.5.2    | Some Specific Nonexperimental Designs, 38                                     |           |
| 2.5.3    | Causal Relationships, 40                                                      |           |
| 2.5.4    | Will Genetic Determinism Replace Design?, 41                                  |           |
| 2.6      | Summary, 41                                                                   |           |
| 2.7      | Questions for Discussion, 41                                                  |           |
| <b>3</b> | <b>Why Clinical Trials Are Ethical</b>                                        | <b>43</b> |
| 3.1      | Introduction, 43                                                              |           |
| 3.1.1    | Science and Ethics Share Objectives, 44                                       |           |
| 3.1.2    | Equipoise and Uncertainty, 46                                                 |           |
| 3.2      | Duality, 47                                                                   |           |
| 3.2.1    | Clinical Trials Sharpen, But Do Not Create, Duality, 47                       |           |
| 3.2.2    | A Gene Therapy Tragedy Illustrates Duality, 48                                |           |
| 3.2.3    | Research and Practice Are Convergent, 48                                      |           |
| 3.2.4    | Hippocratic Tradition Does Not Proscribe Clinical Trials, 52                  |           |
| 3.2.5    | Physicians Always Have Multiple Roles, 54                                     |           |
| 3.3      | Historically Derived Principles of Ethics, 57                                 |           |
| 3.3.1    | Nuremberg Contributed an Awareness of the Worst Problems, 57                  |           |
| 3.3.2    | High-Profile Mistakes Were Made in the United States, 58                      |           |
| 3.3.3    | The Helsinki Declaration Was Widely Adopted, 58                               |           |
| 3.3.4    | Other International Guidelines Have Been Proposed, 61                         |           |

- 3.3.5 Institutional Review Boards Provide Ethics Oversight, 62
- 3.3.6 Ethics Principles Relevant to Clinical Trials, 63
- 3.4 Contemporary Foundational Principles, 65
  - 3.4.1 Collaborative Partnership, 66
  - 3.4.2 Scientific Value, 66
  - 3.4.3 Scientific Validity, 66
  - 3.4.4 Fair Subject Selection, 67
  - 3.4.5 Favorable Risk–Benefit, 67
  - 3.4.6 Independent Review, 68
  - 3.4.7 Informed Consent, 68
  - 3.4.8 Respect for Subjects, 71
- 3.5 Methodologic Reflections, 72
  - 3.5.1 Practice Based on Unproven Treatments Is Not Ethical, 72
  - 3.5.2 Ethics Considerations Are Important Determinants of Design, 74
  - 3.5.3 Specific Methods Have Justification, 75
- 3.6 Professional Conduct, 79
  - 3.6.1 Advocacy, 79
  - 3.6.2 Physician to Physician Communication Is Not Research, 81
  - 3.6.3 Investigator Responsibilities, 82
  - 3.6.4 Professional Ethics, 83
- 3.7 Summary, 85
- 3.8 Questions for Discussion, 86

## **4 Contexts for Clinical Trials**

**87**

- 4.1 Introduction, 87
  - 4.1.1 Clinical Trial Registries, 88
  - 4.1.2 Public Perception Versus Science, 90
- 4.2 Drugs, 91
  - 4.2.1 Are Drugs Special?, 92
  - 4.2.2 Why Trials Are Used Extensively for Drugs, 93
- 4.3 Devices, 95
  - 4.3.1 Use of Trials for Medical Devices, 95
  - 4.3.2 Are Devices Different from Drugs?, 97
  - 4.3.3 Case Study, 98
- 4.4 Prevention, 99
  - 4.4.1 The Prevention versus Therapy Dichotomy Is Over-worked, 100
  - 4.4.2 Vaccines and Biologicals, 101
  - 4.4.3 Ebola 2014 and Beyond, 102
  - 4.4.4 A Perspective on Risk–Benefit, 103
  - 4.4.5 Methodology and Framework for Prevention Trials, 105

- 4.5 Complementary and Alternative Medicine, 106
  - 4.5.1 Science Is the Study of Natural Phenomena, 108
  - 4.5.2 Ignorance Is Important, 109
  - 4.5.3 The Essential Paradox of CAM and Clinical Trials, 110
  - 4.5.4 Why Trials Have Not Been Used Extensively in CAM, 111
  - 4.5.5 Some Principles for Rigorous Evaluation, 113
  - 4.5.6 Historic Examples, 115
- 4.6 Surgery and Skill-Dependent Therapies, 116
  - 4.6.1 Why Trials Have Been Used Less Extensively in Surgery, 118
  - 4.6.2 Reasons Why Some Surgical Therapies Require Less Rigorous Study Designs, 120
  - 4.6.3 Sources of Variation, 121
  - 4.6.4 Difficulties of Inference, 121
  - 4.6.5 Control of Observer Bias Is Possible, 122
  - 4.6.6 Illustrations from an Emphysema Surgery Trial, 124
- 4.7 A Brief View of Some Other Contexts, 130
  - 4.7.1 Screening Trials, 130
  - 4.7.2 Diagnostic Trials, 134
  - 4.7.3 Radiation Therapy, 134
- 4.8 Summary, 135
- 4.9 Questions for Discussion, 136

## 5 Measurement

137

- 5.1 Introduction, 137
  - 5.1.1 Types of Uncertainty, 138
- 5.2 Objectives, 140
  - 5.2.1 Estimation Is The Most Common Objective, 141
  - 5.2.2 Selection Can Also Be an Objective, 141
  - 5.2.3 Objectives Require Various Scales of Measurement, 142
- 5.3 Measurement Design, 143
  - 5.3.1 Mixed Outcomes and Predictors, 143
  - 5.3.2 Criteria for Evaluating Outcomes, 144
  - 5.3.3 Prefer Hard or Objective Outcomes, 145
  - 5.3.4 Outcomes Can Be Quantitative or Qualitative, 146
  - 5.3.5 Measures Are Useful and Efficient Outcomes, 146
  - 5.3.6 Some Outcomes Are Summarized as Counts, 147
  - 5.3.7 Ordered Categories Are Commonly Used for Severity or Toxicity, 147
  - 5.3.8 Unordered Categories Are Sometimes Used, 148
  - 5.3.9 Dichotomies Are Simple Summaries, 148
  - 5.3.10 Measures of Risk, 149

- 5.3.11 Primary and Others, 153
- 5.3.12 Composites, 154
- 5.3.13 Event Times and Censoring, 155
- 5.3.14 Longitudinal Measures, 160
- 5.3.15 Central Review, 161
- 5.3.16 Patient Reported Outcomes, 161
- 5.4 Surrogate Outcomes, 162
  - 5.4.1 Surrogate Outcomes Are Disease-Specific, 164
  - 5.4.2 Surrogate Outcomes Can Make Trials More Efficient, 167
  - 5.4.3 Surrogate Outcomes Have Significant Limitations, 168
- 5.5 Summary, 170
- 5.6 Questions for Discussion, 171

## **6 Random Error and Bias 172**

- 6.1 Introduction, 172
  - 6.1.1 The Effects of Random and Systematic Errors Are Distinct, 173
  - 6.1.2 Hypothesis Tests versus Significance Tests, 174
  - 6.1.3 Hypothesis Tests Are Subject to Two Types of Random Error, 175
  - 6.1.4 Type I Errors Are Relatively Easy to Control, 176
  - 6.1.5 The Properties of Confidence Intervals Are Similar to Hypothesis Tests, 176
  - 6.1.6 Using a one- or two-sided hypothesis test is not the right question, 177
  - 6.1.7 *P*-Values Quantify the Type I Error, 178
  - 6.1.8 Type II Errors Depend on the Clinical Difference of Interest, 178
  - 6.1.9 Post Hoc Power Calculations Are Useless, 180
- 6.2 Clinical Bias, 181
  - 6.2.1 Relative Size of Random Error and Bias is Important, 182
  - 6.2.2 Bias Arises from Numerous Sources, 182
  - 6.2.3 Controlling Structural Bias is Conceptually Simple, 185
- 6.3 Statistical Bias, 188
  - 6.3.1 Selection Bias, 188
  - 6.3.2 Some Statistical Bias Can Be Corrected, 192
  - 6.3.3 Unbiasedness is Not the Only Desirable Attribute of an Estimator, 192
- 6.4 Summary, 194
- 6.5 Questions for Discussion, 194

## **7 Statistical Perspectives 196**

- 7.1 Introduction, 196

- 7.2 Differences in Statistical Perspectives, 197
  - 7.2.1 Models and Parameters, 197
  - 7.2.2 Philosophy of Inference Divides Statisticians, 198
  - 7.2.3 Resolution, 199
  - 7.2.4 Points of Agreement, 199
- 7.3 Frequentist, 202
  - 7.3.1 Binomial Case Study, 203
  - 7.3.2 Other Issues, 204
- 7.4 Bayesian, 204
  - 7.4.1 Choice of a Prior Distribution Is a Source of Contention, 205
  - 7.4.2 Binomial Case Study, 206
  - 7.4.3 Bayesian Inference Is Different, 209
- 7.5 Likelihood, 210
  - 7.5.1 Binomial Case Study, 211
  - 7.5.2 Likelihood-Based Design, 211
- 7.6 Statistics Issues, 212
  - 7.6.1 Perspective, 212
  - 7.6.2 Statistical Procedures Are Not Standardized, 213
  - 7.6.3 Practical Controversies Related to Statistics Exist, 214
- 7.7 Summary, 215
- 7.8 Questions for Discussion, 216

## 8 Experiment Design in Clinical Trials

217

- 8.1 Introduction, 217
- 8.2 Trials As Simple Experiment Designs, 218
  - 8.2.1 Design Space Is Chaotic, 219
  - 8.2.2 Design Is Critical for Inference, 220
  - 8.2.3 The Question Drives the Design, 220
  - 8.2.4 Design Depends on the Observation Model As Well As the Biological Question, 221
  - 8.2.5 Comparing Designs, 222
- 8.3 Goals of Experiment Design, 223
  - 8.3.1 Control of Random Error and Bias Is the Goal, 223
  - 8.3.2 Conceptual Simplicity Is Also a Goal, 223
  - 8.3.3 Encapsulation of Subjectivity, 224
  - 8.3.4 Leech Case Study, 225
- 8.4 Design Concepts, 225
  - 8.4.1 The Foundations of Design Are Observation and Theory, 226
  - 8.4.2 A Lesson from the Women's Health Initiative, 227
  - 8.4.3 Experiments Use Three Components of Design, 229

- 8.5 Design Features, 230
  - 8.5.1 Enrichment, 231
  - 8.5.2 Replication, 232
  - 8.5.3 Experimental and Observational Units, 232
  - 8.5.4 Treatments and Factors, 233
  - 8.5.5 Nesting, 233
  - 8.5.6 Randomization, 234
  - 8.5.7 Blocking, 234
  - 8.5.8 Stratification, 235
  - 8.5.9 Masking, 236
- 8.6 Special Design Issues, 237
  - 8.6.1 Placebos, 237
  - 8.6.2 Equivalence and Noninferiority, 240
  - 8.6.3 Randomized Discontinuation, 241
  - 8.6.4 Hybrid Designs May Be Needed for Resolving Special Questions, 242
  - 8.6.5 Clinical Trials Cannot Meet Certain Objectives, 242
- 8.7 Importance of the Protocol Document, 244
  - 8.7.1 Protocols Have Many Functions, 244
  - 8.7.2 Deviations from Protocol Specifications are Common, 245
  - 8.7.3 Protocols Are Structured, Logical, and Complete, 246
- 8.8 Summary, 252
- 8.9 Questions for Discussion, 253

## 9 The Trial Cohort

254

- 9.1 Introduction, 254
- 9.2 Cohort Definition and Selection, 255
  - 9.2.1 Eligibility and Exclusions, 255
  - 9.2.2 Active Sampling and Enrichment, 257
  - 9.2.3 Participation may select subjects with better prognosis, 258
  - 9.2.4 Quantitative Selection Criteria Versus False Precision, 262
  - 9.2.5 Comparative Trials Are Not Sensitive to Selection, 263
- 9.3 Modeling Accrual, 264
  - 9.3.1 Using a Run-In Period, 264
  - 9.3.2 Estimate Accrual Quantitatively, 265
- 9.4 Inclusiveness, Representation, and Interactions, 267
  - 9.4.1 Inclusiveness Is a Worthy Goal, 267
  - 9.4.2 Barriers Can Hinder Trial Participation, 268
  - 9.4.3 Efficacy versus Effectiveness Trials, 269
  - 9.4.4 Representation: Politics Blunders into Science, 270
- 9.5 Summary, 275
- 9.6 Questions for Discussion, 275

**10 Development Paradigms**

- 10.1 Introduction, 277
  - 10.1.1 Stages of Development, 278
  - 10.1.2 Trial Design versus Development Design, 280
  - 10.1.3 Companion Diagnostics in Cancer, 281
- 10.2 Pipeline Principles and Problems, 281
  - 10.2.1 The Paradigm Is Not Linear, 282
  - 10.2.2 Staging Allows Efficiency, 282
  - 10.2.3 The Pipeline Impacts Study Design, 283
  - 10.2.4 Specificity and Pressures Shape the Pipeline, 283
  - 10.2.5 Problems with Trials, 284
  - 10.2.6 Problems in the Pipeline, 286
- 10.3 A Simple Quantitative Pipeline, 286
  - 10.3.1 Pipeline Operating Characteristics Can Be Derived, 286
  - 10.3.2 Implications May Be Counterintuitive, 288
  - 10.3.3 Optimization Yields Insights, 288
  - 10.3.4 Overall Implications for the Pipeline, 291
- 10.4 Late Failures, 292
  - 10.4.1 Generic Mistakes in Evaluating Evidence, 293
  - 10.4.2 “Safety” Begets Efficacy Testing, 293
  - 10.4.3 Pressure to Advance Ideas Is Unprecedented, 294
  - 10.4.4 Scientists Believe Weird Things, 294
  - 10.4.5 Confirmation Bias, 295
  - 10.4.6 Many Biological Endpoints Are Neither Predictive nor Prognostic, 296
  - 10.4.7 Disbelief Is Easier to Suspend Than Belief, 296
  - 10.4.8 Publication Bias, 297
  - 10.4.9 Intellectual Conflicts of Interest, 297
  - 10.4.10 Many Preclinical Models Are Invalid, 298
  - 10.4.11 Variation Despite Genomic Determinism, 299
  - 10.4.12 Weak Evidence Is Likely to Mislead, 300
- 10.5 Summary, 300
- 10.6 Questions for Discussion, 301

**11 Translational Clinical Trials**

- 11.1 Introduction, 302
  - 11.1.1 Therapeutic Intent or Not?, 303
  - 11.1.2 Mechanistic Trials, 304
  - 11.1.3 Marker Threshold Designs Are Strongly Biased, 305

- 11.2 Inferential Paradigms, 308
  - 11.2.1 Biologic Paradigm, 308
  - 11.2.2 Clinical Paradigm, 310
  - 11.2.3 Surrogate Paradigm, 311
- 11.3 Evidence and Theory, 312
  - 11.3.1 Biological Models Are a Key to Translational Trials, 313
- 11.4 Translational Trials Defined, 313
  - 11.4.1 Translational Paradigm, 313
  - 11.4.2 Character and Definition, 315
  - 11.4.3 Small or “Pilot” Does Not Mean Translational, 316
  - 11.4.4 Hypothetical Example, 316
  - 11.4.5 Nesting Translational Studies, 317
- 11.5 Information From Translational Trials, 317
  - 11.5.1 Surprise Can Be Defined Mathematically, 318
  - 11.5.2 Parameter Uncertainty Versus Outcome Uncertainty, 318
  - 11.5.3 Expected Surprise and Entropy, 319
  - 11.5.4 Information/Entropy Calculated From Small Samples Is Biased, 321
  - 11.5.5 Variance of Information/Entropy, 322
  - 11.5.6 Sample Size for Translational Trials, 324
  - 11.5.7 Validity, 327
- 11.6 Summary, 328
- 11.7 Questions for Discussion, 328

## **12 Early Development and Dose-Finding**

**329**

- 12.1 Introduction, 329
- 12.2 Basic Concepts, 330
  - 12.2.1 Therapeutic Intent, 330
  - 12.2.2 Feasibility, 331
  - 12.2.3 Dose versus Efficacy, 332
- 12.3 Essential Concepts for Dose versus Risk, 333
  - 12.3.1 What Does the Terminology Mean?, 333
  - 12.3.2 Distinguish Dose–Risk From Dose–Efficacy, 334
  - 12.3.3 Dose Optimality Is a Design Definition, 335
  - 12.3.4 Unavoidable Subjectivity, 335
  - 12.3.5 Sample Size Is an Outcome of Dose-Finding Studies, 336
  - 12.3.6 Idealized Dose-Finding Design, 336
- 12.4 Dose-Ranging, 338
  - 12.4.1 Some Historical Designs, 338
  - 12.4.2 Typical Dose-Ranging Design, 339

|           |                                                                             |            |
|-----------|-----------------------------------------------------------------------------|------------|
| 12.4.3    | Operating Characteristics Can Be Calculated,                                | 340        |
| 12.4.4    | Modifications, Strengths, and Weaknesses,                                   | 343        |
| 12.5      | Dose-Finding Is Model Based,                                                | 344        |
| 12.5.1    | Mathematical Models Facilitate Inferences,                                  | 345        |
| 12.5.2    | Continual Reassessment Method,                                              | 345        |
| 12.5.3    | Pharmacokinetic Measurements Might Be Used to Improve CRM Dose Escalations, | 349        |
| 12.5.4    | The CRM Is an Attractive Design to Criticize,                               | 350        |
| 12.5.5    | CRM Clinical Examples,                                                      | 350        |
| 12.5.6    | Dose Distributions,                                                         | 351        |
| 12.5.7    | Estimation with Overdose Control (EWOC),                                    | 351        |
| 12.5.8    | Randomization in Early Development?,                                        | 353        |
| 12.5.9    | Phase I Data Have Other Uses,                                               | 353        |
| 12.6      | General Dose-Finding Issues,                                                | 354        |
| 12.6.1    | The General Dose-Finding Problem Is Unsolved,                               | 354        |
| 12.6.2    | More than One Drug,                                                         | 356        |
| 12.6.3    | More than One Outcome,                                                      | 361        |
| 12.6.4    | Envelope Simulation,                                                        | 363        |
| 12.7      | Summary,                                                                    | 366        |
| 12.8      | Questions for Discussion,                                                   | 368        |
| <b>13</b> | <b>Middle Development</b>                                                   | <b>370</b> |
| 13.1      | Introduction,                                                               | 370        |
| 13.1.1    | Estimate Treatment Effects,                                                 | 371        |
| 13.2      | Characteristics of Middle Development,                                      | 372        |
| 13.2.1    | Constraints,                                                                | 373        |
| 13.2.2    | Outcomes,                                                                   | 374        |
| 13.2.3    | Focus,                                                                      | 375        |
| 13.3      | Design Issues,                                                              | 375        |
| 13.3.1    | Choices in Middle Development,                                              | 375        |
| 13.3.2    | When to Skip Middle Development,                                            | 376        |
| 13.3.3    | Randomization,                                                              | 377        |
| 13.3.4    | Other Design Issues,                                                        | 378        |
| 13.4      | Middle Development Distills True Positives,                                 | 379        |
| 13.5      | Futility and Nonsuperiority Designs,                                        | 381        |
| 13.5.1    | Asymmetry in Error Control,                                                 | 382        |
| 13.5.2    | Should We Control False Positives or False Negatives?,                      | 383        |
| 13.5.3    | Futility Design Example,                                                    | 384        |
| 13.5.4    | A Conventional Approach to Futility,                                        | 385        |
| 13.6      | Dose-Efficacy Questions,                                                    | 385        |

- 13.7 Randomized Comparisons, 386
  - 13.7.1 When to Perform an Error-Prone Comparative Trial, 387
  - 13.7.2 Examples, 388
  - 13.7.3 Randomized Selection, 389
- 13.8 Cohort Mixtures, 392
- 13.9 Summary, 395
- 13.10 Questions for Discussion, 396

## **14 Comparative Trials 397**

- 14.1 Introduction, 397
- 14.2 Elements of Reliability, 398
  - 14.2.1 Key Features, 399
  - 14.2.2 Flexibilities, 400
  - 14.2.3 Other Design Issues, 400
- 14.3 Biomarker-Based Comparative Designs, 402
  - 14.3.1 Biomarkers Are Diverse, 402
  - 14.3.2 Enrichment, 404
  - 14.3.3 Biomarker-Stratified, 404
  - 14.3.4 Biomarker-Strategy, 405
  - 14.3.5 Multiple-Biomarker Signal-Finding, 406
  - 14.3.6 Prospective–Retrospective Evaluation of a Biomarker, 407
  - 14.3.7 Master Protocols, 407
- 14.4 Some Special Comparative Designs, 408
  - 14.4.1 Randomized Discontinuation, 408
  - 14.4.2 Delayed Start, 409
  - 14.4.3 Cluster Randomization, 410
  - 14.4.4 Non Inferiority, 410
  - 14.4.5 Multiple Agents versus Control, 410
- 14.5 Summary, 411
- 14.6 Questions for Discussion, 412

## **15 Adaptive Design Features 413**

- 15.1 Introduction, 413
  - 15.1.1 Advantages and Disadvantages of AD, 414
  - 15.1.2 Design Adaptations Are Tools, Not a Class, 416
  - 15.1.3 Perspective on Bayesian Methods, 417
  - 15.1.4 The Pipeline Is the Main Adaptive Tool, 417
- 15.2 Some Familiar Adaptations, 418
  - 15.2.1 Dose-Finding Is Adaptive, 418
  - 15.2.2 Adaptive Randomization, 418
  - 15.2.3 Staging is Adaptive, 422

- 15.2.4 Dropping a Treatment Arm or Subset, 423
- 15.3 Biomarker Adaptive Trials, 423
- 15.4 Re-Designs, 425
  - 15.4.1 Sample Size Re-Estimation Requires Caution, 425
- 15.5 Seamless Designs, 427
- 15.6 Barriers to the Use of AD, 428
- 15.7 Adaptive Design Case Study, 428
- 15.8 Summary, 429
- 15.9 Questions for Discussion, 429

## 16 Sample Size and Power

430

- 16.1 Introduction, 430
- 16.2 Principles, 431
  - 16.2.1 What Is Precision?, 432
  - 16.2.2 What Is Power?, 433
  - 16.2.3 What Is Evidence?, 434
  - 16.2.4 Sample Size and Power Calculations Are Approximations, 435
  - 16.2.5 The Relationship between Power/Precision and Sample Size Is Quadratic, 435
- 16.3 Early Developmental Trials, 436
  - 16.3.1 Translational Trials, 436
  - 16.3.2 Dose-Finding Trials, 437
- 16.4 Simple Estimation Designs, 438
  - 16.4.1 Confidence Intervals for a Mean Provide a Sample Size Approach, 438
  - 16.4.2 Estimating Proportions Accurately, 440
  - 16.4.3 Exact Binomial Confidence Limits Are Helpful, 441
  - 16.4.4 Precision Helps Detect Improvement, 444
  - 16.4.5 Bayesian Binomial Confidence Intervals, 446
  - 16.4.6 A Bayesian Approach Can Use Prior Information, 447
  - 16.4.7 Likelihood-Based Approach for Proportions, 450
- 16.5 Event Rates, 451
  - 16.5.1 Confidence Intervals for Event Rates Can Determine Sample Size, 451
  - 16.5.2 Likelihood-Based Approach for Event Rates, 454
- 16.6 Staged Studies, 455
  - 16.6.1 Ineffective or Unsafe Treatments Should Be Discarded Early, 455
  - 16.6.2 Two-Stage Designs Increase Efficiency, 456
- 16.7 Comparative Trials, 457
  - 16.7.1 How to Choose Type I and II Error Rates?, 459
  - 16.7.2 Comparisons Using the *t*-Test Are a Good Learning Example, 459

- 16.7.3 Likelihood-Based Approach, 462
- 16.7.4 Dichotomous Responses Are More Complex, 463
- 16.7.5 Hazard Comparisons Yield Similar Equations, 464
- 16.7.6 Parametric and Nonparametric Equations Are Connected, 467
- 16.7.7 Accommodating Unbalanced Treatment Assignments, 467
- 16.7.8 A Simple Accrual Model Can Also Be Incorporated, 469
- 16.7.9 Stratification, 471
- 16.7.10 Noninferiority, 472
- 16.8 Expanded Safety Trials, 478
  - 16.8.1 Model Rare Events with the Poisson Distribution, 479
  - 16.8.2 Likelihood Approach for Poisson Rates, 479
- 16.9 Other Considerations, 481
  - 16.9.1 Cluster Randomization Requires Increased Sample Size, 481
  - 16.9.2 Simple Cost Optimization, 482
  - 16.9.3 Increase the Sample Size for Nonadherence, 482
  - 16.9.4 Simulated Lifetables Can Be a Simple Design Tool, 485
  - 16.9.5 Sample Size for Prognostic Factor Studies, 486
  - 16.9.6 Computer Programs Simplify Calculations, 487
  - 16.9.7 Simulation Is a Powerful and Flexible Design Alternative, 487
  - 16.9.8 Power Curves Are Sigmoid Shaped, 488
- 16.10 Summary, 489
- 16.11 Questions for Discussion, 490

## **17 Treatment Allocation**

492

- 17.1 Introduction, 492
  - 17.1.1 Balance and Bias Are Independent, 493
- 17.2 Randomization, 494
  - 17.2.1 Heuristic Proof of the Value of Randomization, 495
  - 17.2.2 Control the Influence of Unknown Factors, 497
  - 17.2.3 Haphazard Assignments Are Not Random, 498
  - 17.2.4 Simple Randomization Can Yield Imbalances, 499
- 17.3 Constrained Randomization, 500
  - 17.3.1 Blocking Improves Balance, 500
  - 17.3.2 Blocking and Stratifying Balances Prognostic Factors, 501
  - 17.3.3 Other Considerations Regarding Blocking, 503
- 17.4 Adaptive Allocation, 504
  - 17.4.1 Urn Designs Also Improve Balance, 504
  - 17.4.2 Minimization Yields Tight Balance, 504
  - 17.4.3 Play the Winner, 505
- 17.5 Other Issues Regarding Randomization, 507

- 17.5.1 Administration of the Randomization, 507
- 17.5.2 Computers Generate Pseudorandom Numbers, 508
- 17.5.3 Randomized Treatment Assignment Justifies Type I Errors, 509
- 17.6 Unequal Treatment Allocation, 514
  - 17.6.1 Subsets May Be of Interest, 514
  - 17.6.2 Treatments May Differ Greatly in Cost, 515
  - 17.6.3 Variances May Be Different, 515
  - 17.6.4 Multiarm Trials May Require Asymmetric Allocation, 516
  - 17.6.5 Generalization, 517
  - 17.6.6 Failed Randomization?, 518
- 17.7 Randomization Before Consent, 519
- 17.8 Summary, 520
- 17.9 Questions for Discussion, 520

## 18 Treatment Effects Monitoring

522

- 18.1 Introduction, 522
  - 18.1.1 Motives for Monitoring, 523
  - 18.1.2 Components of Responsible Monitoring, 524
  - 18.1.3 Trials Can Be Stopped for a Variety of Reasons, 524
  - 18.1.4 There Is Tension in the Decision to Stop, 526
- 18.2 Administrative Issues in Trial Monitoring, 527
  - 18.2.1 Monitoring of Single-Center Studies Relies on Periodic Investigator Reporting, 527
  - 18.2.2 Composition and Organization of the TEMC, 528
  - 18.2.3 Complete Objectivity Is Not Ethical, 535
  - 18.2.4 Independent Experts in Monitoring, 537
- 18.3 Organizational Issues Related to Monitoring, 537
  - 18.3.1 Initial TEMC Meeting, 538
  - 18.3.2 The TEMC Assesses Baseline Comparability, 538
  - 18.3.3 The TEMC Reviews Accrual and Expected Time to Study Completion, 539
  - 18.3.4 Timeliness of Data and Reporting Lags, 539
  - 18.3.5 Data Quality Is a Major Focus of the TEMC, 540
  - 18.3.6 The TEMC Reviews Safety and Toxicity Data, 541
  - 18.3.7 Efficacy Differences Are Assessed by the TEMC, 541
  - 18.3.8 The TEMC Should Address Some Practical Questions Specifically, 541
  - 18.3.9 The TEMC Mechanism Has Potential Weaknesses, 544
- 18.4 Statistical Methods for Monitoring, 545
  - 18.4.1 There Are Several Approaches to Evaluating Incomplete Evidence, 545

- 18.4.2 Monitoring Developmental Trials for Risk, 547
- 18.4.3 Likelihood-Based Methods, 551
- 18.4.4 Bayesian Methods, 557
- 18.4.5 Decision-Theoretic Methods, 559
- 18.4.6 Frequentist Methods, 560
- 18.4.7 Other Monitoring Tools, 566
- 18.4.8 Some Software, 570
- 18.5 Summary, 570
- 18.6 Questions for Discussion, 572

## **19 Counting Subjects and Events**

**573**

- 19.1 Introduction, 573
- 19.2 Imperfection and Validity, 574
- 19.3 Treatment Nonadherence, 575
  - 19.3.1 Intention to Treat Is a Policy of Inclusion, 575
  - 19.3.2 Coronary Drug Project Results Illustrate the Pitfalls of Exclusions Based on Nonadherence, 576
  - 19.3.3 Statistical Studies Support the ITT Approach, 577
  - 19.3.4 Trials Are Tests of Treatment Policy, 577
  - 19.3.5 ITT Analyses Cannot Always Be Applied, 578
  - 19.3.6 Trial Inferences Depend on the Experiment Design, 579
- 19.4 Protocol Nonadherence, 580
  - 19.4.1 Eligibility, 580
  - 19.4.2 Treatment, 581
  - 19.4.3 Defects in Retrospect, 582
- 19.5 Data Imperfections, 583
  - 19.5.1 Evaluability Criteria Are a Methodologic Error, 583
  - 19.5.2 Statistical Methods Can Cope with Some Types of Missing Data, 584
- 19.6 Summary, 588
- 19.7 Questions for Discussion, 589

## **20 Estimating Clinical Effects**

**590**

- 20.1 Introduction, 590
  - 20.1.1 Invisibility Works Against Validity, 591
  - 20.1.2 Structure Aids Internal and External Validity, 591
  - 20.1.3 Estimates of Risk Are Natural and Useful, 592
- 20.2 Dose-Finding and Pharmacokinetic Trials, 594
  - 20.2.1 Pharmacokinetic Models Are Essential for Analyzing DF Trials, 594
  - 20.2.2 A Two-Compartment Model Is Simple but Realistic, 595

- 20.2.3 PK Models Are Used By “Model Fitting”, 598
- 20.3 Middle Development Studies, 599
  - 20.3.1 Mesothelioma Clinical Trial Example, 599
  - 20.3.2 Summarize Risk for Dichotomous Factors, 600
  - 20.3.3 Nonparametric Estimates of Survival Are Robust, 601
  - 20.3.4 Parametric (Exponential) Summaries of Survival Are Efficient, 603
  - 20.3.5 Percent Change and Waterfall Plots, 605
- 20.4 Randomized Comparative Trials, 606
  - 20.4.1 Examples of Comparative Trials Used in This Section, 607
  - 20.4.2 Continuous Measures Estimate Treatment Differences, 608
  - 20.4.3 Baseline Measurements Can Increase Precision, 609
  - 20.4.4 Comparing Counts, 610
  - 20.4.5 Nonparametric Survival Comparisons, 612
  - 20.4.6 Risk (Hazard) Ratios and Confidence Intervals Are Clinically Useful Data Summaries, 614
  - 20.4.7 Statistical Models Are Necessary Tools, 615
- 20.5 Problems With *P*-Values, 616
  - 20.5.1 *P*-Values Do Not Represent Treatment Effects, 618
  - 20.5.2 *P*-Values Do Not Imply Reproducibility, 618
  - 20.5.3 *P*-Values Do Not Measure Evidence, 619
- 20.6 Strength of Evidence Through Support Intervals, 620
  - 20.6.1 Support Intervals Are Based on the Likelihood Function, 620
  - 20.6.2 Support Intervals Can Be Used with Any Outcome, 621
- 20.7 Special Methods of Analysis, 622
  - 20.7.1 The Bootstrap Is Based on Resampling, 623
  - 20.7.2 Some Clinical Questions Require Other Special Methods of Analysis, 623
- 20.8 Exploratory Analyses, 628
  - 20.8.1 Clinical Trial Data Lend Themselves to Exploratory Analyses, 628
  - 20.8.2 Multiple Tests Multiply Type I Errors, 629
  - 20.8.3 Kinds of Multiplicity, 630
  - 20.8.4 Inevitable Risks from Subgroups, 630
  - 20.8.5 Tale of a Subset Analysis Gone Wrong, 632
  - 20.8.6 Perspective on Subgroup Analyses, 635
  - 20.8.7 Effects the Trial Was Not Designed to Detect, 636
  - 20.8.8 Safety Signals, 637
  - 20.8.9 Subsets, 637
  - 20.8.10 Interactions, 638
- 20.9 Summary, 639
- 20.10 Questions for Discussion, 640

## 21 Prognostic Factor Analyses 644

- 21.1 Introduction, 644
  - 21.1.1 Studying Prognostic Factors is Broadly Useful, 645
  - 21.1.2 Prognostic Factors Can Be Constant or Time-Varying, 646
- 21.2 Model-Based Methods, 647
  - 21.2.1 Models Combine Theory and Data, 647
  - 21.2.2 Scale and Coding May Be Important, 648
  - 21.2.3 Use Flexible Covariate Models, 648
  - 21.2.4 Building Parsimonious Models Is the Next Step, 650
  - 21.2.5 Incompletely Specified Models May Yield Biased Estimates, 655
  - 21.2.6 Study Second-Order Effects (Interactions), 656
  - 21.2.7 PFAs Can Help Describe Risk Groups, 656
  - 21.2.8 Power and Sample Size for PFAs, 660
- 21.3 Adjusted Analyses of Comparative Trials, 661
  - 21.3.1 What Should We Adjust For?, 662
  - 21.3.2 What Can Happen?, 663
  - 21.3.3 Brain Tumor Case Study, 664
- 21.4 PFAs Without Models, 666
  - 21.4.1 Recursive Partitioning Uses Dichotomies, 666
  - 21.4.2 Neural Networks Are Used for Pattern Recognition, 667
- 21.5 Summary, 669
- 21.6 Questions for Discussion, 669

## 22 Factorial Designs 671

- 22.1 Introduction, 671
- 22.2 Characteristics of Factorial Designs, 672
  - 22.2.1 Interactions or Efficiency, But Not Both Simultaneously, 672
  - 22.2.2 Factorial Designs Are Defined by Their Structure, 672
  - 22.2.3 Factorial Designs Can Be Made Efficient, 674
- 22.3 Treatment Interactions, 675
  - 22.3.1 Factorial Designs Are the Only Way to Study Interactions, 675
  - 22.3.2 Interactions Depend on the Scale of Measurement, 677
  - 22.3.3 The Interpretation of Main Effects Depends on Interactions, 677
  - 22.3.4 Analyses Can Employ Linear Models, 678
- 22.4 Examples of Factorial Designs, 680
- 22.5 Partial, Fractional, and Incomplete Factorials, 682
  - 22.5.1 Use Partial Factorial Designs When Interactions Are Absent, 682
  - 22.5.2 Incomplete Designs Present Special Problems, 682
- 22.6 Summary, 683
- 22.7 Questions for Discussion, 683

**23 Crossover Designs**

- 23.1 Introduction, 684
  - 23.1.1 Other Ways of Giving Multiple Treatments Are Not Crossovers, 685
  - 23.1.2 Treatment Periods May Be Randomly Assigned, 686
- 23.2 Advantages and Disadvantages, 686
  - 23.2.1 Crossover Designs Can Increase Precision, 687
  - 23.2.2 A Crossover Design Might Improve Recruitment, 687
  - 23.2.3 Carryover Effects Are a Potential Problem, 688
  - 23.2.4 Dropouts Have Strong Effects, 689
  - 23.2.5 Analysis is More Complex Than for a Parallel-Group Design, 689
  - 23.2.6 Prerequisites Are Needed to Apply Crossover Designs, 689
  - 23.2.7 Other Uses for the Design, 690
- 23.3 Analysis, 691
  - 23.3.1 Simple Approaches, 691
  - 23.3.2 Analysis Can Be Based on a Cell Means Model, 692
  - 23.3.3 Other Issues in Analysis, 696
- 23.4 Classic Case Study, 696
- 23.5 Summary, 696
- 23.6 Questions for Discussion, 697

**24 Meta-Analyses**

- 24.1 Introduction, 698
  - 24.1.1 Meta-Analyses Formalize Synthesis and Increase Precision, 699
- 24.2 A Sketch of Meta-Analysis Methods, 700
  - 24.2.1 Meta-Analysis Necessitates Prerequisites, 700
  - 24.2.2 Many Studies Are Potentially Relevant, 701
  - 24.2.3 Select Studies, 702
  - 24.2.4 Plan the Statistical Analysis, 703
  - 24.2.5 Summarize the Data Using Observed and Expected, 703
- 24.3 Other Issues, 705
  - 24.3.1 Cumulative Meta-Analyses, 705
  - 24.3.2 Meta-Analyses Have Practical and Theoretical Limitations, 706
  - 24.3.3 Meta-Analysis Has Taught Useful Lessons, 707
- 24.4 Summary, 707
- 24.5 Questions for Discussion, 708

**25 Reporting and Authorship**

- 25.1 Introduction, 709

- 25.2 General Issues in Reporting, 710
  - 25.2.1 Uniformity Improves Comprehension, 711
  - 25.2.2 Quality of the Literature, 712
  - 25.2.3 Peer Review Is the Only Game in Town, 712
  - 25.2.4 Publication Bias Can Distort Impressions Based on the Literature, 713
- 25.3 Clinical Trial Reports, 715
  - 25.3.1 General Considerations, 716
  - 25.3.2 Employ a Complete Outline for Comparative Trial Reporting, 721
- 25.4 Authorship, 726
  - 25.4.1 Inclusion and Ordering, 727
  - 25.4.2 Responsibility of Authorship, 727
  - 25.4.3 Authorship Models, 728
  - 25.4.4 Some Other Practicalities, 730
- 25.5 Other Issues in Disseminating Results, 731
  - 25.5.1 Open Access, 731
  - 25.5.2 Clinical Alerts, 731
  - 25.5.3 Retractions, 732
- 25.6 Summary, 732
- 25.7 Questions for Discussion, 733

## **26 Misconduct and Fraud in Clinical Research**

734

- 26.1 Introduction, 734
  - 26.1.1 Integrity and Accountability Are Critically Important, 736
  - 26.1.2 Fraud and Misconduct Are Difficult to Define, 738
- 26.2 Research Practices, 741
  - 26.2.1 Misconduct May Be Increasing in Frequency, 741
  - 26.2.2 Causes of Misconduct, 742
- 26.3 Approach to Allegations of Misconduct, 743
  - 26.3.1 Institutions, 744
  - 26.3.2 Problem Areas, 746
- 26.4 Characteristics of Some Misconduct Cases, 747
  - 26.4.1 Darsee Case, 747
  - 26.4.2 Poisson (NSABP) Case, 749
  - 26.4.3 Two Recent Cases from Germany, 752
  - 26.4.4 Fiddes Case, 753
  - 26.4.5 Potti Case, 754
- 26.5 Lessons, 754
  - 26.5.1 Recognizing Fraud or Misconduct, 754
  - 26.5.2 Misconduct Cases Yield Other Lessons, 756

- 26.6 Clinical Investigators' Responsibilities, 757
  - 26.6.1 General Responsibilities, 757
  - 26.6.2 Additional Responsibilities Related to INDs, 758
  - 26.6.3 Sponsor Responsibilities, 759
- 26.7 Summary, 759
- 26.8 Questions for Discussion, 760

## Appendix A Data and Programs

761

- A.1 Introduction, 761
- A.2 Design Programs, 761
  - A.2.1 Power and Sample Size Program, 761
  - A.2.2 Blocked Stratified Randomization, 763
  - A.2.3 Continual Reassessment Method, 763
  - A.2.4 Envelope Simulation, 763
- A.3 *Mathematica* Code, 763

## Appendix B Abbreviations

764

## Appendix C Notation and Terminology

769

- C.1 Introduction, 769
- C.2 Notation, 769
  - C.2.1 Greek Letters, 770
  - C.2.2 Roman Letters, 771
  - C.2.3 Other Symbols, 772
- C.3 Terminology and Concepts, 772

## Appendix D Nuremberg Code

788

- D.1 Permissible Medical Experiments, 788

## References

790

## Index

871