

# **Applied Longitudinal Analysis**

## WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Garrett M. Fitzmaurice, Harvey Goldstein, Iain M. Johnstone, Geert Molenberghs, David W. Scott, Adrian F. M. Smith, Ruey S. Tsay, Sanford Weisberg*

Editors Emeriti: *Vic Barnett, J. Stuart Hunter, Joseph B. Kadane, Jozef L. Teugels*

A complete list of the titles in this series appears at the end of this volume.

# Applied Longitudinal Analysis

Second Edition

**Garrett M. Fitzmaurice**

**Nan M. Laird**

**James H. Ware**

*Department of Biostatistics  
Harvard University  
Boston, MA*

 **WILEY**

**A JOHN WILEY & SONS, INC., PUBLICATION**

Copyright © 2011 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.  
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representation or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data:***

Fitzmaurice, Garrett M., 1962–

Applied longitudinal analysis / Garrett M. Fitzmaurice, Nan M. Laird, James H. Ware. — 2nd ed.  
p. cm.

ISBN 978-0-470-38027-7 (hardback)

1. Longitudinal method. 2. Regression analysis. 3. Multivariate analysis. 4. Medical statistics. I. Laird, Nan M., 1943– II. Ware, James H., 1941– III. Title.

QA278.F575 2011

519.5'3—dc22

2011012197

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

*To Laura, Kieran, and Aidan*

— G.M.F.

*To Joel, Richard, and Lily*

— N.M.L.

*To Janice, Cameron, and Jake*

— J.H.W.

# Contents

<b>Preface</b>	<b>xvii</b>
<b>Preface to First Edition</b>	<b>xxi</b>
<b>Acknowledgments</b>	<b>xxv</b>

## Part I Introduction to Longitudinal and Clustered Data

<b>1 Longitudinal and Clustered Data</b>	<b>1</b>
1.1 <i>Introduction</i>	1
1.2 <i>Longitudinal and Clustered Data</i>	2
1.3 <i>Examples</i>	5
1.4 <i>Regression Models for Correlated Responses</i>	13
1.5 <i>Organization of the Book</i>	16
1.6 <i>Further Reading</i>	18
<b>2 Longitudinal Data: Basic Concepts</b>	<b>19</b>
2.1 <i>Introduction</i>	19
2.2 <i>Objectives of Longitudinal Analysis</i>	19
2.3 <i>Defining Features of Longitudinal Data</i>	22
	<b>vii</b>

2.4	<i>Example: Treatment of Lead-Exposed Children Trial</i>	31
2.5	<i>Sources of Correlation in Longitudinal Data</i>	36
2.6	<i>Further Reading</i>	44
	<i>Problems</i>	44

## Part II Linear Models for Longitudinal Continuous Data

<b>3</b>	<b>Overview of Linear Models for Longitudinal Data</b>	<b>49</b>
3.1	<i>Introduction</i>	49
3.2	<i>Notation and Distributional Assumptions</i>	50
3.3	<i>Simple Descriptive Methods of Analysis</i>	62
3.4	<i>Modeling the Mean</i>	72
3.5	<i>Modeling the Covariance</i>	74
3.6	<i>Historical Approaches</i>	76
3.7	<i>Further Reading</i>	86
<b>4</b>	<b>Estimation and Statistical Inference</b>	<b>89</b>
4.1	<i>Introduction</i>	89
4.2	<i>Estimation: Maximum Likelihood</i>	90
4.3	<i>Missing Data Issues</i>	94
4.4	<i>Statistical Inference</i>	96
4.5	<i>Restricted Maximum Likelihood (REML) Estimation</i>	101
4.6	<i>Further Reading</i>	104
<b>5</b>	<b>Modeling the Mean: Analyzing Response Profiles</b>	<b>105</b>
5.1	<i>Introduction</i>	105
5.2	<i>Hypotheses Concerning Response Profiles</i>	107
5.3	<i>General Linear Model Formulation</i>	112
5.4	<i>Case Study</i>	117
5.5	<i>One-Degree-of-Freedom Tests for Group by Time Interaction</i>	120
5.6	<i>Adjustment for Baseline Response</i>	124
5.7	<i>Alternative Methods of Adjusting for Baseline Response*</i>	128

5.8	<i>Strengths and Weaknesses of Analyzing Response Profiles</i>	134
5.9	<i>Computing: Analyzing Response Profiles Using PROC MIXED in SAS</i>	136
5.10	<i>Further Reading Problems</i>	140
<b>6</b>	<b>Modeling the Mean: Parametric Curves</b>	<b>143</b>
6.1	<i>Introduction</i>	143
6.2	<i>Polynomial Trends in Time</i>	144
6.3	<i>Linear Splines</i>	149
6.4	<i>General Linear Model Formulation</i>	152
6.5	<i>Case Studies</i>	154
6.6	<i>Computing: Fitting Parametric Curves Using PROC MIXED in SAS</i>	161
6.7	<i>Further Reading Problems</i>	162
<b>7</b>	<b>Modeling the Covariance</b>	<b>165</b>
7.1	<i>Introduction</i>	165
7.2	<i>Implications of Correlation among Longitudinal Data</i>	166
7.3	<i>Unstructured Covariance</i>	168
7.4	<i>Covariance Pattern Models</i>	169
7.5	<i>Choice among Covariance Pattern Models</i>	175
7.6	<i>Case Study</i>	180
7.7	<i>Discussion: Strengths and Weaknesses of Covariance Pattern Models</i>	183
7.8	<i>Computing: Fitting Covariance Pattern Models Using PROC MIXED in SAS</i>	184
7.9	<i>Further Reading Problems</i>	186
<b>8</b>	<b>Linear Mixed Effects Models</b>	<b>189</b>
8.1	<i>Introduction</i>	189
8.2	<i>Linear Mixed Effects Models</i>	194
8.3	<i>Random Effects Covariance Structure</i>	201
8.4	<i>Two-Stage Random Effects Formulation</i>	203
8.5	<i>Choice among Random Effects Covariance Models</i>	208
8.6	<i>Prediction of Random Effects</i>	209



8.7	<i>Prediction and Shrinkage*</i>	211
8.8	<i>Case Studies</i>	213
8.9	<i>Computing: Fitting Linear Mixed Effects Models Using PROC MIXED in SAS</i>	234
8.10	<i>Further Reading</i>	237
	<i>Problems</i>	237
<b>9</b>	<b>Fixed Effects versus Random Effects Models</b>	<b>241</b>
9.1	<i>Introduction</i>	241
9.2	<i>Linear Fixed Effects Models</i>	241
9.3	<i>Fixed Effects versus Random Effects: Bias-Variance Trade-off</i>	246
9.4	<i>Resolving the Dilemma of Choosing Between Fixed and Random Effects Models</i>	249
9.5	<i>Longitudinal and Cross-sectional Information</i>	252
9.6	<i>Case Study</i>	255
9.7	<i>Computing: Fitting Linear Fixed Effects Models Using PROC GLM in SAS</i>	258
9.8	<i>Computing: Decomposition of Between-Subject and Within-Subject Effects Using PROC MIXED in SAS</i>	260
9.9	<i>Further Reading</i>	262
	<i>Problems</i>	262
<b>10</b>	<b>Residual Analyses and Diagnostics</b>	<b>265</b>
10.1	<i>Introduction</i>	265
10.2	<i>Residuals</i>	265
10.3	<i>Transformed Residuals</i>	266
10.4	<i>Aggregating Residuals</i>	269
10.5	<i>Semi-Variogram</i>	272
10.6	<i>Case Study</i>	273
10.7	<i>Summary</i>	285
10.8	<i>Further Reading</i>	286
	<i>Problems</i>	287

**Part III Generalized Linear Models for Longitudinal Data**

<b>11 Review of Generalized Linear Models</b>	<b>291</b>
11.1 <i>Introduction</i>	291
11.2 <i>Salient Features of Generalized Linear Models</i>	292
11.3 <i>Illustrative Examples</i>	297
11.4 <i>Ordinal Regression Models</i>	310
11.5 <i>Overdispersion</i>	319
11.6 <i>Computing: Fitting Generalized Linear Models         Using PROC GENMOD in SAS</i>	324
11.7 <i>Overview of Generalized Linear Models*</i>	327
11.8 <i>Further Reading         Problems</i>	335 336
<b>12 Marginal Models: Introduction and Overview</b>	<b>341</b>
12.1 <i>Introduction</i>	341
12.2 <i>Marginal Models for Longitudinal Data</i>	342
12.3 <i>Illustrative Examples of Marginal Models</i>	346
12.4 <i>Distributional Assumptions for Marginal Models*</i>	351
12.5 <i>Further Reading</i>	352
<b>13 Marginal Models: Generalized Estimating Equations (GEE)</b>	<b>353</b>
13.1 <i>Introduction</i>	353
13.2 <i>Estimation of Marginal Models: Generalized         Estimating Equations</i>	354
13.3 <i>Residual Analyses and Diagnostics</i>	361
13.4 <i>Case Studies</i>	364
13.5 <i>Marginal Models and Time-Varying Covariates</i>	381
13.6 <i>Computing: Generalized Estimating Equations         Using PROC GENMOD in SAS</i>	385
13.7 <i>Further Reading         Problems</i>	390 391
<b>14 Generalized Linear Mixed Effects Models</b>	<b>395</b>
14.1 <i>Introduction</i>	395
14.2 <i>Incorporating Random Effects in Generalized         Linear Models</i>	396

14.3	<i>Interpretation of Regression Parameters</i>	402
14.4	<i>Overdispersion</i>	409
14.5	<i>Estimation and Inference</i>	410
14.6	<i>A Note on Conditional Maximum Likelihood</i>	412
14.7	<i>Case Studies</i>	414
14.8	<i>Computing: Fitting Generalized Linear Mixed Models Using PROC GLIMMIX in SAS</i>	429
14.9	<i>Further Reading</i>	433
	<i>Problems</i>	434
<b>15</b>	<b>Generalized Linear Mixed Effects Models: Approximate Methods of Estimation</b>	<b>441</b>
15.1	<i>Introduction</i>	441
15.2	<i>Penalized Quasi-Likelihood</i>	443
15.3	<i>Marginal Quasi-Likelihood</i>	445
15.4	<i>Cautionary Remarks on the Use of PQL and MQL</i>	446
15.5	<i>Case Studies</i>	452
15.6	<i>Computing: Fitting GLMMs Using PROC GLIMMIX in SAS</i>	459
15.7	<i>Basis of PQL and MQL Approximations*</i>	466
15.8	<i>Further Reading</i>	470
	<i>Problems</i>	471
<b>16</b>	<b>Contrasting Marginal and Mixed Effects Models</b>	<b>473</b>
16.1	<i>Introduction</i>	473
16.2	<i>Linear Models: A Special Case</i>	473
16.3	<i>Generalized Linear Models</i>	474
16.4	<i>Simple Numerical Illustration</i>	479
16.5	<i>Case Study</i>	480
16.6	<i>Conclusion</i>	484
16.7	<i>Further Reading</i>	486

**Part IV Missing Data and Dropout**

<b>17 Missing Data and Dropout: Overview of Concepts and Methods</b>	<b>489</b>
17.1 <i>Introduction</i>	489
17.2 <i>Hierarchy of Missing Data Mechanisms</i>	491
17.3 <i>Implications for Longitudinal Analysis</i>	499
17.4 <i>Dropout</i>	500
17.5 <i>Common Approaches for Handling Dropout</i>	506
17.6 <i>Bias of Last Value Carried Forward Imputation*</i>	511
17.7 <i>Further Reading</i>	513
<b>18 Missing Data and Dropout: Multiple Imputation and Weighting Methods</b>	<b>515</b>
18.1 <i>Introduction</i>	515
18.2 <i>Multiple Imputation</i>	516
18.3 <i>Inverse Probability Weighted Methods</i>	526
18.4 <i>Case Studies</i>	531
18.5 <i>“Sandwich” Variance Estimator Adjusting for Estimation of Weights*</i>	541
18.6 <i>Computing: Multiple Imputation Using PROC MI in SAS</i>	542
18.7 <i>Computing: Inverse Probability Weighted (IPW) Methods in SAS</i>	547
18.8 <i>Further Reading</i>	550

**Part V Advanced Topics for Longitudinal and Clustered Data**

<b>19 Smoothing Longitudinal Data: Semiparametric Regression Models</b>	<b>553</b>
19.1 <i>Introduction</i>	553
19.2 <i>Penalized Splines for a Univariate Response</i>	554
19.3 <i>Case Study</i>	558
19.4 <i>Penalized Splines for Longitudinal Data</i>	563
19.5 <i>Case Study</i>	565

19.6	<i>Fitting Smooth Curves to Individual Longitudinal Data</i>	570
19.7	<i>Case Study</i>	572
19.8	<i>Computing: Fitting Smooth Curves Using PROC MIXED in SAS</i>	576
19.9	<i>Further Reading</i>	579
<b>20</b>	<b>Sample Size and Power</b>	<b>581</b>
20.1	<i>Introduction</i>	581
20.2	<i>Sample Size for a Univariate Continuous Response</i>	582
20.3	<i>Sample Size for a Longitudinal Continuous Response</i>	584
20.4	<i>Sample Size for a Longitudinal Binary Response</i>	598
20.5	<i>Summary</i>	604
20.6	<i>Computing: Sample Size Calculation Using Pseudo-Data</i>	605
20.7	<i>Further Reading</i>	609
<b>21</b>	<b>Repeated Measures and Related Designs</b>	<b>611</b>
21.1	<i>Introduction</i>	611
21.2	<i>Repeated Measures Designs</i>	612
21.3	<i>Multiple Source Data</i>	616
21.4	<i>Case Study 1: Repeated Measures Experiment</i>	617
21.5	<i>Case Study 2: Multiple Source Data</i>	620
21.6	<i>Summary</i>	625
21.7	<i>Further Reading</i>	626
<b>22</b>	<b>Multilevel Models</b>	<b>627</b>
22.1	<i>Introduction</i>	627
22.2	<i>Multilevel Data</i>	628
22.3	<i>Multilevel Linear Models</i>	630
22.4	<i>Multilevel Generalized Linear Models</i>	641
22.5	<i>Summary</i>	651
22.6	<i>Further Reading</i>	652

<b>Appendix A</b>	<b>Gentle Introduction to Vectors and Matrices</b>	<b>655</b>
<b>Appendix B</b>	<b>Properties of Expectations and Variances</b>	<b>665</b>
<b>Appendix C</b>	<b>Critical Points for a 50:50 Mixture of Chi-Squared Distributions</b>	<b>669</b>
<b>References</b>		<b>671</b>
<b>Index</b>		<b>695</b>

# Preface

The first edition of *Applied Longitudinal Analysis* was designed to serve as a textbook for a course on modern statistical methods for longitudinal data analysis, and subsequently, as a reference resource for students and researchers. The book was targeted at a broad audience: graduate students in statistics, statisticians working in the health sciences, pharmaceutical industry, and governmental health-related agencies, as well as researchers and graduate students from a variety of substantive fields. In the seven years that have elapsed since publication of the first edition, *Applied Longitudinal Analysis* has been used extensively in university classrooms throughout the United States and abroad. We are grateful to many colleagues, course instructors, students, and readers who have offered constructive suggestions on how the book could be improved. This feedback has been invaluable and helped shape the content of the second edition.

The feedback we received has encouraged us to retain the general structure and format of the first edition while taking the opportunity to introduce a number of new and important topics. Although there is much new material in this second edition, the principles that guided us in writing the first edition have not changed. Our primary goal is to present a rigorous and comprehensive description of modern statistical methods for the analysis of longitudinal data that is accessible to a wide range of readers. A strong emphasis is placed on the application of these methods to longitudinal data and the interpretation of results. Although the methods are presented in the setting of numerous applications to actual data sets drawn from studies in health-related fields, reflecting our own research interests in the health sciences, they apply equally to other areas of application, for example, education, psychology, and other branches of the behavioral and social sciences.

How does this edition differ from its predecessor? The major changes in this edition have resulted from the addition of six new chapters:

1. A chapter (Chapter 9) on “fixed effects models,” in which subject-specific effects are treated as fixed rather than random, has been added. This chapter complements the existing chapter on mixed effects models (Chapter 8) and includes a discussion of the relative advantages of these two classes of models.
2. In the first edition, a single chapter was devoted to marginal models and generalized estimating equations (GEE) that focused exclusively on binary and count data. We now devote two chapters (Chapters 12 and 13) to marginal models and GEE, with new material on models for ordinal responses, residual diagnostics, and issues that arise when modeling time-varying covariates.
3. A chapter (Chapter 15) on approximate methods for generalized linear mixed effects models discusses penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) methods. We highlight settings where these approximations are unlikely to be accurate and can yield biased estimates of effects.
4. A second chapter (Chapter 18) on missing data and dropout, focusing on multiple imputation and inverse probability weighting (IPW) methods, has been added. To give greater prominence to methods for accounting for missing data and dropout in longitudinal analyses, the two companion chapters (Chapters 17 and 18) now appear before the *Advanced Topics* part of the book.
5. A chapter (Chapter 19) on smoothing longitudinal data has been added to the *Advanced Topics*. This chapter focuses on the connection between penalized splines and linear mixed effects models.
6. A chapter on sample size and power (Chapter 20) has been added to the *Advanced Topics*. This chapter considers issues of sample size, power, number of repeated measurements, and study duration for longitudinal study designs.

In addition the chapter on residual analyses and diagnostics (Chapter 10) has been revised to include material on recently developed model-checking techniques based on cumulative sums of residuals and the chapters that review generalized linear models (Chapter 11) and generalized linear mixed effects models (Chapter 14) have been updated to include new material on models for ordinal data and on methods for handling overdispersion. Finally, extra problem sets have been added to many of the chapters.

As in the first edition, the prerequisites for a course based on this book are an introductory course in statistics and a strong background in regression analysis. Some previous exposure to generalized linear models (e.g., logistic regression) would be helpful, although these models are reviewed in detail in the text. An understanding of matrix algebra or calculus is not assumed. Although we do not assume a high level of mathematical preparation, we have written this book for the motivated reader who is willing to consider mathematical ideas. The more technical or mathematical sections of the book are signposted with asterisks and may be omitted at first reading without loss of continuity.



The methods described in this book require the use of appropriate statistical software. As before, we include illustrative SAS commands for performing the analyses presented throughout the text at the end of many chapters, with basic descriptions of their usage. Because many of the analyses we discuss can be performed using alternative software packages (e.g., *R*, *S-Plus*, *Stata*, and *SPSS*), this book can be supplemented with any one of them. Readers are encouraged to perform and verify the results of analyses using statistical software of their choice. Programming statements and computer output for selected examples, prepared using *SAS*, *Stata*, and *R*, can be downloaded from the website: [www.biostat.harvard.edu/~fitzmaur/ala2e](http://www.biostat.harvard.edu/~fitzmaur/ala2e). Because statistical software is constantly evolving, we will endeavor to update the website as new procedures become available in the major statistical software packages. The thirty-two real data sets used throughout the text and problem sets to illustrate the applications of longitudinal methods also can be downloaded from the website.

We hope this second edition of *Applied Longitudinal Analysis* provides a broader foundation in modern methods for the analysis of longitudinal data and will prove a worthy successor to the first edition. The original impetus for writing this book arose from teaching a graduate-level course on “Applied Longitudinal Analysis” at the Harvard School of Public Health. We are especially grateful to the students who have participated in the course since its inception almost twenty years ago; we have learned much from these extraordinary students. The collection of individuals who gave us useful feedback on the first edition is far too long to list. However, we would like to thank the many friends and colleagues who have helped us with this project. A special word of thanks to Amy Herring and Russell Localio. We thank Amy for her many helpful and constructive suggestions on how the book could be improved. We thank Russell for reading a draft of the new chapters and for providing invaluable feedback and suggestions that improved their content. Thanks also to Nick Horton, Stu Lipsitz, and Caitlin Ravichandran for their helpful suggestions and insightful comments on several chapters. Finally, we thank Steve Quigley and Susanne Steitz-Filler of Wiley, for their advice and encouragement during all stages of this project.

GARRETT M. FITZMAURICE  
NAN M. LAIRD  
JAMES H. WARE

*Boston, Massachusetts*  
*May, 2011*

# *Preface to First Edition*

Our goal in writing this book is to provide a rigorous and systematic description of modern methods for analyzing data from longitudinal studies. In recent years there have been remarkable developments in methods for longitudinal analysis. Despite these important advances, the methods have been somewhat slow to move into the mainstream. *Applied Longitudinal Analysis* bridges the gap between theory and application by presenting a comprehensive account of these methods in a way that is accessible to a wide range of readers.

The impetus for this book arose from teaching a graduate-level course on “Applied Longitudinal Analysis” at the Harvard School of Public Health. As course instructors, we were frustrated by the lack of a suitable textbook that adequately covered modern statistical methods for longitudinal analysis at a level accessible to a broad audience of researchers and graduate students in the health and medical sciences. We envision this book as a textbook for such a course and, subsequently, as a reference resource for researchers and graduate students. It is also suitable for graduate students in statistics and for statisticians already working in the health sciences, governmental health-related agencies, and the pharmaceutical industry. It is intended to allow a diverse group of statisticians, researchers, and graduate students in substantive fields to master modern methods for longitudinal data analysis.

The scope of this book is broad, covering methods for the analysis of diverse types of longitudinal data arising in the health sciences. The methods are presented in the setting of numerous applications to real data sets. Our main emphasis is on the practical rather than the theoretical aspects of longitudinal analysis. Twenty-five real data sets, drawn from studies in health-related fields, are

used throughout the text and problem sets to illustrate the applications of longitudinal methods. These data sets can be downloaded from the website for the book: [www.biostat.harvard.edu/~fitzmaur/ala](http://www.biostat.harvard.edu/~fitzmaur/ala). Although the methods are applied to data sets drawn from the health sciences, they apply equally to other areas of application, for example, education, psychology, and other branches of the behavioral and social sciences.

Because longitudinal data are a special case of clustered data, albeit with a natural ordering of the measurements within a cluster, we include also a description of modern methods for analyzing clustered data, more broadly defined. Indeed, one of our goals is to demonstrate that methods for longitudinal analysis are, more or less, special cases of more general regression methods for clustered data. As a result a comprehensive understanding of longitudinal data analysis provides the basis for a broader understanding of methods for analyzing the wide range of clustered data that commonly arises in studies in the biomedical and health sciences.

The prerequisites for a course based on this book are an introductory course in statistics and a strong background in regression analysis. Some previous exposure to generalized linear models (e.g., logistic regression) would be helpful, although these models are reviewed in the text. An understanding of matrix algebra or calculus is not assumed; the reader will be gently introduced to only those aspects of vector and matrix notation necessary for understanding the matrix representation of regression models for longitudinal data. Because vectors and matrices are used to simplify notation, the reader is required to attain some basic facility with the addition and multiplication of vectors and matrices. Although we do not assume a high level of mathematical preparation, a willingness to read and consider mathematical ideas is required. More technical or mathematical sections of the book are marked with asterisks and may be omitted at first reading without loss of continuity.

To use the methods described in this book, appropriate statistical software is required. In general, the methods available via commercially available software lag behind the recent advances in statistical methods; longitudinal data analysis is not exceptional in this regard. Recently the introduction of new programs for analyzing multivariate and longitudinal data has made these methods far more accessible to practitioners and students. We use *SAS*, which is widely available, to perform the analyses presented throughout the text. Illustrative *SAS* commands are included at the end of many of the chapters, with basic descriptions of their usage. Programming statements and computer output for the examples, prepared using *SAS*, can be downloaded from the website: [www.biostat.harvard.edu/~fitzmaur/ala](http://www.biostat.harvard.edu/~fitzmaur/ala). We selected *SAS* because all of the analyses we discuss can be performed using its procedures. Many of the methods can be carried out using alternative software packages (e.g., *S-Plus* and *Stata*) or special purpose programs (e.g., *BMDP5-V*) and this book can be supplemented with any one of them. Readers are encouraged to perform and verify the results of analyses using software of their choice. Because statistical software is constantly evolving, we anticipate that all of the methods we discuss will soon be available within most of the major statistical packages.

Throughout the text references have been kept to an absolute minimum. Instead, at the end of each chapter we include suggestions for further readings that provide

more in-depth coverage of certain topics. We also include “bibliographic notes” that highlight key references in the mainstream statistical literature. Although many of our readers may find the latter references to be too technical, they are included to give due credit to those who have contributed to the statistical methods described in each chapter.

Finally, we would like to thank the many friends and colleagues who have helped us to write this book. A special word of thanks to Misha Salganik, for preparation of the diagrams and many helpful suggestions for improvement of graphical displays. We are especially grateful to Joe Hogan and Russell Localio, for reading a first draft and providing invaluable feedback, comments, and suggestions that improved the book. We would also like to thank Rino Bellocco, Brent Coull, Nick Horton, Sharon-Lise Normand, Misha Salganik, Judy Singer, S. V. Subramanian, and Florin Vaida, for their insightful comments on several chapters. We are grateful to the students who have participated in the course on “Applied Longitudinal Analysis” at the Harvard School of Public Health since its inception; they have provided the impetus and motivation for writing this book. We gratefully acknowledge support from grant GM 29745 from the National Institutes of Health. The first author gratefully acknowledges support from the Junior Faculty Sabbatical Program at the Harvard School of Public Health; the support provided by a sabbatical created a unique opportunity to begin writing this book. Last, but not least, we thank Steve Quigley and Susanne Steitz of Wiley, for their advice and encouragement during all stages of this project.

GARRETT M. FITZMAURICE

NAN M. LAIRD

JAMES H. WARE

*Boston, Massachusetts*

*March, 2004*

# *Acknowledgments*

Throughout this book we have used data sets drawn from published studies in health-related fields to exemplify important concepts in the analysis of longitudinal and clustered data. We are grateful to the following investigators for sharing their data with us: Graham Benthall, Doug Dockery, Brian Flay, Robert Greenberg, Keith Henry, Aviva Must, Elena Naumova, George Rhoads, Jan Schouten, Linda Van Marter, and Gwen Zahner.

We also thank the following publishers for permission to reproduce published data sets in print and electronic format: The American Statistical Association, Blackwell Publishing, Brooks/Cole (a division of Thomson Learning), CRC Press, Elsevier, Iowa State Press, Oxford University Press, and SAS Institute, Inc.

Finally, in all data sets used throughout this book, the original subject identification (ID) numbers have been deleted and replaced with new subject ID numbers, to ensure that the data sets cannot be linked to the original records.