

The background of the cover features a large iceberg floating in a blue ocean. The tip of the iceberg is visible above the water line, while the much larger, submerged portion is visible below. The text is overlaid on the submerged part of the iceberg.

# Clinical Trials with Missing Data

## A Guide for Practitioners

MICHAEL O'KELLY  
BOHDANA RATITCH

STATISTICS IN PRACTICE

WILEY

# Contents

[Cover](#)

[Series](#)

[Title Page](#)

[Copyright](#)

[Dedication](#)

[Preface](#)

[References](#)

[Acknowledgments](#)

[Notation](#)

[Table of SAS code fragments](#)

[Contributors](#)

[Chapter 1: What's the problem with missing data?](#)

[1.1 What do we mean by missing data?](#)

[1.2 An illustration](#)

[1.3 Why can't I use only the available primary endpoint data?](#)

[1.4 What's the problem with using last observation carried forward?](#)

[1.5 Can we just assume that data are missing at random?](#)

[1.6 What can be done if data may be missing not at random?](#)

[1.7 Stress-testing study results for robustness to missing data](#)

[1.8 How the pattern of dropouts can bias the outcome](#)

[1.9 How do we formulate a strategy for missing data?](#)

[1.10 Description of example datasets](#)

[Appendix 1.A: Formal definitions of MCAR, MAR and MNAR](#)

[References](#)

## [Chapter 2: The prevention of missing data](#)

[2.1 Introduction](#)

[2.2 The impact of “too much” missing data](#)

[2.3 The role of the statistician in the prevention of missing data](#)

[2.4 Methods for increasing subject retention](#)

[2.5 Improving understanding of reasons for subject withdrawal](#)

[Acknowledgments](#)

[Appendix 2.A: Example protocol text for missing data prevention](#)

## References

### Chapter 3: Regulatory guidance – a quick tour

3.1 International conference on harmonization guideline: Statistical principles for clinical trials: E9

3.2 The US and EU regulatory documents

3.3 Key points in the regulatory documents on missing data

3.4 Regulatory guidance on particular statistical approaches

3.5 Guidance about how to plan for missing data in a study

3.6 Differences in emphasis between the NRC report and EU guidance documents

3.7 Other technical points from the NRC report

3.8 Other US/EU/international guidance documents that refer to missing data

3.9 And in practice?

## References

### Chapter 4: A guide to planning for missing data

4.1 Introduction

4.2 Planning for missing data

4.3 Exploring and presenting missingness

4.4 Model checking

[4.5 Interpreting model results when there is missing data](#)

[4.6 Sample size and missing data](#)

[Appendix 4.A: Sample protocol/SAP text for study in Parkinson's disease](#)

[Appendix 4.B: A formal definition of a sensitivity parameter](#)

[References](#)

## [Chapter 5: Mixed models for repeated measures using categorical time effects \(MMRM\)](#)

[5.1 Introduction](#)

[5.2 Specifying the mixed model for repeated measures](#)

[5.3 Understanding the data](#)

[5.4 Applying the mixed model for repeated measures](#)

[5.5 Additional mixed model for repeated measures topics](#)

[5.6 Logistic regression mixed model for repeated measures using the generalized linear mixed model](#)

[References](#)

[Table of SAS Code Fragments](#)

## [Chapter 6: Multiple imputation](#)

[6.1 Introduction](#)

[6.2 Imputation phase](#)

[6.3 Analysis phase: Analyzing multiple imputed datasets](#)

[6.4 Pooling phase: Combining results from multiple datasets](#)

[6.5 Required number of imputations](#)

[6.6 Some practical considerations](#)

[6.7 Pre-specifying details of analysis with multiple imputation](#)

[Appendix 6.A: Additional methods for multiple imputation](#)

[References](#)

[Table of SAS Code Fragments](#)

## [Chapter 7: Analyses under missing-not-at-random assumptions](#)

[7.1 Introduction](#)

[7.2 Background to sensitivity analyses and pattern-mixture models](#)

[7.3 Two methods of implementing sensitivity analyses via pattern-mixture models](#)

[7.4 A “toolkit”: Implementing sensitivity analyses via SAS](#)

[7.5 Examples of realistic strategies and results for illustrative datasets of three indications](#)

[Appendix 7.A How one could implement the neighboring case missing value assumption using visit-by-visit multiple imputation](#)

[Appendix 7.B SAS code to model withdrawals from the experimental arm, using observed data from the control arm](#)

[Appendix 7.C SAS code to model early withdrawals from the experimental arm, using the last-observation-carried-forward-like values](#)  
[Appendix 7.D SAS macro to impose delta adjustment on a responder variable in the mania dataset](#)  
[Appendix 7.E SAS code to implement tipping point via exhaustive scenarios for withdrawals in the mania dataset](#)  
[Appendix 7.F SAS code to perform sensitivity analyses for the Parkinson's disease dataset](#)  
[Appendix 7.G SAS code to perform sensitivity analyses for the insomnia dataset](#)  
[Appendix 7.H SAS code to perform sensitivity analyses for the mania dataset](#)  
[Appendix 7.I Selection models](#)  
[Appendix 7.J Shared parameter models](#)  
[References](#)  
[Table of SAS Code Fragments](#)

## [Chapter 8: Doubly robust estimation](#)

[8.1 Introduction](#)  
[8.2 Inverse probability weighted estimation](#)  
[8.3 Doubly robust estimation](#)  
[8.4 Vansteelandt \*et al.\* method for doubly robust estimation](#)  
[8.5 Implementing the Vansteelandt \*et al.\* method via SAS](#)  
[Appendix 8.A How to implement Vansteelandt \*et al.\* method for mania dataset \(binary response\)](#)

[Appendix 8.B SAS code to calculate estimates from the bootstrapped datasets](#)

[Appendix 8.C How to implement Vansteelandt \*et al.\* method for insomnia dataset](#)

[References](#)

[Table of SAS Code Fragments](#)

[Bibliography](#)

[Index](#)

[Statistics in Practice](#)

# STATISTICS IN PRACTICE

*Series Advisors*

## **Human and Biological Sciences**

Stephen Senn

*CRP-Santé, Luxembourg*

## **Earth and Environmental Sciences**

Marian Scott

*University of Glasgow, UK*

## **Industry, Commerce and Finance**

Wolfgang Jank

*University of Maryland, USA*

## **Founding Editor**

Vic Barnett

*Nottingham Trent University, UK*

---

*Statistics in Practice* is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceuticals; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges. It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

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

# Clinical Trials with Missing Data

A Guide for Practitioners

**Michael O'Kelly**

*Senior Strategic Biostatistics Director  
Quintiles Ireland Ltd, Ireland*

**Bohdana Ratitch**

*Statistical Scientist, Quintiles, Montreal, Canada*

**WILEY**

This edition first published 2014

© 2014 John Wiley & Sons, Ltd

*Registered office*

John Wiley & Sons Ltd, The Atrium, Southern Gate,  
Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at [www.wiley.com](http://www.wiley.com).

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. 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 or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations 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. It is sold

on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

*Library of Congress Cataloging-in-Publication Data*

O'Kelly, Michael, author.

Clinical trials with missing data : a guide for practitioners / Michael O'Kelly, Bohdana Ratitch.

p. ; cm. – (Statistics in practice)

Includes bibliographical references and index.

ISBN 978-1-118-46070-2 (hardback)

I. Ratitch, Bohdana, author. II. Title. III. Series: Statistics in practice.

[DNLM: 1. Clinical Trials as Topic. 2. Bias (Epidemiology) 3. Models, Statistical. 4. Research Design. QV 771.4]

R853.C55

610.72'4-dc23

2013041088

A catalogue record for this book is available from the British Library.

ISBN: 978-1-118-46070-2

*To Raymond Kearns, teacher and Linda O’Nolan, partner.*

*—Michael O’Kelly*

*To my family, with love and gratitude for inspiration and  
support.*

*—Bohdana Ratitch*

# Preface

The aim of this book is to explain the difficulties that arise with the credibility and interpretability of clinical study results when there is missing data; and to provide practical strategies to deal with these difficulties. We try to do this in straightforward language, using realistic clinical trial examples.

This book is written to serve the needs of a broad audience of pharmaceutical industry professionals and regulators, including statisticians and non-statisticians, as well as academics with an interest in or need to understand the practical side of handling missing data. This book could also be used for a practical course in methods for handling missing data. For statisticians, this book provides mathematical background for a wide spectrum of statistical methodologies that are currently recommended to deal with missing data, avoiding unnecessary complexity. We also present a variety of examples and discussions on how these methods can be implemented using mainstream statistical software. The book includes a framework in which the entire clinical study team can contribute to a sound design of a strategy to deal with missing data, from prevention, to formulating clinically plausible assumptions about unobserved data, to statistical analysis and interpretation.

In the past, missing data was sometimes viewed as a problem that can be taken care of within statistical methodology without burdening others with the technicalities of it. While it is true that sophisticated statistical methods can and should be used to conduct sound analyses in the presence of missing data, all these methods make assumptions about missing data that clinical experts should help to formulate – assumptions that should

be clinically interpretable and plausible. Moreover, it is important to understand that some assumptions about missing data are *a/ways* being made, be it explicitly or implicitly. Even a strategy using only observed data for analysis carries within it certain implicit assumptions about subjects with missing data, and these assumptions are being implicitly made part of study conclusions. Clinicians fully participate in the effort to select carefully the type of data (clinical endpoints) that could best serve as evidence for efficacy and safety of a treatment. Their clinical expertise is invaluable for the choice of data that is collected in a clinical trial and subsequently used as observed data. Similarly, it is only natural to expect that the same level of clinical expertise would be provided to make choices for “hidden data” – the assumptions that would be used in place of missing data as an integral part of the overall body of evidence. Parts of this book (Chapters 1–4) contain non-technical material that can be easily understood by non-statisticians, and we hope that it will help clinicians and statisticians to build a common ground and a common language in order to tackle appropriately the problem of missing data together. Chapter 2 is dedicated entirely to prevention of missing data, which is the best way to deal with the problem, albeit not sufficient by itself in reality. Everyone involved in the planning and conduct of clinical trials would benefit from the ideas presented in this chapter.

Chapters 5 through 8 are aimed primarily at statisticians and cover well-understood methods that are presently regarded as statistically sound ways of conducting analyses in the presence of missing data and which can provide clinically meaningful estimands of treatment effect. In particular, this book covers direct likelihood methodology for longitudinal data with repeated correlated measurements; multiple imputation; pattern-mixture models; and inverse

weighting and doubly robust methods. We discuss in detail how these methodologies can be applied under a variety of clinical assumptions about unobserved data, both in the context of primary and sensitivity analyses. Aspects that are covered more briefly include selection models and non-parametric approaches. Examples cover both continuous outcomes and binary responses (e.g., treatment success/failure). Missing data problems in other contexts, such as time-to-event analyses, are not covered in this book.

Along with algebraic basics and plain language explanations of statistical methodology, this book contains numerous examples of practical implementations using SAS®. Throughout the book, as well as in supplemental material, we provide fragments of SAS code that would be sufficient for readers to use as templates or at least good starting points to implement all analyses mentioned in this book. We also provide pointers and explanations for a number of SAS macros publicly available at [www.missingdata.org.uk](http://www.missingdata.org.uk), developed by members of the Drug Information Association Scientific Working Group on Missing Data. Both authors of this book are members of this Working Group. We note that alternative software solutions exist in other programming environments, including free packages such as R. Other authors, for example, Carpenter and Kenward (2013) and van Buuren (2012), have provided tools that the readers would be able to use in order to implement general analysis principles discussed in this book.

Examples of realistic clinical trial data featured in this book provide illustrations of how reasonable missing data strategies can be designed in several different clinical indications, each with some specific challenges and characteristics. All examples have two treatment arms – experimental and control – but the methodology discussed

in this book can be applied in more general settings with more than two arms in a straightforward manner.

We have also endeavored to make the book suitable for casual use, allowing the professional statistician with a particular need to use a particular section without having to be familiar with the whole book. Therefore, each chapter begins with a list of key points covered; abbreviations are expanded on first appearance in each chapter; references are listed at the end of each chapter; explanations of particular points may be repeated if it helps to make a passage readable (although there are many cross-references between chapters too); where a book is referenced, we try to give page numbers if we think this might be helpful; and for some references to journal papers we also give web links to enable fast reference to abstracts and to enable downloading for those who may have electronic subscriptions.

Finally, we would like to stress that the problem of missing data unfortunately does not have a one-fits-all solution. A clinical research team must evaluate their strategy for missing data in the context of a specific clinical indication, subject population, expected mechanism of action of the experimental treatment, control treatment used in the study, and standards of care that would be available to subjects once they leave the trial. This book aims at providing the reader with a good general understanding of the issues involved and a tool box of methods from which to select the ones that would be the most appropriate for a study at hand.

## References

Carpenter JR, Kenward MG (2013) *Multiple Imputation and its Application*. John Wiley & Sons Ltd, West Sussex.

Van Buuren S (2012) *Flexible Imputation of Missing Data*. Chapman & Hall/CRC Press, Boca Raton, FL.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the United States and other countries.<sup>®</sup> indicates USA registration.

# Acknowledgments

We thank the contributors to this book, Sonia Davis, Sara Hughes, Belinda Hernández and Ilya Lipkovich, for their clear contributions and constant helpfulness.

We have found the scholars and experts on missing data to be friendly, approachable and willing to share ideas and expertise. As many who learn from him will tell you, James Roger epitomizes this spirit of willingness to spark ideas off others and to share the fruits of applied mathematics and elegant programming. It is likely that much of our work on sensitivity analyses in this book would not have been done without Roger's inspiration and example. It seems typical of those who work on missing data that much material from one of our favorite books on the subject was made freely available on the internet by authors James Carpenter and Mike Kenward. We thank these two scholars. Gary Koch first pointed us to Roger's ideas on sensitivity analyses; Gary also suggested the usefulness of sequential multiple imputation; over many years he has answered our questions on the direction of our work. Craig Mallinckrodt has genially chaired the Scientific Working Group for Missing Data and we thank him for fostering co-operation between pharmaceutical companies and academia, all with the aim of improving our handling of missing data. Geert Molenberghs gave thought-provoking answers to our queries; he also reviewed this book and we thank him for his helpful comments. Thank you to John Preisser, Willem Koetse, Forrest DeMarcus, and David Couper for reviews and contributions to Chapter 5. We thank Quintiles' Judith Beach for her legal review and general advice; we thank Quintiles' Kevin Nash for his review also. The errors that remain are ours. From our employers, Quintiles, we thank especially

Olga Marchenko, who championed the research; Tom Pike and King Jolly for their encouragement; and Andy Garrett and Yves Lalonde who supported Bohdana Ratitch in making time for the book. We also pay tribute to Quintiles for its remarkable support for the development of its employees – Michael O'Kelly owes his entire post-graduate education to the support of Quintiles, and in particular to the support of Imelda Parker when she was head of the Quintiles Dublin statistics department. Thanks to Ilaria Meliconi from Wiley who first encouraged us to think of writing the book; and to Wiley's Debbie Jupe who guided us as the book progressed, with help from Richard Davies and Heather Kay. Finally, we thank our spouses and family for their support.

# Notation

Throughout this book, algebraic notation will be as described below. Occasionally, the same symbol may be used for different purposes in different chapters following well-established conventions in respective domains. We will define specific meanings of such symbols in each relevant chapter; we note some variants in use here.

## Notation conventions

| Index letter    | Index range       | Description   |
|-----------------|-------------------|---|
| <b><i>i</i></b> | $1, \dots, N$     | Subject   |
| <b><i>t</i></b> | $1, \dots, T$     | Treatment allocated to the individual subject   |
| <b><i>j</i></b> | $1, \dots, J$     | Time point or visit number  |
| <b><i>c</i></b> | $1, \dots, T$     | Reference or control treatment, used when necessary to identify control vs. experimental treatment arm  |
| <b><i>l</i></b> | $0, \dots, L$     | Withdrawal visit: the visit at which the last observation is made for the individual subject (0 for those with no observed post-baseline data and $L = J$ for those with complete data) |
| <b><i>p</i></b> | $1, \dots, P$     | Pattern: group of subjects; can be defined in many ways, depending on analysis  |
| <b><i>s</i></b> | $0, \dots, S$     | Time-invariant (baseline) covariate; may be extended to include auxiliary covariables in multiple imputation, depending on context  |
| <b><i>v</i></b> | $1, \dots, V$     | Post-baseline covariate <b><i>V</i></b> is also used for variance-covariance matrix in Chapter 5  |
| <b><i>m</i></b> | $1, \dots, M$     | Imputation number for multiple imputation   |
| <b><i>k</i></b> | Context dependent | Indexes model parameters; range may be $0, \dots, S$ , $0, \dots, V$ or $0, \dots, S+V$ , depending on the context  |

## Notation conventions

| Variable letter | Highest index number(s) | Description |
|-----------------|-------------------------|-------------|
|                 |                         |             |

| Variable letter       | Highest index number(s) | Description  |
|-----------------------|-------------------------|--|
| <b><math>Y</math></b> | $N$ and $J$             | A set of outcome variables, e.g., $Y_j$ represents outcome at time point $j$ ; in the context of the imputation model for multiple imputation, may refer to both primary outcome and post-baseline auxiliary variables                           |
| <b><math>X</math></b> | $S$                     | A set of covariates; usually baseline covariates, but may also include observed post-baseline covariates, depending on the context   |
| <b><math>W</math></b> | Context dependent       | Effects included in a statistical model; may include $X$ , <b><math>Y</math></b> and their interactions  |
| $\beta$               | Context dependent       | Regression model coefficients  |
| $\theta$              | Context dependent       | Parameters of a statistical model, usually in a joint probability distribution   |
| <b><math>R</math></b> | $N$ and $J$             | A set of missingness indicators, e.g., $R_j = 0$ represents observation missing at time point $j$ , $R_j = 1$ represents observation available at time point $j$ <b><math>R</math></b> is also used for residuals covariance matrix in Chapter 5 |
| <b><math>F</math></b> | None                    | A set of missingness model covariates  |
| $\psi$                | None                    | Missingness model regression coefficients  |

### Additional conventions:

- Letter in upper case with an index refers to a variable at an individual visit (not its value), for example, in a context of a model,  $Y_1, \dots, Y_J$ .
- Letter in upper case without indices refers to a set of variables (not their values), for example,  $Y = (Y_1, \dots, Y_J)$ .
- Letter in lower case with indices refers to a value of an individual variable for an individual subject, for example,  $y_{ij}$  – value of a variable  $Y_j$  for subject  $i$ .
- Letter in lower case, with an index, and bolded refers either to
  - values of an individual variable for a set of subjects, for example,  $y_j = (y_{1j}, \dots, y_{Nj})$  – values of a variable  $Y_j$  for subjects  $i = 1, \dots, N$  or

- values of a set of variables for a subject, for example,  $y_i = (y_{i1}, \dots, y_{iJ})$  – values at all time points  $j = 1, \dots, J$  for subject  $i$ .  
Specific meaning is described in each context where this notation is used.
- Letter in either upper or lower case, without indices, and bolded refers to values of a set of variables for a set of subjects, for example,  **$Y$**  or  $y = (y_1, \dots, y_J)$  – data matrix, values of all variables  $Y$  for all subjects.
- **$Y_{obs}$**  and  **$Y_{mis}$**  refer to the set of observed values and the set of missing values of a data matrix  **$Y$** , respectively.

# Table of SAS code fragments

[5.1](#) Standard MMRM SAS code for Parkinson's disease example.

[5.2](#) MMRM with spatial power, random subject effect and sandwich estimator.

[5.3](#) GEE with compound symmetry working correlation matrix.

[5.4](#) MMRM SAS code for Parkinson's disease treatment × gender interaction at Visit 8.

[5.5](#) MMRM SAS code for Parkinson's disease cLDA model with baseline as a repeated measurement.

[5.6](#) Logistic GLMM SAS code for the mania study.

[5.7](#) Logistic GEE SAS code for the mania study.

[6.1](#) Using PROC MI to examine patterns of missingness of TST in the insomnia example.

[6.2](#) Imputation using monotone regression with a default model for each imputed variable corresponding to TST assessments in the insomnia example.

[6.3](#) Imputation using monotone regression with a user-specified model for each imputed variable corresponding to a TST assessment in the insomnia example.

[6.4](#) Imputation of TST using MSS assessments as ancillary variables in the insomnia example using monotone regression with a default model.

[6.5](#) Full imputation with the MCMC method for the UPDRS score in the Parkinson's disease example.

[6.6](#) Partial imputation using the MCMC method with the remaining imputations done by monotone regression for the UPDRS score in the Parkinson's disease example.

[6.7](#) An ANCOVA analysis by time point on multiply-imputed data for the insomnia example.

**6.8** MMRM analysis on multiply-imputed data for the insomnia example.

**6.9** Combining results from ANCOVA analysis of multiply-imputed data for the insomnia example.

**6.10** Combining results from ANCOVA analysis of multiply-imputed data using adjusted degrees of freedom for the insomnia example.

**6.11** Combining results from an MMRM analysis of multiply-imputed data for the insomnia example.

**6.12** Combining results from a logistic regression analysis on multiply-imputed responder status at Visit 5 (study Day 28) for the mania example.

**6.13** Combining results from a CMH test on multiply-imputed responder status at Visit 5 (study Day 28) for the mania example.

**6.14** Combining Mantel-Haenszel estimates of the common odds ratio for multiply-imputed responder status at Visit 5 (study Day 28) in the mania example.

**6.15** Combining estimates of responder proportions in each treatment arm and difference in proportions from multiply-imputed responder status at Visit 5 (study Day 28) in the mania example.

**6.16** Tests for no difference in baseline characteristics between completers and dropouts for the insomnia example.

**6.17** Longitudinal logistic model of dropout to explore the effects of baseline and post-baseline values on discontinuation in the insomnia example.

**6.18** Imputation using monotone regression with a model for each imputed variable corresponding to TST assessments in the insomnia example including treatment by outcome interactions for post-baseline visits used as predictors.

**6.19** Diagnostic plots of MCMC convergence for the insomnia example.

**6.20** Imputation using monotone regression and predictive mean matching methods for TST and MSS in the insomnia example.

**7.1** Simplified MI model to illustrate interaction by visit.

**7.2** Simplified direct likelihood MMRM model to illustrate interaction by visit.

**7.3** Use MCMC to impute non-monotone missing values.

**7.4** Use regression imputation to complete the MAR imputation, to match an MMRM where covariances are estimated across all treatment groups.

**7.5** Use regression imputation to complete the MAR imputation; the inclusion of interactions with treatment is consistent with an MMRM where covariances are estimated separately for each treatment group (as with the `GROUP=trt` option in PROC MIXED).

**7.6** Select observations so as to impute Visit 1 using model based on control arm.

**7.7** Impute Visit 1 using model based on control arm.

**7.8** Re-assemble the dataset to impute missing values at the next visit.

**7.9** Impute Visit 2 using model-based control arm.

**7.10** Create dataset with missing values for reference treatment imputed assuming MAR, as preparation for J2R.

**7.11** Impute distribution of control to Visit 2 value in experimental arm with baseline as the only covariable.

**7.12** Perform the MI assuming MAR for the control arm.

**7.13** Prepare to model the distribution of baseline values for BOCF-like imputation.

**7.14** Impute baseline-like values for the subject visits with missing outcomes.

**7.15** Transpose imputed values to obtain a dataset with one record per subject with all imputed values for a subject on the same record.

**7.16** Use baseline-distributed values for imputing the subject visits with missing post-baseline assessments.

**7.17** Select subjects with last observation at Visit 1 to impute LOCF values for Visits 2, 3, 4 and 5 for these subjects.

**7.18** Create LOCF-like values for subjects whose last observation was at Visit 1.

**7.19** Imputing last observation-distributed values, adapting the code from SAS Code Fragment 7.14.

**7.20** Imputing baseline-like values for subjects with reason for discontinuation = "AE."

**7.21** Imputing a base imputation before imposing a delta adjustment.

**7.22** Adjust imputed value by an amount  $\delta$ .

**7.23** Imposing a delta adjustment after all subjects have been imputed based on MAR assumption via MI with joint modeling as implemented by Roger.

**7.24** Imposing an adjustment to imputed binary response with a probability  $\delta$ .

**7.25** Call the *run\_tdelta* macro repeatedly to implement a tipping point analysis.

**8.1** Take a bootstrap sample.

**8.2** Simple macro to create missingness indicators.

**8.3** Code to calculate the probability of being observed at the first visit.

**8.4** Code to calculate the probability of being observed at the first or the second visit.

**8.5** Code to calculate the unconditional probability of being observed at each visit.

**8.6** Calculate the inverse probability of being observed.

**8.7** Convert the dataset to vertical format and run the weighted imputation model.

**8.8** Set the predicted outcome from the imputation model as the new response and run the final analysis model.

**8.9** Use LSMEANS to get the treatment effect at the final visit for each bootstrap sample and PROC MEANS to get the overall estimate of treatment effect.

**8.10** Imputation model for each bootstrap sample.

**8.11** Final analysis model to calculate the treatment change from baseline.

# Contributors

Chapter 5 by:

**Sonia M. Davis**, Collaborative Studies Coordinating Center, Professor of the Practice, Department of Biostatistics, University of North Carolina, USA.

Chapter 8 by:

**Belinda Hernández**, School of Mathematical Sciences (Discipline of Statistics) and the School of Medicine and Medical Science, University College Dublin, Ireland.

Chapter 2 by:

**Sara Hughes**, Head of Clinical Statistics, GlaxoSmithKline, UK.

Contributions to and review of Chapter 8 by:

**Ilya Lipkovich**, Center for Statistics in Drug Development, Quintiles Innovation, Morrisville, North Carolina, USA.

---

**Michael O'Kelly**: authored chapters 1, 3, 4 and 7, contributed research for Chapter 8, and reviewed all chapters.