

Geert Verbeke • Geert Molenberghs Linear Mixed Models for Longitudinal Data

This paperback edition is a reprint of the 2000 edition.

This book provides a comprehensive treatment of linear mixed models for continuous longitudinal data. Next to model formulation, this edition puts major emphasis on exploratory data analysis for all aspects of the model, such as the marginal model, subject-specific profiles, and residual covariance structure. Further, model diagnostics and missing data receive extensive treatment. Sensitivity analysis for incomplete data is given a prominent place. Several variations to the conventional linear mixed model are discussed (a heterogeneity model, conditional linear mixed models). This book will be of interest to applied statisticians and biomedical researchers in industry, public health organizations, contract research organizations, and academia. The book is explanatory rather than mathematically rigorous. Most analyses were done with the MIXED procedure of the SAS software package, and many of its features are clearly elucidated. However, some other commercially available packages are discussed as well. Great care has been taken in presenting the data analyses in a software-independent fashion.

Geert Verbeke is Professor in Biostatistics at the Biostatistical Centre of the Katholieke Universiteit Leuven in Belgium. He is Past President of the Belgian Region of the International Biometric Society, a Board Member of the American Statistical Association, and past Joint Editor of the *Journal of the Royal Statistical Society, Series A* (2005–2008). He is the director of the Leuven Center for Biostatistics and statistical Bioinformatics (L-BioStat), and vice-director of the Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), a joint initiative of the Hasselt and Leuven universities in Belgium.

Geert Molenberghs is Professor of Biostatistics at Universiteit Hasselt and Katholieke Universiteit Leuven in Belgium. He was Joint Editor of *Applied Statistics* (2001–2004) and Co-Editor of *Biometrics* (2007–2009). He was President of the International Biometric Society (2004–2005), and has received the Guy Medal in Bronze from the Royal Statistical Society and the Myrto Lefkopoulou award from the Harvard School of Public Health. He is founding director of the Center for Statistics and also the director of the Interuniversity Institute for Biostatistics and statistical Bioinformatics.

Both authors have received the American Statistical Association's Excellence in Continuing Education Award in 2002, 2004, 2005, and 2008. Both are elected Fellows of the American Statistical Association and elected members of the International Statistical Institute.

STATISTICS

ISBN 978-1-4419-0299-3



Preface	vii
Acknowledgments	ix
1 Introduction	1
2 Examples	7
2.1 The Rat Data	7
2.2 The Toenail Data (TDO)	9
2.3 The Baltimore Longitudinal Study of Aging (BLSA)	10
2.3.1 The Prostate Data	11
2.3.2 The Hearing Data	14
2.4 The Vorozole Study	15
2.5 Heights of Schoolgirls	16
2.6 Growth Data	16

2.7	Mastitis in Dairy Cattle	18
3	A Model for Longitudinal Data	19
3.1	Introduction	19
3.2	A Two-Stage Analysis	20
3.2.1	Stage 1	20
3.2.2	Stage 2	20
3.2.3	Example: The Rat Data	21
3.2.4	Example: The Prostate Data	21
3.2.5	Two-Stage Analysis	22
3.3	The General Linear Mixed-Effects Model	23
3.3.1	The Model	23
3.3.2	Example: The Rat Data	25
3.3.3	Example: The Prostate Data	26
3.3.4	A Model for the Residual Covariance Structure	26
4	Exploratory Data Analysis	31
4.1	Introduction	31
4.2	Exploring the Marginal Distribution	31
4.2.1	The Average Evolution	31
4.2.2	The Variance Structure	33
4.2.3	The Correlation Structure	34
4.3	Exploring Subject-Specific Profiles	35
4.3.1	Measuring the Overall Goodness-of-Fit	35
4.3.2	Testing for the Need of a Model Extension	37
4.3.3	Example: The Rat Data	38
4.3.4	Example: The Prostate Data	39

5	Estimation of the Marginal Model	41
5.1	Introduction	41
5.2	Maximum Likelihood Estimation	42
5.3	Restricted Maximum Likelihood Estimation	43
5.3.1	Variance Estimation in Normal Populations	43
5.3.2	Estimation of Residual Variance in Linear Regression	43
5.3.3	REML Estimation for the Linear Mixed Model . . .	44
5.3.4	Justification of REML Estimation	46
5.3.5	Comparison Between ML and REML Estimation .	46
5.4	Model-Fitting Procedures	47
5.5	Example: The Prostate Data	48
5.6	Estimation Problems	50
5.6.1	Small Variance Components	50
5.6.2	Model Misspecifications	52
6	Inference for the Marginal Model	55
6.1	Introduction	55
6.2	Inference for the Fixed Effects	55
6.2.1	Approximate Wald Tests	56
6.2.2	Approximate t -Tests and F -Tests	56
6.2.3	Example: The Prostate Data	57
6.2.4	Robust Inference	61
6.2.5	Likelihood Ratio Tests	62
6.3	Inference for the Variance Components	64
6.3.1	Approximate Wald Tests	64
6.3.2	Likelihood Ratio Tests	65
6.3.3	Example: The Rat Data	66

6.3.4	Marginal Testing for the Need of Random Effects . . .	69
6.3.5	Example: The Prostate Data	72
6.4	Information Criteria	74
7	Inference for the Random Effects	77
7.1	Introduction	77
7.2	Empirical Bayes Inference	78
7.3	Henderson's Mixed-Model Equations	79
7.4	Best Linear Unbiased Prediction (BLUP)	80
7.5	Shrinkage	80
7.6	Example: The Random-Intercepts Model	81
7.7	Example: The Prostate Data	82
7.8	The Normality Assumption for Random Effects	83
7.8.1	Introduction	83
7.8.2	Impact on EB Estimates	85
7.8.3	Impact on the Estimation of the Marginal Model . .	87
7.8.4	Checking the Normality Assumption	89
8	Fitting Linear Mixed Models with SAS	93
8.1	Introduction	93
8.2	The SAS Program	94
8.2.1	The PROC MIXED Statement	95
8.2.2	The CLASS Statement	96
8.2.3	The MODEL Statement	96
8.2.4	The ID Statement	97
8.2.5	The RANDOM Statement	97
8.2.6	The REPEATED Statement	98

8.2.7	The CONTRAST Statement	101
8.2.8	The ESTIMATE Statement	101
8.2.9	The MAKE Statement	102
8.2.10	Some Additional Statements and Options	102
8.3	The SAS Output	104
8.3.1	Information on the Iteration Procedure	104
8.3.2	Information on the Model Fit	105
8.3.3	Information Criteria	107
8.3.4	Inference for the Variance Components	107
8.3.5	Inference for the Fixed Effects	111
8.3.6	Inference for the Random Effects	113
8.4	Note on the Mean Parameterization	114
8.5	The RANDOM and REPEATED Statements	117
8.6	PROC MIXED versus PROC GLM	119
9	General Guidelines for Model Building	121
9.1	Introduction	121
9.2	Selection of a Preliminary Mean Structure	123
9.3	Selection of a Preliminary Random-Effects Structure	125
9.4	Selection of a Residual Covariance Structure	128
9.5	Model Reduction	132
10	Exploring Serial Correlation	135
10.1	Introduction	135
10.2	An Informal Check for Serial Correlation	136
10.3	Flexible Models for Serial Correlation	137
10.3.1	Introduction	137

10.3.2	Fractional Polynomials	137
10.3.3	Example: The Prostate Data	138
10.4	The Semi-Variogram	141
10.4.1	Introduction	141
10.4.2	The Semi-Variogram for Random-Intercepts Models	142
10.4.3	Example: The Vorozole Study	144
10.4.4	The Semi-Variogram for Random-Effects Models	144
10.4.5	Example: The Prostate Data	147
10.5	Some Remarks	148
11	Local Influence for the Linear Mixed Model	151
11.1	Introduction	151
11.2	Local Influence	153
11.3	The Detection of Influential Subjects	158
11.4	Example: The Prostate Data	162
11.5	Local Influence Under REML Estimation	167
12	The Heterogeneity Model	169
12.1	Introduction	169
12.2	The Heterogeneity Model	171
12.3	Estimation of the Heterogeneity Model	173
12.4	Classification of Longitudinal Profiles	177
12.5	Goodness-of-Fit Checks	178
12.6	Example: The Prostate Data	180
12.7	Example: The Heights of Schoolgirls	183
13	Conditional Linear Mixed Models	189
13.1	Introduction	189

13.2	A Linear Mixed Model for the Hearing Data	190
13.3	Conditional Linear Mixed Models	194
13.4	Applied to the Hearing Data	197
13.5	Relation with Fixed-Effects Models	198
14	Exploring Incomplete Data	201
15	Joint Modeling of Measurements and Missingness	209
15.1	Introduction	209
15.2	The Impact of Incompleteness	210
15.3	Simple ad hoc Methods	211
15.4	Modeling Incompleteness	212
15.5	Terminology	214
15.6	Missing Data Patterns	215
15.7	Missing Data Mechanisms	215
15.8	Ignorability	217
15.9	A Special Case: Dropout	218
16	Simple Missing Data Methods	221
16.1	Introduction	221
16.2	Complete Case Analysis	223
16.3	Simple Forms of Imputation	223
16.3.1	Last Observation Carried Forward	224
16.3.2	Imputing Unconditional Means	225
16.3.3	Buck's Method: Conditional Mean Imputation	225
16.3.4	Discussion of Imputation Techniques	226
16.4	Available Case Methods	227
16.5	MCAR Analysis of Toenail Data	227

17 Selection Models	231
17.1 Introduction	231
17.2 A Selection Model for the Toenail Data	233
17.2.1 MAR Analysis	233
17.2.2 MNAR analysis	234
17.3 Scope of Ignorability	239
17.4 Growth Data	240
17.4.1 Analysis of Complete Growth Data	240
17.4.2 Frequentist Analysis of Incomplete Growth Data	256
17.4.3 Likelihood Analysis of Incomplete Growth Data	257
17.4.4 Missingness Process for the Growth Data	267
17.5 A Selection Model for Nonrandom Dropout	269
17.6 A Selection Model for the Vorozole Study	270
18 Pattern-Mixture Models	275
18.1 Introduction	275
18.1.1 A Simple Illustration	275
18.1.2 A Paradox	278
18.2 Pattern-Mixture Models	280
18.3 Pattern-Mixture Model for the Toenail Data	281
18.4 A Pattern-Mixture Model for the Vorozole Study	287
18.5 Some Reflections	291
19 Sensitivity Analysis for Selection Models	295
19.1 Introduction	295
19.2 A Modified Selection Model for Nonrandom Dropout	297
19.3 Local Influence	298

19.3.1	Review of the Theory	299
19.3.2	Applied to the Model of Diggle and Kenward	300
19.3.3	Special Case: Compound Symmetry	302
19.3.4	Serial Correlation	306
19.4	Analysis of the Rat Data	307
19.5	Mastitis in Dairy Cattle	312
19.5.1	Informal Sensitivity Analysis	312
19.5.2	Local Influence Approach	319
19.6	Alternative Local Influence Approaches	326
19.7	Random-coefficient-based Models	328
19.8	Concluding Remarks	330
20	Sensitivity Analysis for Pattern-Mixture Models	331
20.1	Introduction	331
20.2	Pattern-Mixture Models and MAR	332
20.2.1	MAR and ACMV	333
20.2.2	Nonmonotone Patterns: A Counterexample	335
20.3	Multiple Imputation	336
20.3.1	Parameter and Precision Estimation	338
20.3.2	Hypothesis Testing	338
20.4	Pattern-Mixture Models and Sensitivity Analysis	339
20.5	Identifying Restrictions Strategies	343
20.5.1	Strategy Outline	343
20.5.2	Identifying Restrictions	344
20.5.3	ACMV Restrictions	347
20.5.4	Drawing from the Conditional Densities	350
20.6	Analysis of the Vorozole Study	352

20.6.1	Fitting a Model	352
20.6.2	Hypothesis Testing	366
20.6.3	Model Reduction	371
20.7	Thoughts	373
21	How Ignorable Is Missing At Random ?	375
21.1	Introduction	375
21.2	Information and Sampling Distributions	377
21.3	Illustration	379
21.4	Example	383
21.5	Implications for PROC MIXED	385
22	The Expectation-Maximization Algorithm	387
23	Design Considerations	391
23.1	Introduction	391
23.2	Power Calculations Under Linear Mixed Models	392
23.3	Example: The Rat Data	393
23.4	Power Calculations When Dropout Is to Be Expected	394
23.5	Example: The Rat Data	397
23.5.1	Constant $p_{j,k \geq k}$, Varying n_j	399
23.5.2	Constant $p_{j,k \geq k}$, Constant n_j	401
23.5.3	Increasing $p_{j,k \geq k}$ over Time, Constant n_j	402
24	Case Studies	405
24.1	Blood Pressures	405
24.2	The Heat Shock Study	411
24.2.1	Introduction	411

24.2.2	Analysis of Heat Shock Data	415
24.3	The Validation of Surrogate Endpoints from Multiple Trials	420
24.3.1	Introduction	420
24.3.2	Validation Criteria	421
24.3.3	Notation and Motivating Examples	424
24.3.4	A Meta-Analytic Approach	429
24.3.5	Data Analysis	434
24.3.6	Computational Issues	439
24.3.7	Extensions	442
24.3.8	Reflections on Surrogacy	443
24.3.9	Prediction Intervals	444
24.3.10	SAS Code for Random-Effects Model	445
24.4	The Milk Protein Content Trial	446
24.4.1	Introduction	446
24.4.2	Informal Sensitivity Analysis	448
24.4.3	Formal Sensitivity Analysis	457
24.5	Hepatitis B Vaccination	470
24.5.1	Time Evolution of Antibodies	472
24.5.2	Prediction at Year 12	481
24.5.3	SAS Code for Vaccination Models	482

Appendix

A	Software	485
A.1	The SAS System	485
A.1.1	Standard Applications	485
A.1.2	New Features in SAS Version 7.0	485

A.2	Fitting Mixed Models Using MLwiN	489
A.3	Fitting Mixed Models Using SPlus	493
A.3.1	Standard SPlus Functions	494
A.3.2	OSWALD for Nonrandom Nonresponse	497
B	Technical Details for Sensitivity Analysis	515
B.1	Local Influence: Derivation of Components of Δ_i	515
B.2	Proof of Theorem 20.1	518
	References	523
	Index	554