Preface

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 web site for the book: 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 web site: 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