

Monographs on Statistics and Applied Probability 158

Multistate Models for the Analysis of Life History Data

Richard J. Cook

University of Waterloo
Canada

Jerald F. Lawless

University of Waterloo
Canada



CRC Press

Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Contents

Preface	xiii
Glossary	xvii
Abbreviations	xxv
1 Introduction to Life History Processes and Multistate Models	1
1.1 Life History Analysis with Multistate Models	1
1.2 Some Illustrative Studies	3
1.2.1 Disease Recurrence Following Treatment in a Clinical Trial	3
1.2.2 Complications from Type 1 Diabetes	4
1.2.3 Joint Damage in Psoriatic Arthritis	5
1.2.4 Viral Load Dynamics in Individuals with HIV Infection	7
1.3 Introduction to Multistate Processes	8
1.3.1 Counting Processes and Multistate Models	8
1.3.2 Features of Multistate Processes	9
1.3.3 Marginal Features and Partial Models	11
1.4 Some Aspects of Modeling, Analysis and Design	12
1.4.1 Objectives	12
1.4.2 Components of a Model	13
1.4.3 Study Design and Data	14
1.5 Software	16
1.6 Introduction to Some Studies and Dataframes	17
1.6.1 A Trial of Breast Cancer Patients with Skeletal Metastases	17
1.6.2 An International Breast Cancer Trial	19
1.6.3 Viral Rebounds in HIV-Positive Individuals	20
1.6.4 Viral Shedding in HIV Patients with CMV Infection	21
1.7 Bibliographic Notes	22
1.8 Problems	24
2 Event History Processes and Multistate Models	27
2.1 Intensity Functions and Counting Processes	27
2.2 Likelihood for Multistate Analyses	29
2.2.1 Product Integration and Sample Path Probabilities	29
2.2.2 Time-Dependent Covariates and Random Censoring	31
2.3 Some Important Multistate Models	35
2.3.1 Modulated Markov Models	35
2.3.2 Modulated Semi-Markov Models	37

2.3.3	Models with Dual Time Scales	37
2.3.4	Models Accommodating Heterogeneity	38
2.3.5	Linked Models and Local Dependence	39
2.4	Process Features of Interest	41
2.5	Simulation of Multistate Processes	42
2.6	Bibliographic Notes	44
2.7	Problems	45
3	Multistate Analysis Based on Continuous Observation	49
3.1	Maximum Likelihood Methods for Parametric Models	49
3.1.1	Markov Models	50
3.1.2	Semi-Markov Models	54
3.1.3	Multistate Processes with Hybrid Time Scales	55
3.1.4	Comments on Parametric Models	56
3.2	Nonparametric Estimation	56
3.2.1	Markov Models	56
3.2.2	An Illness-Death Analysis of a Metastatic Breast Cancer Trial	58
3.2.3	Semi-Markov Models	63
3.2.4	Recurrent Outbreaks of Symptoms from Herpes Simplex Virus	64
3.3	Semiparametric Regression Models	68
3.3.1	Multiplicative Modulated Markov Models	68
3.3.2	Regression Analysis of a Palliative Breast Cancer Trial	73
3.3.3	Multiplicative Modulated Semi-Markov Models	76
3.3.4	Regression Analysis of Outbreaks from Herpes Simplex Virus	77
3.3.5	Additive Markov and Semi-Markov Models	84
3.3.6	Analysis of Herpes Outbreaks with Additive Models	86
3.4	Nonparametric Estimation of State Occupancy Probabilities	89
3.4.1	Aalen-Johansen Estimates	89
3.4.2	Adjustment for Process-Dependent Censoring	93
3.4.3	Skeletal Complications and Mortality in Cancer Metastatic to Bone	95
3.5	Model Assessment	101
3.5.1	Checking Parametric Models	101
3.5.2	Semiparametric Models	104
3.5.3	Predictive Performance of Models	105
3.5.4	Consequences of Model Misspecification and Robustness	107
3.6	Design Issues	109
3.7	Bibliographic Notes	110
3.8	Problems	113
4	Additional Applications of Multistate Models	117
4.1	Competing Risks Analysis	117
4.1.1	Model Features and Intensity-Based Analysis	117
4.1.2	Methods Based on Cumulative Incidence Functions	120

4.1.3	Methods Based on Direct Binomial Regression	123
4.1.4	Models for State Occupancy Based on Pseudo-Values	125
4.1.5	A Competing Risks Analysis of Shunts in Hydrocephalus	127
4.2	Alternative Methods for State Occupancy Probabilities	138
4.2.1	Estimation Based on State Entry Time Distributions	138
4.2.2	Estimation Based on Binomial Data	139
4.2.3	A Utility-Based Analysis of a Therapeutic Breast Cancer Clinical Trial	140
4.3	Analysis of State Sojourn Time Distributions	143
4.4	Bibliographic Notes	147
4.5	Problems	149
5	Studies with Intermittent Observation of Individuals	153
5.1	Introduction	153
5.2	Estimation and Analysis for Markov Models	155
5.2.1	Model Fitting	155
5.2.2	Parametric Information and Study Design	156
5.2.3	Model Checking	159
5.2.4	Illustration: Progression of Diabetic Retinopathy	159
5.3	Nonparametric Estimation of State Occupancy Probabilities	168
5.4	Process-Dependent Observation Times	170
5.4.1	Further Remarks on Dependent Visit Processes	170
5.4.2	Marginal Features and Inverse Intensity of Visit Weighting	172
5.4.3	Estimation of Visit Process Intensities	174
5.4.4	Nonparametric Estimation of Occupancy Probabilities	175
5.4.5	Progression to Arthritis Mutilans	177
5.5	Intermittent Observation and Non-Markov Models	181
5.6	Mixed Observation Schemes	183
5.6.1	Illness-Death Models	183
5.6.2	General Models	184
5.6.3	Progression-Free Survival in Cancer Trials	185
5.7	Bibliographic Notes	192
5.8	Problems	194
6	Heterogeneity and Dependence in Multistate Processes	201
6.1	Accommodating Heterogeneity in Life History Processes	201
6.1.1	Frailty Models in Survival Analysis	201
6.1.2	A Progressive Multistate Model with Random Effects	204
6.1.3	Random Effect Models with Recurrent States	207
6.1.4	Analysis of Exacerbations in Chronic Bronchitis	210
6.2	Modeling Correlated Multistate Processes	213
6.2.1	Dependence Models Based on Random Effects	214
6.2.2	Intensity-Based Models for Local Dependence	215
6.2.3	Dependence Models Retaining Simple Marginal Properties	217
6.2.4	The Development of Axial Involvement in Psoriatic Arthritis	220
6.3	Finite Mixture Models	226

6.3.1	Notation and Maximum Likelihood Estimation	226
6.3.2	Modeling Variation in Disease Activity in Lupus	229
6.4	Hidden Markov Models	232
6.4.1	Models and Estimation	232
6.4.2	A Hidden Markov Model for Retinopathy in the DCCT	235
6.5	Bibliographic Notes	236
6.6	Problems	240
7	Process-Dependent Sampling Schemes	245
7.1	History- and State-Dependent Selection	245
7.1.1	Types of Selection Schemes and Likelihoods	245
7.1.2	Empirical Studies of Design Efficiencies for Markov Processes	251
7.1.3	Prevalent Cohort Sampling and Failure Times	254
7.1.4	Design Based on Probabilistic State-Dependent Sampling	261
7.1.5	Selection and Initial Conditions with Heterogeneous Processes	263
7.1.6	Initial Conditions with a Finite Mixture Model	265
7.2	Outcome-Dependent Subsampling and Two-Phase Studies	267
7.2.1	Two-Phase Studies	267
7.2.2	Multistate Processes	270
7.2.3	Inference for Models with Semiparametric Multiplicative Intensities	273
7.2.4	Design Issues	275
7.2.5	Checks on Independent Follow-up Assumptions	278
7.3	Bibliographic Notes	284
7.4	Problems	286
8	Additional Topics	291
8.1	Analysis of Process-Related Costs and Benefits	291
8.1.1	Individual-Level Models	291
8.1.2	Quality of Life Analysis and Breast Cancer Treatment	292
8.1.3	Individual-Level Decision-Making	293
8.1.4	Population-Level Cost Analysis	294
8.2	Prediction	295
8.2.1	Viral Rebounds among Persons with HIV	297
8.3	Joint Modeling of Marker and Event Processes	299
8.3.1	Roles of Markers in Disease Modeling	299
8.3.2	Models for Markers and Life History Processes	302
8.3.3	Intermittent Measurement of Markers and Censoring	303
8.3.4	A Joint Multistate and Discrete Marker Process Model	306
8.4	Remarks on Causal Inference with Life History Processes	312
8.5	Bibliographic Notes	317
8.6	Problems	319

Appendix A Selected Software Packages	325
A.1 Software for Time to Event Data	325
A.1.1 Parametric Analyses	325
A.1.2 Semiparametric Analyses	325
A.2 Selected Software for Multistate Analyses	326
A.2.1 Multistate Software	326
A.2.2 Methods Based on Marginal Features	327
A.2.3 Dataframe Construction with the <code>mstate</code> Package	327
A.3 Software for Intermittently Observed Multistate Processes	329
A.4 Miscellaneous Functions Useful for Multistate Analysis	329
A.4.1 Timeline Plots	329
A.4.2 Lexis Diagrams	331
A.5 Drawing Multistate Diagrams with the <code>Epi</code> R Package	333
Appendix B Simulation of Multistate Processes	335
B.1 Generating a 3-State Time-Nonhomogeneous Markov Process	335
B.1.1 Intensities Featuring Smooth Time Trends	335
B.1.2 Processes with Piecewise-Constant Intensities	338
B.2 Simulating Multistate Processes under Intermittent Inspection	341
Appendix C Code and Output for Illustrative Analyses	343
C.1 Illustrative Analysis of Diabetic Retinopathy	343
C.1.1 Fitting the Reversible Markov Model M1B with <code>msm</code>	344
C.1.2 Fitting the Progressive Markov Model M2B with <code>msm</code>	346
C.1.3 Fitting the Hidden Markov Model with <code>msm</code>	347
C.2 Code for the Onset of Arthritis Mutilans in PsA	348
C.2.1 Dataframe and Fit of Intensity-Based Model	348
C.2.2 Marginal Model for Time to Entry to the Absorbing State	349
C.2.3 Inverse Intensity Weighted Nonparametric Estimation	351
Appendix D Datasets	355
D.1 Mechanical Ventilation in an Intensive Care Unit	355
D.2 Outcomes in Blood and Marrow Transplantation (EBMT)	356
D.3 A Trial of Platelet Dose and Bleeding Outcomes	358
D.4 Shedding of Cytomegalovirus in HIV-Infected Individuals	360
D.5 Micronutrient Powder and Infection in Malnourished Children	361
D.6 The Dynamics of <i>Giardia lamblia</i> Infection in Children	362
D.7 The Development of Arthritis Mutilans in Psoriatic Arthritis	364
D.8 Damage of the Sacroiliac Joints in Psoriatic Arthritis	365
D.9 The Incidence of PsA in Individuals with Psoriasis	366
Bibliography	369
Author Index	393
Subject Index	401