Journal of Biopharmaceutical Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597263

A Review of: "Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis, by M. J. Daniels and J. W. Hogan"

David Ohlssen [a]
[a] Novartis Pharmaceuticals, One Health Plaza, East Hanover, NJ 07936, USA

Online Publication Date: 01 May 2009

To cite this Article Ohlssen, David(2009)'A Review of: "Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis, by M. J. Daniels and J. W. Hogan"',Journal of Biopharmaceutical Statistics,19:3,563 — 565
To link to this Article: DOI: 10.1080/10543400902802482
URL: http://dx.doi.org/10.1080/10543400902802482

Journal of Biopharmaceutical Statistics, 19: 563–565, 2009

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543400902802482

Taylor & Francis
Taylor & Francis Group

BOOK REVIEW

Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis, by M. J. Daniels and J. W. Hogan, Boca Raton: Chapman & Hall/CRC, 2008, ISBN 1-58488-609-9, xx + 303 pp., $79.95.

The appropriate way of handling missing data in longitudinal studies remains a topic of great interest in statistical research (e.g., Diggle et al., 2007). There has also been increased interest in how missing data should be handled in practice. For example, Carpenter and Kenward (2007) recently produced a freely available monograph, with accessible explanations and software, focusing on improving the handling of missing data in clinical trials. Some of the techniques described by Carpenter and Kenward, such as Rubin's multiple imputation method and the use of the computational power of Markov Chain Monte Carlo to fit selection models, certainly draw on the Bayesian paradigm. Nonetheless, the book under review appears to be the first reference that solely focuses on Bayesian approaches to handle missing data in longitudinal studies.

In the preface of the book, the authors argue that the subjective assumptions required when dealing with missing data provide strong motivation for the use of Bayesian approaches. They also emphasize eleven worked examples, which are used to illustrate the main concepts provided in the book. From a practical perspective, WinBUGS code is provided for some of these examples (although this was not available at the time of writing). The remainder of the book is split into two parts: the first covering the preliminary material (Chapters 1–4) and the second focusing on approaches to handling missing data (Chapters 5–10).

The first chapter describes a series of running examples. Most of the examples are from clinical trials, with the exception of one observational study examining HIV epidemiology. These examples seem to be well chosen and cover a wide range of situations (e.g., different outcome scales, dropout rates, interventions, and trial designs). One difficulty the authors must have faced is in deciding how much background material to include. Longitudinal modeling, covered by Chapter 2, and Bayesian statistics and computation, described in Chapter 3, are complex topics that have been covered in many other books. However, these chapters were clearly needed in order to state the notation used throughout the book and also to provide detail on key modeling techniques. As with all technical sections in the book, the style used is succinct mathematical notation, which allows fairly general models and concepts to be expressed in a small amount of space. I found this style helpful when reviewing the approaches that I was fairly familiar with (e.g., standard hierarchical modeling and the background on Bayesian statistics and computation) but more challenging when reviewing some of the more innovative modeling (e.g., marginal transition models).

Chapter 4, the final chapter in the first part of the book, focused on illustrating some of the models and concepts developed in Chapters 2 and 3 using the examples introduced in Chapter 1. Although, this chapter was fairly brief, it certainly helped my understanding of the modeling. The importance of model selection techniques, particularly for covariance structures, was rightly emphasized. Clearly, selecting appropriate covariance structures has a big influence on the robustness of missing data assumptions, such as missing at random (MAR). Overall, I think the background chapters (1–4) struck a good balance between providing enough material to allow the reader to use the book as a standalone reference and not overloading the book with preliminary material.

The second part of the book begins with Chapter 5, a technical introduction to missing data mechanisms that can be used for longitudinal data. Key concepts such as MAR, missing not at random, ignorability, and nonignorability are introduced as well as selection and pattern mixture models. Speaking from the perspective of an applied statistician, I found this chapter rather heavy going. However, it was clearly necessary to cover this material and the approach used was in keeping with the style of the monograph.

Chapter 6 focuses on inference under an ignorability assumption. A large emphasis is placed on the need for carefully tailored modeling of covariance structure. Motivation is provided by theory showing the consequences of misspecification. To me the highlight of this chapter is the work on serial correlation models and how the structure presented can be extended to covariate depended covariance structures. This approach is based on a series of research papers from the last 8 years and could well become a much more frequently used technique in applied statistics.

Chapter 7 compliments the previous chapter by applying a number of the techniques, under the assumption of ignorable missingness, to the running examples. This was certainly helpful in understanding how the various approaches could be applied in practice. One area for further research may be how to conduct model selection when various assumptions are under consideration. The authors generally applied the standard deviance information criterion. However, it is not entirely clear that this is the most appropriate approach and this could provide an area for further research.

The final three chapters of the book focus on the assumption of nonignorable missingness. In Chapter 8, selection models and pattern mixture models are considered in some detail. Whereas, Chapter 9 focuses on the use of informative priors for sensitivity analysis. The final chapter considers the application of methods for ignorable missingness to the running examples. These chapters are of a similar standard to the material on ignorable missingness with a mixture of mathematical model descriptions, areas of current research, and examples that give a good idea of how to use the methods in practice.

Overall I think this is a well written technical monograph. The preliminary sections on longitudinal data analysis, Bayesian statistics, and missing data have been covered elsewhere but are well written and serve to make this book a self-contained reference. The models presented to analyze missing data in longitudinal studies cover many ideas from the current literature, and some of the methods are at the cutting edge of research. The book will probably have greatest appeal to statisticians with a research interest in missing data. Although I also think applied

biostatisticians who like to use Bayesian approaches and in particular WinBUGS will find this book very useful.

REFERENCES

Carpenter, J., Kenward, M. G. (2007). *Missing Data in Clinical Trials—A Practical Guide*. Birmingham: National Health Service Coordinating Centre for Research Methodology. (Available from www.missingdata.org.uk.)

Diggle, P., Farewell, D., Henderson, R. (2007). Analysis of longitudinal data with drop-out: objectives, assumptions, and a proposal. *Applied Statistics* 56:499–550.

David Ohlssen
Novartis Pharmaceuticals,
One Health Plaza,
East Hanover, NJ 07936, USA