Preventing and Treating Missing Data in Longitudinal Clinical Trials

A Practical Guide

Recent decades have brought advances in statistical theory for missing data, which, combined with advances in computing ability, have allowed implementation of a wide array of analyses. In fact, so many methods are available that it can be difficult to ascertain when to use which method. This book focuses on the prevention and treatment of missing data in longitudinal clinical trials. Based on his extensive experience with missing data, the author offers advice on choosing analysis methods and on ways to prevent missing data through appropriate trial design and conduct. He offers a practical guide to key principles and explains analytic methods for the non-statistician using limited statistical notation and jargon. The book's goal is to present a comprehensive strategy for preventing and treating missing data, and to make available the programs used to conduct the analyses of the example dataset.

Craig H. Mallinckrodt is Research Fellow in the Decision Sciences and Strategy Group at Eli Lilly and Company. Dr. Mallinckrodt has supported drug development in all four clinical phases and in several therapeutic areas. He currently leads Lilly's Advanced Analytics hub for missing data and their Placebo Response Task Force, and is a member of a number of other scientific work groups. He has authored more than 170 papers, book chapters, and texts, including extensive works on missing data and longitudinal data analysis in journals such as *Statistics in Medicine*, *Pharmaceutical Statistics*, the *Journal of Biopharmaceutical Statistics*, the *Journal of Psychiatric Research*, the *Archives of General Psychiatry*, and *Nature*. He currently chairs the Drug Information Association's Scientific Working Group on Missing Data.

Practical Guides to Biostatistics and Epidemiology

This series of short and practical but authoritative books is for biomedical researchers, clinical investigators, public health researchers, epidemiologists, and non-academic and consulting biostatisticians who work with data from biomedical, epidemiological, and genetic studies. Some books explore a modern statistical method and its applications, others may focus on a particular disease or condition and the statistical techniques most commonly used in studying it.

The series is for people who use statistics to answer specific research questions. Books will explain the application of techniques, specifically the use of computational tools, and emphasize the interpretation of results, not the underlying mathematical and statistical theory.

Published in the series

Applied Multilevel Analysis, by **Jos W. R. Twisk**
Secondary Data Sources for Public Health, by **Sarah Boslaugh**
Survival Analysis for Epidemiologic and Medical Research, by **Steve Selvin**
Statistical Learning for Biomedical Data, by **James D. Malley**, **Karen G. Malley**, and **Sinisa Pajevic**
Measurement in Medicine, by **Henrica C.W. deVet**, **Caroline B. Terwee**, **Lidwine B. Mokkink, and Dirk L. Knol**
Genomic Clinical Trials and Predictive Medicine, by **Richard M. Simon**

Preventing and Treating Missing Data in Longitudinal Clinical Trials

A Practical Guide

Craig H. Mallinckrodt

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9781107679153

First published 2013

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Mallinckrodt, Craig, 1958–
Preventing and treating missing data in longitudinal clinical trials:
A practical guide / Craig Mallinckrodt.
pages cm – (Practical guides to biostatistics and epidemiology)
Includes bibliographical references and index.
ISBN 978-1-107-03138-8 (hardback) – ISBN 978-1-107-67915-3 (paperback)
1. Clinical trials – Longitudinal studies. 2. Medical sciences – Statistical methods.
3. Regression analysis – Data processing. I. Title.
R853.C55M3374 2013
610.72′4–dc23 2012038442

ISBN 978-1-107-03138-8 Hardback
ISBN 978-1-107-67915-3 Paperback

Contents

List of Figures

List of Tables

Acknowledgments

It has been my good fortune to collaborate with many excellent researchers in the field of missing data. These collaborations were of great benefit to this book. First, many thanks to those who were forced to read early versions of this book and provided valuable feedback: Geert Molenberghs (Universiteit Hasselt, Diepenbeek), Lei Xu (Eli Lilly, Indianapolis), and Adam Meyers (BioGen Idec, Boston).

For assistance in developing the programs used to analyze example data: Ilya Lipkovich (Quintiles, Indianapolis), Hank Wei (Eli Lilly, Indianapolis), Qun Lin (Eli Lilly, Indianapolis), and Dustin Ruff (Eli Lilly, Indianapolis).

For collaborations over the years that significantly influenced the content of this book: Caroline Beunckens (Universiteit Hasselt, Diepenbeek), James Carpenter (London School of Hygiene and Tropical Medicine), Raymond Carroll (Texas A&M University, College Station), Christy Chuang-Stein (Pfizer, New York), Scott Clark (Eli Lilly, Indianapolis), Mike Detke (MedAvante, Hamilton), Ivy Jansen (Universiteit Hasselt, Diepenbeek), Chris Kaiser (Eli Lilly, Indianapolis), Mike Kenward (London School of Hygiene and Tropical Medicine), Peter Lane (Glaxosmithkline, Harlow), Andy Leon (Weill Medical College, Cornell, New York), Stacy Lindborg (BioGen Idec, Boston), Rod Little (University of Michigan, Ann Arbor), James Roger (London School of Hygiene and Tropical Medicine), Steve Ruberg (Eli Lilly, Indianapolis), Shuyi Shen (Genentech, Ocenside), Cristina Sotto (Universiteit Hasselt, Diepenbeek), Birhanu Teshome (Universiteit Hasselt, Diepenbeek),

Herbert Thijs (Universiteit Hasselt, Diepenbeek), and Russ Wolfinger (SAS, Cary).

To Donna and Marissa, for your understanding and for helping make the time possible to work on this book, and for the encouragement and support needed to finish it!

Preface

This book focuses on the prevention and treatment of missing data in longitudinal clinical trials with repeated measures, such as are common in later phases of medical research and drug development. Recent decades have brought advances in statistical theory, which, combined with advances in computing ability, have allowed implementation of a wide array of analyses. In fact, so many methods are available that it can be difficult to ascertain when to use which method. A danger in such circumstances is to blindly use newer methods without proper understanding of their strengths and limitations, or to disregard all newer methods in favor of familiar approaches.

Moreover, the complex discussions on how to analyze incomplete data have overshadowed discussions on ways to prevent missing data, which would of course be the preferred solution. Therefore, preventing missing data through appropriate trial design and conduct is given significant attention in this book. Nevertheless, despite all efforts at prevention, missing data will remain an ever-present problem and analytic approaches will continue to be an important consideration.

Recent research has fostered an emerging consensus regarding the analysis of incomplete longitudinal data. Key principles and analytic methods are explained in terms non-statisticians can understand. Although the use of equations, symbols, and Greek letters to describe the analyses is largely avoided, sufficient technical detail is provided so readers can take away more than a peripheral understanding of the methods and issues. For those with in-depth statistical interests, reference to more technical literature is provided.

Part I begins with illustrations of how missing data erode the reliability and credibility of medical research. Subsequent chapters discuss missing

data mechanisms and the estimands (what is to be estimated) of interest in longitudinal trials with incomplete data. Part II covers trial design and conduct features that can help prevent missing data. Part III includes chapters on common methods of estimation, data and modeling considerations, and means of dealing with missing data (e.g., imputation). Part IV ties together the topics covered in Part III to illustrate various analyses applicable to incomplete longitudinal data. Small example data sets are used to illustrate and explain key analyses. An actual clinical trial data set is the focal point for proposing and implementing an overall analytic strategy that includes sensitivity analyses for assessing the impact of missing data.

This strategy is referred to as the analytic road map. A road map is different from driving instructions. Unlike driving directions, a road map does not chart a specific course, with instructions on exactly how far to go and when to turn. Instead, the road map lays out the alternatives so that the best route for a particular situation can be chosen.

The concluding chapter refocuses on the key issues covered throughout the book, presents a comprehensive strategy for preventing and treating missing data, and makes available the programs used to conduct the analyses of the example dataset.