

# Epidemiology and Medical Statistics

Edited by

**C.R. Rao**

Center for Multivariate Analysis  
Department of Statistics, The Pennsylvania State University  
University Park, PA, USA

**J.P. Miller**

Division of Biostatistics  
School of Medicine, Washington University in St. Louis  
St. Louis, MO, USA

**D.C. Rao**

Division of Biostatistics  
School of Medicine, Washington University in St. Louis  
St. Louis, MO, USA

**C**

Amsterdam • Boston • Heidelberg • London • New York • Oxford  
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo  
North-Holland is an imprint of Elsevier

**ELSEVIER**

# Table of contents

Preface    xiii

Contributors    xv

Ch. 1.    Statistical Methods and Challenges in Epidemiology and  
          Biomedical Research    I  
          *Ross L. Prentice*

- 1. Introduction    1
- 2. Characterizing the study cohort    3
- 3. Observational study methods and challenges    6
- 4. Randomized controlled trials    12
- 5. Intermediate, surrogate, and auxiliary outcomes    17
- 6. Multiple testing issues and high-dimensional biomarkers    18
- 7. Further discussion and the Women's Health Initiative example    21
- References    22

Ch. 2.    Statistical Inference for Causal Effects, With Emphasis on  
          Applications in Epidemiology and Medical Statistics    2f  
          *Donald B. Rubin*

- 1. Causal inference primitives    28
- 2. The assignment mechanism    36
- 3. Assignment-based modes of causal inference    41
- 4. Posterior predictive causal inference    47
- 5. Complications    55
- References    58

Ch. 3.    Epidemiologic Study Designs    64  
          *Kenneth J. Rothman, Sander Greenland and Timothy L. Lash*

- 1. Introduction    64
- 2. Experimental studies    65
- 3. Nonexperimental studies    73
- 4. Cohort studies    73

5. Case-control studies	84
6. Variants of the case-control design	97
7. Conclusion	104
References	104

Ch. 4. Statistical Methods for Assessing Biomarkers and Analyzing Biomarker Data 109

*Stephen W. Looney and Joseph L. Hagan*

1. Introduction	109
2. Statistical methods for assessing biomarkers	110
3. Statistical methods for analyzing biomarker data	126
4. Concluding remarks	143
References	144

Ch. 5. Linear and Non-Linear Regression Methods in Epidemiology and Biostatistics 148

*Eric Vittinghoff, Charles E. McCulloch, David V. Glidden and Stephen C. Shiboski*

1. Introduction	148
2. Linear models	151
3. Non-linear models	167
4. Special topics	176
References	182

Ch. 6. Logistic Regression 187

*Edward L. Spitznael Jr.*

1. Introduction	187
2. Estimation of a simple logistic regression model	188
3. Two measures of model fit	191
4. Multiple logistic regression	192
5. Testing for interaction	194
6. Testing goodness of fit: Two measures for lack of fit	195
7. Exact logistic regression	196
8. Ordinal logistic regression	201
9. Multinomial logistic regression	204
10. Probit regression	206
1.1. Logistic regression in case-control studies	207
References	209

Ch. 7. Count Response Regression Models 210

*Joseph M. Hilbe and William H. Greene*

1. Introduction	210
2. The Poisson regression model	212
3. Heterogeneity and overdispersion	224
4. Important extensions of the models for counts	230

*Table of contents*

5. Software	247
6. Summary and conclusions	250
References	251

Ch. 8. Mixed Models 253  
*Matthew J. Gurka and Lloyd J. Edwards*

1. Introduction	253
2. Estimation for the linear mixed model	259
3. Inference for the mixed model	261
4. Selecting the best mixed model	264
5. Diagnostics for the mixed model	268
6. Outliers	270
7. Missing data	270
8. Power and sample size	272
9. Generalized linear mixed models	273
10. Nonlinear mixed models	274
11. Mixed models for survival data	275
12. Software	276
13. Conclusions	276
References	277

Ch. 9. Survival Analysis 281  
*John P. Klein and Mei-Jie Zhang*

1. Introduction	281
2. Univariate analysis	282
3. Hypothesis testing	288
4. Regression models	295
5. Regression models for competing risks	310
References	317

Ch. 10. A Review of Statistical Analyses for Competing Risks 321  
*Melvin L. Moeschberger, Kevin P. Tordoff and Nidhi Kochar*

1. Introduction	321
2. Approaches to the statistical analysis of competing risks	324
3. Example	327
4. Conclusion	339
References	340

Ch. 11. Cluster Analysis 342  
*William D. Shannon*

1. Introduction	342
2. Proximity measures	344
3. Hierarchical clustering	350
4. Partitioning	355
5. Ordination (scaling)	358

*Table of contents*

6. How many clusters?	361
7. Applications in medicine	364
8. Conclusion	364
References	365

Ch. 12. Factor Analysis and Related Methods 367

*Carol M. Woods and Michael C. Edwards*

1. Introduction	367
2. Exploratory factor analysis (EFA)	368
3. Principle components analysis (PCA)	375
4. Confirmatory factor analysis (CFA)	375
5. FA with non-normal continuous variables	379
6. FA with categorical variables	380
7. Sample size in FA	382
8. Examples of EFA and CFA	383
9. Additional resources	389
Appendix A	391
Appendix B	391
References	391

Ch. 13. Structural Equation Modeling 395

*Kentaro Hayashi, Peter M. Benller and Ke-Hai Yuan*

1. Models and identification	395
2. Estimation and evaluation	399
3. Extensions of SEM	410
4. Some practical issues	415
References	418

Ch. 14. Statistical Modeling in Biomedical Research: Longitudinal Data Analysis 429

*Chengjie Xiong, Kejun Zhu, Kai Yu and J. Philip Miller*

1. Introduction	429
2. Analysis of longitudinal data	431
3. Design issues of a longitudinal study	456
References	460

Ch. 15. Design and Analysis of Cross-Over Trials 464

*Michael G. Ken ward and Byron Jones*

1. Introduction	464
2. The two-period two-treatment cross-over trial	467
3. Higher-order designs	476
4. Analysis with non-normal data	482
5. Other application areas	485
6. Computer software	488
References	489

Ch. 16. Sequential and Group Sequential Designs in Clinical Trials: Guidelines for Practitioners	491
<i>Mad/tu Mazumdar and Heejung Bang</i>	
1. Introduction	492
2. Historical background of sequential procedures	493
3. Group sequential procedures for randomized trials	494
4. Steps for GSD design and analysis	507
5. Discussion	508
References	509
Ch. 17. Early Phase Clinical Trials: Phases I and II	513
<i>Feng Gao, Kalhryn Trinkaus and J. Philip Miller</i>	
1. Introduction	513
2. Phase I designs	514
3. Phase II designs	526
4. Summary	539
References	541
Ch. 18. Definitive Phase III and Phase IV Clinical Trials	546
<i>Barry R. Davis and Sarah Baraniuk</i>	
1. Introduction	546
2. Questions	548
3. Randomization	550
4. Recruitment	551
5. Adherence/sample size/power	552
6. Data analysis	554
7. Data quality and control/data management	558
8. Data monitoring	558
9. Phase IV trials	563
10. Dissemination - trial reporting and beyond	564
11. Conclusions	565
References	565
Ch. 19. Incomplete Data in Epidemiology and Medical Statistics	569
<i>Susanne Rdssler, Donald B. Rubin and Elizabeth R. Zell</i>	
1. Introduction	569
2. Missing-data mechanisms and ignorability	571
3. Simple approaches to handling missing data	573
4. Single imputation	574
5. Multiple imputation	578
6. Direct analysis using model-based procedures	581
7. Examples	584
8. Literature review for epidemiology and medical studies	586
9. Summary and discussion	587
Appendix A	588
Appendix B	592
References	598

Ch. 20. Meta-Analysis	602
<i>Edward L. Spilznagel Jr.</i>	
1. Introduction	602
2. History	603
3. The Cochran-Mantel-Haenszel test	604
4. Glass's proposal for meta-analysis	606
5. Random effects models	607
6. The forest plot	609
7. Publication bias	610
8. The Cochrane Collaboration	614
References	614
Ch. 21. The Multiple Comparison Issue in Health Care Research	616
<i>Lemuel A. Moye</i>	
1. Introduction	616
2. Concerns for significance testing	617
3. Appropriate use of significance testing	618
4. Definition of multiple comparisons	619
5. Rational for multiple comparisons	620
6. Multiple comparisons and analysis triage	621
7. Significance testing and multiple comparisons	623
8. Familywise error rate	625
9. The Bonferroni inequality	626
10. Alternative approaches	629
11. Dependent testing	631
12. Multiple comparisons and combined endpoints	635
13. Multiple comparisons and subgroup analyses	641
14. Data dredging	651
References	651
Ch. 22. Power: Establishing the Optimum Sample Size	656
<i>Richard A. Zeller and Yan Yan</i>	
1. Introduction	656
2. Illustrating power	658
3. Comparing simulation and software approaches to power	663
4. Using power to decrease sample size	672
5. Discussion	677
References	677
Ch. 23. Statistical Learning in Medical Data Analysis	679
<i>Grace Wahba</i>	
1. Introduction	679
2. Risk factor estimation: penalized likelihood estimates	681
3. Risk factor estimation: likelihood basis pursuit and the LASSO	690

4. Classification: support vector machines and related estimates	693
5. Dissimilarity data and kernel estimates	700
6. Tuning methods	704
7. Regularization, empirical Bayes. Gaussian processes priors, and reproducing kernels	707
References	708
Ch. 24. Evidence Based Medicine and Medical Decision Making	712
<i>Dan Mayer</i>	
1. The definition and history of evidence based medicine	712
2. Sources and levels of evidence	715
3. The five stage process of EBM	717
4. The hierarchy of evidence: study design and minimizing bias	718
5. Assessing the significance or impact of study results: Statistical significance and confidence intervals	721
6. Meta-analysis and systematic reviews	722
7. The value of clinical information and assessing the usefulness of a diagnostic test	722
8. Expected values decision making and the threshold approach to diagnostic testing	726
9. Summary	727
10. Basic principles	727
References	728
Ch. 25. Estimation of Marginal Regression Models with Multiple Source Predictors	730
<i>Heather J. Litman, Nicholas J. Norton, Bernardo Hernandez and Nan M. Laird</i>	
1. Introduction	730
2. Review of the generalized estimating equations approach	732
3. Maximum likelihood estimation	735
4. Simulations	737
5. Efficiency calculations	740
6. Illustration	741
7. Conclusion	743
References	745
Ch. 26. Difference Equations with Public Health Applications	747
<i>Asha Seth Kapadia and Lemuel A. Moyer</i>	
1. Introduction	747
2. Generating functions	748
3. Second-order nonhomogeneous equations and generating functions	750
4. Example in rhythm disturbances	752
5. Follow-up losses in clinical trials	758
6. Applications in epidemiology	765
References	773

Ch. 27. The Bayesian Approach to Experimental Data Analysis	775
<i>Bruno Lecoutre</i>	
Preamble: and if you were a Bayesian without knowing it?	775
1. Introduction	776
2. Frequentist and Bayesian inference	778
3. An illustrative example	783
4. Other examples of inferences about proportions	795
5. Concluding remarks and some further topics	803
References	808
Subject Index	813
Handbook of Statistics Contents of Previous Volumes	823