STATISTICS FOR INDUSTRY AND TECHNOLOGY

Jean-Louis Auget
N. Balakrishnan
Mounir Mesbah
Geert Molenberghs
Editors

Advances in Statistical Methods for the Health Sciences

Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis

Birkhäuser

Contents

Preface xix
Contributors xxi
List of Tables xxix
List of Figures xxxv

PART I: PROGNOSTIC STUDIES AND GENERAL EPIDEMIOLOGY

1 Systematic Review of Multiple Studies of Prognosis: The Feasibility of Obtaining Individual Patient Data 3
D. G. Altman, M. Trivella, F. Pezzella, A. L. Harris, and U. Pastorino

1.1 Introduction 3
1.2 Systematic Review Based on Individual Patient Data 5
1.3 A Case Study: Microvessel Density in Non-Small Cell Lung Cancer 6
1.3.1 Identifying studies (data sets) and obtaining the data 7
1.3.2 Checking the data 9
1.3.3 MVD measurements 11
1.3.4 Meta-analysis 12
1.4 Discussion 12
1.4.1 Systematic review of prognostic studies using individual patient data 13
1.4.2 The need for higher-quality prognostic studies 14
References 16

2 On Statistical Approaches for the Multivariable Analysis of Prognostic Marker Studies 19
N. Holländer and W. Sauerbrei

2.1 Introduction 19

2.2 Examples: Two Prognostic Studies in Breast Cancer 20
2.3 Statistical Methods 21
2.3.1 Regression models 21
2.3.2 Classification and regression trees (CART) 22
2.3.3 Formation of risk groups 23
2.4 Results in the GBSG-2 Study 23
2.4.1 Regression models – standard applications 23
2.4.2 Regression models – the MFP-approach 25
2.4.3 Summary assessment – implication of the modelling strategy 25
2.4.4 Application of classification and regression trees 28
2.5 Formation and Validation of Risk Groups 30
2.6 Discussion 32
References 35

3 Where Next for Evidence Synthesis of Prognostic Marker Studies? Improving the Quality and Reporting of Primary Studies to Facilitate Clinically Relevant Evidence-Based Results 39
R. D. Riley, K. R. Abrams, P. C. Lambert, A. J. Sutton, and D. G. Altman

3.1 Introduction and Aims 40
3.1.1 Prognostic markers and prognostic marker studies 40
3.1.2 The need for formal evidence syntheses of prognostic marker studies 40
3.1.3 Aims of this chapter 41
3.2 Difficulties of an Evidence Synthesis of Prognostic Marker Studies 42
3.2.1 Poor and heterogeneous reporting 42
3.2.2 Poor study design and problems clarifying study purpose 44
3.2.3 Little indication of how to implement markers in clinical practice 45
3.2.4 Small sample sizes 46
3.2.5 Publication bias, selective within-study reporting, and selective analyses 46
3.2.6 Lack of appreciation or validation of previous findings 48
3.3 What Improvements Are Needed in Primary Prognostic Marker Studies? 49
3.4 Evidence Synthesis Using Individual Patient Data Rather than Summary Statistics 51

3.5 Discussion 54
References 55

PART II: PHARMACOVIGILANCE

4 Sentinel Event Methods for Monitoring Unanticipated Adverse Events **61**
P. A. Lachenbruch and J. Wittes

4.1 Introduction 61
4.2 Examples 63
4.3 Usual Approaches to Monitoring Safety 64
4.4 Methods for Sentinel Events 66
4.4.1 Constant follow-up time 66
4.4.2 Censoring at a fixed calendar time 68
4.5 Methods for Sentinel Event Rates 71
4.6 Bayesian Models 72
4.7 Summary 73
References 74

5 Spontaneous Reporting System Modelling for the Evaluation of Automatic Signal Generation Methods in Pharmacovigilance **75**
E. Roux, F. Thiessard, A. Fourrier, B. Bégaud, and P. Tubert-Bitter

5.1 Introduction 76
5.2 Methods 78
5.2.1 Spontaneous reporting system modelling 78
5.2.2 Exposure to the drug (T_i) 79
5.2.3 Events' relative risk (RR_{ij}) 79
5.2.4 Reporting probability (p_{ij}) 80
5.2.5 Data generation process 84
5.3 Application 84
5.3.1 Values of the model parameters 84
5.3.2 Application of the empirical Bayes method 86
5.4 Results 86
5.5 Discussion 88
5.6 Conclusion 89
References 90

PART III: QUALITY OF LIFE

6 Latent Covariates in Generalized Linear Models: A Rasch Model Approach **95**
K. B. Christensen

6.1 Introduction 95
6.2 Generalized Linear Mixed Models 96
6.2.1 Latent regression models 97
6.3 Interpretation of Parameters 98
6.4 Generalized Linear Models with a Latent Covariate 98
6.4.1 Model 99
6.4.2 Interpretation of parameters 99
6.4.3 Parameter estimation 100
6.5 Example 101
6.5.1 Latent covariate 102
6.5.2 Job group level effect of the latent covariate 103
6.6 Discussion 104
References 105

7 Sequential Analysis of Quality of Life Measurements with the Mixed Partial Credit Model 109
V. Sébille, T. Challa, and M. Mesbah

7.1 Introduction 110
7.2 Methods 111
7.2.1 The partial credit model 111
7.2.2 Estimation of the parameters 112
7.2.3 Sequential analysis 112
7.2.4 The Z and V statistics 112
7.2.5 Traditional sequential analysis 113
7.2.6 Sequential analysis based on partial credit measurements 113
7.2.7 The sequential probability ratio test and the triangular test 115
7.2.8 Simulation design 115
7.3 Results 116
7.4 Discussion 121
7.5 Conclusion 122
Appendix 122
References 123

8 A Parametric Degradation Model Used in Reliability, Survival Analysis, and Quality of Life 127
M. Nikulin, L. Gerville-Réache, and S. Orazio

8.1 Introduction 127
8.2 Degradation Process 129
8.3 Estimation Problem 130
8.4 Linear MVUE for a 131
8.5 Solution of the Optimizatio Problem 132

8.6 Estimation of σ^2 and θ_0 132
References 134

9 Agreement Between Two Ratings with Different Ordinal Scales **139**
S. Natarajan, M. B. McHenry, S. Lipsitz, N. Klar, and S. Lipshultz

9.1 Introduction 139
9.2 Notation and Model 142
9.3 Examples and Interpretation 144
9.4 Discussion 147
References 147

PART IV: SURVIVAL ANALYSIS

10 The Role of Correlated Frailty Models in Studies of Human Health, Ageing, and Longevity **151**
A. Wienke, P. Lichtenstein, K. Czene, and A.I. Yashin

10.1 Introduction 151
10.2 Shared Frailty Model 153
10.3 Correlated Frailty Model 155
10.4 Correlated Gamma Frailty Model 156
10.4.1 Swedish breast cancer twin data 157
10.4.2 Parametric and semiparametric models 158
10.4.3 Correlated gamma frailty model with covariates 159
10.4.4 Cure-mixture models 160
10.5 Discussion 163
References 164

11 Prognostic Factors and Prediction of Residual Survival for Hospitalized Elderly Patients **167**
M. L. Calle, P. Roura, A. Arnau, A. Yáñez, and A. Leiva

11.1 Introduction 167
11.2 Cohort Description and Follow-Up 168
11.3 Multistate Survival Model 171
11.3.1 Predictive process 173
11.4 Discussion 176
References 177

12 New Models and Methods for Survival Analysis of Experimental Data **179**
G. V. Semenchenko, A. I. Yashin, T. E. Johnson, and J. W. Cypser

12.1 Introduction 179
12.2 Semiparametric Model of Mortality in Heterogeneous Populations Influenced by Exogenous Interventions 180
12.2.1 The heterogeneous mortality model 180
12.2.2 Changes in the baseline hazard and frailty distribution 181
12.2.3 Semiparametric representation of the model 183
12.2.4 Interpretation of parameters 184
12.3 One Example of the Model Application to the Analysis of Poststress Survival Data 187
12.3.1 Heat shock of different duration applied to populations of nematodes *Caenorhabditis elegans* 187
12.3.2 Technical details 188
12.3.3 Results of fitting the model to the data 188
12.4 Discussion 190
References 191

13 Uniform Consistency for Conditional Lifetime Distribution Estimators Under Random Right-Censorship **195**
P. Deheuvels and G. Derzko

13.1 Introduction and Results 195
13.1.1 Notation and statement of the problem 195
13.1.2 Definition of the estimators and uniform consistency 198
13.2 Proofs 203
13.2.1 Construction of the estimators 203
13.2.2 A useful auxiliary result 206
13.2.3 Concluding comments 208
References 208

14 Sequential Estimation for the Semiparametric Additive Hazard Model **211**
L. Bordes and C. Breuils

14.1 Introduction 211
14.2 Example and Numerical Study 213
14.3 Assumptions and Theoretical Results 215
14.4 Technical Tools 220
14.4.1 Complete convergence, Anscombe condition, and exponential inequality 220
14.4.2 Technical results 221
References 222

15 Variance Estimation of a Survival Function with Doubly Censored Failure Time Data **225**
C. Zhu and J. Sun

15.1 Introduction 225
15.2 Preliminaries 227
15.3 Variance Estimation 228
15.3.1 Simple bootstrap method 228
15.3.2 Generalized Greenwood formula 229
15.3.3 Imputation methods I and II 230
15.4 Numerical Results 231
15.5 Concluding Remarks 233
References 234

PART V: CLUSTERING

16 Statistical Models and Artificial Neural Networks: Supervised Classification and Prediction Via Soft Trees **239**
A. Ciampi and Y. Lechevallier

16.1 Introduction 240
16.2 The Prediction Problem: Statistical Models and ANNs 241
16.2.1 Generalized linear models as ANNs 243
16.2.2 Generalized additive models as ANNs 245
16.2.3 Classification and regression trees as ANNs 246
16.3 Combining Prediction Models: Hierarchy of Experts 248
16.4 The Soft Tree 250
16.4.1 General concepts in tree construction 251
16.4.2 Constructing a soft tree from data 252
16.4.3 An example of data analysis 253
16.4.4 An evaluation study 256
16.5 Extending the Soft Tree 257
16.6 Conclusions 259
References 260

17 Multilevel Clustering for Large Databases **263**
Y. Lechevallier and A. Ciampi

17.1 Introduction 263
17.2 Data Reduction by Kohonen SOMs 265
17.2.1 Kohonen SOMs and PCA initialization 266
17.2.2 Binning of the original data matrix using a Kohonen map 266
17.2.3 Dissimilarity for microregimens 268

17.3 Clustering Multilevel Systems 269
17.3.1 A two-level statistical model 270
17.3.2 Estimating parameters by the EM algorithm 270
17.4 Extracting Dietary Patterns from the Nutritional Data 271
References 273

18 Neural Networks: An Application for Predicting Smear Negative Pulmonary Tuberculosis 275
A. M. Santos, B. B. Pereira, J. M. Seixas, F. C. Q. Mello, and A. L. Kritski

18.1 Introduction 276
18.2 Materials and Methods 277
18.2.1 Data set 277
18.2.2 Neural network design 278
18.2.3 Data selection for network design 279
18.3 Relevance of Explanatory Variables 283
18.4 Results 283
18.5 Conclusions 284
References 285

19 Assessing Drug Resistance in HIV Infection Using Viral Load Using Segmented Regression 289
H. Liang, W.-Y. Tan, and X. Xiong

19.1 Introduction 289
19.2 The Model 292
19.2.1 The likelihood function 294
19.2.2 The prior distribution 294
19.2.3 The posterior distributions 295
19.3 The Gibbs Sampling Procedure 297
19.4 Analysis of the ACTG 315 Data 298
19.5 Conclusion and Dicussion 301
References 303

20 Assessment of Treatment Effects on HIV Pathogenesis Under Treatment By State Space Models 305
W.-Y. Tan, P. Zhang, X. Xiong, and P. Flynn

20.1 Introduction 305
20.2 A Stochastic Model for HIV Pathogenesis Under Treatment 306
20.2.1 Stochastic differential equations of state variables 307
20.2.2 The probability distribution of state variables 308

20.3 A State Space Model for HIV Pathogenesis Under Antiretroviral Drugs 309
20.4 Estimation of Unknown Parameters and State Variables 310
20.5 An Illustrative Example 311
20.6 Conclusions and Discussion 314
References 317

PART VI: SAFETY AND EFFICACY ASSESSMENT

21 Safety Assessment Versus Efficacy Assessment **323**
M. A. Foulkes

21.1 Introduction 323
21.2 Design Issues 324
21.2.1 Outcomes 324
21.2.2 Power 325
21.2.3 Population 326
21.2.4 Comparison 326
21.3 Analytic Issues 326
21.3.1 Compliance 326
21.3.2 Missing data 327
21.3.3 Confounding 327
21.3.4 Bias 328
21.3.5 Misclassification 328
21.3.6 Multiplicity 328
21.4 Analytic Approaches 329
21.5 Inferences 330
21.6 Conclusions 330
References 332

22 Cancer Clinical Trials with Efficacy and Toxicity Endpoints: A Simulation Study to Compare Two Nonparametric Methods **335**
A. Letierce and P. Tubert-Bitter

22.1 Introduction 336
22.2 Setting 337
22.3 Method for the Simulation Study 340
22.4 Results 342
22.5 Conclusion 346
References 347

23 Safety Assessment in Pilot Studies When Zero Events Are Observed **349**
R. E. Carter and R. F. Woolson

23.1 Introduction 349
23.2 Clinical Setting 350
23.3 Notation 351
23.4 Binomial Setting 351
23.5 Geometric Setting 352
23.6 Bayesian Credible Interval 354
23.7 Clinical Setting: Revisited 356
23.8 Summary 357
References 357

PART VII: CLINICAL DESIGNS

24 An Assessment of Up-and-Down Designs and Associated Estimators in Phase I Trials **361**
H. K. T. Ng, S. G. Mohanty, and N. Balakrishnan

24.1 Introduction 361
24.2 Notation and Designs 362
24.2.1 The biased coin design (BCD) 363
24.2.2 The k-in-a-row rule (KROW) 364
24.2.3 The Narayana rule (NAR) 364
24.2.4 Continual reassessment method (CRM) 364
24.3 Estimation of Maximum Tolerated Dose 365
24.4 Simulation Setting 367
24.5 Comparison of Estimators 368
24.6 Comparison of Designs 369
References 385

25 Design of Multicentre Clinical Trials with Random Enrolment **387**
V. V. Anisimov and V. V. Fedorov

25.1 Introduction 387
25.2 Recruitment Time Analysis 389
25.3 Analysis of Variance of the Estimated ECRT 393
25.4 Inflation of the Variance Due to Random Enrolment 395
25.5 Solution of the Optimization Problem 396
25.6 Conclusions 398
Appendix 398
References 400

26 Statistical Methods for Combining Clinical Trial Phases II and III **401**
N. Stallard and S. Todd

26.1 Introduction 401
26.2 Background 402
26.2.1 The clinical evaluation programme for new drugs 402
26.2.2 Combining clinical trial phases II and III 403
26.2.3 Background to sequential clinical trials 404
26.3 Methods for Combining Phases II and III 407
26.3.1 Two-stage methods 408
26.3.2 A multistage group-sequential method 408
26.3.3 A multistage adaptive design method 409
26.3.4 Example—a multistage trial comparing three doses of a new drug for the treatment of Alzheimer's disease 410
26.4 Discussion and Future Directions 413
References 415

27 SCPRT: A Sequential Procedure That Gives Another Reason to Stop Clinical Trials Early **419**
X. Xiong, M. Tan, and J. Boyett

27.1 Introduction 420
27.2 The SCPRT Procedure 421
27.2.1 Controlling the boundary 423
27.2.2 Boundaries in terms of P-values 423
27.3 An Example 424
27.4 SCPRT with Unknown Variance 426
27.5 Clinical Trials with Survival Data 428
27.6 Conclusion 432
References 432

PART VIII: MODELS FOR THE ENVIRONMENT

28 Seasonality Assessment for Biosurveillance Systems **437**
E. N. Naumova and I. B. MacNeill

28.1 Introduction 438
28.1.1 Conceptual framework for seasonality assessment 438
28.2 δ-Method in Application to a Seasonality Model 440
28.2.1 Single-variable case 440
28.2.2 Two-variables case 441

28.2.3 Application to a seasonality model 442
28.2.4 Potential model extension 443
28.2.5 Additional considerations 444
28.3 Application to Temperature and Infection Incidence Analysis 446
28.4 Conclusion 449
References 450

29 Comparison of Three Convolution Prior Spatial Models for Cancer Incidence 451
E.-A. Sauleau, M. Musio, A. Etienne, and A. Buemi

29.1 Introduction 451
29.2 Materials and Methods 453
29.2.1 ICAR model 455
29.2.2 Distance model based on the exponential function 455
29.2.3 Two-dimensional P-splines model 456
29.2.4 Implementation of the models 456
29.3 Results 456
29.4 Discussion 459
References 463

30 Longitudinal Analysis of Short-Term Bronchiolitis Air Pollution Association Using Semiparametric Models 467
S. Willems, C. Segala, M. Maidenberg, and M. Mesbah

30.1 Introduction 467
30.2 Data 469
30.2.1 Sanitary data 469
30.2.2 Environmental data 469
30.3 Methods 470
30.3.1 Generalized additive models 470
30.3.2 Air pollution time series studies strategy 471
30.3.3 Criticism about the use of standard statistical software to fit GAM to epidemiological time series data 474
30.4 Results 475
30.4.1 Series of number of hospital consultations: Results with S-Plus 475
30.4.2 Series of number of hospital consultations: Results with R 482
30.4.3 SAS results 483

30.5 Discussion 484
References 486

PART IX: GENOMIC ANALYSIS

31 Are There Correlated Genomic Substitutions? 491
M. Karnoub, P. K. Sen, and F. Seillier-Moiseiwitsch

31.1 Introduction 491
31.2 The Probabilistic Model 492
31.3 Parameter Estimation 496
31.4 New Test Statistic 497
31.5 Power Studies 500
31.6 Numerical Studies 501
31.7 Data Analysis 504
31.8 Discussion 507
Appendix 31.1 509
Appendix 31.2 510
References 513

PART X: ANIMAL HEALTH

32 Swiss Federal Veterinary Office Risk Assessments: Advantages and Limitations of the Qualitative Method 519
R. Hauser, E. Breidenbach, and K. D. C. Stärk

32.1 Introduction 520
32.2 Health Risks from Consumption of Milk and Dairy Products: An Example of a Qualitative Risk Assessment 521
32.2.1 Risk profile 521
32.2.2 Hazard identification 521
32.2.3 Risk network 522
32.2.4 Risk estimation 523
32.2.5 Recommendations for random sample planning and risk managers 524
32.3 Advantages and Disadvantages of Qualitative Risk Assessment 525
32.4 Statistics' Part in Qualitative Risk Assessment 526
References 526

33 Qualitative Risk Analysis in Animal Health: A Methodological Example **527**
B. Dufour and F. Moutou

33.1 Introduction 528
33.2 Global Presentation of the Method 528
33.3 Qualitative Appreciation of the Probability of Each Event 530
33.4 Qualitative Risk Appreciation 531
33.5 Qualitative Appreciation Examples 532
33.6 Discussion 534
References 536

Index **539**