

Contents

Preface	xxi
Acknowledgments	xxv
Contributors	xxvii
PART I CLINICAL TRIALS	1
1. Phase I Clinical Trials	3
<i>Anastasis Ivanova and Nancy Flournoy</i>	
1.1 Introduction, 3	
1.2 Phase I Trials in Healthy Volunteers, 3	
1.3 Phase I Trials with Toxic Outcomes Enrolling Patients, 5	
1.3.1 Parametric versus Nonparametric designs, 6	
1.3.2 Markovian-Motivated Up-and-Down Designs, 6	
1.3.3 Isotonic Designs, 9	
1.3.4 Bayesian Designs, 9	
1.3.5 Time-to-Event Design Modifications, 10	
1.4 Other Design Problems in Dose Finding, 11	
1.5 Concluding Remarks, 12	
2. Phase II Clinical Trials	15
<i>Nigel Stallard</i>	
2.1 Introduction, 15	
2.1.1 Background, 15	
2.1.2 The Role of Phase II Clinical Trials in Clinical Evaluation of a Novel Therapeutic Agent, 16	
2.1.3 Phase II Clinical Trial Designs, 17	
2.2 Frequentist Methods in Phase II Clinical Trials, 18	

- 2.2.1 Review of Frequentist Methods and Their Applications in Phase II Clinical Trials, 18
- 2.2.2 Frequentist Methods for Single-Treatment Pilot Studies, 19
- 2.2.3 Frequentist Methods for Comparative Studies, 21
- 2.2.4 Frequentist Methods for Screening Studies, 22
- 2.3 Bayesian Methods in Phase II Clinical Trials, 22
 - 2.3.1 Review of Bayesian Methods and Their Application in Phase II Clinical Trials, 22
 - 2.3.2 Bayesian Methods for Single-Treatment Pilot Studies, Comparative Studies and Selection Screens, 24
- 2.4 Decision-Theoretic Methods in Phase II Clinical Trials, 25
- 2.5 Analysis of Multiple Endpoints in Phase II Clinical Trials, 26
- 2.6 Outstanding Issues in Phase II Clinical Trials, 27

3. Response-Adaptive Designs in Phase III Clinical Trials

33

Atanu Biswas, Uttam Bandyopadhyay, and Rahul Bhattacharya

- 3.1 Introduction, 33
- 3.2 Adaptive Designs for Binary Treatment Responses, 34
 - 3.2.1 Play-the-Winner Design, 34
 - 3.2.2 Randomized Play-the-Winner Design, 34
 - 3.2.3 Generalized Pólya's Urn (GPU), 35
 - 3.2.4 Randomized Pólya Urn Design, 36
 - 3.2.5 Birth-and-Death Urn Design, 37
 - 3.2.6 Birth-and-Death Urn with Immigration Design, 37
 - 3.2.7 Drop-the-Loser Urn Design, 37
 - 3.2.8 Sequential Estimation-Adjusted Urn Design, 38
 - 3.2.9 Doubly Adaptive Biased Coin Design, 39
- 3.3 Adaptive Designs for Binary Treatment Responses Incorporating Covariates, 40
 - 3.3.1 Covariate-Adaptive Randomized Play-the-Winner Design, 40
 - 3.3.2 Treatment Effect Mappings, 41
 - 3.3.3 Drop-the-Loser Design with Covariate, 41
- 3.4 Adaptive Designs for Categorical Responses, 41
- 3.5 Adaptive Designs for Continuous Responses, 42
 - 3.5.1 Nonparametric-Score-Based Allocation Designs, 42
 - 3.5.2 Link-Function-Based Allocation Designs, 43
 - 3.5.3 Continuous Drop-the-Loser Design, 43
- 3.6 Optimal Adaptive Designs, 43
- 3.7 Delayed Responses in Adaptive Designs, 44
- 3.8 Biased Coin Designs, 45
- 3.9 Real Adaptive Clinical Trials, 45

3.10	Data Study for Different Adaptive Schemes, 46	
3.10.1	Fluoxetine Trial, 46	
3.10.2	Pregabalin Trial, 47	
3.10.3	Simulated Trial, 48	
3.11	Concluding Remarks, 49	
4.	Inverse Sampling for Clinical Trials: A Brief Review of Theory and Practice	55
	<i>Atanu Biswas and Uttam Bandyopadhyay</i>	
4.1	Introduction, 55	
4.1.1	Inverse Binomial Sampling, 56	
4.1.2	Partial Sequential Sampling, 58	
4.2	Two-Sample Randomized Inverse Sampling for Clinical Trials, 59	
4.2.1	Use of Mann–Whitney Statistics, 59	
4.2.2	Fixed-Width Confidence Interval Estimation, 60	
4.2.3	Fixed-Width Confidence Interval for Partial Sequential Sampling, 61	
4.3	An Example of Inverse Sampling: Boston ECMO, 62	
4.4	Inverse Sampling in Adaptive Designs, 62	
4.5	Concluding Remarks, 63	
5.	The Design and Analysis Aspects of Cluster Randomized Trials	67
	<i>Hrishikesh Chakraborty</i>	
5.1	Introduction: Cluster Randomized Trials, 67	
5.2	Intracluster Correlation Coefficient and Confidence Interval, 69	
5.3	Sample Size Calculation for Cluster Randomized Trials, 71	
5.4	Analysis of Cluster Randomized Trial Data, 73	
5.5	Concluding Remarks, 75	
PART II	EPIDEMIOLOGY	81
6.	HIV Dynamics Modeling and Prediction of Clinical Outcomes in AIDS Clinical Research	83
	<i>Yangxin Huang and Hulin Wu</i>	
6.1	Introduction, 83	
6.2	HIV Dynamic Model and Treatment Effect Models, 84	
6.2.1	HIV Dynamic Model, 84	
6.2.2	Treatment Effect Models, 85	
6.3	Statistical Methods for Predictions of Clinical Outcomes, 87	
6.3.1	Bayesian Nonlinear Mixed-Effects Model, 87	
6.3.2	Predictions Using the Bayesian Mixed-Effects Modeling Approach, 89	

- 6.4 Simulation Study, 90
- 6.5 Clinical Data Analysis, 91
- 6.6 Concluding remarks, 92

7. Spatial Epidemiology **97**

Lance A. Waller

- 7.1 Space and Disease, 97
- 7.2 Basic Spatial Questions and Related Data, 98
- 7.3 Quantifying Pattern in Point Data, 99
- 7.4 Predicting Spatial Observations, 107
- 7.5 Concluding Remarks, 118

8. Modeling Disease Dynamics: Cholera as a Case Study **123**

Edward L. Ionides, Carles Bretó, and Aaron A. King

- 8.1 Introduction, 123
- 8.2 Data Analysis via Population Models, 124
- 8.3 Sequential Monte Carlo, 126
- 8.4 Modeling Cholera, 130
 - 8.4.1 Fitting Structural Models to Cholera Data, 132
- 8.5 Concluding Remarks, 136

9. Misclassification and Measurement Error Models in Epidemiologic Studies **141**

Surupa Roy and Tathagata Banerjee

- 9.1 Introduction, 141
- 9.2 A Few Examples, 143
 - 9.2.1 Atom Bomb Survivors Data, 143
 - 9.2.2 Coalminers Data, 143
 - 9.2.3 Effect of Maternal Dietary Habits on Low Birth Weight in Babies, 143
- 9.3 Binary Regression Models with Two Types of Error, 144
- 9.4 Bivariate Binary Regression Models with Two Types of Error, 146
- 9.5 Models for Analyzing Mixed Misclassified Binary and Continuous Responses, 149
- 9.6 Atom Bomb Data Analysis, 151
- 9.7 Concluding Remarks, 152

PART III SURVIVAL ANALYSIS **157**

10. Semiparametric Maximum-Likelihood Inference in Survival Analysis **159**

Michael R. Kosorok

10.1	Introduction, 159	
10.2	Examples of Survival Models, 160	
10.3	Basic Estimation and Limit Theory, 162	
10.4	The Bootstrap, 163	
10.4.1	The Regular Case, 165	
10.4.2	When Slowly Converging Nuisance Parameters are Present, 166	
10.5	The Profile Sampler, 166	
10.6	The Piggyback Bootstrap, 168	
10.7	Other Approaches, 170	
10.8	Concluding Remarks, 171	
11.	An Overview of the Semi-Competing Risks Problem	177
	<i>Limin Peng, Hongyu Jiang, Rick J. Chappell, and Jason P. Fine</i>	
11.1	Introduction, 177	
11.2	Nonparametric Inferences, 179	
11.3	Semiparametric One-Sample Inference, 181	
11.4	Semiparametric Regression Method, 184	
11.4.1	Functional Regression Modeling, 185	
11.4.2	A Bivariate Accelerated Lifetime Model, 187	
11.5	Concluding Remarks, 189	
12.	Tests for Time-Varying Covariate Effects within Aalen's Additive Hazards Model	193
	<i>Torben Martinussen and Thomas H. Scheike</i>	
12.1	Introduction, 193	
12.2	Model Specification and Inferential Procedures, 194	
12.2.1	A Pseudo-Score Test, 197	
12.3	Numerical Results, 199	
12.3.1	Simulation Studies, 199	
12.3.2	Trace Data, 202	
12.4	Concluding Remarks, 204	
12.5	Summary, 204	
	Appendix 12A, 205	
13.	Analysis of Outcomes Subject to Induced Dependent Censoring: A Marked Point Process Perspective	209
	<i>Yijian Huang</i>	
13.1	Introduction, 209	
13.2	Induced Dependent Censoring and Associated Identifiability Issues, 210	
13.3	Marked Point Process, 212	

13.3.1	Hazard Functions with Marked Point Process, 212	
13.3.2	Identifiability, 213	
13.3.3	Nonparametric Estimation, 213	
13.3.4	Martingales, 214	
13.4	Modeling Strategy for Testing and Regression, 215	
13.4.1	Two-Sample Test for Lifetime Utility or Cost, 216	
13.4.2	Calibration Regression for Lifetime Medical Cost, 216	
13.4.3	Two-Sample Multistate Accelerated Sojourn-Time Model, 217	
13.5	Concluding Remarks, 218	
14.	Analysis of Dependence in Multivariate Failure-Time Data	221
	<i>Li Hsu and Zoe Moodie</i>	
14.1	Introduction, 221	
14.2	Nonparametric Bivariate Survivor Function Estimation, 223	
14.2.1	Path-Dependent Estimators, 223	
14.2.2	Inverse Censoring Probability Weighted Estimators, 223	
14.2.3	NPMLE-Type Estimators, 224	
14.2.4	Data Application to Danish Twin Data, 226	
14.3	Non- and Semiparametric Estimation of Dependence Measures, 230	
14.3.1	Nonparametric Dependence Estimation, 230	
14.3.2	Semiparametric Dependence Estimation, 232	
14.3.3	An Application to a Case–Control Family Study of Breast Cancer, 236	
14.4	Concluding Remarks, 239	
15.	Robust Estimation for Analyzing Recurrent-Event Data in the Presence of Terminal Events	245
	<i>Rajeshwari Sundaram</i>	
15.1	Introduction, 245	
15.2	Inference Procedures, 247	
15.2.1	Estimation in the Presence of Only Independent Censoring (with All Censoring Variables Observable), 247	
15.2.2	Estimation in the Presence of Terminal Events, 248	
15.3	Large-Sample Properties, 249	
15.4	Numerical Results, 252	
15.4.1	Simulation Studies, 252	
15.4.2	rhDNase Data, 257	
15.4.3	Bladder Tumor Data, 259	
15.5	Concluding Remarks, 259	
	Appendix 15A, 260	

16. Tree-Based Methods for Survival Data **265***Mousumi Banerjee and Anne-Michelle Noone*

- 16.1 Introduction, 265
- 16.2 Review of CART, 266
- 16.3 Trees for Survival Data, 268
 - 16.3.1 Methods Based on Measure of Within-Node Homogeneity, 268
 - 16.3.2 Methods Based on Between-Node Separation, 271
 - 16.3.3 Pruning and Tree Selection, 272
- 16.4 Simulations for Comparison of Different Splitting Methods, 272
- 16.5 Example: Breast Cancer Prognostic Study, 274
- 16.6 Random Forest for Survival Data, 278
 - 16.6.1 Breast Cancer Study: Results from Random Forest Analysis, 280
- 16.7 Concluding Remarks, 281

17. Bayesian Estimation of the Hazard Function with Randomly Right-Censored Data **287***Jean-François Angers and Brenda MacGibbon*

- 17.1 Introduction, 287
 - 17.1.1 The Random Right-Censorship Model, 289
 - 17.1.2 The Bayesian Model, 290
- 17.2 Bayesian Functional Model Using Monotone Wavelet Approximation, 292
- 17.3 Estimation of the Subdensity F^* , 295
- 17.4 Simulations, 296
- 17.5 Examples, 298
- 17.6 Concluding Remarks, 300
- Appendix 17A, 301

PART IV BIOINFORMATICS **307****18. The Effects of Intergene Associations on Statistical Inferences from Microarray Data** **309***Kerby Shedden*

- 18.1 Introduction, 309
- 18.2 Intergene Correlation, 310
- 18.3 Differential Expression, 314
- 18.4 Timecourse Experiments, 315
- 18.5 Meta-Analysis, 319
- 18.6 Concluding Remarks, 321

19. A Comparison of Methods for Meta-Analysis of Gene Expression Data	325
<i>Hyungwon Choi and Debashis Ghosh</i>	
19.1 Introduction, 325	
19.2 Background, 326	
19.2.1 Technology Details and Gene Identification, 326	
19.2.2 Analysis Methods, 327	
19.3 Example, 328	
19.4 Cross-Comparison of Gene Signatures, 329	
19.5 Best Common Mean Difference Method, 329	
19.6 Effect Size Method, 331	
19.7 POE Assimilation Method, 332	
19.8 Comparison of Three Methods, 334	
19.8.1 Signatures, 335	
19.8.2 Classification Performance, 335	
19.8.3 Directionality of Differential Expression, 336	
19.9 Conclusions, 336	
20. Statistical Methods for Identifying Differentially Expressed Genes in Replicated Microarray Experiments: A Review	341
<i>Lynn kuo, Fang Yu, and Yifang Zhao</i>	
20.1 Introduction, 341	
20.2 Normalization, 344	
20.3 Methods for Selecting Differentially Expressed Genes, 349	
20.3.1 BH-T, 350	
20.3.2 SAM, 351	
20.3.3 SPH, 352	
20.3.4 LIMMA, 354	
20.3.5 MAANOVA, 355	
20.4 Simulation Study, 357	
20.4.1 Results of Simulation Studies, 359	
20.4.2 Other Considerations, 360	
20.5 Concluding Remarks, 360	
21. Clustering of Microarray Data via Mixture Models	365
<i>Geoffrey J. McLachlan, Richard W. Bean, and Angus Ng</i>	
21.1 Introduction, 365	
21.2 Clustering of Microarray Data, 367	
21.3 Notation, 367	
21.4 Clustering of Tissue Samples, 369	

- 21.5 The EMMIX-GENE Clustering Procedure, 370
 - 21.5.1 Step 1. Screening of Genes, 370
 - 21.5.2 Step 2. Clustering of Genes: Formation of Metagenes, 371
 - 21.5.3 Step 3. Clustering of Tissues, 372
- 21.6 Clustering of Gene Profiles, 372
- 21.7 EMMIX-WIRE, 373
- 21.8 Maximum-Likelihood Estimation via the EM Algorithm, 374
- 21.9 Model Selection, 376
- 21.10 Example: Clustering of Timecourse Data, 377
- 21.11 Concluding Remarks, 379

22. Censored Data Regression in High-Dimensional and Low-Sample-Size Settings for Genomic Applications

385

Hongzhe Li

- 22.1 Introduction, 385
- 22.2 Censored Data Regression Models, 386
 - 22.2.1 The Cox Proportional Hazards Model, 386
 - 22.2.2 Accelerated Failure-Time Model, 387
 - 22.2.3 Additive Hazard Regression Models, 388
- 22.3 Regularized Estimation for Censored Data Regression Models, 388
 - 22.3.1 L_2 Penalized Estimation of the Cox Model Using Kernels, 389
 - 22.3.2 L_1 Penalized Estimation of the Cox Model Using Least-Angle Regression, 390
 - 22.3.3 Threshold Gradient Descent Procedure for the Cox Model, 391
 - 22.3.4 Regularized Buckley–James Estimation for the AFT Model, 391
 - 22.3.5 Regularization Based on Inverse Probability of Censoring Weighted Loss Function for the AFT Model, 392
 - 22.3.6 Penalized Estimation for the Additive Hazard Model, 393
 - 22.3.7 Use of Other Penalty Functions, 394
- 22.4 Survival Ensemble Methods, 394
 - 22.4.1 The Smoothing-Spline-Based Boosting Algorithm for the Nonparametric Additive Cox Model, 394
 - 22.4.2 Random Forests and Gradient Boosting Procedure for the AFT Model, 395
- 22.5 Nonparametric-Pathway-Based Regression Models, 395
- 22.6 Dimension-Reduction-Based Methods and Bayesian Variable Selection Methods, 396
- 22.7 Criteria for Evaluating Different Procedures, 397
- 22.8 Application to a Real Dataset and Comparisons, 397
- 22.9 Discussion and Future Research Topics, 398

22.9.1	Test of Treatment Effect Adjusting for High-Dimensional Genomic Data, 399	
22.9.2	Development of Flexible Models for Gene–Gene and Gene–Environment Interactions, 399	
22.9.3	Methods for Other Types of Genomic Data, 400	
22.9.4	Development of Pathway- and Network-Based Regression Models for Censored Survival Phenotypes, 400	
22.10	Concluding Remarks, 400	
23.	Analysis of Case–Control Studies in Genetic Epidemiology	405
	<i>Nilanjan Chatterjee</i>	
23.1	Introduction, 405	
23.2	Maximum-Likelihood Analysis of Case–Control Data with Complete Information, 406	
23.2.1	Background, 407	
23.2.2	Maximum-Likelihood Estimation under HWE and Gene–Environment Independence, 408	
23.2.3	An Example, 410	
23.3	Haplotype-based Genetic Analysis with Missing Phase Information, 410	
23.3.1	Background, 410	
23.3.2	Methods, 412	
23.3.3	Application, 414	
23.4	Concluding Remarks, 415	
24.	Assessing Network Structure in the Presence of Measurement Error	419
	<i>Denise Scholtens, Raji Balasubramanian, and Robert Gentleman</i>	
24.1	Introduction, 419	
24.2	Graphs of Biological Data, 420	
24.2.1	Integrating Multiple Data Types, 421	
24.3	Statistics on Graphs, 421	
24.4	Graph-Theoretic Models, 422	
24.5	Types of Measurement Error, 425	
24.5.1	Stochastic Error, 426	
24.5.2	Systematic Error, 426	
24.5.3	Sampling, 426	
24.6	Exploratory Data Analysis, 426	
24.6.1	Reciprocity, 427	
24.6.2	Sampling, 427	
24.6.3	Underlying Network Structure, 428	
24.7	Influence of Measurement Error on Graph Statistics, 429	
24.7.1	Path Length: L , 429	
24.7.2	Clustering Coefficient: C , 431	
24.7.3	Node Degree Distribution, 436	
24.8	Biological Implications, 436	

- 24.8.1 Experimental Data, 436
- 24.8.2 Simulation Data, 439
- 24.9 Conclusions, 439

25. Prediction of RNA Splicing Signals **443**

Mark R. Segal

- 25.1 Introduction, 443
 - 25.1.1 Biologic Overview of Splicing, 444
- 25.2 Existing Approaches to Splice Site Identification, 445
 - 25.2.1 Maximum-Entropy Models, 445
 - 25.2.2 Permuted Variable-Length Markov Models, 447
 - 25.2.3 Bayesian Network Approaches, 448
- 25.3 Splice Site Recognition via Contemporary Classifiers, 450
 - 25.3.1 Random Forests, 451
 - 25.3.2 Support Vector Machines, 454
 - 25.3.3 Boosting, 454
- 25.4 Results, 455
 - 25.4.1 Data Generation, 455
 - 25.4.2 Predictive Performance, 455
 - 25.4.3 Interpretational Yield, 457
 - 25.4.4 Computational Considerations, 458
- 25.5 Concluding Remarks, 459

26. Statistical Methods for Biomarker Discovery Using Mass Spectrometry **465**

Bradley M. Broom and Kim-Anh Do

- 26.1. Introduction, 465
 - 26.1.1 Sample Ionization, 467
 - 26.1.2 Mass Analysis, 468
- 26.2 Biomarker Discovery, 470
- 26.3 Statistical Methods for Preprocessing, 470
- 26.4 Statistical Methods for Multiple Testing, Classification, and Applications, 473
 - 26.4.1 Multiple Testing and Identification of Differentially Expressed Peaks, 473
 - 26.4.2 A Peak Probability Contrast (PPC) Procedure for Sample Classification, 474
 - 26.4.3 A Semiparametric Model for Protein Mass Spectroscopy, 474
 - 26.4.4 Smoothed Principal-Component Analysis (PCA) for Proteomic Spectra, 476
 - 26.4.5 Wavelet-Based Functional Mixed Model and Application, 477

26.4.6	A Nonparametric Bayesian Model Based on Kernel Functions,	479
26.5	Potential Statistical Developments,	481
26.6	Concluding Remarks,	483
27.	Genetic Mapping of Quantitative Traits: Model-Free Sib-Pair Linkage Approaches	487
	<i>Saurabh Ghosh and Partha P. Majumder</i>	
27.1	Introduction,	487
27.2	The Basic QTL Framework For Sib-Pairs,	488
27.3	The Haseman–Elston Regression Framework,	489
27.4	Nonparametric Alternatives,	489
27.5	The Modified Nonparametric Regression,	490
27.5.1	Evaluation of Significance Levels,	491
27.6	Comparison With Linear Regression Methods,	492
27.7	Significance Levels and Empirical Power,	493
27.8	An Application to Real Data,	495
27.9	Concluding Remarks,	496
PART V	MISCELLANEOUS TOPICS	499
28.	Robustness Issues in Biomedical Studies	501
	<i>Ayanendranath Basu</i>	
28.1	Introduction: The Need for Robust Procedures,	501
28.2	Standard Tools for Robustness,	502
28.2.1	M-Estimators,	503
28.2.2	Influence Function,	503
28.2.3	Breakdown Point,	504
28.2.4	Basic Miscellaneous Procedures,	504
28.2.5	Alternative Approaches,	505
28.3	The Robustness Question in Biomedical Studies,	506
28.4	Robust Estimation in the Logistic Regression Model,	508
28.5	Robust Estimation for Censored Survival Data,	513
28.6	Adaptive Robust Methods in Clinical Trials,	518
28.7	Concluding Remarks,	521
29.	Recent Advances in the Analysis of Episodic Hormone Data	527
	<i>Timothy D. Johnson and Yuedong Wang</i>	
29.1	Introduction,	527
29.2	A General Biophysical Model,	530
29.3	Bayesian deconvolution model (BDM),	531

29.3.1	Posterior Processing, 534	
29.3.2	An Example, 536	
29.4	Nonlinear Mixed-Effects Partial-Splines Models, 537	
29.5	Concluding Remarks, 542	
30.	Models for Carcinogenesis	547
	<i>Anup Dewanji</i>	
30.1	Introduction, 547	
30.2	Statistical Models, 549	
30.3	Multistage Models, 552	
30.4	Two-Stage Clonal Expansion Model, 555	
30.5	Physiologically Based Pharmacokinetic Models, 560	
30.6	Statistical Methods, 562	
30.7	Concluding Remarks, 564	
	Index	569

