

# Preface

The linear mixed model has become the main parametric tool for the analysis of continuous longitudinal data. Verbeke and Molenberghs (2000) devoted an entire text to the model, a number of its extensions, and how to deal with incompletely observed longitudinal profiles. The model can be fitted in a wide variety of commercially available software packages, such as the SAS procedure MIXED, the SPlus function lme, the MLwiN package, etc. Although the model can be interpreted as a natural hierarchical extension of linear regression and analysis of variance, it is our experience from courses, scientific collaboration, and statistical consultancy that the model remains surrounded with non-trivial issues such as the difference between a hierarchical and a marginal interpretation, complexities arising with inference for variance components, assessing goodness-of-fit, the effect of (mis-)specifying the random-effects distribution, etc.

Our courses, consultancy, and research in the area of longitudinal data analysis have included the non-Gaussian setting as well, including binary, ordinal repeated measures, as well as counts measured repeatedly over time. Our experience has been that the issues in this field are a multiple of those in the continuous case, predominantly due to the lack of an unambiguous counterpart of the multivariate normal distribution. Almost all models exhibit a certain amount of non-linearity. Even when attention is restricted to the special non-linear models of the generalized linear type, important differences between the classes of marginal, conditional, and subject-specific models arise. Within each of these, subfamilies can be identified within which, in turn, many different models can be placed. Different problems may call for different solutions and hence different modeling strategies.

In addition, due to computational complexity, many models require the use of approximate numeric methods, each one with its advantages and disadvantages. The issues are further compounded when planned measurement sequences are incompletely observed and strategies to deal with such incompleteness may depend in important ways on the inferential framework within which a particular model is framed. Fortunately, a variety of standard statistical software tools is now available to handle, possibly incomplete, non-Gaussian repeated measures, including the SAS procedures GENMOD, GLIMMIX, NLMIXED, MI, and MIANALYZE.

Verbeke and Molenberghs (2000) have not dealt with the non-Gaussian case, and we aim to fill this gap with the current text. Regular and short courses have helped shape our thinking regarding the selection of material and the emphasis to put on various model families, models, and inferential aspects. We mention in particular the regular courses on Correlated and Multivariate Data in the Master of Science in Applied Statistics Programme of the Limburgs Universitair Centrum, the Longitudinal Data Analysis and Advanced Modeling Techniques courses of the Master of Science in Biostatistics Programme of the Limburgs Universitair Centrum, and the Repeated Measures course in the International Study Programme in Statistics of the Katholieke Universiteit Leuven. We further learned a lot from teaching short courses to audiences with various backgrounds at numerous locations in Europe, North and South America, the Caribbean, Australia, and Asia.

Just as with Verbeke and Molenberghs (2000), we hope this book will be of value to a wide audience, including applied statisticians and biomedical researchers, particularly in the biopharmaceutical industry, medical and public health research organizations, contract research organizations, and academic departments. The majority of the chapters are explanatory rather than research oriented, although some chapters contain advanced material. A perspective is given in Chapter 1. Practice is emphasized rather than mathematical rigor. In this respect, guidance and advice on practical issues are important focuses of the text, and numerous extensively analyzed examples are included, many running across several chapters.

Virtually all of the statistical analyses were performed using SAS procedures such as MIXED, GENMOD, GLIMMIX, NLMIXED, MI, and MIANALYZE, as well as the SAS macro GLIMMIX. Almost all analyses were done using the SAS Version 9.1. The GLIMMIX procedure used here is experimental. Nevertheless, both the methodological development and the analysis of the case studies are presented in a software-independent fashion. Illustration of how to use SAS for the various model strategies is concentrated in a small number of chapters and sections, and the text can be read without any problem if these software excursions are ignored. Selected programs, macros, output, and publicly available datasets can be found at Springer-Verlag's URL: [www.springer-ny.com](http://www.springer-ny.com), as well as at the authors' web site.

Geert Molenberghs (Diepenbeek) and Geert Verbeke (Leuven)

# Acknowledgments

This text has benefited from the help of a large number of people. A lot of the more advanced chapters are based on joint research with many colleagues, as well as with current and past doctoral students. We gratefully acknowledge the support of these co-authors: Marc Aerts (Limburgs Universitair Centrum), Ariel Alonso (Limburgs Universitair Centrum), Caroline Beunckens (Limburgs Universitair Centrum), Larry Brant (National Institute of Aging and The Johns Hopkins University, Baltimore), Luc Bijmens (Johnson & Johnson Pharmaceutical Research and Development, Beerse), Tomasz Burzykowski (Limburgs Universitair Centrum), Marc Buyse (International Institute for Drug Development, Brussels), Raymond J. Carroll (Texas A& M University, College Station), Paul Catalano (Harvard School of Public Health, Boston), José Cortiñas Abrahantes (Limburgs Universitair Centrum), Linda Danielson (UCB, Braine-l'Alleud), Paul De Boeck (Katholieke Universiteit Leuven), Steffen Fieuws (Katholieke Universiteit Leuven), Krista Fisher-Lapp (Tartu University), Helena Geys (Johnson & Johnson Pharmaceutical Research and Development, Beerse), Els Goetghebeur (Universiteit Gent), Niel Hens (Limburgs Universitair Centrum), Ivy Jansen (Limburgs Universitair Centrum), Michael G. Kenward (London School of Hygiene and Tropical Medicine), Emmanuel Lesaffre (Katholieke Universiteit Leuven), Stuart Lipsitz (Medical University of South Carolina, Charleston), Craig Mallinckrodt (Eli Lilly and Company, Indianapolis), Bart Michiels (Janssen Research Foundation, Beerse), Christopher Morrell (National Institute of Aging and Loyola College, Baltimore), Meredith Regan (Harvard School of Public Health, Boston), Didier Renard (Eli Lilly and Company, Mont-Saint-Guibert),

Louise Ryan (Harvard School of Public Health, Boston), Jan Serroyen (Limburgs Universitair Centrum), Bart Spiessens (Glaxo Smith Kline Belgium), Herbert Thijs (Limburgs Universitair Centrum), Fabián Tibaldi (Eli Lilly and Company, Mont-Saint-Guibert), Tony Vangeneugden (Virco-Tibotec, Mechelen), Kristel Van Steen (Harvard School of Public Health, Boston), and Paige Williams (Harvard School of Public Health, Boston).

Several people have helped us with the computational side of the models presented. We mention in particular Caroline Beunckens, Steffen Fieuwis, and Oliver Schabenberger (SAS Institute, Cary, North Carolina).

We gratefully acknowledge support from Research Project Fonds voor Wetenschappelijk Onderzoek Vlaanderen G.0002.98, “Sensitivity Analysis for Incomplete Data”; NATO Collaborative Research Grant CRG950648, “Statistical Research for Environmental Risk Assessment”; Onderzoeksfonds K.U.Leuven grant PDM/96/105, and Belgian IUAP/PAI network “Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data.”

The feedback we received from our regular and short course audiences has been invaluable. We are grateful for such interactions in Australia (Cairns, Coolangatta), Belgium (Beerse, Braine-l’Alleud, Brussels, Diepenbeek, Gent, Leuven, Wavre), Brasil (Piracicaba), Canada (Toronto), Cuba (La Habana, Varadero), Finland (Turku), France (Marseille, Toulouse, Vannes), Germany (Freiburg, Heidelberg), Ireland (Dublin), Spain (Barcelona, Pamplona, Santiago de Compostela), and the United States of America (Ann Arbor, Arlington, Atlanta, New York City, Rockville, San Francisco, Tampa, Washington, DC).

As always, it has been a pleasure to work with John Kimmel at Springer and his colleagues from the production department, in particular C. Curioli and M. Koy.

We apologize to our wives, daughters, and sons for the time not spent with them during the preparation of this book, and we are very grateful for their understanding. The preparation of this book has been a period of close collaboration and stimulating exchange, of which we will keep good memories.

Geert and Geert  
Kessel-Lo and Herent, Belgium, February 2005



<http://www.springer.com/978-0-387-25144-8>

Models for Discrete Longitudinal Data

Molenberghs, G.; Verbeke, G.

2005, XXII, 687 p., Hardcover

ISBN: 978-0-387-25144-8