Missing Data in Clinical Studies

Geert Molenberghs
Center for Statistics
Hasselt University
Diepenbeek, Belgium

Michael G. Kenward
Medical Statistics Unit
London School of Hygiene and Tropical Medicine
London, UK

John Wiley & Sons, Ltd
Missing Data in Clinical Studies
Statistics in Practice

Advisory Editor

Stephen Senn
University of Glasgow, UK

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 pharmaceutics; 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.
Missing Data in Clinical Studies

Geert Molenberghs
Center for Statistics
Hasselt University
Diepenbeek, Belgium

Michael G. Kenward
Medical Statistics Unit
London School of Hygiene and Tropical Medicine
London, UK
To Conny, An, and Jasper

To Pirkko
Contents

Preface xv

Acknowledgements xix

I Preliminaries 1

1 Introduction 3

1.1 From Imbalance to the Field of Missing Data Research 3
1.2 Incomplete Data in Clinical Studies 5
1.3 MAR, MNAR, and Sensitivity Analysis 8
1.4 Outline of the Book 9

2 Key Examples 11

2.1 Introduction 11
2.2 The Vorozole Study 12
2.3 The Orthodontic Growth Data 12
2.4 Mastitis in Dairy Cattle 14
2.5 The Depression Trials 14
2.6 The Fluvoxamine Trial 17
2.7 The Toenail Data 18
2.8 Age-Related Macular Degeneration Trial 20
2.9 The Analgesic Trial 22
2.10 The Slovenian Public Opinion Survey 24

3 Terminology and Framework 27

3.1 Modelling Incompleteness 27
3.2 Terminology 29
3.3 Missing Data Frameworks 30
3.4 Missing Data Mechanisms 31
3.5 Ignorability 33
3.6 Pattern-Mixture Models 34
## II Classical Techniques and the Need for Modelling 39

### 4 A Perspective on Simple Methods 41

4.1 Introduction 41

4.1.1 Measurement model 41

4.1.2 Method for handling missingness 42

4.2 Simple Methods 42

4.2.1 Complete case analysis 42

4.2.2 Imputation methods 43

4.2.3 Last observation carried forward 45

4.3 Problems with Complete Case Analysis and Last Observation Carried Forward 47

4.4 Using the Available Cases: a Frequentist versus a Likelihood Perspective 50

4.4.1 A bivariate normal population 50

4.4.2 An incomplete contingency table 52

4.5 Intention to Treat 53

4.6 Concluding Remarks 54

### 5 Analysis of the Orthodontic Growth Data 55

5.1 Introduction and Models 55

5.2 The Original, Complete Data 56

5.3 Direct Likelihood 57

5.4 Comparison of Analyses 59

5.5 Example SAS Code for Multivariate Linear Models 62

5.6 Comparative Power under Different Covariance Structures 63

5.7 Concluding Remarks 65

### 6 Analysis of the Depression Trials 67

6.1 View 1: Longitudinal Analysis 68

6.2 Views 2a and 2b and All versus Two Treatment Arms 72

## III Missing at Random and Ignorability 75

### 7 The Direct Likelihood Method 77

7.1 Introduction 77

7.2 Ignorable Analyses in Practice 78

7.3 The Linear Mixed Model 79

7.4 Analysis of the Toenail Data 82

7.5 The Generalized Linear Mixed Model 85

7.6 The Depression Trials 90

7.7 The Analgesic Trial 91

### 8 The Expectation–Maximization Algorithm 93

8.1 Introduction 93

8.2 The Algorithm 94

8.2.1 The initial step 94
8.2.2 The E step 95
8.2.3 The M step 95
8.3 Missing Information 95
8.4 Rate of Convergence 96
8.5 EM Acceleration 97
8.6 Calculation of Precision Estimates 98
8.7 A Simple Illustration 98
8.8 Concluding Remarks 103

9 Multiple Imputation 105
9.1 Introduction 105
9.2 The Basic Procedure 105
9.3 Theoretical Justification 107
9.4 Inference under Multiple Imputation 108
9.5 Efficiency 109
9.6 Making Proper Imputations 110
9.7 Some Roles for Multiple Imputation 115
9.8 Concluding Remarks 117

10 Weighted Estimating Equations 119
10.1 Introduction 119
10.2 Inverse Probability Weighting 120
10.3 Generalized Estimating Equations for Marginal Models 123
10.3.1 Marginal models for non-normal data 123
10.3.2 Generalized estimating equations 123
10.3.3 A method based on linearization 124
10.4 Weighted Generalized Estimating Equations 126
10.5 The Depression Trials 126
10.6 The Analgesic Trial 128
10.7 Double Robustness 130
10.8 Concluding Remarks 133

11 Combining GEE and MI 135
11.1 Introduction 135
11.2 Data Generation and Fitting 136
11.2.1 The Bahadur model 136
11.2.2 A transition model 137
11.3 MI-GEE and MI-Transition 137
11.4 An Asymptotic Simulation Study 137
11.4.1 Design 138
11.4.2 Results 139
11.5 Concluding Remarks 142

12 Likelihood-Based Frequentist Inference 145
12.1 Introduction 145
12.2 Information and Sampling Distributions 147
12.3 Bivariate Normal Data 149
12.4 Bivariate Binary Data 153
12.5 Implications for Standard Software 156
16.6.1 Derivations 223
16.6.2 Application to the vorozole study 224
16.7 A Clinical Trial in Alzheimer’s Disease 237
16.8 Analysis of the Fluvoxamine Trial 242
16.8.1 Selection modelling 242
16.8.2 Pattern-mixture modelling 243
16.8.3 Comparison 246
16.9 Concluding Remarks 246

17 Shared-Parameter Models 249

18 Protective Estimation 253
18.1 Introduction 253
18.2 Brown’s Protective Estimator for Gaussian Data 254
18.3 A Protective Estimator for Categorical Data 256
18.3.1 Likelihood estimation 260
18.3.2 Pseudo-likelihood estimation 263
18.3.3 Variance estimation 264
18.3.4 Analysis of artificial data 269
18.3.5 Analysis of the fluvoxamine trial 270
18.3.6 Presence or absence of colds 274
18.4 A Protective Estimator for Gaussian Data 275
18.4.1 Notation and maximum likelihood 275
18.4.2 Protective estimator 277
18.4.3 The six cities study 279
18.5 Concluding Remarks 282

V Sensitivity Analysis 283

19 MNAR, MAR, and the Nature of Sensitivity 285
19.1 Introduction 285
19.2 Every MNAR Model Has an MAR Bodyguard 286
19.2.1 A bivariate outcome with dropout 289
19.2.2 A trivariate outcome with dropout 290
19.2.3 A bivariate outcome with non-monotone missingness 291
19.3 The General Case of Incomplete Contingency Tables 292
19.3.1 A bivariate contingency table with dropout 293
19.3.2 A bivariate contingency table with non-monotone missingness 294
19.4 The Slovenian Public Opinion Survey 295
19.4.1 The BRD models 296
19.4.2 Initial analysis 296
19.4.3 BRD analysis 299
19.5 Implications for Formal and Informal Model Selection 302
19.6 Behaviour of the Likelihood Ratio Test for MAR versus MNAR 305
19.6.1 Simulated null distributions 306
19.6.2 Performance of bootstrap approaches 307
19.7 Concluding Remarks 311
## 20 Sensitivity Happens 313

20.1 Introduction 313
20.2 A Range of MNAR Models 314
20.3 Identifiability Problems 320
20.4 Analysis of the Fluvoxamine Trial 322
20.5 Concluding Remarks 327

## 21 Regions of Ignorance and Uncertainty 329

21.1 Introduction 329
21.2 Prevalence of HIV in Kenya 330
21.3 Uncertainty and Sensitivity 330
21.4 Models for Monotone Patterns 331
21.5 Models for Non-monotone Patterns 332
21.6 Formalizing Ignorance and Uncertainty 333
21.7 Analysis of the Fluvoxamine Trial 338
21.7.1 Identified models 339
21.7.2 Sensitivity analysis 341
21.8 Artificial Examples 345
21.9 The Slovenian Public Opinion Survey 348
21.10 Some Theoretical Considerations 351
21.11 Concluding Remarks 351

## 22 Local and Global Influence Methods 353

22.1 Introduction 353
22.2 Gaussian Outcomes 354
22.2.1 Application to the Diggle–Kenward model 356
22.2.2 The special case of three measurements 359
22.3 Mastitis in Dairy Cattle 360
22.3.1 Informal sensitivity analysis 361
22.3.2 Local influence approach 367
22.4 Alternative Local Influence Approaches 373
22.5 The Milk Protein Content Trial 375
22.5.1 Informal sensitivity analysis 377
22.5.2 Formal sensitivity analysis 386
22.6 Analysis of the Depression Trials 398
22.7 A Local Influence Approach for Ordinal Data with Dropout 405
22.8 Analysis of the Fluvoxamine Data 406
22.9 A Local Influence Approach for Incomplete Binary Data 410
22.10 Analysis of the Fluvoxamine Data 411
22.11 Concluding Remarks 415

## 23 The Nature of Local Influence 417

23.1 Introduction 417
23.2 The Rats Data 418
23.3 Analysis and Sensitivity Analysis of the Rats Data 419
23.4 Local Influence Methods and Their Behaviour 422
23.4.1 Effect of sample size 423
23.4.2 Pointwise confidence limits and simultaneous confidence bounds for the local influence measure 424
23.4.3 Anomalies in the missingness mechanism 425
23.4.4 Anomalies in the measurement model 428
23.5 Concluding Remarks 430

24 A Latent-Class Mixture Model for Incomplete Longitudinal Gaussian Data 431

24.1 Introduction 431
24.2 Latent-Class Mixture Models 431
24.3 The Likelihood Function and Estimation 434
  24.3.1 Likelihood function 434
  24.3.2 Estimation using the EM algorithm 436
  24.3.3 The E step 437
  24.3.4 The M step 438
  24.3.5 Some remarks regarding the EM algorithm 439
24.4 Classification 440
24.5 Simulation Study 441
  24.5.1 A simplification of the latent-class mixture model 441
  24.5.2 Design 442
  24.5.3 Results 443
24.6 Analysis of the Depression Trials 446
  24.6.1 Formulating a latent-class mixture model 446
  24.6.2 A sensitivity analysis 449
24.7 Concluding Remarks 450

VI Case Studies 451

25 The Age-Related Macular Degeneration Trial 453

25.1 Selection Models and Local Influence 453
25.2 Local Influence Analysis 455
25.3 Pattern-Mixture Models 458
25.4 Concluding Remarks 459

26 The Vorozole Study 461

26.1 Introduction 461
26.2 Exploring the Vorozole Data 461
  26.2.1 Average evolution 461
  26.2.2 Variance structure 464
  26.2.3 Correlation structure 465
  26.2.4 Missing data aspects 466
26.3 A Selection Model for the Vorozole Study 471
26.4 A Pattern-Mixture Model for the Vorozole Study 475
26.5 Concluding Remarks 481

References 483

Index 497
Preface

Three quarters of a century ago, Karl Pearson, Sir Ronald Fisher, Egon Pearson, and a good number of their contemporaries were concerned about the loss of balance that could arise if data collected from carefully designed experiments turned out to be incomplete. Their concern was wholly appropriate in the pre-automated computing era, when balance was an integral part of manageable statistical analyses. Following the Second World War there was a great increase in the development and practice of experimental and observational studies in human subjects. At the same time the regulation of such studies became – quite understandably, with the aberrations of the war fresh in mind – much more strict. Some decades later, the key paper of Rubin (1976) established incomplete data as a proper field of study in its own right within the domain of statistics. Subsequent major contributions in this field were the expectation–maximization algorithm (Dempster et al. 1977) and multiple imputation (Rubin 1987), to name but two. It can fairly be said that Little and Rubin’s (1987) book, the first monograph encompassing the subject, marked the coming of age of the analysis of incomplete data as a field of research.

The intervening two decades have seen a tremendous amount of research output on the problem of missing data. It is possible to distinguish from this output several different strands. First, considerable attention has been paid to the accommodation of incompleteness in epidemiological and clinical studies in a practically accessible manner that provides due protection to the interests of both industry and the general public. This is particularly true for clinical trials within regulatory processes, with a variety of interested parties, including the regulatory authorities, the biopharmaceutical companies, and the clinical investigators, to name but a few, actively contributing to the debate. Second, methodological developments have continued apace, some of a highly sophisticated nature. We can usefully distinguish between a parametric school, essentially of a likelihood and Bayesian nature, with landmark contributions from Don Rubin, Rod Little, and many of their students and collaborators, and a semi-parametric school, built around Horvitz–Thompson and estimating equations ideas, with seminal contributions from Jamie Robins, Andrea Rotnitzky, and a variety of their students and co-workers.
There is noticeable divergence between the various lines of thinking. Ideally, there should be continuing dialogue, ranging from the very applied to the most theoretical researcher in the field, regardless of modelling and inferential preferences. At the same time, there is a broad consensus that no single modelling approach, no matter how sophisticated, can overcome the limitation of simply not having access to the missing data. All parties – academia, industry, and regulatory authorities – stress the need for, and importance of, sensitivity analysis. By this we mean, in a broad sense, the formal or informal assessment of the impact of incompleteness on key statistical inferences. Scientific research, standard operating procedures, and guidelines, such as International Conference on Harmonisation Guideline E9, all point in this same direction. There is considerably less agreement, however, as to how such sensitivity analyses should be conceived, conducted and presented in practice, and this is not surprising given both the early stage of, and feverish activity within, this particular area of missing data research.

A key prerequisite for a method to be embraced, no matter how important, is the availability of trustworthy and easy-to-use software, preferably in commercially or otherwise generally accessible *bona fide* packages. It is not an exaggeration to claim that the last decade has seen tremendous progress on this front. Such methods and techniques as direct likelihood and Bayesian analysis, multiple imputation, the expectation–maximization algorithm, and weighted estimating equations are appearing in SAS, Stata, S-Plus, R, and various other mainstream statistical packages.

These observations, developed through numerous lectures, shorter and longer courses, and through statistical consultancy with academic, biopharmaceutical, and governmental partners, led to the conception of this book. The opportunity to interact with audiences of various backgrounds and with a broad range of interests, across the globe, has been an invaluable educational experience for us. In turn, we hope the book will be of value to a wide audience, including applied statisticians and biomedical researchers, in particular in the biopharmaceutical industry, medical and public health research organizations, contract research organizations, and academic departments. We have chosen an explanatory rather than a research-oriented style of exposition, even though some chapters contain advanced material. A perspective is given in Chapter 1. We focus on practice rather than on mathematical rigour, and this is reflected in the large number of worked examples dotted throughout the text.

Many of the statistical analyses have been performed using such SAS procedures as MIXED, GENMOD, GLIMMIX, NLMIXED, MI, and MIANALYZE, with the addition of user-defined SAS macros, as well as functions in other packages, such as GAUSS, where the necessary facilities have not yet been implemented elsewhere. In spite of this, the methodological development and the analyses of the case studies alike are presented in a software-independent fashion. Illustrations on how to use SAS for a selected collection of model strategies are confined to a small number of chapters and sections, implying that
the text can be read without problem if these software excursions are ignored. Selected programs, macros, output, and publicly available data sets can be found on Wiley’s website, as well as on the authors’ site (www.uhasselt.be/censtat).

Geert Molenberghs (Diepenbeek) and Michael G. Kenward (London)
Acknowledgements

This text has benefited tremendously from stimulating interaction and joint work with a number of colleagues. We are grateful to many for their kind permission to use their data. We gratefully acknowledge the support of: Marc Aerts (Universiteit Hasselt, Diepenbeek), Caroline Beunckens (Universiteit Hasselt, Diepenbeek), Luc Bijnens (Johnson & Johnson Pharmaceutical Research and Development, Beerse), Marc Buyse (International Drug Development Institute, Ottignies-Louvain-la-Neuve), James Carpenter (London School of Hygiene and Tropical Medicine), Raymond Carroll (Texas A&M University, College Station), Desmond Curran (Icon Clinical Research, Dublin), Stephen Evans (London School of Hygiene and Tropical Medicine), Garrett Fitzmaurice (Harvard School of Public Health, Boston), Els Goetghebeur (Universiteit Gent), Niel Hens (Universiteit Hasselt, Diepenbeek), Joseph Ibrahim (University of North Carolina, Chapel Hill), Ivy Jansen (Universiteit Hasselt, Diepenbeek), Emmanuel Lesaffre (Katholieke Universiteit Leuven), Stuart Lipsitz (Harvard School of Public Health, Boston), Rod Little (University of Michigan, Ann Arbor), Craig Mallinckrodt (Eli Lilly & Company, Indianapolis), Bart Michiels (Johnson & Johnson Pharmaceutical Research and Development, Beerse), James Roger (GlaxoSmithKline), Cristina Sotto (Universiteit Hasselt, Diepenbeek), Herbert Thijs (Universiteit Hasselt, Diepenbeek), Butch Tsiatis (North Carolina State University, Raleigh), Tony Vangeneugden (Tibotec, Mechelen), Stijn Vansteelandt (Universiteit Gent), Kristel Van Steen (Universiteit Gent), and Geert Verbeke (Katholieke Universiteit Leuven).

The feedback we received from our regular and short course audiences has been invaluable. We are grateful for such interactions in Argentina (Corrientes), Australia (Cairns, Coolangatta), Belgium (Beerse, Braine-l’Alleud, Brussels, Diepenbeek, Gent, Leuven, Wavre), Brazil (Londrina, Piracicaba), Canada (Toronto), Cuba (Havana, Varadero), Denmark (Copenhagen), Finland (Jokioinen, Tampere, Turku), France (Paris, Marseille, Toulouse, Vannes), Germany (Freiburg, Heidelberg), Greece (Athens), Ireland (Dublin), Korea (Seoul), the Netherlands (Rotterdam), New Zealand (Auckland, Christchurch, Hamilton), Spain (Barcelona, Pamplona, Santiago de Compostela), South Africa (Stellenbosch), Switzerland (Neuchâtel, Basle), the United Kingdom
Acknowledgements

(London, Manchester, Harlow, Sandwich, Stevenage, Sunningdale), and the United States of America (Ann Arbor, Arlington, Atlanta, Atlantic City, Minneapolis, New Jersey, New York City, Rockville, San Francisco, Seattle, Tampa, Washington, DC).

Several people have helped us with the computational side of the models presented. We mention in particular Caroline Beunckens, Ivy Jansen, Bart Michiels, Oliver Schabenberger (SAS Institute, Cary, North Carolina), Cristina Sotto, Herbert Thijs, and Kristel Van Steen.

We gratefully acknowledge support from Research Project Fonds voor We-tenschappelijk Onderzoek Vlaanderen G.0002.98, ‘Sensitivity Analysis for Incomplete Data’, NATO Collaborative Research Grant CRG 950648, ‘Statistical Research for Environmental Risk Assessment’, Belgian IUAP/PAI network ‘Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data’, and US grants HL 69800, AHRQ 10871, HL 52329, HL 61769, GM 29745, MH 54693, CA 57030 and CA 70101 from the US National Institutes of Health, the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P3–ES09106), from the UK Economic and Social Research Council and National Health Service Research and Development Methodology Programme.

All along, it has been a fine experience working with our colleagues at John Wiley.

We are indebted to Conny, An, Jasper, and Pirkko, for their understanding and for time not spent with them while preparing this volume. Working on this book has been a period of close collaboration and stimulating exchange, which we will remember with affection for years to come.

Geert and Mike
Kessel-Lo, Belgium and Luton, England
September 2006
Part I

Preliminaries
In this chapter we give a broad introduction to the problem of missing data. We provide a perspective on the topic, reviewing the main developments of the last century (Section 1.1), in the process paying special attention to the setting of clinical studies (Section 1.2). We examine the move towards more principled approaches, more elaborate modelling strategies and, most recently, the important role of sensitivity analysis (Section 1.3). Finally, we map out the developments and material that make up rest of the book (Section 1.4). In the next chapter we introduce the key sets of data that will be used throughout the book to illustrate the analyses.

1.1 FROM IMBALANCE TO THE FIELD OF MISSING DATA RESEARCH

It is very common for sets of quantitative data to be incomplete, in the sense that not all planned observations are actually made. This is especially true when studies are conducted on human subjects. Examples abound in epidemiologic studies (Piantadosi 1997; Clayton and Hills 1993; Green et al. 1997; Friedman et al. 1998), in clinical trials (Kahn and Sempos 1989; Lilienfeld and Stolley 1994; Selvin 1996), and in the social sciences, especially in sample surveys, psychometry, and econometrics (Fowler 1988, Schafer et al. 1993; Rubin 1987; Rubin et al. 1995), to name but a few areas.

Our focus in this book is on intervention-based clinical studies. We mean this in an inclusive sense, however, implying that the methodology presented may be appropriate outside this setting, for example in the context of epidemiological studies as well as experimental and observational data in non-human life sciences, including agricultural, biological, and environmental research.
Early work on the problem of missing data, especially during the 1920s and 1930s, was largely confined to algorithmic and computational solutions to the induced lack of balance or deviations from the intended study design. See, for example, the reviews by Afifi and Elashoff (1966) and Hartley and Hocking (1971). In the last quarter of the twentieth century, general algorithms, such as the expectation–maximization (EM: Dempster et al. 1977), and data imputation and augmentation procedures (Rubin 1987; Tanner and Wong 1987), combined with powerful computing resources, largely provided a solution to this aspect of the problem.

Rubin (1976) provided a formal framework for the field of incomplete data by introducing the important taxonomy of missing data mechanisms, consisting of missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). An MCAR mechanism potentially depends on observed covariates, but not on observed or unobserved outcomes. An MAR mechanism depends on the observed outcomes and perhaps also on the covariates, but not further on unobserved measurements. Finally, when an MNAR mechanism is operating, missingness does depend on unobserved measurements, perhaps in addition to dependencies on covariates and/or on observed outcomes. During the same era, the selection model, pattern-mixture model, and shared-parameter model frameworks were established. These are depicted schematically in Figure 1.1. In a selection model, the joint distribution of the $i$th subject’s outcomes, denoted $Y_i$, and vector of missingness indicators, written $R_i$, is factored as the marginal outcome distribution and the conditional distribution of $R_i$ given $Y_i$. A pattern-mixture approach starts from the reverse factorization. In a shared-parameter model, a set of latent variables, latent classes, and/or random effects is assumed to drive both the $Y_i$ and $R_i$ processes. An important version of such a model further asserts that, conditional on the latent variables, $Y_i$ and $R_i$ exhibit no further dependence. Rubin (1976) contributed the concept of ignorability, stating that under precise conditions, the missing data mechanism can be ignored when interest lies in inferences about the measurement process. Combined with regularity conditions, ignorability applies to MCAR and MAR combined, when likelihood or Bayesian inference routes are chosen, but the stricter MCAR condition is required for frequentist inferences to be generally valid. A final distinction is made between missingness patterns. Dropout or attrition refers to the specific situation, arising in longitudinal studies, where subjects are observed without uninterruption from the beginning of the study until a given point in time, perhaps prior to the scheduled end of the study, when they drop out and do not return to the study. Given a rather strong focus in this book on longitudinal studies, dropout, an indicator of which is denoted by $D_i$, will occupy a prominent position. The general mechanism, where subjects can be observed and missing on any partition of the set of planned measurement occasions, is often called non-monotone


\[ f(Y_i, R_i | X_i, \theta, \psi) \]

Selection models: \( f(Y_i | X_i, \theta) f(R_i | X_i, Y_i^0, Y_i^m, \psi) \)

\[
\begin{array}{ccc}
\text{MCAR} & \rightarrow & \text{MAR} & \rightarrow & \text{MNAR} \\
\text{LOCF} & \rightarrow & \text{EM} & \rightarrow & \text{MI} \\
\text{WGEE} & \rightarrow & \text{pattern-mixture models: } f(Y_i | X_i, R_i, \theta) f(R_i | X_i, \psi) \\
\text{shared-parameter models: } f(Y_i | X_i, b_i, \theta) f(R_i | X_i, b_i, \psi) \\
\end{array}
\]

Figure 1.1 Schematic representation of the missing data frameworks and mechanisms, together with simple and more advanced methods, as well as sensitivity analysis. (MCAR, missing completely at random; MAR, missing at random; MNAR, missing not at random; CC, complete case analysis; LOCF, last observation carried forward; EM, expectation–maximization algorithm; MI, multiple imputation; WGEE, weighted generalized estimating equations.)

missingness. These and additional concepts are formalized and expanded upon in Chapter 3.

1.2 INCOMPLETE DATA IN CLINICAL STUDIES

In clinical trials, dropout is not only a common occurrence, there are also specific procedures for reporting and subsequently dealing with it. Patients who drop out of a clinical trial are usually listed on a separate withdrawal sheet of the case record form, with the reasons for withdrawal entered by the authorized investigator. Reasons frequently encountered are adverse events, illness not related to study medication, an uncooperative patient, protocol violation, and ineffective study medication. Further specifications may include so-called loss
to follow-up. Based on this medically inspired typology, Gould (1980) proposed specific methods to handle this type of incompleteness.

Even though the primary focus of such trials is often on a specific time of measurement, usually the last, the outcome of interest is recorded in a longitudinal fashion, and dropout is a common occurrence. While dropout, in contrast to non-monotone missingness, may simplify model formulation and manipulation, the causes behind it can be more problematic. For example, dropout may derive from lack of efficacy, or from potentially serious and possible treatment-related side effects. In contrast, an intermittently missing endpoint value may be due more plausibly to the patient skipping a visit for practical or administrative reasons, to measurement equipment failure, and so on. In addition, one often sees that incomplete sequences in clinical trials are, for the vast majority, of a dropout type, with a relatively minor fraction of incompletely observed patients producing non-monotone sequences. For all of these reasons we will put major emphasis on the problem of dropout, although not entirely neglecting non-monotone missingness in the process.

In a strict sense the conventional justification for the analysis of data from a randomized trial is removed when data are missing for reasons outside the control of the investigator. Before one can address this problem, however, it is necessary to establish clearly the purpose of the study (Heyting et al. 1992). If one is working within a pragmatic setting, the event of dropout, for example, may well be a legitimate component of the response. It may make little sense to ask what response the subject would have shown had they remained on study, and the investigator may then require a description of the response conditional on a subject remaining in the trial. This, together with the pattern of missingness encountered, may then be the appropriate and valid summary of the outcome. We might call this a conditional description. Shih and Quan (1997) argue that such a description will be of more relevance in many clinical trials. On the other hand, from a more explanatory perspective, one might be interested in the behaviour of the responses that occurred irrespective of whether we were able to record them or not. This might be termed a marginal description of the response. For a further discussion of intention to treat and explanatory analyses in the context of dropout see Heyting et al. (1992) and Little and Yau (1996), as well as Section 4.5 of this volume. It is commonly suggested (Shih and Quan 1997) that such a marginal representation is not meaningful when the nature of dropout (e.g., death) means that the response cannot subsequently exist, irrespective of whether it is measured. While such dropout may in any particular setting imply that a marginal model is not helpful, it does not imply that it necessarily has no meaning. Provided that the underlying model does not attach a probability of one to dropout for a particular patient, then non-dropout and subsequent observations are an outcome consistent with the model and logically no different from any other event in a probability model. Such distinctions, particularly with respect to the conditional analysis, are complicated by the inevitable mixture of causes behind missing values. The conditional description
Incomplete Data in Clinical Studies

is a mirror of what has been observed, and so its validity is less of an issue than its interpretation. In contrast, other methods of handling incompleteness make some correction or adjustment to what has been directly observed, and therefore address questions other than those corresponding to the conditional setting. In seeking to understand the validity of these analyses we need to compare their consequences with their aims.

Two simple, common approaches to analysis are (1) to discard subjects with incomplete sequences and (2) simple imputation. The first approach has the advantage of simplicity, although the wide availability of more sophisticated methods of analysis minimizes the significance of this. It is also an inefficient use of information. In a trivial sense it provides a description of the response conditional on a subject remaining in the trial. Whether this reflects a response of interest depends entirely on the mechanism(s) generating the missing values and the aims of the trial. It is not difficult to envisage situations where it can be very misleading, and examples of this exist in the literature (Kenward et al. 1994, Wang-Clow et al. 1995). Such imputation methods share the same drawbacks, although not all to the same degree. The data set that results will mimic a sample from the population of interest, itself determined by the aims of the analysis, only under particular and potentially unrealistic assumptions. Further, these assumptions depend critically on the missing value mechanism(s). For example, under certain dropout mechanisms the process of imputation may recover the actual marginal behaviour required while under other mechanisms it may be wildly misleading, and it is only under the simplest and most ignorable mechanisms that the relationship between imputation procedure and assumption is easily deduced. Little (1994a) gives two simple examples where the relationship is clear.

We therefore see that when there are missing values, simple methods of analysis do not necessarily imply simple, or even accessible, assumptions, and without understanding properly the assumptions being made in an analysis we are not in a position to judge its validity or value. It has been argued that while any particular ad hoc analysis may not represent the true picture behind the data, a collection of such analyses should provide a reasonable envelope within which the truth should lie. Even this claim is open to major criticisms, however, and we return to such ideas when sensitivity analyses are considered in Part III. In Chapter 4, after formally introducing of terminology and the necessary frameworks in Chapter 3, we provide a detailed examination of the advantages and drawbacks of simple methods, especially with a view to clinical trial practice.

As we explain in Chapter 4, it is unfortunate that so much emphasis has been given to methods such as last observation carried forward (LOCF), complete case analysis (CC), or simple forms of imputation. These are ad hoc methods defined procedurally in terms of manipulation of the data, rather than derived in a statistically principled way from the design of the trial and the aims of the analysis. As a consequence the relationship between their validity
Introduction

and underlying assumptions can be far from clear and, when the relevant assumptions can be identified, they are seen to be very strong and unrealistic. In the LOCF procedure the missing measurements are replaced by the last one available. In particular, even the strong MCAR assumption does not suffice to guarantee that an LOCF analysis is valid. On the other hand, under MAR, valid inferences can be obtained through a likelihood-based or Bayesian analysis, without the need for modelling the dropout process. As a consequence, one can simply use, for example, linear or generalized linear mixed models (Verbeke and Molenberghs 2000; Molenberghs and Verbeke 2005), without additional complication or effort. This does not imply that these particular analyses are appropriate for all questions that might be asked of trial data, but the clarity of the underlying assumptions means that appropriate modifications can be readily identified when non-MAR analyses are called for, for example with intention to treat (ITT) analyses when dropout is associated with termination of treatment.

We will argue in Chapter 4, through the cases studies in Chapters 5 and 6, and then further throughout Part III, that such MAR-based likelihood analyses not only enjoy much wider validity than the simple methods but, moreover, are simple to conduct, without additional data manipulation, using such tools as the SAS procedures MIXED, GLIMMIX, or NLMIXED. Thus, clinical trial practice should shift away from the ad hoc methods and focus on likelihood-based ignorable primary analyses instead. As will be argued further, the cost involved in having to specify a model will arguably be mild to moderate in realistic clinical trial settings. Thus, we promote the use of direct likelihood ignorable methods and demote the use of the LOCF and CC approaches. Mallinckrodt et al. (2003a, 2003b), Molenberghs et al. (2004), and Lavori et al. (1995) propose direct likelihood and multiple imputation methods, respectively, to deal with incomplete longitudinal data. Siddiqui and Ali (1998) compare direct likelihood and LOCF methods.

1.3 MAR, MNAR, AND SENSITIVITY ANALYSIS

From the previous section, it is clear that not only is it advisable to avoid simple ad hoc methods such as complete case analysis and last observation carried forward, but there exists more appropriate flexible, broadly valid and widely implemented methodology. Principled methods and techniques such as direct likelihood and Bayesian analyses, the EM algorithm, multiple imputation, and weighted generalized estimating equations are systematically reviewed in Part IV. All of these methods are valid under the relatively relaxed assumption of MAR.

At the same time, it is important to consider reasons for departures from MAR, and the possible consequences of this for the conclusions reached. One obvious example, mentioned above, concerns treatment termination among dropouts in an ITT analysis. More generally, the reasons for, and implications