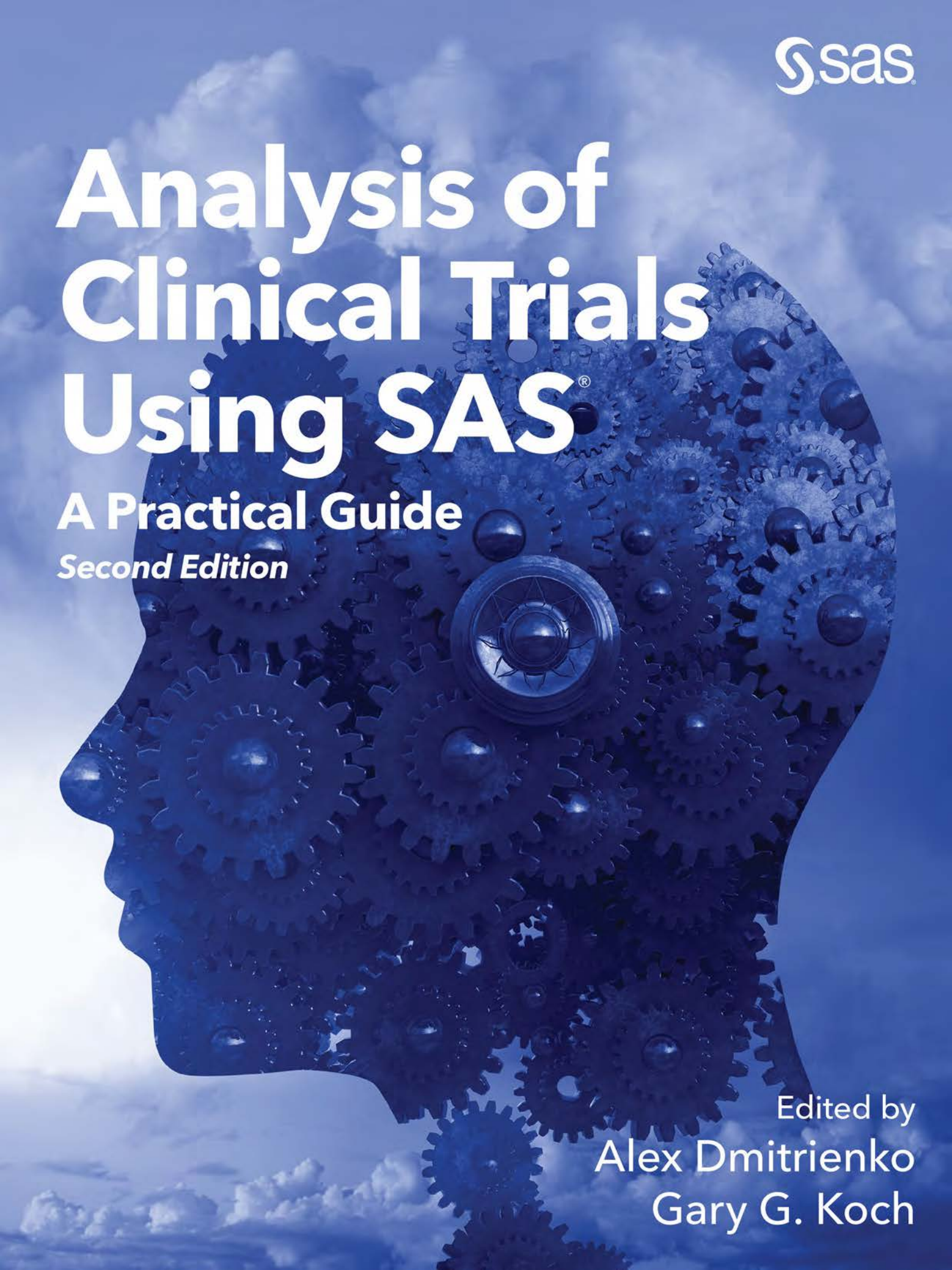
Analysis of Clinical Trials Using SAS®

A Practical Guide

Second Edition



Edited by
Alex Dmitrienko
Gary G. Koch

Contents

Preface v

About This Book xi

About These Authors xii

1 Model-based and Randomization-based Methods 1

By Alex Dmitrienko and Gary G. Koch

1.1 Introduction 1
1.2 Analysis of continuous endpoints 4
1.3 Analysis of categorical endpoints 20
1.4 Analysis of time-to-event endpoints 41
1.5 Qualitative interaction tests 56
References 61

2 Advanced Randomization-based Methods 67

By Richard C. Zink, Gary G. Koch, Yunro Chung and Laura Elizabeth Wiener

2.1 Introduction 67
2.2 Case studies 70
2.3 `%NParCov4` macro 73
2.4 Analysis of ordinal endpoints using a linear model 74
2.5 Analysis of binary endpoints 78
2.6 Analysis of ordinal endpoints using a proportional odds model 79
2.7 Analysis of continuous endpoints using the log-ratio of two means 80
2.8 Analysis of count endpoints using log-incidence density ratios 81
2.9 Analysis of time-to-event endpoints 82
2.10 Summary 86

3 Dose-Escalation Methods 101

By Guochen Song, Zoe Zhang, Nolan Wages, Anastasia Ivanova, Olga Marchenko and Alex Dmitrienko

3.1 Introduction 101
3.2 Rule-based methods 103
3.3 Continual reassessment method 107
3.4 Partial order continual reassessment method 116
3.5 Summary 123
References 123

4 Dose-finding Methods **127**
By Srinand Nandakumar, Alex Dmitrienko and Ilya Lipkovich

4.1 Introduction 127
4.2 Case studies 128
4.3 Dose-response assessment and dose-finding methods 132
4.4 Dose finding in Case study 1 145
4.5 Dose finding in Case study 2 160
References 176

5 Multiplicity Adjustment Methods **179**
By Thomas Brechenmacher and Alex Dmitrienko

5.1 Introduction 179
5.2 Single-step procedures 184
5.3 Procedures with a data-driven hypothesis ordering 189
5.4 Procedures with a prespecified hypothesis ordering 202
5.5 Parametric procedures 212
5.6 Gatekeeping procedures 221
References 241
Appendix 244

6 Interim Data Monitoring **251**
By Alex Dmitrienko and Yang Yuan

6.1 Introduction 251
6.2 Repeated significance tests 253
6.3 Stochastic curtailment tests 292
References 315

7 Analysis of Incomplete Data **319**
By Geert Molenberghs and Michael G. Kenward

7.1 Introduction 319
7.2 Case Study 322
7.3 Data Setting and Methodology 324
7.4 Simple Methods and MCAR 334
7.5 Ignorable Likelihood (Direct Likelihood) 338
7.6 Direct Bayesian Analysis (Ignorable Bayesian Analysis) 341
7.7 Weighted Generalized Estimating Equations 344
7.8 Multiple Imputation 349
7.9 An Overview of Sensitivity Analysis 362
7.10 Sensitivity Analysis Using Local Influence 363
7.11 Sensitivity Analysis Based on Multiple Imputation and Pattern-Mixture Models 371
7.12 Concluding Remarks 378
References 378

Index **385**

From *Analysis of Clinical Trials Using SAS®: A Practical Guide, Second Edition* by Alex Dmitrienko and Gary G. Koch.
Copyright © 2017, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.

Chapter 7

Analysis of Incomplete Data

Geert Molenberghs (Universiteit Hasselt and KU Leuven)
Michael G. Kenward

7.1 Introduction 319
7.2 Case Study 322
7.3 Data Setting and Methodology 324
7.4 Simple Methods and MCAR 334
7.5 Ignorable Likelihood (Direct Likelihood) 338
7.6 Direct Bayesian Analysis (Ignorable Bayesian Analysis) 341
7.7 Weighted Generalized Estimating Equations 344
7.8 Multiple Imputation 349
7.9 An Overview of Sensitivity Analysis 362
7.10 Sensitivity Analysis Using Local Influence 363
7.11 Sensitivity Analysis Based on Multiple Imputation and Pattern-Mixture Models 371
7.12 Concluding Remarks 378
7.13 References 378

More often than not, empirical studies are prone to incompleteness. For about half a century, methods have been developed to address this issue in data analysis. Older methods are relatively simple to use, but their validity is rightly called into question. With increasing computational power and software tools available, more flexible methods have come within reach. This chapter sketches a general taxonomy (Rubin, 1976) within which incomplete data methods can be placed. It then focuses on broadly valid methods that can be implemented within the SAS environment, thereby commenting on their relative advantages and disadvantages. All methods are illustrated using real data, and sufficiently generic SAS code is offered. Both Gaussian and non-Gaussian outcomes are given treatment. Apart from standard analysis tools, sensitivity analysis to examine the impact of non-verifiable model assumptions is addressed.

7.1 Introduction

In a longitudinal study, each unit is measured on several occasions. It is not unusual in practice for some sequences of measurements to terminate early for reasons outside the control of the investigator, and any unit so affected is called a dropout. It might, therefore, be necessary to accommodate dropout in the modeling process.

Early work on missing values was largely concerned with algorithmic and computational solutions to the induced lack of balance or deviations from the intended

study design (Afifi and Elashoff, 1966; Hartley and Hocking, 1971). This was followed by the development of general algorithms such as expectation-maximization (EM) (Dempster, Laird, and Rubin, 1977), and data imputation and augmentation procedures (Rubin 1987). These methods, combined with contemporary powerful computing resources and the progressive implementation of advanced methods in the SAS system, have addressed the problem in important ways. There is also the very difficult and important question of assessing the impact of missing data on subsequent statistical inference. This has received attention in particular in the setting of clinical trials (Little et al., 2010). Several authors give practical advice regarding the use of incomplete data methods (Mallinckrodt, 2013; O'Kelly and Ratitch, 2014), while others focus on the broad methodological underpinnings (Molenberghs and Kenward, 2007), or on specific methods, such as multiple imputation (MI; van Buuren, 2012; Carpenter and Kenward, 2013). The edited volumes by Fitzmaurice et al. (2009) and Molenberghs et al. (2015) present overviews of the longitudinal data and incomplete data state of research, respectively.

When referring to the missing-value, or non-response, process, we will use terminology of Little and Rubin (2014, Chapter 6). A non-response process is said to be *missing completely at random* (MCAR) if missingness is independent of both unobserved and observed data and *missing at random* (MAR) if, conditional on the observed data, missingness is independent of the unobserved measurements. A process that is neither MCAR nor MAR is termed *non-random* (MNAR). In the context of likelihood or Bayesian inferences, when the parameters describing the measurement process are functionally independent of the parameters describing the missingness process, and provided some mild regularity conditions hold, MCAR and MAR are *ignorable*, while a non-random process is non-ignorable. In the same vein, MI is valid under MAR. The method offers an attractive Monte Carlo-based alternative to direct likelihood and Bayesian inferences. For frequentist inferences, only a strong MCAR assumption is a sufficient condition for ignorability. This is relevant when discussing such methods as *generalized estimating equations* (GEE; Liang and Zeger, 1986).

We will pay particular attention to these methods, because of their relevance and the ease with which they can be implemented in SAS, thanks to the availability of a suite of SAS procedures. This implies that historic methods such as *complete case analysis* (CC) and *last observation carried forward* (LOCF) will be de-emphasized, in line with Little et al. (2010). Indeed, valid inference can be obtained through a likelihood-based analysis, a Bayesian analysis, or multiple imputation, without the need for modeling the dropout or missingness process. Likelihood-based analyses of longitudinal data can easily be conducted *without additional data manipulation* using, for example, the SAS procedures MIXED, GLIMMIX, NLMIXED, or related procedures (Verbeke and Molenberghs, 2000), without additional complication or effort. Thanks to the availability and flexibility of the procedures MI and MIANALYZE, multiple imputation is also rather straightforward to conduct. Furthermore, whereas a proper GEE analysis (i.e., valid under MAR) requires substantial additional programming with PROC GENMOD, the newer GEE procedure has made so-called *weighted* GEE (WGEE) particularly easy.

At the same time, we cannot avoid reflecting on the status of MNAR-based approaches. In realistic settings, the reasons for missingness or dropout are varied and hard to know with sufficient certainty. It is, therefore, difficult to fully justify on *a priori* grounds the assumption of MAR. At first sight, this calls for a further shift towards MNAR models. However, careful considerations have to be made, the most important of which is that no modeling approach, whether MAR or MNAR, can recover the lack of information that occurs due to incompleteness of the data.

First, under MAR, a standard analysis would follow, if we would be entirely sure of the MAR nature of the mechanism. However, it is only rarely the case that such an assumption is known to hold (Murray and Findlay, 1988). Nevertheless, ignorable

analyses may provide reasonably stable results, even when the assumption of MAR is violated, in the sense that such analyses constrain the behavior of the unseen data to be similar to that of the observed data (Mallinckrodt et al., 2001ab). A discussion of this phenomenon in the survey context can be found in Rubin, Stern, and Vehovar (1995). These authors argue that, in well conducted experiments (some surveys and many confirmatory clinical trials), the assumption of MAR is often to be regarded as a realistic one. Second, and very important for confirmatory trials, an MAR analysis can be specified *a priori* without additional work relative to a situation with complete data. Third, while MNAR models are more general and explicitly incorporate the dropout mechanism, the inferences they produce are typically highly dependent on untestable and often implicit assumptions built in regarding the distribution of the unobserved measurements given the observed ones. The quality of the fit to the observed data need not reflect at all the appropriateness of the implied structure governing the unobserved data. This point is irrespective of the MNAR route taken, whether a parametric model of the type of Diggle and Kenward (1994) is chosen, or a semi-parametric approach such as in Robins, Rotnitzky, and Scharfstein (1998). Hence in any incomplete-data setting there cannot be anything like a definitive analysis.

Thus, arguably, in the presence of MNAR missingness, a wholly satisfactory analysis of the data is not feasible. In fact, modeling in this context often rests on strong (untestable) assumptions and relatively little evidence from the data themselves. Glynn, Laird, and Rubin (1986) indicated that this is typical for selection models. It is somewhat less the case for pattern-mixture models (Little 1993, 1994; Hogan and Laird 1997), although caution should be used (Thijs, Molenberghs, and Verbeke, 2000). This awareness and the resulting skepticsm about fitting MNAR models initiated the search for methods to investigate the results with respect to model assumptions and for methods allowing to assess influences in the parameters describing the measurement process, as well as the parameters describing the non-random part of the dropout mechanism. Several authors have suggested various types of sensitivity analyses to address this issue (Molenberghs, Kenward, and Goetghebeur, 2001; Scharfstein, Rotnitzky, and Robins, 1999; Van Steen et al., 2001; and Verbeke et al., 2001). Verbeke et al. (2001) and Thijs, Molenberghs, and Verbeke (2000) developed a local influence-based approach for the detection of subjects that strongly influence the conclusions. These authors focused on the Diggle and Kenward (1994) model for continuous outcomes. Van Steen et al. (2001) adapted these ideas to the model of Molenberghs, Kenward and Lesaffre (1997), for monotone repeated ordinal data. Jansen et al. (2003) focused on the model family proposed by Baker, Rosenberger, and DerSimonian (1992). Recently, considerable research attention has been devoted to the use of pattern-mixture models, combined with multiple imputation, as a viable route for sensitivity analysis (Carpenter and Kenward, 2013; Carpenter, Roger, and Kenward, 2013). In summary, to explore the impact of deviations from the MAR assumption on the conclusions, we should ideally conduct a sensitivity analysis, within which MNAR models can play a major role.

The rest of the chapter is organized as follows. The clinical trial that will be used throughout the chapter is introduced in Section 7.2. The general datasetting is introduced in Section 7.3, as well as a formal framework for incomplete longitudinal data. A brief overview on the problems associated with simple methods is presented in Section 7.4. In subsequent sections, key methods are examined: ignorable likelihood (Section 7.5); ignorable Bayesian analysis (Section 7.6); generalized estimating equations (Section 7.7); and multiple imputation (Section 7.8). A brief introduction to sensitivity analysis is given in Section 7.9. Generally sensitivity analysis tools are discussed in Section 7.10, while in Section 7.11 we focus on sensitivity analysis tools that make use of multiple imputation.

The SAS code and data sets included in this chapter are available on the book's website at http://support.sas.com/publishing/authors/dmitrienko.html.

7.2 Case Study

EXAMPLE: Age-related macular degeneration trial

These data arise from a randomized multi-center clinical trial comparing an experimental treatment (interferon-α) to a corresponding placebo in the treatment of patients with age-related macular degeneration. In this book, we focus on the comparison between placebo and the highest dose (6 million units daily) of interferon-α (Z). But the full results of this trial have been reported elsewhere (Pharmacological Therapy for Macular Degeneration Study Group 1997). Patients with macular degeneration progressively lose vision. In the trial, the patients' visual acuity was assessed at different time points (4 weeks, 12 weeks, 24 weeks, and 52 weeks) through their ability to read lines of letters on standardized vision charts. These charts display lines of 5 letters of decreasing size, which the patient must read from top (largest letters) to bottom (smallest letters). The raw patient's visual acuity is the total number of letters correctly read. In addition, we often refer to each line with at least 4 letters correctly read as a "line of vision." The primary endpoint of the trial was the loss of at least 3 lines of vision at 1 year, compared to their baseline performance (a binary endpoint). The secondary endpoint of the trial was the visual acuity at 1 year (treated as a continuous endpoint). Buyse and Molenberghs (1998) examined whether the patient's performance at 6 months could be used as a surrogate for their performance at 1 year with respect to the effect of interferon-α. They looked at whether the loss of 2 lines of vision at 6 months could be used as a surrogate for the loss of at least 3 lines of vision at 1 year (Table 7.1). They also looked at whether visual acuity at 6 months could be used as a surrogate for visual acuity at 1 year.

TABLE 7.1 The Age-related Macular Degeneration Trial. Loss of at least 3 lines of vision at 1 year according to loss of at least 2 lines of vision at 6 months and according to randomized treatment group (placebo versus interferon-α).

	12 months			
	Placebo		Active	
6 months	0	1	0	1
No event (0)	56	9	31	9
Event (1)	8	30	9	38

Table 7.2 shows the visual acuity (mean and standard error) by treatment group at baseline, at 6 months, and at 1 year.

TABLE 7.2 The Age-related Macular Degeneration Trial. Mean (standard error) of visual acuity at baseline, at 6 months and at 1 year according to randomized treatment group (placebo versus interferon-α).

Time point	Placebo	Active	Total
Baseline	55.3 (1.4)	54.6 (1.3)	55.0 (1.0)
6 months	49.3 (1.8)	45.5 (1.8)	47.5 (1.3)
1 year	44.4 (1.8)	39.1 (1.9)	42.0 (1.3)

Visual acuity can be measured in several ways. First, we can record the number of letters read. Alternatively, dichotomized versions (at least 3 lines of vision lost) can

be used as well. Therefore, these data will be useful to illustrate methods for the joint modeling of continuous and binary outcomes, with or without taking the longitudinal nature into account. In addition, though there are 190 subjects with both month 6 and month 12 measurements available, the total number of longitudinal profiles is 240, but for only 188 of these have the four follow-up measurements been made.

Thus, indeed, 50 incomplete subjects could be considered for analysis as well. Both intermittent missingness as well as dropout occurs. An overview is given in Table 7.3.

TABLE 7.3 The Age-related Macular Degeneration Trial. Overview of missingness patterns and the frequencies with which they occur. 'O' indicates observed and 'M' indicates missing.

Measurement occasion					
4 wks	12 wks	24 wks	52 wks	Number	%
Completers					
O	O	O	O	188	78.33
Dropouts					
O	O	O	M	24	10.00
O	O	M	M	8	3.33
O	M	M	M	6	2.50
M	M	M	M	6	2.50
Non-monotone missingness					
O	O	M	O	4	1.67
O	M	M	O	1	0.42
M	O	O	O	2	0.83
M	O	M	M	1	0.42

Thus, 78.33% of the profiles are complete, while 18.33% exhibit monotone missingness. Out of the latter group, 2.5% or 6 subjects have no follow-up measurements. The remaining 3.33%, representing 8 subjects, have intermittent missing values. Thus, as in many of the examples seen already, dropout dominates intermediate patterns as the source of missing data.

Age-related Macular Degeneration Trial. Partial printout.

```
CRF     TRT    VISUAL0    VISUAL4    VISUAL12    VISUAL24    VISUAL52    lesion

1002     4       59         55         45           .           .          3
1003     4       65         70         65          65          55          1
1006     1       40         40         37          17           .          4
1007     1       67         64         64          64          68          2
1010     4       70          .          .           .           .          1
1110     4       59         53         52          53          42          3
1111     1       64         68         74          72          65          1
1112     1       39         37         43          37          37          3
1115     4       59         58         49          54          58          2
1803     1       49         51         71          71           .          1
1805     4       58         50          .           .           .          1
...
```

The original outcome (number of letters correctly read on a vision chart or its difference with the baseline reading) can be considered continuous for practical purposes. The derived dichotomous outcome (defined as number of letters read, has increased versus decreased when compared with baseline) will be considered as well.

Note that of the 52 subjects with incomplete follow-up, 8 exhibit a non-monotone pattern. While this will not hamper direct-likelihood analyses, Bayesian analyses, or multiple imputation, it is a challenge for weighted GEE and will need to be addressed.

7.3 Data Setting and Methodology

Assume that for subject $i = 1, \ldots, N$ in the study a sequence of responses, Y_{ij} is designed to be measured at occasions $j = 1, \ldots, n_i$. The outcomes are grouped into a vector of random variables $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$. In addition, define a dropout indicator D_i for the occasion at which dropout occurs and make the convention that $D_i = n_i + 1$ for a complete sequence. It is often handy to split the vector $\boldsymbol{Y}_i$ into observed ($\boldsymbol{Y}_i^o$) and missing ($\boldsymbol{Y}_i^m$) components, respectively. Dropout is a particular case of monotone missingness. To have a monotone pattern, there has to exist a permutation of the components of $\boldsymbol{Y}_i$ for all i simultaneously, such that, if a component is missing, then all later components are missing as well. For example, consider a vector of length four: $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})'$, with all but the second component Y_{i2} fully observed. Then the ordering $(Y_{i1}, Y_{i3}, Y_{i4}, Y_{i2})'$ satisfies the definition of monotone missingness. A counterexample is when, for every i, either Y_{i1} or Y_{i2} is observed. In that case, no monotone re-ordering is possible. For this definition to be meaningful, we need to have a balanced design in the sense of a common set of measurement occasions across all study subjects. Other patterns are referred to as non-monotone or intermittent missingness.

In principle, we would like to consider the density of the full data $f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi})$, where the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describe the measurement and missingness processes, respectively. Covariates are assumed to be measured but, for notational simplicity, suppressed from notation unless strictly needed.

The taxonomy, constructed by Rubin (1976), further developed in Little and Rubin (1987, with later editions in 2002 and 2014) and informally sketched in Section 7.1, is based on the factorization

$$f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|\boldsymbol{\theta}) f(d_i|\boldsymbol{y}_i, \boldsymbol{\psi}), \tag{7.3.1}$$

where the first factor is the marginal density of the measurement process, and the second one is the density of the missingness process, conditional on the outcomes. Factorization (7.3.1) forms the basis of *selection modeling* as the second factor corresponds to the (self-)selection of individuals into ''observed'' and '''missing'' groups. An alternative taxonomy can be built based on so-called *pattern-mixture models* (Little, 1993, 1994). These are based on the factorization

$$f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|d_i, \boldsymbol{\theta}) f(d_i|\boldsymbol{\psi}). \tag{7.3.2}$$

Indeed, (7.3.2) can be seen as a mixture of different populations, characterized by the observed pattern of missingness.

In the selection modeling framework, let us first describe a measurement and missingness model in turn, and then formally introduce and comment on ignorability.

7.3.1 Linear Mixed Models

Assume that we want to perform a longitudinal analysis of a continuous outcome. We then often assume a linear mixed-effects model, sometimes with an additional serial correlation:

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_{(1)i} + \boldsymbol{\varepsilon}_{(2)i}, \tag{7.3.3}$$

(Verbeke and Molenberghs, 2000) where $\boldsymbol{Y}_i$ is the n-dimensional response vector for subject i, $1 \leq i \leq N$; N is the number of subjects; X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ known design matrices; $\boldsymbol{\beta}$ is the p dimensional vector containing the fixed effects; and $\boldsymbol{b}_i \sim N(\mathbf{0}, G)$ is the q dimensional vector containing the random effects. The residual components $\boldsymbol{\varepsilon_i}$ are decomposed as $\boldsymbol{\varepsilon_i} = \boldsymbol{\varepsilon}_{(1)\boldsymbol{i}} + \boldsymbol{\varepsilon}_{(2)\boldsymbol{i}}$ in which $\boldsymbol{\varepsilon}_{(2)\boldsymbol{i}}$ is a component of serial correlation and $\boldsymbol{\varepsilon}_{(1)\boldsymbol{i}} \sim N(\mathbf{0}, \sigma^2 I_{n_i})$ is an extra component of measurement error. Thus, serial correlation is captured by the realization of a Gaussian stochastic process, $\boldsymbol{\varepsilon}_{(2)\boldsymbol{i}}$, which is assumed to follow a $N(\mathbf{0}, \tau^2 H_i)$ law. The serial covariance matrix H_i only depends on i through the number n of observations and through the time points t_{ij} at which measurements are taken. The structure of the matrix H_i is determined through the autocorrelation function $\rho(t_{ij} - t_{ik})$. This function decreases such that $\rho(0) = 1$ and $\rho(+\infty) = 0$. Further, G is a general $(q \times q)$ covariance matrix with (i, j) element $d_{ij} = d_{ji}$. Finally, $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\varepsilon}_{(1)1}, \ldots, \boldsymbol{\varepsilon}_{(2)N}, \boldsymbol{\varepsilon}_{(2)1}, \ldots, \boldsymbol{\varepsilon}_{(2)N}$ are assumed to be independent. Inference is based on the marginal distribution of the response $\boldsymbol{Y}_i$ which, after integrating over random effects, can be expressed as

$$\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, Z_i G Z_i' + \Sigma_i). \tag{7.3.4}$$

Here, $\Sigma_i = \sigma^2 I_{n_i} + \tau^2 H_i$ is a $(n \times n)$ covariance matrix that groups the measurement error and serial components. Further, we define $V_i = Z_i G Z_i' + \Sigma_i$ as the general covariance matrix of $\boldsymbol{Y}_i$.

The most commonly used SAS procedure to fit linear mixed models is PROC MIXED. The fixed-effect structure is specified via the MODEL statement, while the random-effects structure is entered using the RANDOM statement. If, in addition, serial correlation is assumed to be present, the REPEATED statement can be added. Also, several marginal models derived from a linear mixed-effects model can be specified directly using the REPEATED statement. For details, we refer to Verbeke and Molenberghs (2000).

7.3.2 Generalized Linear Mixed Models

Perhaps the most commonly encountered subject-specific (or random-effects) model for arbitrary outcome data type is the generalized linear mixed model (GLMM). A general framework for mixed-effects models can be expressed as follows.

It is assumed that, conditionally on q-dimensional random effects $\boldsymbol{b}_i$ that are drawn independently from $N(\mathbf{0}, G)$, the outcomes Y_{ij} are independent with densities of the form

$$f_i(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) = \exp\left\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi)\right\},$$

with $\eta(\mu_{ij}) = \eta(E(Y_{ij}|\boldsymbol{b}_i)) = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i$ for a known link function $\eta(\cdot)$, with $\boldsymbol{x_{ij}}$ and $\boldsymbol{z_{ij}}$ p-dimensional and q-dimensional vectors of known covariate values; with $\boldsymbol{\beta}$ a p-dimensional vector of unknown fixed regression coefficients; with ϕ a scale parameter; and with θ_{ij} the natural (or canonical) parameter. Further, let $f(\boldsymbol{b}_i|G)$ be the density of the $N(\mathbf{0}, G)$ distribution for the random effects $\boldsymbol{b}_i$.

Due to the above independence assumption, this model is often referred to as a *conditional independence* model. This assumption is the basis of the implementation in the NLMIXED procedure. Just as in the linear mixed model case, the model can be extended with residual correlation, in addition to the one induced by the random effects. Such an extension can be implemented in the SAS procedure GLIMMIX, and its predecessor the GLIMMIX macro. It is relevant to realize that GLIMMIX can be used without random effects as well, thus effectively producing a marginal model, with estimates and standard errors similar to the ones obtained with GEE (see Section 7.3.4).

In general, unless a fully Bayesian approach is followed, inference is based on the marginal model for $\boldsymbol{Y}_i$ which is obtained from integrating out the random effects. The likelihood contribution of subject i then becomes

$$f_i(\boldsymbol{y}_i|\boldsymbol{\beta}, G, \phi) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi)\ f(\boldsymbol{b}_i|G)\ d\boldsymbol{b}_i$$

from which the likelihood for $\boldsymbol{\beta}$, D, and ϕ is derived as

$$\begin{aligned} L(\boldsymbol{\beta}, G, \phi) &= \prod_{i=1}^{N} f_i(\boldsymbol{y}_i|\boldsymbol{\beta}, G, \phi) \\ &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi)\ f(\boldsymbol{b}_i|G)\ d\boldsymbol{b}_i. \end{aligned} \tag{7.3.5}$$

The key problem in maximizing the obtained likelihood is the presence of N integrals over the q-dimensional random effects. In some special cases, these integrals can be worked out analytically. However, since no analytic expressions are available for these integrals, numerical approximations are needed. Here, we will focus on the most frequently used methods to do so. In general, the numerical approximations can be subdivided into those that are based on the approximation of the integrand; those based on an approximation of the data; and those that are based on the approximation of the integral itself. An extensive overview of a number of available approximations can be found in Tuerlinckx et al. (2004), Pinheiro and Bates (2000), and Skrondal and Rabe-Hesketh (2004). Finally, to simplify notation, it will be assumed that natural link functions are used, but straightforward extensions can be applied.

When integrands are approximated, the goal is to obtain a tractable integral such that closed-form expressions can be obtained, making the numerical maximization of the approximated likelihood feasible. Several methods have been proposed, but basically all come down to Laplace-type approximations of the function to be integrated (Tierney and Kadane 1986).

A second class of approaches is based on a decomposition of the data into the mean and an appropriate error term, with a Taylor series expansion of the mean, which is a nonlinear function of the linear predictor. All methods in this class differ in the order of the Taylor approximation and/or the point around which the approximation is expanded. More specifically, we consider the decomposition

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij} \ = \ h(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i) + \varepsilon_{ij}, \tag{7.3.6}$$

in which $h(\cdot)$ equals the inverse link function $\eta^{-1}(\cdot)$, and where the error terms have the appropriate distribution with variance equal to $\text{Var}(Y_{ij}|\boldsymbol{b}_i) = \phi v(\mu_{ij})$ for $v(\cdot)$, the usual variance function in the exponential family. Note that, with the natural link function,

$$v(\mu_{ij}) = \frac{\partial h}{\partial \eta}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i).$$

Several approximations of the mean μ_{ij} in (7.3.6) can be considered. One possibility is to derive a linear Taylor expansion of (7.3.6) around current estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}_i$ of the fixed effects and random effects, respectively. This will result in the expression

$$\boldsymbol{Y}_i^* \equiv \widehat{W}_i^{-1}(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i) + X_i\widehat{\boldsymbol{\beta}} + Z_i\widehat{\boldsymbol{b}}_i \ \approx \ X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i^*, \tag{7.3.7}$$

with $\widehat{W}_i$ equal to the diagonal matrix with diagonal entries equal to $v(\widehat{\mu_{ij}})$, and for $\boldsymbol{\varepsilon}_i^*$ equal to $\widehat{W}_i^{-1}\boldsymbol{\varepsilon}_i$, which still has mean zero. Note that (7.3.7) can be viewed as a

linear mixed model for the pseudo data $\boldsymbol{Y}_i^*$, with fixed effects $\boldsymbol{\beta}$, random effects $\boldsymbol{b}_i$, and error terms $\boldsymbol{\varepsilon}_i^*$.

This immediately yields an algorithm for fitting the original generalized linear mixed model. Given starting values for the parameters $\boldsymbol{\beta}$, G, and ϕ in the marginal likelihood, empirical Bayes estimates are calculated for $\boldsymbol{b}_i$, and pseudo data $\boldsymbol{Y}_i^*$ are computed. Then, the approximate linear mixed model (7.3.7) is fitted, yielding updated estimates for $\boldsymbol{\beta}$, G, and ϕ. These are then used to update the pseudo data, and this whole scheme is iterated until convergence is reached.

The resulting estimates are called *penalized quasi-likelihood* estimates (PQL) in the literature (e.g., Molenberghs and Verbeke, 2005), or *pseudo-quasi-likelihood* in the documentation of the GLIMMIX procedure because they can be obtained from optimizing a quasi-likelihood function that only involves first and second-order conditional moments, augmented with a penalty term on the random effects. The pseudo-likelihood terminology derives from the fact that the estimates are obtained by (restricted) maximum likelihood of the pseudo-response or working variable.

An alternative approximation is very similar to the PQL method, but is based on a linear Taylor expansion of the mean μ_{ij} in (7.3.6) around the current estimates $\widehat{\boldsymbol{\beta}}$ for the fixed effects and around $\boldsymbol{b}_i = \boldsymbol{0}$ for the random effects. The resulting estimates are called *marginal quasi-likelihood* estimates (MQL). We refer to Breslow and Clayton (1993) and Wolfinger and O'Connell (1993) for more details. Since the linearizations in the PQL and the MQL methods lead to linear mixed models, the implementation of these procedures is often based on feeding updated pseudo data into software for the fitting of linear mixed models. However, it should be emphasized that the results from these fittings, which are often reported intermediately, should be interpreted with great care. For example, reported (log)likelihood values correspond to the assumed normal model for the pseudo data and should not be confused with (log-)likelihood for the generalized linear mixed model for the actual data at hand. Further, fitting of linear mixed models can be based on maximum likelihood (ML) as well as restricted maximum likelihood (REML) estimation. Hence, within the PQL and MQL frameworks, both methods can be used for the fitting of the linear model to the pseudo data, yielding (slightly) different results. Finally, the quasi-likelihood methods discussed here are very similar to the method of linearization for fitting generalized estimating equations (GEE). The difference is that here, the correlation between repeated measurements is modelled through the inclusion of random effects, conditionally on which repeated measures are assumed independent. But in the GEE approach, this association is modelled through a marginal working correlation matrix.

Note that, when there are no random effects, both this method and GEE reduce to a marginal model, the difference being in the way that the correlation parameters are estimated. In both cases, it is possible to allow for misspecification of the association structure by resorting to empirically corrected standard errors. When this is done, the methods are valid under MCAR. In case we would have confidence in the specified correlation structure, purely model-based inference can be conducted, and, hence, the methods are valid when missing data are MAR.

A third method of numerical approximation is based on the approximation of the integral itself. Especially in cases where the above two approximation methods fail, this numerical integration turns out to be very useful. Of course, a wide toolkit of numerical integration tools, available from the optimization literature, can be applied. Several of those have been implemented in various software tools for generalized linear mixed models. A general class of quadrature rules selects a set of abscissas and constructs a weighted sum of function evaluations over those. In the particular context of random-effects models, so-called *adaptive* quadrature rules can be used (Pinheiro and Bates, 1995, 2000), where the numerical integration is centered around the EB estimates of the random effects. The number of quadrature points is then selected in terms of the desired accuracy.

To illustrate the main ideas, we consider Gaussian and adaptive Gaussian quadrature, designed for the approximation of integrals of the form $\int f(z)\phi(z)dz$, for an known function $f(z)$ and for $\phi(z)$ the density of the (multivariate) standard normal distribution. We will first standardize the random effects such that they get the identity covariance matrix. Let $\boldsymbol{\delta_i}$ be equal to $\boldsymbol{\delta}_i = G^{-1/2}\boldsymbol{b}_i$. We then have that $\boldsymbol{\delta}_i$ is normally distributed with mean $\mathbf{0}$ and covariance I. The linear predictor then becomes $\theta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}G^{1/2}\boldsymbol{\delta_i}$, so the variance components in G have been moved to the linear predictor. The likelihood contribution for subject i, expressed in the original parameters, then equals

$$f_i(\boldsymbol{y}_i|\boldsymbol{\beta}, G, \phi) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi)\ f(\boldsymbol{b}_i|G)\ d\boldsymbol{b}_i. \tag{7.3.8}$$

Obviously, (7.3.8) is of the form $\int f(z)\phi(z)dz$ as required to apply (adaptive) Gaussian quadrature.

In Gaussian quadrature, $\int f(z)\phi(z)dz$ is approximated by the weighted sum

$$\int f(z)\phi(z)dz \approx \sum_{q=1}^{Q} w_q f(z_q).$$

Q is the order of the approximation. The higher Q, the more accurate the approximation will be. Further, the so-called nodes (or quadrature points) z_q are solutions to the Qth order Hermite polynomial, while the w_q are well-chosen weights. The nodes z_q and weights w_q are reported in tables. Alternatively, an algorithm is available for calculating all z_q and w_q for any value Q (Press et al., 1992). In case of univariate integration, the approximation consists of subdividing the integration region in intervals, and approximating the surface under the integrand by the sum of surfaces of the so-obtained approximating rectangles. An example is given in the left window of Figure 7.1, for the case of $Q = 10$ quadrature points. A similar interpretation is possible for the approximation of multivariate integrals. Note that the figure immediately highlights one of the main disadvantages of (non-adaptive) Gaussian quadrature, i.e., the fact that the quadrature points z_q are chosen based on $\phi(z)$, independent of the function $f(z)$ in the integrand. Depending on the support of

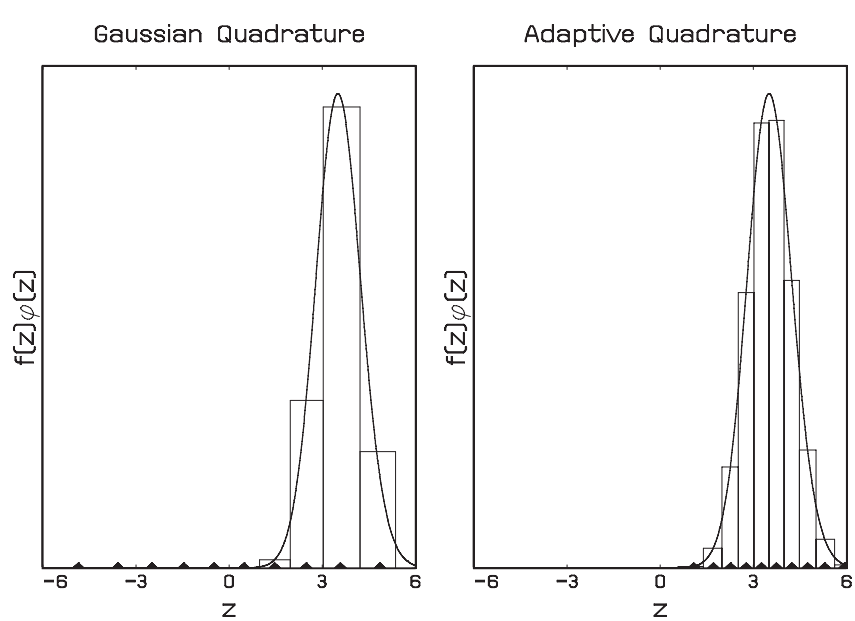


Figure 7.1 Graphical illustration of Gaussian (left window) and adaptive Gaussian (right window) quadrature of order $Q = 10$. The black triangles indicate the position of the quadrature points, while the rectangles indicate the contribution of each point to the integral.

$f(z)$, the z_q will or will not lie in the region of interest. Indeed, the quadrature points are selected to perform well in case $f(z)\phi(z)$ approximately behaves like $\phi(z)$, i.e., like a standard normal density function. This will be the case, for example, if $f(z)$ is a polynomial of a sufficiently low order. In our applications, however, the function $f(z)$ will take the form of a density from the exponential family---hence, an exponential function. It might then be helpful to re-scale and shift the quadrature points such that more quadrature points lie in the region of interest. This is shown in the right window of Figure 7.1, and is called adaptive Gaussian quadrature.

In general, the higher the order Q, the better the approximation will be of the N integrals in the likelihood. Typically, adaptive Gaussian quadrature needs (many) fewer quadrature points than classical Gaussian quadrature. On the other hand, adaptive Gaussian quadrature requires for each unit the numerical maximization of a function of the form $\ln(f(z)\phi(z))$ for the calculation of $\widehat{z}$. This implies that adaptive Gaussian quadrature is much more time consuming.

Since fitting of GLMMs is based on maximum likelihood principles, inferences for the parameters are readily obtained from classical maximum likelihood theory.

The Laplace method (Tierny and Kadane, 1986) has been designed to approximate integrals of the form

$$I = \int e^{Q(\boldsymbol{b})} d\boldsymbol{b}, \tag{7.3.9}$$

where $Q(\boldsymbol{b})$ is a known, unimodal, and bounded function of a q-dimensional variable $\boldsymbol{b}$. Let $\widehat{\boldsymbol{b}}$ be the value of $\boldsymbol{b}$ for which Q is maximized. We then have that the second-order Taylor expansion of $Q(\boldsymbol{b})$, which is of the form

$$Q(\boldsymbol{b}) \approx Q(\widehat{\boldsymbol{b}}) + \frac{1}{2}(\boldsymbol{b} - \widehat{\boldsymbol{b}})' Q''(\widehat{\boldsymbol{b}})(\boldsymbol{b} - \widehat{\boldsymbol{b}}), \tag{7.3.10}$$

for $Q''(\widehat{\boldsymbol{b}})$ equal to the Hessian of Q, i.e., the matrix of second-order derivative of Q, evaluated at $\widehat{\boldsymbol{b}}$. Replacing $Q(\boldsymbol{b})$ in (7.3.9) by its approximation in (7.3.10), we obtain

$$I \approx (2\pi)^{q/2} \left| -Q''(\widehat{\boldsymbol{b}}) \right|^{-1/2} e^{Q(\widehat{\boldsymbol{b}})}.$$

Clearly, each integral in (7.3.5) is proportional to an integral of the form (7.3.9), for functions $Q(\boldsymbol{b})$ given by

$$Q(\boldsymbol{b}) = \phi^{-1} \sum_{j=1}^{n_i} [y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}) - \psi(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b})] - \frac{1}{2}\boldsymbol{b}' D^{-1} \boldsymbol{b},$$

such that Laplace's method can be applied here. Note that the mode $\widehat{\boldsymbol{b}}$ of Q depends on the unknown parameters $\boldsymbol{\beta}$, ϕ, and D, such that in each iteration of the numerical maximization of the likelihood, $\widehat{\boldsymbol{b}}$ will be re-calculated conditionally on the current values for the estimates for these parameters.

The Laplace approximation is exact when $Q(\boldsymbol{b})$ is a quadratic function of $\boldsymbol{b}$, i.e., if the integrands in (7.3.5) are exactly equal to normal kernels. Interpreting these integrands as unnormalized posterior distributions of the random effects $\boldsymbol{b}_i$, it is known from the Bayesian literature (Gelman et al., 1995) that this will be the case only in very special examples such as linear models, or provided that the number n_i of repeated measurements for all subjects are sufficiently large.

To fit GLMMs, the SAS procedures GLIMMIX and NLMIXED are obvious choices. While a variety of GLMMs can be fitted using both procedures, there are fundamental differences. GLIMMIX is restricted to generalized linear mixed models, whereas NLMIXED allows for fully nonlinear (mixed) models. For this

reason, GLIMMIX models are specified in a conventional, symbolic way (e.g., using syntax of the form `Y=X1 X2 X1*X2`), whereas in NLMIXED the user programs the mean and, where appropriate, variance functions, including fixed and random effects. GLIMMIX allows for serial correlation, using a REPEATED statement. Both procedures allow for multiple RANDOM statements. The integration options in GLIMMIX include PQL, MQL, Laplace approximation, and adaptive Gaussian quadrature. The corresponding options for NLMIXED are adaptive and non-adaptive Gaussian quadrature. Both allow for a variety of updating algorithms and tuning parameters. Because GLIMMIX is restricted to GLMM and, hence, can efficiently make use of generalized linear model features (exponential family results, the use of linear predictors, etc.), it is generally faster and stabler when both procedures can be used. However, NLMIXED offers additional flexibility thanks to the open programming abilities.

Using NLMIXED, the conditional distribution of the data, given the random effects, is specified in the MODEL statement. Valid distributions are:

- normal(m, v): Normal with mean m and variance v,
- binary(p): Bernoulli with probability p,
- binomial(n, p): Binomial with count n and probability p,
- gamma(a, b): Gamma with shape a and scale b,
- negbin(n, p): Negative binomial with count n and probability p,
- poisson(m): Poisson with mean m,
- general$(\ell\ell)$: General model with log-likelihood $\ell\ell$.

The `general` structure is especially convenient when a non-convential model is fitted. The RANDOM statement defines the random effects and their distribution. The procedure requires the data to be ordered by subject.

The valid distributions for the GLIMMIX procedure are: beta, binary, binomial, exponential, gamma, Gaussian (normal), geometric, inverse Gaussian, lognormal, multinomial, negative binomial, Poisson, and central t. Each one of them has a default link function attached to them. Users have the ability to deviate from these, but should check whether an alternative choice is coherent with the natural range of the outcome type. For example, a probit link instead of a logit link is also a sensible choice for binary outcomes, while a log link would usually be problematic for interval-type data.

7.3.3 Likelihood-based Approaches

Consider, for the sake of argument, a continuous longitudinal outcome. Assume that incompleteness is due to dropout only, and that the first measurement Y_{i1} is obtained for everyone. The model for the dropout process can be based on, for example, a logistic regression for the probability of dropout at occasion j, given the subject is still in the study. We denote this probability by $g(\boldsymbol{h}_{ij}, y_{ij})$ in which $\boldsymbol{h}_{ij}$ is a vector containing all responses observed up to but not including occasion j, as well as relevant covariates. We then assume that $g(\boldsymbol{h}_{ij}, y_{ij})$ satisfies

$$\text{logit}[g(\boldsymbol{h}_{ij}, y_{ij})] = \text{logit}\left[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)\right] = \boldsymbol{h}_{ij}\boldsymbol{\psi} + \omega y_{ij}, \tag{7.3.11}$$

$i = 1, \ldots, N$. When ω equals zero, the posited dropout model is MAR, and all parameters can be estimated easily using SAS since the measurement model for which we use a linear mixed model and the dropout model, assumed to follow a logistic regression, can then be fitted separately. If $\omega \neq 0$, the posited dropout process is MNAR. Model (7.3.11) provides the building blocks for the dropout

process $f(d_i|\boldsymbol{y}_i, \boldsymbol{\psi})$. As a cautionary note, we should not lose sight of the fact that the true nature of the dropout mechanism cannot be determined based on observed data alone (Molenberghs et al., 2008), pointing to the need for sensitvity analysis.

Rubin (1976) and Little and Rubin (2014) have shown that, under MAR and mild regularity conditions (parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are functionally independent), likelihood-based and Bayesian inferences are valid when the missing data mechanism is ignored (see also Verbeke and Molenberghs, 2000). Practically speaking, the likelihood of interest is then based upon the factor $f(\boldsymbol{y}_i^o|\boldsymbol{\theta})$. This is called *ignorability*. We return to this in more detail in Sections 7.5 and 7.6.

The practical implication is that a software module with likelihood estimation facilities and with the ability to handle incompletely observed subjects manipulates the correct likelihood, providing valid parameter estimates and likelihood ratio values. A few cautionary remarks are in place. First, when at least part of the scientific interest is directed towards the nonresponse process, obviously both processes need to be considered. Still, under MAR, both processes can be modeled and parameters estimated separately. Second, likelihood inference is often surrounded with references to the sampling distribution (e.g., to construct precision estimators and for statistical hypothesis tests; Kenward and Molenberghs, 1998). However, the practical implication is that standard errors and associated tests, when based on the observed rather than the expected information matrix and given that the parametric assumptions are correct, are valid. Third, it may be hard to fully rule out the operation of an MNAR mechanism. This point was brought up in the introduction and will be discussed further in Sections 7.9--7.10. Fourth, a full longitudinal analysis is necessary, even when interest lies, for example, in a comparison between the two treatment groups at the last occasion. In the latter case, the fitted model can be used as the basis for inference at the last occasion. A common criticism is that a model needs to be considered. However, it should be noted that, in many clinical trial settings, the repeated measures are balanced in the sense that a common (and often limited) set of measurement times is considered for all subjects, allowing the a priori specification of a saturated model (e.g., full group by time interaction model for the fixed effects and unstructured variance-covariance matrix). Such an ignorable linear mixed model specification is given in Mallinckrodt et al. (2001ab).

7.3.4 Generalized Estimating Equations

Overview

Two sometimes quoted issues with full likelihood approaches are the computational complexity they entail and their vulnerability to model assumptions. When we are mainly interested in first-order marginal mean parameters and pairwise association parameters, i.e., second-order moments, a full likelihood procedure can be replaced by quasi-likelihood methods (McCullagh and Nelder, 1989). In quasi-likelihood, the mean response is expressed as a parametric function of covariates; and the variance is assumed to be a function of the mean up to possibly unknown scale parameters. Wedderburn (1974) first noted that likelihood and quasi-likelihood theories coincide for exponential families and that the quasi-likelihood ''estimating equations'' provide consistent estimates of the regression parameters $\boldsymbol{\beta}$ in any generalized linear model, even for choices of link and variance functions that do not correspond to exponential families.

For clustered and repeated data, Liang and Zeger (1986) proposed so-called *generalized estimating equations* (GEE or GEE1), which require only the correct specification of the univariate marginal distributions provided we are willing to adopt ''working'' assumptions about the association structure. They estimate the parameters associated with the expected value of an individual's vector of binary

responses and phrase the working assumptions about the association between pairs of outcomes in terms of marginal correlations. The method combines estimating equations for the regression parameters $\boldsymbol{\beta}$ with moment-based estimating for the correlation parameters entering the working assumptions.

Prentice (1988) extended their results to allow joint estimation of probabilities and pairwise correlations. Lipsitz, Laird, and Harrington (1991) modified the estimating equations of Prentice (1988) to allow modeling of the association through marginal odds ratios rather than marginal correlations. When adopting GEE1 we do not use information of the association structure to estimate the main effect parameters. As a result, it can be shown that GEE1 yields consistent main effect estimators, even when the association structure is misspecified. However, severe misspecification can seriously affect the efficiency of the GEE1 estimators. In addition, GEE1 should be avoided when some scientific interest is placed on the association parameters.

A second-order extension of these estimating equations (GEE2) that include the marginal pairwise association as well has been studied by Liang, Zeger, and Qaqish (1992). They note that GEE2 is nearly fully efficient though bias might occur in the estimation of the main effect parameters when the association structure is misspecified.

Carey, Zeger, and Diggle (1993) proposed so-called *alternating logistic regressions*(ALR), applicable to repeated binary data with logit link and the association modeled using odds ratios. See also Molenberghs and Verbeke (2005). While they allow for association-modeling, they are computationally simpler than GEE2.

Some Methodological Detail

After this short overview of the GEE approach, the GEE methodology will now be explained a little further. We start by recalling the score equations, to be solved when computing maximum likelihood estimates under a marginal normal model $\boldsymbol{y}_i \sim N(X_i\boldsymbol{\beta}, V_i)$:

$$\sum_{i=1}^{N} X_i'(A_i^{1/2}R_iA_i^{1/2})^{-1}(\boldsymbol{y}_i - X_i\boldsymbol{\beta}) = \mathbf{0}, \tag{7.3.12}$$

in which the marginal covariance matrix V_i has been decomposed in the form $V_i = A_i^{1/2}R_iA_i^{1/2}$, with A_i the diagonal matrix; with the marginal variances along the main diagonal; and with R_i equal to the marginal correlation matrix. Second, the score equations to be solved when computing maximum likelihood estimates under a marginal generalized linear model, assuming independence of the responses within units (i.e., ignoring the repeated measures structure), are given by:

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}(A_i^{1/2}I_{n_i}A_i^{1/2})^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \tag{7.3.13}$$

Note that (7.3.12) is of the form (7.3.13) but with the correlations between repeated measures taken into account. A straightforward extension of (7.3.13) that accounts for the correlation structure is

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'}(A_i^{1/2}R_iA_i^{1/2})^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) \ = \ \mathbf{0}, \tag{7.3.14}$$

which is obtained from replacing the identity matrix I_{n_i} by a correlation matrix $R_i = R_i(\boldsymbol{\alpha})$, often referred to as the *working* correlation matrix. Usually, the marginal covariance matrix $V_i = A_i^{1/2}R_iA_i^{1/2}$ contains a vector $\boldsymbol{\alpha}$ of unknown parameters---leading to $V_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = A_i^{1/2}(\boldsymbol{\beta})R_i(\boldsymbol{\alpha})A_i^{1/2}(\boldsymbol{\beta})$---which is replaced for practical purposes by a consistent estimate.

Assuming that the marginal mean $\boldsymbol{\mu}_i$ has been correctly specified as $h(\boldsymbol{\mu}_i) = X_i\boldsymbol{\beta}$, it can be shown that, under mild regularity conditions, the estimator $\widehat{\boldsymbol{\beta}}$ obtained from solving (7.3.14) is asymptotically normally distributed with mean $\boldsymbol{\beta}$ and with covariance matrix

$$I_0^{-1} I_1 I_0^{-1}, \tag{7.3.15}$$

where

$$I_0 = \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'} \right), \qquad I_1 = \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\beta}} V_i^{-1} \mathrm{Var}(\boldsymbol{y}_i) V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}'} \right).$$

In practice, $\mathrm{Var}(\boldsymbol{y}_i)$ in (7.3.15) is replaced by $(\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'$, which is unbiased on the sole condition that the mean was again correctly specified.

Note that valid inferences can now be obtained for the mean structure, only assuming that the model assumptions with respect to the first-order moments are correct. Note also that, although arising from a likelihood approach, the GEE equations in (7.3.14) cannot be interpreted as score equations corresponding to some full likelihood for the data vector $\boldsymbol{y}_i$.

Liang and Zeger (1986) proposed moment-based estimates for the working correlation. To this end, first define deviations:

$$e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}$$

and decompose the variance slightly more generally as above in the following way:

$$V_i = \phi A_i^{1/2} R_i A_i^{1/2},$$

where ϕ is an overdispersion parameter.

Some of the more popular choices for the working correlations are independence ($\mathrm{Corr}(Y_{ij}, Y_{ik}) = 0,\ j \neq k$); exchangeability ($\mathrm{Corr}(Y_{ij}, Y_{ik}) = \alpha,\ j \neq k$); AR(1) ($\mathrm{Corr}(Y_{ij}, Y_{i,j+t}) = \alpha^t,\ t = 0, 1, \ldots, n_i - j$); and unstructured ($\mathrm{Corr}(Y_{ij}, Y_{ik}) = \alpha_{jk},\ j \neq k$). Typically, moment-based estimation methods are used to estimate these parameters, as part of an integrated iterative estimation procedure (Aerts, Geys, Molenberghs, and Ryan, 2002). The overdispersion parameter is approached in a similar fashion. The standard iterative procedure to fit GEE, based on Liang and Zeger (1986), is then as follows: (1) compute initial estimates for $\boldsymbol{\beta}$, using a univariate GLM (i.e., assuming independence); (2) compute the quantities needed in the estimating equation, such as means and variances; (3) compute Pearson residuals e_{ij}; (4) compute estimates for $\boldsymbol{\alpha}$; (5) compute $R_i(\boldsymbol{\alpha})$; (6) compute an estimate for ϕ; (7) compute $V_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \phi A_i^{1/2}(\boldsymbol{\beta}) R_i(\boldsymbol{\alpha}) A_i^{1/2}(\boldsymbol{\beta})$; and (8) update the estimate for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \left[\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right]^{-1} \left[\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\beta}} V_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \right].$$

Steps (2)--(8) are iterated until convergence.

In SAS, three procedures can be used for GEE. First, there is the GENMOD procedure. In its basic form, it fits generalized linear models to univariate data. Adding the REPEATED statement, repeated measures can be analyzed using GEE or ALR, with a suite of working correlation structures available. As of SAS 9.4, the GEE procedure is available. It is essentially a ''synonym'' to GENMOD for standard GEE, but its main attraction lies in the use of weighted GEE, for which we refer to Section 7.7. As mentioned earlier, the GLIMMIX procedure can also be used, provided no random effects are included, but merely "serial correlation," using the `RANDOM _residual_ /` syntax, combined with the use of empirically corrected standard errors using the `empirical` option in the `PROC GLIMMIX` statement.

7.4 Simple Methods and MCAR

We will briefly review a number of relatively simple methods that have been and are still in extensive use. For a number of them, MCAR is required, while for others, such as LOCF, the conditions for validity are wholly different. A detailed account is given in Verbeke and Molenberghs (1997, 2000) and Molenberghs and Kenward (2007). The case of clinical trials received specific attention in Molenberghs et al. (2003). The focus will be on the complete case method, where data are removed, on the one hand, and on imputation strategies and where data are filled in on the other hand. Regarding imputation, we distinguish between single and multiple imputation. In the first case, a single value is substituted for every ''hole'' in the data set, and the resulting data set is analyzed as if it represented the true complete data. Multiple imputation properly acknowledges the uncertainty stemming from filling in missing values rather than observing them (Rubin, 1987; Schafer, 1997), and is deferred to Section 7.8. LOCF will be discussed within the context of imputation strategies, although not every author classifies the method as belonging to the imputation family.

7.4.1 Complete Case Analysis

A complete case analysis includes only those cases for analysis for which all n_i planned measurements were actually recorded. This method has obvious advantages. It is very simple to describe, and, since the data structure is as would have resulted from a complete experiment, standard statistical software can be used. Further, since the complete estimation is done on the same subset of completers, there is a common basis for inference, unlike with the available case methods.

Unfortunately, the method suffers from severe drawbacks. First, there is nearly always a substantial loss of information. For example, suppose there are 20 measurements, with 10% of missing data on each measurement. Suppose further that missingness on the different measurements is independent. Then, the estimated percentage of incomplete observations is as high as 87%. The impact on precision and power is dramatic. Even though the reduction of the number of complete cases will be less dramatic in realistic settings where the missingness indicators $\boldsymbol{R}_i$ are correlated, the effect just sketched will often undermine a lot of complete case analyses. In addition, severe bias can result when the missingness mechanism is MAR but not MCAR. Indeed, should an estimator be consistent in the complete data problem, then the derived complete case analysis is consistent only if the missingness process is MCAR. Unfortunately, the MCAR assumption is much more restrictive than the MAR assumption.

Complete Case Analysis and SAS

The only step required to perform a complete case analysis is deletion of subjects for which not all designed measurements have been obtained. When the data are organized ''horizontally,'' i.e., one record per subject, this is particularly easy. With ''vertically'' organized data, slightly more data manipulation is needed, and the SAS macro, discussed below, can be used.

For example, for the age related macular degeneration trial, running the next statement produces the complete case CC data set, for the continuous outcome ('diff' is the difference of number of letters correctly read versus, baseline'):

PROGRAM 7.1 Preparing the data for complete case analysis (continuous outcome)

```
%cc(data=armd155,id=subject,time=time,response=diff,out=armdcc2);
```

and for the binary outcome ('bindif' is a discretization of 'diff', with 1 for nonnegative values and 0 otherwise). See Program 7.2.

PROGRAM 7.2 Preparing the data for complete case analysis (discrete outcome)

```
%cc(data=armd111,id=subject,time=time,response=bindif,out=armdcc);
```

Clearly, the CC macro requires four arguments. The `data=` argument is the data set to be analyzed. If not specified, the most recent data set is used. The name of the variable in the data set that contains the identification variable is specified by `id=`, and `time=` specifies the variable indicating the time ordering within a subject. The outcome variable is passed on by means of the `response=` argument, and the name of the output data set, created with the macro, is defined through `out=`.

After performing this data preprocessing, a complete case analysis follows of any type requested by the user, including, but not limited to, longitudinal analysis.

The macro requires records, corresponding to missing values, to be present in the data set. Otherwise, it is assumed that a measurement occasion not included is missing by design.

Upon creation of the new data set, the code for Model (7.5.16), to be presented in Section 7.5 on ignorable likelihood, is given by Program 7.3.

PROGRAM 7.3 Complete case analysis (continuous outcome)

```
proc mixed data=armdcc2 method=ml;
title 'CC - continuous';
class time treat subject;
model diff = time treat*time / noint solution ddfm=kr;
repeated time / subject=subject type=un;
run;
```

When, in contrast, GEE of the form (7.7.28) is applied to the completers, the following code can be used for standard GEE. See Program 7.4.

PROGRAM 7.4 Complete case analysis (binary outcome, GEE, PROC GENMOD)

```
proc genmod data=armdcc;
title 'CC - GEE';
class time treat subject;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelse;
run;
```

or see Program 7.5.

PROGRAM 7.5 Complete case analysis (binary outcome, GEE, PROC GEE)

```
proc gee data=armdcc;
title 'CC - GEE';
class time treat subject;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelse;
run;
```

Alternatively, for the linearization-based version of GEE, with empirically corrected standard errors, we can use Program 7.6.

PROGRAM 7.6 Complete case analysis (binary outcome, GEE, linearized version, PROC GLIMMIX)

```
proc glimmix data=armdcc empirical;
title 'CC - GEE - linearized version - empirical';
nloptions maxiter=50 technique=newrap;
class time treat subject;
model bindif = time treat*time /  noint solution dist=binary;
random _residual_ / subject=subject type=cs;
run;
```

For the generalized linear mixed model (7.5.17), with numerical quadrature, the following code is useful. See Program 7.7.

PROGRAM 7.7 Complete case analysis (binary outcome, GLMM, PROC GLIMMIX)

```
proc glimmix data=armdcc method=gauss(q=20);
title 'CC - mixed - quadrature';
nloptions maxiter=50 technique=newrap;
class time treat subject;
model bindif = time treat*time / noint solution dist=binary;
random intercept  / subject=subject type=un g gcorr;
run;
```

With NLMIXED, we could use Program 7.8.

PROGRAM 7.8 Complete case analysis (binary outcome, GLMM, PROC NLMIXED)

```
data help;
set armdcc;
time1=0; if time=1 then time1=1;
time2=0; if time=2 then time2=1;
time3=0; if time=3 then time3=1;
time4=0; if time=4 then time4=1;
run;

proc nlmixed data=help  qpoints=20 maxiter=100 technique=newrap;
title 'CC - mixed - numerical integration';
eta = beta11*time1+beta12*time2+beta13*time3+beta14*time4
      +b
      +(beta21*time1+beta22*time2+beta23*time3+beta24*time4)
      *(2-treat);
p = exp(eta)/(1+exp(eta));
model bindif ~ binary(p);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau^2' tau*tau;
run;
```

Note that the DATA step in Program 7.8 merely creates dummy variables for each of the four measurement times. The `ESTIMATE` statement allows for the easy estimation of the random-effects variance and its standard error, because the model parameter τ is the corresponding standard deviation.

None of the above programs is specific to CC. Only the data preprocessing using the `%cc(...)` macro defines it as CC.

7.4.2 Simple Imputation Methods

An alternative way to obtain a data set on which complete data methods can be used is filling in the missing values, instead of deleting subjects with incomplete sequences. The principle of imputation is particularly easy. The observed values are used to impute values for the missing observations. There are several ways to use the observed information. First, we can use information on the same subject (e.g., last observation carried forward). Second, information can be borrowed from other subjects (e.g., mean imputation). Finally, both within and between subject information can be used (e.g., conditional mean imputation, hot deck imputation). Standard references are Little and Rubin (2014) and Rubin (1987). Imputation strategies have historically been very popular in sample survey methods.

However, great care has to be taken with imputation strategies. Dempster and Rubin (1983) write

> The idea of imputation is both seductive and dangerous. It is seductive because it can lull the user into the pleasurable state of believing that the data are complete after all, and it is dangerous because it lumps together situations where the problem is sufficiently minor that it can be legitimately handled in this way and situations where standard estimators applied to the real and imputed data have substantial biases.

For example, Little and Rubin (2014) show that the method could work for a linear model with one fixed effect and one error term, but that it generally does not for hierarchical models, split-plot designs, repeated measures (with a complicated error structure), random-effects, and mixed-effects models. At the very least, different imputations for different effects would be necessary.

The user of imputation strategies faces several dangers. First, the imputation model could be wrong, and, hence, the point estimates would be biased. Second, even for a correct imputation model, the uncertainty resulting from incompleteness is masked. Indeed, even when we are reasonably sure about the mean value that the unknown observation would have, the actual stochastic realization, depending on both the mean structure as well as on the error distribution, is still unknown.

Last Observation Carried Forward

In this case, whenever a value is missing, the last observed value is substituted. It is typically applied to settings where incompleteness is due to attrition.

Very strong and often unrealistic assumptions have to be made to ensure validity of this method. First, either when we consider a longitudinal analysis or when the scientific question is in terms of the last planned occasion, we have to believe that a subjects' measurement stays at the same level from the moment of dropout onwards (or during the period they are unobserved in the case of intermittent missingness). In a clinical trial setting, we might believe that the response profile *changes* as soon as a patient goes off treatment and even that it would flatten. However, the constant profile assumption is even stronger. Second, this method shares with other single imputation methods that it overestimates the precision by treating imputed and actually observed values on equal footing.

The situation, in which the scientific question is in terms of the last observed measurement, is often considered to be the real motivation for LOCF. However in some cases, the question, defined as such, has a very unrealistic and ad hoc flavor. Clearly, measurements at (self-selected) dropout times are lumped together with measurements made at the (investigator defined) end of the study.

Last Observation Carried Forward and SAS

Similar steps as needed for a complete case analysis need to be performed when LOCF is the goal. For a vertically organized data set, the following macro, also written by Caroline Beunckens, can be used, in the continuous case. See Program 7.9.

PROGRAM 7.9 Preparing for LOCF analysis

```
%locf(data=armd155,id=subject,time=time,response=diff,out=armdlocf2);
```

or in the dichotomous case:

```
%locf(data=armd111,id=subject,time=time,response=bindif,out=armdlocf);
```

The arguments are exactly the same and have the same meaning as in the `%cc(...)` macro of the previous section. Note that there is now a *new* response variable created, named 'locf', which should be used in the corresponding analysis programs. Thus, all SAS procedure MIXED, GENMOD, GEE, GLIMMIX, and NLMIXED code of the previous section remains valid, upon replacing the response variables 'diff' and 'bindiff' by 'locf' and, of course, by appropriately changing the names of the data sets.

7.5 Ignorable Likelihood (Direct Likelihood)

As discussed in Section 7.3, likelihood based inference is valid whenever the mechanism is MAR and provided the technical condition holds that the parameters describing the nonresponse mechanism are distinct from the measurement model parameters (Little and Rubin, 2014). In other words, the missing data process should be ignorable in the likelihood inference sense, since then the log-likelihood partitions into two functionally independent component. As a consequence, a software module for likelihood estimation can be used, provided it can handle incompletely observed subjects. In other words, it should be able to handle subjects (or: blocks) of varying lengths, which virtually all longitudinal procedures do. The ensuing parameter estimates, standard errors, likelihood ratio values, etc. are valid.

In conclusion, a likelihood-based ignorable analysis (referred to for short as ignorable likelihood or direct likelihood) is preferable since it uses all available information, without the need to delete or to impute measurements or entire subjects. It is theoretically justified whenever the missing data mechanism is MAR. There is no statistical information distortion, given that observations are neither removed (such as in complete case analysis) nor added (such as in single imputation). There is no additional programming involved to implement an ignorable analysis in the MIXED, GLIMMIX, or NLMIXED procedures, provided the order of the measurements is correctly specified. This can be done either by supplying records with missing data in the input data set or by properly indicating the order of the measurement in the REPEATED and/or RANDOM statements.

7.5.1 Normally Distributed Outcomes

EXAMPLE: Age-related macular degeneration trial

We consider first a simple multivariate normal model, with unconstrained time trend under placebo, an occasion-specific treatment effect, and a 4×4 unstructured

variance-covariance matrix. Thus,

$$Y_{ij} = \beta_{j1} + \beta_{j2} T_i + \varepsilon_{ij}, \tag{7.5.16}$$

where $T_i = 0$ for placebo and $T_1 = 1$ for interferon-α. The direct-likelihood analysis is contrasted with CC and LOCF, and parameter estimates (standard errors) for the eight mean model parameters are presented in Table 7.4.

TABLE 7.4 The Age Related Macular Degeneration Trial. Parameter estimates (standard errors) for the linear mixed models, fitted to the continuous outcome 'difference of the number of letters read versus baseline'. CC, LOCF, and direct likelihood. p values are presented for treatment effect at each of the four times separately, as well as for all four times jointly.

Effect	Parameter	CC	LOCF	direct lik.
		Parameter estimates (standard errors)		
Intercept 4	β_{11}	-3.24(0.77)	-3.48(0.77)	-3.48(0.77)
Intercept 12	β_{21}	-4.66(1.14)	-5.72(1.09)	-5.85(1.11)
Intercept 24	β_{31}	-8.33(1.39)	-8.34(1.30)	-9.05(1.36)
Intercept 52	β_{41}	-15.13(1.73)	-14.16(1.53)	-16.21(1.67)
Treatm. eff. 4	β_{12}	2.32(1.05)	2.20(1.08)	2.20(1.08)
Treatm. eff. 12	β_{22}	2.35(1.55)	3.38(1.53)	3.51(1.55)
Treatm. eff. 24	β_{32}	2.73(1.88)	2.41(1.83)	3.03(1.89)
Treatm. eff. 52	β_{42}	4.17(2.35)	3.43(2.15)	4.86(2.31)
		p-values		
Treatm. eff. 4	β_{12}	0.0282	0.0432	0.0435
Treatm. eff. 12	β_{22}	0.1312	0.0287	0.0246
Treatm. eff. 24	β_{32}	0.1491	0.1891	0.1096
Treatm. eff. 52	β_{42}	0.0772	0.1119	0.0366
Treatm. eff. (overall)		0.1914	0.1699	0.1234

While there is no overall treatment effect, and the p-values between the three methods do not vary too much, the picture is different for the occasion-specific treatment effects. At week 4, all three p-values indicate significance. While this is the only significant effect when only the completers are analyzed, there is one more significant effect with LOCF (week 12) and two more when direct likelihood is employed (weeks 12 and 52). Once more, CC and LOCF miss important treatment differences, the most important one being the one at week 52, the end of the study.

7.5.2 Non-Gaussian Outcomes

EXAMPLE: Age-related macular degeneration trial

Let us now turn to a random-intercept logistic model, similar in spirit to (7.7.28):

$$\text{logit}[P(Y_{ij} = 1 | T_i, t_j, b_i)] = \beta_{j1} + b_i + \beta_{j2} T_i, \tag{7.5.17}$$

with notation as before and $b_i \sim N(0, \tau^2)$. Both PQL and numerical integration are used for model fitting. The results for this model are given in Table 7.5.

We observe the usual downward bias in the PQL versus numerical integration analysis, as well as the usual relationship between the marginal parameters of Table 7.6 and their random-effects counterparts. Note also that the random-intercepts variance is largest under LOCF, underscoring again that this method artificially increases the association between measurements on the same subject. In this case, in contrast to marginal models, LOCF and, in fact, also CC considerably overestimate the treatment effect at certain times, by varying degrees ranging from trivial to important, in particular at 4 and 24 weeks.

TABLE 7.5 **The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) for the random-intercept models: PQL and numerical-integration based fits on the CC and LOCF population, and on the observed data (direct-likelihood).**

Effect	Parameter	CC	LOCF	direct lik.
		PQL		
Int.4	β_{11}	-1.19(0.31)	-1.05(0.28)	-1.00(0.26)
Int.12	β_{21}	-1.05(0.31)	-1.18(0.28)	-1.19(0.28)
Int.24	β_{31}	-1.35(0.32)	-1.30(0.28)	-1.26(0.29)
Int.52	β_{41}	-1.97(0.36)	-1.89(0.31)	-2.02(0.35)
Trt.4	β_{12}	0.45(0.42)	0.24(0.39)	0.22(0.37)
Trt.12	β_{22}	0.58(0.41)	0.68(0.38)	0.71(0.37)
Trt.24	β_{32}	0.55(0.42)	0.50(0.39)	0.49(0.39)
Trt.52	β_{42}	0.44(0.47)	0.39(0.42)	0.46(0.46)
R.I. s.d.	τ	1.42(0.14)	1.53(0.13)	1.40(0.13)
R.I. var.	τ^2	2.03(0.39)	2.34(0.39)	1.95(0.35)
		Numerical integration		
Int.4	β_{11}	-1.73(0.42)	-1.63(0.39)	-1.50(0.36)
Int.12	β_{21}	-1.53(0.41)	-1.80(0.39)	-1.73(0.37)
Int.24	β_{31}	-1.93(0.43)	-1.96(0.40)	-1.83(0.39)
Int.52	β_{41}	-2.74(0.48)	-2.76(0.44)	-2.85(0.47)
Trt.4	β_{12}	0.64(0.54)	0.38(0.52)	0.34(0.48)
Trt.12	β_{22}	0.81(0.53)	0.98(0.52)	1.00(0.49)
Trt.24	β_{32}	0.77(0.55)	0.74(0.52)	0.69(0.50)
Trt.52	β_{42}	0.60(0.59)	0.57(0.56)	0.64(0.58)
R.I. s.d.	τ	2.19(0.27)	2.47(0.27)	2.20(0.25)
R.I. var.	τ^2	4.80(1.17)	6.08(1.32)	4.83(1.11)

TABLE 7.6 **The Age-related Macular Degeneration Trial. Parameter estimates (model-based standard errors; empirically corrected standard errors) for the marginal models: standard and linearization-based GEE on the CC and LOCF population, and on the observed data. In the latter case, also WGEE is used. All analyses based on PROC GENMOD.**

Effect	Par.	CC	LOCF	Observed data	
				Unweighted	WGEE
			Standard GEE		
Int.4	β_{11}	-1.01(0.24;0.24)	-0.87(0.20;0.21)	-0.87(0.21;0.21)	-0.98(0.10;0.44)
Int.12	β_{21}	-0.89(0.24;0.24)	-0.97(0.21;0.21)	-1.01(0.21;0.21)	-1.78(0.15;0.38)
Int.24	β_{31}	-1.13(0.25;0.25)	-1.05(0.21;0.21)	-1.07(0.22;0.22)	-1.11(0.15;0.33)
Int.52	β_{41}	-1.64(0.29;0.29)	-1.51(0.24;0.24)	-1.71(0.29;0.29)	-1.72(0.25;0.39)
Tr.4	β_{12}	0.40(0.32;0.32)	0.22(0.28;0.28)	0.22(0.28;0.28)	0.80(0.15;0.67)
Tr.12	β_{22}	0.49(0.31;0.31)	0.55(0.28;0.28)	0.61(0.29;0.29)	1.87(0.19;0.61)
Tr.24	β_{32}	0.48(0.33;0.33)	0.42(0.29;0.29)	0.44(0.30;0.30)	0.73(0.20;0.52)
Tr.52	β_{42}	0.40(0.38;0.38)	0.34(0.32;0.32)	0.44(0.37;0.37)	0.74(0.31;0.52)
Corr.	ρ	0.39	0.44	0.39	0.33
			Linearization-based GEE		
Int.4	β_{11}	-1.01(0.24;0.24)	-0.87(0.21;0.21)	-0.87(0.21;0.21)	-0.98(0.18;0.44)
Int.12	β_{21}	-0.89(0.24;0.24)	-0.97(0.21;0.21)	-1.01(0.22;0.21)	-1.78(0.26;0.42)
Int.24	β_{31}	-1.13(0.25;0.25)	-1.05(0.21;0.21)	-1.07(0.23;0.22)	-1.19(0.25;0.38)
Int.52	β_{41}	-1.64(0.29;0.29)	-1.51(0.24;0.24)	-1.71(0.29;0.29)	-1.81(0.39;0.48)
Tr.4	β_{12}	0.40(0.32;0.32)	0.22(0.28;0.28)	0.22(0.29;0.29)	0.80(0.26;0.67)
Tr.12	β_{22}	0.49(0.31;0.31)	0.55(0.28;0.28)	0.61(0.28;0.29)	1.85(0.32;0.64)
Tr.24	β_{32}	0.48(0.33;0.33)	0.42(0.29;0.29)	0.44(0.30;0.30)	0.98(0.33;0.60)
Tr.52	β_{42}	0.40(0.38;0.38)	0.34(0.32;0.32)	0.44(0.37;0.37)	0.97(0.49;0.65)
	σ^2	0.62	0.57	0.62	1.29
	τ^2	0.39	0.44	0.39	1.85
Corr.	ρ	0.39	0.44	0.39	0.59

7.5.3 Direct Likelihood and SAS

In contrast to CC and LOCF, no extra data processing is necessary when a direct likelihood analysis is envisaged, provided the software tool used for analysis can handle measurement sequences of unequal length. This is the case for virtually all longitudinal data analysis tools, including the SAS procedures MIXED, NLMIXED, and GLIMMIX.

One note of caution is relevant, however. When residual correlation structures are used for which the order of the measurements within a sequence is important, such as unstructured and AR(1), but not simple or compound symmetry, and intermittent missingness occurs, care has to be taken to ensure that the *design* order within the sequence, and not the *apparent* order, is passed on. In the SAS procedure MIXED, a statement such as

```
repeated / subject=subject type=un;
```

is fine when every subject has, say, four designed measurements. However, when for a particular subject, the second measurement is missing, there is a risk that the remaining measurements are considered the first, second, and third, rather than the first, third, and fourth. Thus, it is sensible to replace the above statement by:

```
repeated time / subject=subject type=un;
```

For the GENMOD and GEE procedures, the option `withinsubject=time` of the REPEATED statement can be used. Note that this produces GEE and not direct likelihood. For the GLIMMIX procedure, there is no such feature. Evidently, we can also avoid the problem by properly sorting the measurements within a subject and at the same time ensuring that for missing values a record is included with, of course, a missing value instead of the actual measurement.

In all cases, especially when GLIMMIX is used, the proper order is passed on when a record is included, even for the missing measurements.

When the NLMIXED procedure is used, only random effects can be included. In such a case, all relevant information is contained in the actual effects that define the random effects structure. For example, the order is immaterial for a random intercepts model, and, for a random slope in time, all information needed about time is passed on, for example, by the RANDOM statement:

```
RANDOM intercept time / subject=subject type=un;
```

Thus, in conclusion, all code for likelihood-based analyses, listed in Section 7.4.1 can be used, provided the original data sets (`armd155.sas7bdat` and `armd111.sas7bdat`) are passed on, and not the derived ones.

We conclude that, with only a minimal amount of care, a direct likelihood analysis is no more complex than the corresponding analysis on a set of data that is free of missingness.

7.6 Direct Bayesian Analysis (Ignorable Bayesian Analysis)

As stated earlier, not only likelihood but also Bayesian analyses are ignorable under MAR and appropriate regularity conditions. This means that, just like with ignorable likelihood, an ignorable Bayesian analysis is as easy to carry out with

complete as well as incomplete data. To illustrate this, consider the following simple linear mixed model for the `diff` outcome:

$$Y_{ij} \sim N\Big((\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_j + (\beta_3 + b_{3i})T_i + (\beta_4 + b_{4i})T_i t_j, \sigma^2\Big), \tag{7.6.18}$$

$$\begin{pmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \\ b_{4i} \end{pmatrix} \sim N(\mathbf{0}, G). \tag{7.6.19}$$

To allow comparison between ignorable likelihood and ignorable Bayesian analysis, we first provide the ignorable likelihood code.

PROGRAM 7.10 Direct likelihood

```
proc mixed data=m.armd13k method=ml nobound covtest;
title 'direct likelihood';
class subject;
model diff = time treat treat*time /  solution ddfm=kr;
random intercept time treat treat*time / subject=subject type=vc;
run;
```

Relevant output is:

Selected direct likelihood output

```
                 Solution for Fixed Effects

                          Standard
Effect        Estimate      Error      DF    t Value    Pr > |t|

Intercept       4.5954     2.0798     316       2.21      0.0279
time           -2.5089     1.0527     286      -2.38      0.0178
treat          -1.5121     1.3584     383      -1.11      0.2664
time*treat     -0.7952     0.6731     350      -1.18      0.2383

                  Covariance Parameter Estimates

                                     Standard         Z
Cov Parm      Subject    Estimate      Error     Value        Pr Z

Intercept     subject     12.3563    12.9356      0.96      0.3395
time          subject     14.3486     3.3803      4.24      <.0001
treat         subject      5.1696     4.8308      1.07      0.2846
time*treat    subject     -0.2543     1.1974     -0.21      0.8318
Residual                  50.7478     3.3151     15.31      <.0001
```

Note that the variance component associated with the time by treatment interaction is negative, though not significantly different from zero. In other words, we have a model that allows a marginal but no hierarchical interpretation. This will be different in the upcoming Bayesian analysis, where a hierarchical interpretation is inherent in the model.

For Bayesian analysis purposes, we rewrite (7.6.18)--(7.6.19) as:

$$Y_{ij} \sim N\Big(\beta_{1i} + \beta_{2i}t_j + \beta_{3i}T_i + \beta_{4i}T_i t_j, \sigma^2\Big), \tag{7.6.20}$$

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \beta_{4i} \end{pmatrix} \sim N(\boldsymbol{\beta}, G)., \tag{7.6.21}$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \Sigma_0), \tag{7.6.22}$$

$$G \sim \text{IWishart}(\rho, S), \tag{7.6.23}$$

$$\sigma^2 \sim \text{IGamma}(\alpha, \gamma). \tag{7.6.24}$$

The corresponding program, incorporating choices for the hyperprior parameters, is Program 7.11.

PROGRAM 7.11 Direct Bayesian analysis

```
proc mcmc data=m.armd13k nmc=10000 outpost=m.armd13l seed=23 init=random;
title 'direct Bayes';
array theta[4] beta1 beta2 beta3 beta4;
array theta_c[4];
array dmat[4,4];
array beta0[4] (0 0 0 0);
array Sig0[4,4] (1000 0 0 0 0 1000 0 0 0 0 1000 0 0 0 0 1000);
array S[4,4] (100 0 0 0 0 100 0 0 0 0 100 0 0 0 0 100);
parms theta_c dmat {10 0 0 0 0 10 0 0 0 0 10 0 0 0 0 10} var_y;
prior theta_c ~ mvn(beta0,Sig0);
prior dmat ~ iwish(4,S);
prior var_y ~ igamma(0.01,scale=0.01);
random theta~mvn(theta_c,dmat) subject=subject;
mu=beta1+beta2*treat+beta3*time+beta4*treat*time;
model diff~normal(mu,var=var_y);
run;
```

The corresponding output is below.

Selected direct Bayesian output

Posterior Summaries and Intervals

Parameter	N	Mean	Standard Deviation	95% HPD Interval	
theta_c1	10000	4.3933	2.9525	-1.8276	8.5445
theta_c2	10000	-1.4268	1.9469	-4.4090	2.5441
theta_c3	10000	-2.2187	1.2372	-4.5489	0.0692
theta_c4	10000	-0.9799	0.8187	-2.5878	0.4703
dmat1	10000	43.8501	25.5374	8.7026	97.8559
dmat2	10000	-14.0943	16.7127	-53.0171	4.6373
dmat3	10000	-13.0795	17.2618	-52.0841	11.3516
dmat4	10000	5.1822	9.6912	-8.5062	28.9990
dmat5	10000	-14.0943	16.7127	-53.0171	4.6373
dmat6	10000	21.8172	12.0049	5.2413	48.3360
dmat7	10000	0.7602	11.0049	-15.7870	28.3676
dmat8	10000	-4.1026	5.9859	-18.0130	5.5606
dmat9	10000	-13.0795	17.2618	-52.0841	11.3516
dmat10	10000	0.7602	11.0049	-15.7870	28.3676
dmat11	10000	36.4025	19.3862	9.0475	76.2131
dmat12	10000	-14.9697	11.7158	-39.6488	-0.1558
dmat13	10000	5.1822	9.6912	-8.5062	28.9990
dmat14	10000	-4.1026	5.9859	-18.0130	5.5606
dmat15	10000	-14.9697	11.7158	-39.6488	-0.1558
dmat16	10000	12.6873	6.8567	3.9227	26.8400
var_y	10000	46.2681	3.0475	40.4443	52.2954

We observe that the results are similar to those of the direct likelihood analysis, but that there are differences as well. This is due to the effect of the prior distributions. Another source of difference is the fact that the likelihood-based model does not impose bounds on the components of G, whereas the Bayesian model is intrinsically hierarchical.

7.7 Weighted Generalized Estimating Equations

7.7.1 Concept

Generalized estimating equations (GEE), as discussed in Section 7.3.4, are appealing to model repeated measures when the research questions are formulated in terms of the marginal mean function, especially but not only when outcomes are of a non-Gaussian type.

However, as Liang and Zeger (1986) pointed out, incomplete-data based inferences with GEE are valid only under the strong assumption that the data are missing completely at random (MCAR). To allow the data to be missing at random (MAR), Robins, Rotnitzky, and Zhao (1995) proposed weighted estimating equations (WGEE). In Section 7.8, we will also discuss the combination of GEE with multiple imputation.

The idea is to weight each subject's contribution to the GEE by the inverse probability, either of being fully observed, or of being observed up to a certain time. In line with Molenberghs et al. (2011), let π_i be the probability for subject i to be completely observed, and π_i' the probability for subject i to drop out on occasion d_i. These can be written as

$$\pi_i = \prod_{\ell=2}^{n_i}(1-p_{i\ell}), \qquad \pi_i' = \left[\prod_{\ell=2}^{d_i-1}(1-p_{i\ell})\right]\cdot p_{id_i}, \tag{7.7.25}$$

where $p_{i\ell} = P\left(D_i = \ell | D_i \geq \ell, Y_{i\,\bar{\ell}}, X_{i\,\bar{\ell}}\right)$ are the component probabilities of dropping out at occasion ℓ, given that the subject is still in the study, the covariate history $X_{i\,\bar{\ell}}$, and the outcome history $Y_{i\,\bar{\ell}}$. In such a case, we can opt either for WGEE based on the completers only:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N}\frac{\widetilde{R}_i}{\pi_i}\,\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\beta}'}V_i^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_i) \;=\; \boldsymbol{0}, \tag{7.7.26}$$

with $\widetilde{R}_i = 1$ if a subject is fully observed and 0 otherwise, or, upon using (7.7.25), for WGEE using all subjects:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N}\frac{1}{\pi_i'}\,\frac{\partial\boldsymbol{\mu}_i^o}{\partial\boldsymbol{\beta}'}(V_i^o)^{-1}(\boldsymbol{y}_i^o-\boldsymbol{\mu}_i^o) \;=\; \boldsymbol{0}. \tag{7.7.27}$$

Here the superscript 'o' indicates the portion corresponding to the observed data in the corresponding matrix or vector. Of course, with (7.7.26), the incomplete subjects also contribute through the model for the dropout probabilities π_i.

EXAMPLE: Age-related macular degeneration trial

Consider the binary outcome that indicates whether the number of letters correctly read at a follow-up occasion is higher or lower than the corresponding number of letters at baseline. A population averaged (or marginal model) is used. We compare analyses performed on the completers only (CC), on the LOCF imputed data, as well as on the observed data. In all cases, standard GEE and the linearization-based version are considered. For the observed, partially incomplete data, GEE is supplemented with WGEE.

The GEE analyses are reported in Table 7.6. In all cases, we use the logit link, and the model takes the form:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i, \tag{7.7.28}$$

similar in spirit to (7.5.16). A working exchangeable correlation matrix is considered. For the WGEE analysis, the following weight model is assumed:

$$\begin{aligned}&\text{logit}[P(D_i = j|D_i \geq j)]\\ &= \psi_0 + \psi_1 y_{i,j-1} + \psi_2 T_i + \psi_{31}L_{1i} + \psi_{32}L_{2i} + \psi_{34}L_{3i}\\ &\quad + \psi_{41}I(t_j = 2) + \psi_{42}I(t_j = 3),\end{aligned} \tag{7.7.29}$$

with $y_{i,j-1}$ the binary outcome at the previous time $t_{i,j-1} = t_{j-1}$, $L_{ki} = 1$ if the patient's eye lesion is of level $k = 1, \ldots, 4$ (since one dummy variable is redundant, only three are used), and $I(\cdot)$ is the indicator function. Parameter estimates and standard errors for the dropout model are given in Table 7.7. Intermittent missingness will be ignored at this time. We return to this point in Section 7.8. Covariates of importance are treatment assignment, the level of lesions at baseline (a four-point categorical variable, for which three indicator variables are needed), and time at which dropout occurs. For the latter covariates, there are three levels, since dropout can occur at times 2, 3, or 4. Hence, two indicator variables are included. Finally, the previous outcome does not have a significant impact, but will be kept in the model nevertheless.

TABLE 7.7 The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) for a logistic regression model to describe dropout.

		Estimate (s.e.)	
Effect	Parameter	GENMOD	GEE
Intercept	ψ_0	0.14 (0.49)	0.17 (0.56)
Previous outcome	ψ_1	0.04 (0.38)	-0.05 (0.38)
Treatment	ψ_2	-0.86 (0.37)	-0.87 (0.37)
Lesion level 1	ψ_{31}	-1.85 (0.49)	-1.82 (0.49)
Lesion level 2	ψ_{32}	-1.91 (0.52)	-1.88 (0.52)
Lesion level 3	ψ_{33}	-2.80 (0.72)	-2.79 (0.72)
Time 2	ψ_{41}	-1.75 (0.49)	-1.73 (0.49)
Time 3	ψ_{42}	-1.38 (0.44)	-1.36 (0.44)

Note: GENMOD is called after the `%dropout` *macro is called. The GEE parameters result from the MISSMODEL statement within the procedure.*

From Table 7.6, it is clear that there is very little difference between the standard GEE and linearization-based GEE results. This is undoubtedly the case for CC, LOCF, and unweighted GEE on the observed data. For these three cases, also, the model-based and empirically corrected standard errors agree extremely well, owing to the unstructured nature of the full time by treatment mean structure. However, we do observe differences in the WGEE analyses. Not only do the parameter estimates

differ a little between the two GEE versions, but there is also a dramatic difference between the model-based and empirically corrected standard errors. This is entirely due to the weighting scheme. The weights were not calibrated to add up to the total sample size, which is reflected in the model-based standard errors. In the linearization-based case, part of the effect is captured as overdispersion. This can be seen from adding the parameters σ^2 and τ^2. In all other analyses, the sum is close to one, as it should be when there is no residual overdispersion, but, in the last column, these add up to 3.14. Nevertheless, the two sets of empirically corrected standard errors agree very closely, which is reassuring.

In spite of there being no strong evidence for MAR, the results between GEE and WGEE differ in a nontrivial way. It is noteworthy that at 12 weeks, a treatment effect is observed with WGEE, which is undetected when using the other marginal analyses. This finding is confirmed to some extent by the subject-specific random-intercept model, presented in the next section, when the data are used as observed.

When comparing parameter estimates across CC, LOCF, and observed data analyses, it is clear that LOCF has the effect of artificially increasing the correlation between measurements. The effect is mild in this case. The parameter estimates of the observed-data GEE are close to the LOCF results for earlier time points and close to CC for later time points. This is to be expected, as at the start of the study the LOCF and observed populations are very similar, with the same holding between CC and observed populations near the end of the study. Note also that the treatment effect under LOCF, especially at 12 weeks and after 1 year, is biased downward in comparison to the GEE analyses.

7.7.2 WGEE and SAS, Using PROC GENMOD

We will first discuss the steps to be taken when using the older SAS procedure GENMOD. Afterwards, we will switch to the more recent and easier to use GEE procedure.

A GENMOD program for the standard GEE analysis is Program 7.12.

PROGRAM 7.12 Standard GEE

```
proc genmod data=armdwgee;
class time treat subject;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelse;
run;
```

Likewise, the linearization-based version can be used without any problem, using Program 7.13:

PROGRAM 7.13 Linearization-based GEE

```
proc glimmix data=armdwgee empirical;
nloptions maxiter=50 technique=newrap;
class time treat subject;
model bindif = time treat*time /  noint solution dist=binary;
random _residual_ / subject=subject type=cs;
run;
```

Note that PROC GENMOD produces empirical as well as a model-based standard errors simultaneously, because of the `modelse` option. In the GLIMMIX code, we merely obtain the empirically corrected standard errors, because of the `empirical` option. Upon omitting this option, the model-based standard errors are obtained.

We now sketch the steps to be taken when conducting a weighted GEE analysis. To compute the weights, we first have to fit the dropout model using, for example, logistic regression. The outcome `dropout` is binary and indicates whether dropout occurs at a given time from the start of the measurement sequence until the time of dropout or the end of the sequence. Covariates in the model are the outcomes at previous occasions (`prev`), supplemented with genuine covariate information. The `%dropout` macro, constructed by Caroline Beunckens, is used to construct the variables `dropout` and `prev`.

Likewise, once a logistic regression has been fitted, these need to be translated into weights. These weights are defined at the individual measurement level and are equal to the product of the probabilities of not dropping out up to the measurement occasion. The last factor is either the probability of dropping out at that time or continuing the study. This task can be performed with the `%dropwgt` macro. The arguments are the same as in the `%dropout` macro, except that now also the predicted values from the logistic regression have to be passed on through the `pred=` argument, and the dropout indicator is passed on through the `dropout=` argument.

Using these macros, Program 7.14 can be used to prepare for a WGEE analysis.

PROGRAM 7.14 Preparing for WGEE (PROC GENMOD)

```
%dropout(data=armd111,id=subject,time=time,response=bindif,out=armdhlp);

proc genmod data=armdhlp descending;
class trt prev lesion time;
model dropout = prev trt lesion time / pred dist=binomial;
ods output obstats=pred;
run;

data pred;
set pred;
keep observation pred;
run;

data armdhlp;
merge pred armdhlp;
run;

%dropwgt(data=armdhlp,id=subject,time=time,pred=pred,
         dropout=dropout,out=armdwgee);
```

To sum up, the dropout indicator and previous outcome variable are defined using the `%dropout` macro, after an ordinary logistic regression is performed. Predicted values are first saved and then merged with the original data. Finally, the predicted values are translated into proper weights using the `%dropwgt` macro. Note that this approach is restricted to subject-level weights.

After these preparatory steps, we need only include the weights through the WEIGHT (or, equivalently, SCWGT) statement within the GENMOD procedure. This statement identifies a variable in the input data set to be used as the exponential family dispersion parameter weight for each observation. The exponential family dispersion parameter is divided by the WEIGHT variable value for each observation. Whereas the inclusion of the REPEATED statement turns a univariate exponential family model into GEE, the addition of WEIGHT further switches to WGEE. In other words, we merely need to add Program 7.15.

PROGRAM 7.15 Additional statement for weighted generalized estimating equations

```
weight wi;
```

Note that the use of the WEIGHT statement can also be used in the GLIMMIX procedure, so implying that a weighted version of the linearization-based GEE method is feasible.

7.7.3 WGEE and SAS, Using PROC GEE

The standard GEE analysis as presented in Program 7.12 can equivalently be implemented using PROC GEE. Literally, all it takes is to replace 'GENMOD' by 'GEE'. The main value of the procedure lies in the ease with which WGEE can be conducted. The preparatory steps are limited to defining the additional variables, needed in the weight model. In this case, this is the variable `prevbindif`, containing the previous value of `bindif` and indicators for the second and third time point.

PROGRAM 7.16 Preparing for WGEE (PROC GEE)

```
data help;
set armdwgee;
by subject;
prevbindif=lag(bindif);
if first.id then prevbindif=1;
time2=0;
if time=2 then time2=1;
time3=0;
if time=3 then time3=1;
run;
```

Upon completing this step, we merely need to add the MISSMODEL statement to the PROC GEE code. There is no need to specify an outcome variable, because this will always be the dropout indicator.

PROGRAM 7.17 WGEE, Using PROC GEE

```
proc gee data=help;
class time treat subject lesion;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch corrw modelse;
missmodel prevbindif treat lesion time2 time3 / type=obslevel;
run;
```

Note that we also include `type=obslevel` as an option to the MISSMODEL statement. It specifies that the weights need to be calculated at the level of the observation, rather than at the level of the subject as a whole. The latter corresponds to the `type=sublevel` option. Observation-level weights are the default. The estimates for the weight model are presented in Table 7.7 (second column). Note that they are similar but not identical to the ones obtained from the earlier analysis. The reason is that in the GENMOD-based analysis, all usable information from the non-monotone sequences is used to compute the weights, whereas in the PROC GEE analysis, non-monotone subjects are removed entirely from analysis. This also explains the small differences between the results presented in Tables 7.6 and 7.8. The Standard GEE (WGEE) analysis in the former is comparable to the subject-level analysis in the latter.

A very striking feature is the difference between observation- and subject-level weighting in Table 7.8. Some standard errors are 50--100% larger when observation-level weights are employed. In other words, such more refined weights overcome one of the main issues with WGEE, i.e., that of reduced precision. To emphasize this,

TABLE 7.8 The Age-related Macular Degeneration Trial. Parameter estimates (empirically corrected standard errors) for WGEE using PROC GEE, with both observation-level weights (observation) and subject-level weights (subject).

		Weights	
Effect	Parameter	observation	subject
Int.4	β_{11}	-0.95 (0.20)	-0.98 (0.35)
Int.12	β_{21}	-1.03 (0.22)	-1.77 (0.30)
Int.24	β_{31}	-1.03 (0.23)	-1.11 (0.29)
Int.52	β_{41}	-1.52 (0.30)	-1.72 (0.37)
Tr.4	β_{12}	0.32 (0.28)	0.78 (0.56)
Tr.12	β_{22}	0.65 (0.29)	1.83 (0.47)
Tr.24	β_{32}	0.39 (0.30)	0.71 (0.49)
Tr.52	β_{42}	0.30 (0.39)	0.72 (0.47)
Corr.	ρ	0.38	0.33

the observation-level analysis produces standard errors in the order of magnitude of unweighted GEE (Table 7.8).

7.7.4 Double Robustness

We finish the section on GEE by referring to more recent developments by Robins and colleagues that are designed to improve the efficiency of WGEE, more generally termed inverse probability weighting (IPW). For overviews, see Carpenter, Kenward, and Vansteelandt (2006); Molenberghs and Kenward (2007); and Molenberghs et al. (2015). Essentially, standard WGEE are supplemented with a second term, which has expectation zero given the observed data, and is most often written in terms of the predictive distribution of the unobserved outcomes given the observed ones. The method is termed doubly robust because it leads to a consistent and asymptotically normal estimator when either the model for the weights or the predictive model is correctly specified, but not necessarily both. Currently, the methodology is not yet available in standard procedures, although various user-defined implementations exist. See, for example (at `www.missingdata.org.uk`.), Mallinckrodt and Lipkovich (2016, Ch. 17), which presents SAS code for double robust estimation.

7.8 Multiple Imputation

Next to the methods already discussed, multiple imputation (MI) is an attractive tool in the modeler's kit. The method is ignorable under MAR. Extensions exist for MNAR mechanisms. These are well suited for sensitivity analyses and will be discussed in Section 7.11.

Multiple imputation (MI) was formally introduced by Rubin (1978). Rubin (1987) provides a comprehensive early treatment. Several other sources, such as Rubin and Schenker (1986); Little and Rubin (2014); Tanner and Wong (1987); Schafer (1997); van Buuren (2012); Carpenter and Kenward (2013); and O'Kelly and Ratitch (2014), offer easy-to-read descriptions of the technique.

The key MI idea is to replace each missing value with a set of M plausible values, i.e., values "drawn" from the distribution of our data, that represent the uncertainty about the right value to impute. The imputed data sets are then analyzed by using standard procedures for complete data and combining the results from these

analyses. Drawing imputations can be done in a large variety of ways, and the most commonly used of them will be discussed in what follows.

An evident question is when to use multiple imputation. This question is relevant because, given the availability of such procedures as MIXED, GLIMMIX, NLMIXED, MCMC, and related software tools, direct likelihood and direct Bayesian analyses are within reach. Also WGEE has become relatively easy to use thanks to the GEE procedures.

That said, we broadly see at least six settings where MI can be of use, without limiting it to other uses. First, when there is a combination of missing covariates and missing outcomes, multiple imputation comes in handy to either handle the incomplete covariates, or the incomplete covariates and outcomes combined. In the former case, a standard missing-data technique can be used on the incomplete outcome data with completed covariates. In the latter one, any complete-data technique can be used. Second, when several analyses are envisaged on the same set of incomplete data, missingness can be handled using MI, after which the various analyses can be undertaken. A simple example is when the same set of incomplete data will be modeled using both GEE and GLMM. Third, when a technique requires the missing data patterns to be monotone, MI can be used, either as an alternative to the technique envisaged (e.g., no direct likelihood or WGEE but rather a complete-data technique after MI), or to monotonize the incomplete data (e.g., enabling WGEE). Fourth, MI is attractive when likelihood-based analyses turn out to be difficult to implement, especially when we would like to use jointly several outcomes (e.g., binary, continuous, and/or count outcomes). See also Mallinckrodt and Lipkovich (2016, Sec. 15.4). Fifth, certain MI extensions are applicable when MNAR-type analyses are considered, in particular in the context of sensitivity analysis (see Section 7.11). Sixth, MI is useful when an analysis is to be conducted based on a discretized version of an incompletely observed continuous outcome or set of outcomes. We can then begin by imputing the original, continuous outcome, followed by discretizing the so-obtained completed data sets.

Technically, MI involves three distinct steps:

Imputation step. The missing values are filled in M times to generate M complete data sets.

Analysis step. The M complete data sets are analyzed by using standard procedures.

Inference step. The results from the M analyses are combined for inference purposes.

The SAS procedure MI creates multiple imputed data sets from incomplete p-dimensional multivariate data. It uses methods that incorporate appropriate variability across the M imputations. Once the M complete data sets are analyzed by using standard procedures, PROC MIANALYZE can be used to generate valid statistical inferences about these parameters by combining results for the M complete data sets. Alternative versions exist to combine, for example, M p-values into a single one. More details on SAS for MI are provided in Sections 7.8.5--7.8.6.

7.8.1 Theoretical Justification

Suppose we have a sample of N, i.i.d. $n \times 1$ random vectors Y_i. In a data set with the number of measurements per subject n_i variable, we can define $n = \max_{i=1}^{N} n_i$. Our interest lies in estimating some parameter vector $\boldsymbol{\theta}$ of the distribution of $\boldsymbol{Y}_i$. Multiple imputation fills in the missing data $\boldsymbol{Y}^m$ using the observed data $\boldsymbol{Y}^o$, several times, and then the completed data are used to estimate $\boldsymbol{\theta}$.

If we knew the distribution of $\boldsymbol{Y}_i = (\boldsymbol{Y}_i^o, \boldsymbol{Y}_i^m)$, with parameter vector $\boldsymbol{\theta}$, then we could impute $\boldsymbol{Y}_i^m$ by drawing a value of $\boldsymbol{Y}_i^m$ from the conditional distribution

$$f(\boldsymbol{y}_i^m | \boldsymbol{y}_i^o, \boldsymbol{\theta}).$$

The objective of the imputation process is to sample from this true predictive distribution. Since we do not know $\boldsymbol{\theta}$, we must estimate it from the data, say $\widehat{\boldsymbol{\theta}}$, and presumably use

$$f(\boldsymbol{y}_i^m | \boldsymbol{y}_i^o, \widehat{\boldsymbol{\theta}})$$

to impute the missing data. Frequentists sometimes favor incorporating uncertainty in $\boldsymbol{\theta}$ in the multiple imputation scheme using bootstrap or other methods. However, in Bayesian terms, $\boldsymbol{\theta}$ is a random variable, in which the posterior distribution is a function of the data, so we must account for its uncertainty. The Bayesian approach relies on integrating out $\boldsymbol{\theta}$, which provides a more natural and unifying framework for accounting for the uncertainty in $\boldsymbol{\theta}$. Thus, $\boldsymbol{\theta}$ is a random variable with mean equal to the estimated $\widehat{\boldsymbol{\theta}}$ from the data. Given this distribution, using multiple imputation, we first draw a random $\boldsymbol{\theta}^*$ from the distribution of $\boldsymbol{\theta}$, and then put this $\boldsymbol{\theta}^*$ in to draw a random $\boldsymbol{Y}_i^m$ from

$$f(\boldsymbol{y}_i^m | \boldsymbol{y}_i^o, \boldsymbol{\theta}^*).$$

The imputation algorithm is as follows:

1. Draw $\boldsymbol{\theta}^*$ from the distribution of $\boldsymbol{\theta}$.
2. Draw $\boldsymbol{Y}_i^{m*}$ from $f(\boldsymbol{y}_i^m | \boldsymbol{y}_i^o, \boldsymbol{\theta}^*)$. This can be done in a variety of ways, including multivariate normal models, log-linear models, a combination thereof, Monte Carlo Markov chain methods, so-called full conditional specification (FCS), etc. (van Buuren, 2012; Carpenter and Kenward, 2013).
3. To estimate $\boldsymbol{\beta}$, we then calculate the estimate of the parameter of interest, and its estimated variance, using the completed data, $(\boldsymbol{Y}^o, \boldsymbol{Y}^{m*})$:

   $$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\boldsymbol{Y}) = \widehat{\boldsymbol{\beta}}(\boldsymbol{Y}^o, \boldsymbol{Y}^{m*}),$$

   and the *within*-imputation variance is $\boldsymbol{U} = \widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}})$.
4. Repeat steps 1, 2, and 3 a number of M times $\Rightarrow \widehat{\boldsymbol{\beta}}^m$ and $\boldsymbol{U}^m$, for $m = 1, \ldots, M$.

Steps 1 and 2 constitute the *Imputation Task*. Step 3 is the *Analysis Task.*

7.8.2 Pooling Information

Of course, we want to combine the M inferences into a single one (the *Inference Task*). In this section, we will discuss parameter and precision estimation.

With no missing data, suppose that inference about the parameter $\boldsymbol{\beta}$ is made by

$$(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \sim N(\boldsymbol{0}, \boldsymbol{U}).$$

The M within-imputation estimates for $\boldsymbol{\beta}$ are pooled to give the multiple imputation estimate

$$\widehat{\boldsymbol{\beta}}^* = \frac{\sum_{m=1}^{M} \widehat{\boldsymbol{\beta}}^m}{M}.$$

Further, we can make normal based inferences for $\boldsymbol{\beta}$ based upon

$$(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^*) \sim N(\boldsymbol{0}, \boldsymbol{V}),$$

where

$$\boldsymbol{V} = \boldsymbol{W} + \left(\frac{M+1}{M}\right)\boldsymbol{B},$$

$$\boldsymbol{W} = \frac{\sum_{m=1}^{M} \boldsymbol{U}^m}{M}$$

is the average within-imputation variance, and

$$\boldsymbol{B} = \frac{\sum_{m=1}^{M} (\widehat{\boldsymbol{\beta}}^m - \widehat{\boldsymbol{\beta}}^*)(\widehat{\boldsymbol{\beta}}^m - \widehat{\boldsymbol{\beta}}^*)'}{M-1}.$$

is the *between*-imputation variance (Rubin, 1987).

7.8.3 Hypothesis Testing

When MI is used, the asymptotic results and, hence, the χ^2 reference distributions do not only depend on the sample size N, but also on the number of imputations M. Therefore, Li, Raghunathan, and Rubin (1991) propose the use of an F reference distribution with appropriate degrees-of-freedom. To test the hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, they advocate the following method to calculate p-values:

$$p = P(F_{k,w} > F),$$

where k is the length of the parameter vector $\boldsymbol{\theta}$, $F_{k,w}$ is an F random variable with k numerator and w denominator degrees of freedom, and

$$\begin{aligned}
F &= \frac{(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)' W^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)}{k(1+r)}, \\
w &= 4 + (\tau - 4)\left[1 + \frac{(1 - 2\tau^{-1})}{r}\right]^2, \\
r &= \frac{1}{k}\left(1 + \frac{1}{M}\right)\mathrm{tr}(BW^{-1}), \\
\tau &= k(M-1).
\end{aligned}$$

Here, r is the average relative increase in variance due to nonresponse across the components of $\boldsymbol{\theta}$. The limiting behavior of this F variable is that if $M \to \infty$, then the reference distribution of F approaches an $F_{k,\infty} = \chi^2/k$ distribution.

Clearly, this procedure is not only applicable when the full vector $\boldsymbol{\theta}$, but also when one component, a sub-vector, or a set of linear contrasts, is the subject of hypothesis testing. In case of a sub-vector, or as a special case one component, we use the corresponding sub-matrices of B and W in the formulas. For a set of linear contrasts $L\boldsymbol{\beta}$, we should use the appropriately transformed covariance matrices: $\tilde{W} = LWL'$, $\tilde{B} = LBL'$, and $\tilde{V} = LVL'$.

7.8.4 Efficiency

Multiple imputation is attractive because it can be highly efficient even for small values of M. Historically, numbers as small as $M = 5$ were often advocated (Rubin, 1987, p. 114). Of course, efficiency depends on a variety of factors, such as the amount of missingness, data type, and whether the inferential goal is estimation or

hypothesis testing. It is prudent to use somewhat higher values. The current SAS default is $M = 25$ (as of SAS/STAT 14.1; formerly $M = 5$ was the default). Users are encouraged to conduct simple numerical sensitivity analyses, by varying the number of imputations over a range of values, until a desired level of precision is attained. Carpenter and Kenward (2013) offer guidelines in this respect.

7.8.5 Imputation Mechanisms

The method of choice to create the imputed data sets depends on the missing data pattern and the type(s) of the outcome variables. Carpenter and Kenward (2013) describe the most commonly available methods for univariate and multivariate outcomes, as well as a number of methods developed for specific cases such as time-to-event data, data with nonlinear relationships, multilevel models, etc.

A data set is said to have a monotone missing data pattern if, a missing outcome Y_{ij} implies that Y_{ik}, $k > j$ are missing for the same individual i, perhaps after permuting the columns of the data matrix with components Y_{ij}.

The widest array of methods is available for monotone data. In the SAS procedure MI, apart from the general MCMC and FCS statements, also the MONOTONE statement can be used. Within these, several options, and hence methods, are available.

For monotone missing data patterns, a parametric *regression method* can be used that assumes multivariate normality, logistic regression, or a combination thereof. When opting for this approach, a regression model is fitted for each variable with missing values, with the previous variables as covariates. Based on the resulting model, a new regression model is then fitted and is used to impute the missing values for each variable (Rubin, 1987), in line with steps 1 and 2 in Section 7.8.1. Since the data set has a monotone missing data pattern, the process can easily be repeated sequentially for variables with missing values. To this end, the options `reg` and `logistic` are available. In addition, for categorical data, a discriminant analysis based method can be used, using the `discrim` option.

When the `logistic` option is used with the FCS or MONOTONE statements, we can make use of the `likelihood=augment` sub-option. This tool is handy when maximum likelihood estimates for logistic regression do not exist (or tend to infinity) because of so-called quasi-complete separation. The methodology was developed by White, Daniel, and Royston (2010).

Alternatively, we can rely on a nonparametric method that uses *propensity scores*, by means of the `propensity` option. The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates (Rosenbaum and Rubin, 1983). In the propensity score method, a propensity score is generated for each variable with missing values to indicate the probability of that observation being missing. The observations are then grouped based on these propensity scores, and an approximate Bayesian bootstrap imputation (Rubin, 1987) is applied to each group. The propensity score method uses only the covariate information that is associated with whether the imputed variable values are missing. It does not use correlations among repeated measures. It is effective for inferences about the distributions of individual imputed variables, but it is not appropriate for analyses involving relationships among variables.

Finally, for monotone data, the so-called *predictive mean matching (PMM)* is also available, using the option `regpmm`. The method is similar to regression imputation, except that the value imputed is not merely taken from the predictive distribution, but rather from a pool of donors with value close to the predictive mean. In some applications, simply the closest value is selected, while in others a random selection still takes place (Heitjan and Litte, 1991; Carpenter and Kenward, 2013, p. 133).

For arbitrary missing data patterns, and, hence, in particular for monotone patterns as well, there are two main methods: *multivariate modeling* (also referred to as joint modeling; Schafer 1997, Carpenter and Kenward 2013) and full conditional specification (FCS). In SAS, the former is implemented using the MCMC statement, while the latter has become available since SAS 9.4 by way of the FCS statement.

In statistical applications, MCMC is used to generate pseudo-random draws from multidimensional and otherwise intractable probability distributions via Markov chains. A Markov chain is a sequence of random variables in which the distribution of each element depends on the value of the previous one(s). In the *MCMC method*, we construct a Markov chain long enough for the distribution of the elements to stabilize to a target distribution. This stationary distribution is the one of interest. By repeatedly simulating steps of the chain, it simulates draws from the distribution of interest.

In more detail, the MCMC method works as follows. We assume that the data are from a multivariate normal distribution, $N(\boldsymbol{\mu}, \Sigma)$, say. We then proceed as follows.

1. In the first step, the **initial step**, we have to choose starting values, $\boldsymbol{Y}_i \sim \boldsymbol{\mu}^{(0)}$ and $\Sigma^{(0)}$, say. This can be done by computing a vector of means and a covariance matrix from the complete data. These are used to estimate the prior distribution. More precisely, this means that the parameters of the prior distributions for means and variances of the multivariate normal distribution are estimated, using the informative prior option.
2. The next step is called the **imputation step**: Values for missing data items are simulated by randomly selecting a value from the available distribution of values, i.e., the predictive distribution of missing values given the observed values. Technically, the predictive distribution makes use of the result:

$$\begin{aligned}&\boldsymbol{Y}_i^{(m)}|\boldsymbol{y}_i^{(o)}\\&\sim N\left(\boldsymbol{\mu}_i^{(m)} + \Sigma^{(mo)}\Sigma^{(mm)-1}\left(\boldsymbol{y}_i^{(o)} - \boldsymbol{\mu}_i^{(o)}\right); \Sigma^{(mm)} - \Sigma^{(mo)}\Sigma^{(mm)-1}\Sigma^{(mo)}\right).\end{aligned}$$

3. In the **posterior step**, the posterior distribution of the mean and covariance parameters is updated, by updating the parameters governing their distribution (e.g., the inverted Wishart distribution for the variance-covariance matrix and the normal distribution for the means). This is then followed by sampling from the posterior distribution of mean and covariance parameters, based on the updated parameters.
4. The previous two steps, i.e., the imputation and the posterior steps, are iterated until the distribution is stationary. This means that the mean vector and covariance matrix are unchanged as we iterate.
5. To conclude, we use the imputations from the final iteration to form a data set that has no missing values.

The MCMC method is implemented in the SAS procedure MI via the MCMC statement. While in principle the method applies to arbitrary data type, its SAS implementation is restricted to multivariate normal data. If needed, the TRANSFORM statement can be used to transform clearly non-normal outcomes to (near) normality. An important feature of the MCMC statement is that it can be used to either impute all missing values or just enough to make non-monotonically missing data patterns monotone. In the latter case, the '`impute=monotone` option has to be used. It is important to note that this option is entirely different, syntactically and semantically, from the MONOTONE statement discussed earlier:

- The MONOTONE statement takes monotone patterns as input and returns completed data.

- The MCMC statement with `impute=monotone` option also takes data with non-monotone patterns as input and returns monotonized data.

The most recent addition to the MI procedure is FCS, through the FCS statement (van Buuren et al., 1999; van Buuren, 2007, 2012; and Carpenter and Kenward, 2013, Sec. 3.3). The monotone regression method is easy, because every missing value in a monotone sequence can be predicted by its predecessors. It is flexible in that it can handle a combination of continuous and categorical outcomes. On the other hand, many application data sets have non-monotone patterns, even though they might not be the majority of the patterns. The method can best be viewed as a modification of the monotone method. The outcomes are first ordered such that the patterns are as close to monotone as possible. As an initial step, the missing values are imputed by drawing with replacement from the observed values of each variable. Then a number of cycles are run, each of which consists of two steps:

- A regression of the observed part of the jth variable, $\boldsymbol{Y}_j$, on the remaining variables is conducted. In this regression, missing values are replaced by the current value of the imputations.
- Given the result of these regressions, and using the same algorithm as with regression imputation, new imputations for the missing values in $\boldsymbol{Y}_j$ are imputed. To account for parameter uncertainty, not just outcomes but also the parameters $\boldsymbol{\theta}$ are sampled.

After a set of burn-in sequences, the first imputation is obtained. After that, a new set of cycles is run to obtain the second imputation, etc. Apart from the regression method, the FCS statement also allows for the logistic, discriminant, and predictive mean matching methods.

EXAMPLE: Age-related macular degeneration trial

When analyzing the data using WGEE in Section 7.7, one complication that arose was that only monotone missingness is allowed. We, therefore, removed the non-monotone sequences. To overcome this, MI is an appealing alternative. We can either monotonize the data and still apply WGEE, or impute the incomplete data altogether, followed by standard GEE. We will refer to the latter methods as MI-GEE.

An appealing feature of MI is that imputation can be based on the continuous outcome (visual acuity in this case) *before* dichotomizing the outcome. See also Mallinckrodt and Lipkovich (2016, Sec. 15.4). In other words, more information can be used during imputation than when analyzing the data. Here, this takes the form of outcomes prior to dichotomization. Additionally, auxiliary covariates can be used in the imputation process as well. We take both of these measures in this analysis. Ten multiply-imputed data sets were created. The imputation model also included, apart from the four continuous outcome variables, the four-point categorical variable 'lesions.' For simplicity, the latter was treated as continuous. Separate imputations were conducted for each of the two treatment groups. These choices imply that the imputed values depend on lesions and treatment assignment, and, hence, analysis models that include one or both of these effects are *proper* in the sense of Rubin (1987). This means, broadly speaking, that the model used for imputation should include all relationships that will be considered later in the analysis and inference steps. The added advantage of including 'lesions' in the imputation model is that even individuals for which none of the four follow-up measurements are available, are still imputed and hence retained for analysis. Using the SAS procedure MI, the MCMC method was used, with EM starting values, and a single chain for all imputations.

Upon imputation, a marginal model (using GEE as in Section 7.7) was fitted, together with a generalized linear mixed model as in Section 7.5. The final results, obtained by making use of Rubin's combination rules, are reported in Table 7.9. The parameter estimates and standard errors are very similar to their counterparts in Table 7.6 and 7.5, respectively. In the GEE case, the parameter estimates are similar to those in Table 7.8, obtained with observation-level weights. Also, the similarity between the direct likelihood method (bottom right column of Table 7.5) is clear, with only a minor deviation in estimate for the treatment effect after one year.

TABLE 7.9 The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) for the standard GEE and numerical-integration based random-intercept models, after generating 10 multiple imputations.

Effect	Par.	GEE	GLMM
Int.4	β_{11}	-0.84(0.20)	-1.46(0.36)
Int.12	β_{21}	-1.02(0.22)	-1.75(0.38)
Int.24	β_{31}	-1.07(0.23)	-1.83(0.38)
Int.52	β_{41}	-1.61(0.27)	-2.69(0.45)
Trt.4	β_{12}	0.21(0.28)	0.32(0.48)
Trt.12	β_{22}	0.60(0.29)	0.99(0.49)
Trt.24	β_{32}	0.43(0.30)	0.67(0.51)
Trt.52	β_{42}	0.37(0.35)	0.52(0.56)
R.I. s.d.	τ		2.20(0.26)
R.I. var.	τ^2		4.85(1.13)

7.8.6 SAS for Multiple Imputation

To conduct multiple imputation in SAS, a sequence of three procedures is used:

Imputation Task: PROC MI: To generate M imputed data sets.

Analysis Task: Data analysis procedure: Using an appropriate procedure or other analysis tool, the M imputed data sets are analyzed. For example, if a GLMM is envisaged, PROC GLIMMIX or PROC NLMIXED can be used. Routinely, parameter estimates and their estimated variance-covariance matrices are saved into data sets.

Inference Task: PROC MIANALYZE: The combination rules are applied to the data sets saved in the previous step and appropriate inferences are drawn.

We discuss each of these in turn.

PROC MI

Some information on PROC MI was already given in Section 7.8.5, in relation to the methodology used for generating imputations.

PROC MI creates M imputed data sets, physically stored in a single data set with indicator `_IMPUTATION_` to separate the various imputed copies from each other. We will describe some options available in the PROC MI statement. The option `simple` displays simple descriptive statistics and pairwise correlations based on available cases in the input data set. The number of imputations is specified by `nimpute` and is by default equal to 25 (as of SAS/STAT 14.1). The option `round` controls the number of decimal places in the imputed values (by default, there is

no rounding). If more than one number is specified, we should also use a VAR statement, and the specified numbers must correspond to variables in the VAR statement. The `seed` option specifies a positive integer, which is used by PROC MI to start the pseudo-random number generator. The default is a value generated from the time of day from the computer's clock.

The imputation step is carried out separately for each level of the BY variables.

As stated in Section 7.8.5, there are three imputation statements. For monotone missingness only, we use the MONOTONE statement, with options: `reg` for the standard regression method, `logistic` for the logistic regression method, `discrim` for the discriminant analysis method, `regpmm` for the predictive mean matching method, and `propensity` for the propensity score method. We can specify more than one method in the MONOTONE statement, and for each imputed variable, the covariates can be specified separately.

For general patterns of missingness, we can use the MCMC statement, which is also the default. Recall that the method uses a multivariate normal model. or the MCMC method. We can give the initial mean and covariance estimates to begin the MCMC process by `initial`. Tools are available to monitor convergence of the MCMM sequence. With `initial=EM` (default), PROC MI uses the means and standard deviations from available cases as the initial estimates for the EM algorithm. The resulting estimates are used to begin the MCMC process. We can also specify `initial=`*input SAS-data-set* to use a SAS data set with the initial estimates of the mean and covariance matrix for each imputation. Further, `niter` specifies the number of iterations between imputations in a single chain (the default is equal to 30).

As already mentioned in Section 7.8.5, for full conditional specification, the FCS statement is available, with the same modeling options as for the MONOTONE statement (`reg`, `logistic`, `discrim`, and `regpmm`), except `propensity`. Using `nbiter`, the number of burn-in iterations can be specified.

The CLASS statement is intended to specify categorical variables. Such classification variables are used as either covariates for imputed variables or as imputed variables for data sets with monotone missingness patterns. When a CLASS statement is included, either MONOTONE or FCS must be used.

Of note is the EM statement. It calculates expectation-maximization (EM) algorithm based parameter estimates for a multivariate normal sample (Dempster, Laird, and Rubin, 1977). When the number of iterations is set equal to zero and the EM statement is invoked, PROC MI in fact calculates EM-based rather than MI-based estimates. The EM estimates are also useful as initial values for the various MI techniques.

When a variable is assumed to be normally distributed but its actual distribution is deviating from it, the TRANSFORM statement can be used to transform the variable, using one of the prescribed transformations, such as Box-Cox, logarithmic, logistic, etc.

When using MI for longitudinal or otherwise hierarchical data, some data analysis is necessary before and after invoking PROC MI. In most hierarchical data sets, there is a single data set line reserved for each measurements. This implies that a subject (block) runs across several lines. We term this the vertical (counting process) layout. However, PROC MI assumes that each line is an independent block, the horizontal (multivariate) layout. Therefore, a hierarchical data set has to be transformed from a vertical to a horizontal format prior to calling PROC MI. Afterwards, the output data set needs to be transformed again to the vertical format, to allow calling one of the hierarchical procedures, such as GENMOD, GEE, MIXED, GLIMMIX, NLMIXED, etc.

Discussion of one further statement available in the MI procedure, the MNAR statement, is deferred to Section 7.11.

Data Analysis Procedure

Next, the imputed data sets are analyzed using a standard procedure. It is important to ensure that the `BY _imputation_` syntax is used to force an analysis for each of the imputed sets of data separately. Appropriate output (estimates and the precision thereof) is stored in output data sets, typically using the generic ODS statement. In most cases, it is also advisable to save parameter names along with their values, to facilitate proper matching between the components of a parameter vector and that of the corresponding variance-covariance matrix.

PROC MIANALYZE

Finally, PROC MIANALYZE combines the M inferences into a single one, by making use of the theory laid out in Section 7.8.2. Appropriate output data sets generated by the analysis procedure and containing parameter estimates, precision estimates, and parameter names, are used as input data sets to PROC MIANALYZE. Depending on the input procedure, such information is passed on using one or more of the following options: `parms=`, `data=`, `parminfo=`, `covb=`, and/or `xpxi=`.

The parameters to be analyzed are passed on via the MODELEFFECTS statement. If some effects correspond to categorical variables, the CLASS statement should be used. Unless a dedicated data structure is used to pass on parameter estimates and the corresponding variance-covariance matrices, the STDERR statement needs to be used to pass on the corresponding standard errors. In that case, both estimates and standard errors come from an ordinary SAS data set.

The TEST statement allows testing of hypotheses about linear combinations of the parameters. The statement is based on Rubin (1987), and uses a t distribution which is the univariate version of the work by Li, Raghunathan, and Rubin (1991), described in Section 7.8.3. Several tests can be combined; each hypothesis testing can be simple or compound.

7.8.7 SAS Code for Age-related Macular Degeneration Trial

The three steps associated with MI, discussed earlier in this section, will be illustrated using the ARMD analysis reported in Section 7.8.5.

The MI Procedure for the Imputation Task

PROC MI is used to generate the imputations. It creates M imputed data sets from an input data set, physically stored in a single data set with indicator variable `_imputation_`, created by the procedure, to separate the imputed copies.

For imputations from a multivariate Gaussian imputation model, the following MI program can be used:

PROGRAM 7.18 PROC MI for the Imputation Task, using MCMC

```
proc mi data=armd13 seed=486048 out=armd13a simple nimpute=10 round=0.1;
var lesion diff4 diff12 diff24 diff52;
by treat;
run;
```

We have chosen to generate $M = 10$ imputed data sets, rather than the default number of 25. Imputed values are rounded to one decimal place (by including the `round=` option). Simple statistics are displayed in the output, because of the inclusion of the `simple` option. As stated before, we allow for the imputation model to depend on `lesion` (treated as a continuous variable) and `treat` (treated as categorical, by including it in the BY statement).

The `seed=` option is useful when a given data analysis needs to be reproducible, because then every time the same seed is used on the same input data and with the same imputation model (and hence the same program), the same random values will be generated.

Observe that no imputation method is specified, i.e., that there is no MONOTONE, MCMC, or FCS statement. As a result, the default MCMC method will be invoked. To use, for example, FCS, Program 7.19 can be used instead.

PROGRAM 7.19 PROC MI for the Imputation Task, using FCS

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13fcs nimpute=30
        round=0.01;
fcs reg(diff4=lesion);
fcs reg(diff12=lesion diff4);
fcs reg(diff24=lesion diff4 diff12);
fcs reg(diff52=lesion diff4 diff12 diff24);
var lesion diff4 diff12 diff24 diff52;
by treat;
run;
```

Note that, after carrying out the imputation step, the data are still in horizontal format and need to put in the longitudinal, or vertical, format again, which will be done at the outset of the analysis step.

The Analysis Task

The imputed data sets are now analyzed using a standard complete data procedure. It is important to include `BY _imputation_` to ensure that a separate analysis be carried out for each completed data set.

Also, parameter estimates and their estimated covariance matrices need to be stored in appropriate output data sets, so they can be passed on to the MIANALYZE procedure. We will return to this when discussing the inference step.

To prepare for the data analysis, indicator variables are created, and then the data are sorted by imputation number. A step, specific for our analysis, is that we need to dichotomize the variables.

PROGRAM 7.20 Dichotomization of imputed data

```
proc sort data=m.armd13a;
by _imputation_ subject;
run;

data m.armd13a;
set m.armd13a;
bindif4=0;  if diff4 <= 0   then bindif4=1;
bindif12=0; if diff12 <= 0  then bindif12=1;
bindif24=0; if diff24 <= 0  then bindif24=1;
bindif52=0; if diff52 <= 0  then bindif52=1;
if diff4=. then bindif4=.;
if diff12=. then bindif12=.;
if diff24=. then bindif24=.;
if diff52=. then bindif52=.;
run;
```

Next, the data are transformed from the horizontal format to a vertical one, to allow for longitudinal analyses.

PROGRAM 7.21 Transforming a horizontal data set in a vertical data set

```
data m.armd13b;
set m.armd13a;
array x (4) bindif4 bindif12 bindif24 bindif52;
array y (4) diff4 diff12 diff24 diff52;
do j=1 to 4;
   bindif=x(j);
   diff=y(j);
   time=j;
   output;
end;
run;
```

While the MIXED, GEE, GENMOD, and GLIMMIX procedures can handle CLASS variables, dummies need to be expressly created for use with the NLMIXED procedure.

PROGRAM 7.22 Creating dummies

```
data m.armd13c;
set m.armd13b;
time1=0;
time2=0;
time3=0;
time4=0;
trttime1=0;
trttime2=0;
trttime3=0;
trttime4=0;
if time=1 then time1=1;
if time=2 then time2=1;
if time=3 then time3=1;
if time=4 then time4=1;
if (time=1 & treat=1) then trttime1=1;
if (time=2 & treat=1) then trttime2=1;
if (time=3 & treat=1) then trttime3=1;
if (time=4 & treat=1) then trttime4=1;
run;

proc sort data=m.armd13cs;
by _imputation_ subject time;
run;
```

The GENMOD or GEE procedures can then be called for a GEE analysis.

PROGRAM 7.23 GEE after multiple imputation

```
proc gee data=armd13c;
class time subject;
by _imputation_;
model bindif = time1 time2 time3 time4 trttime1 trttime2 trttime3 trttime4
      / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelse covb;
ods output GEEEmpPEst=gmparms parminfo=gmpinfo GEERCov=gmcovb;
run;
```

While we could have used the coding `time treat*time`, the already created dummies are used instead. This makes no difference. The BY statement has been added, as well as the ODS statement, to store the parameter estimates and the covariance parameters. For the latter, the `parminfo=` option is used next to the `covb=` option, to ensure that the proper names of the covariate effects are mapped to abbreviations of type `Prm1`, etc. The parameter estimates are generated by default. The output of the GEE procedure will be a GEE analysis for each of the ten imputed data sets. As such, they represent an intermediate step in the full multiple imputation analysis and are of no direct scientific interest. Formal inference needs to be conducted using only the results from the inference step.

Because the `noint` option was included, the effect `Prm1` formally exists when PROC GENMOD is used (not when PROC GEE is used), but is unavailable as a parameter estimate. It is, therefore, necessary to delete it from the parameter information, as in Program 7.24:

PROGRAM 7.24 Deletion of redundant intercept name

```
data gmpinfo;
set gmpinfo;
if parameter='Prm1' then delete;
run;
```

The above program should not be used with PROC GEE.

Analogously, the GLMM analysis can be conducted on the multiple imputed data sets. Evidently, both the procedures GLIMMIX and NLMIXED can be used. It is interesting to illustrate the use of a programming-type procedure, such as NLMIXED, in conjunction with multiple imputation.

PROGRAM 7.25 GLMM after multiple imputation

```
proc nlmixed data=armd13c qpoints=20 maxiter=100 technique=newrap cov ecov;
by _imputation_;
eta = beta11*time1+beta12*time2+beta13*time3+beta14*time4+b
     +beta21*trttime1+beta22*trttime2+beta23*trttime3+beta24*trttime4;
p = exp(eta)/(1+exp(eta));
model bindif ~ binary(p);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau2' tau*tau;
ods output ParameterEstimates=nlparms
           CovMatParmEst=nlcovb
           AdditionalEstimates=nlparmsa
           CovMatAddEst=nlcovba;
run;
```

Apart from adding the BY statement, we now also generate four output data sets using the ODS statement. For the standard model parameters, we only need the `parameterestimates=` and `covmatparmest=` options. If, in addition, multiple imputation inference is requested about additional estimates, then they can be saved as well using the `additionalestimates=` and `covmataddest=` options. However, it is also possible to calculate the additional estimates directly from the results of the inference step, i.e., to conduct multiple imputation inference first and then calculate additional estimates, rather than the other way around. For both covariance matrices to be generated, the options `cov` and `ecov`, respectively, need to be included into the PROC NLMIXED statement.

For both the GEE and GLMM models, we can now conduct multiple imputation inference, following Rubin's combination rules.

The Inference Task

Applying the MIANALYZE procedure to the GEE analysis on the ARMD data, presented in Section 7.8.7, can be done using the code in Program 7.26.

PROGRAM 7.26 Inference step after GEE

```
proc mianalyze parms=gmparms covb=gmcovb parminfo=gmpinfo wcov bcov tcov;
modeleffects time1 time2 time3 time4 trttime1 trttime2 trttime3 trttime4;
run;
```

Conducting multiple imputation inference for the NLMIXED analysis, presented in Section 7.8.7, is done by means of Program 7.27.

PROGRAM 7.27 Inference step after GLMM

```
proc mianalyze parms=nlparms covb=nlcovb wcov bcov tcov;
modeleffects beta11 beta12 beta13 beta14 beta21 beta22 beta23 beta24;
run;
```

7.8.8 Creating Monotone Missingness

When missingness is non-monotone, we might think of several mechanisms operating simultaneously: e.g., a simple (MCAR or MAR) mechanism for the intermediate missing values and a more complex (MNAR) mechanism for the missing data past the moment of dropout. However, analyzing such data is complicated since many model strategies, especially those under the assumption of MNAR, have been developed for dropout only. Therefore, a solution might be to generate multiple imputations that make all patterns monotone, by use of Program 7.28.

PROGRAM 7.28 Creating monotone missingness

```
mcmc impute=monotone;
```

Once done, we can apply a method of choice to the so-completed multiple sets of data. Note that this is different from the monotone method in PROC MI, intended to fully complete already monotone sets of data.

7.9 An Overview of Sensitivity Analysis

All methods considered so far are valid under MAR and then evidently also under MCAR. The only exception is unweighted GEE, for which in general MCAR is required.

We should not lose sight of the fact that an MNAR mechanism might be operating while it is at the same time formally impossible to distinguish between MAR and MNAR mechanisms, based on observed data alone (Molenberghs et al., 2008). Thus, while it is formally possible to fit models under the assumption of MNAR (Diggle and Kenward, 1994; Verbeke and Molenberghs, 2000, Ch. 18), these should not be considered as evidence for or against MNAR. It is a more viable route to explore how sensitive key inferences (e.g., in terms of parameter estimation or hypothesis testing) are to varying assumptions about the missing-data mechanism. This type of sensitivity to non-identifiable assumptions has been reported in various publications (Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005; Molenberghs and Kenward, 2007).

Therefore, a sensible compromise between blindly shifting to MNAR models or ignoring them altogether, is to make them a component of a sensitivity analysis.

Broadly, we could define a sensitivity analysis as one in which several statistical models are considered simultaneously and/or where a statistical model is further scrutinized using specialized tools (such as diagnostic measures). This rather loose and very general definition encompasses a wide variety of approaches. The simplest procedure is to fit a selected number of (MNAR) models that are all deemed plausible, or one in which a preferred (primary) analysis is supplemented with a number of variations. The extent to which conclusions (inferences) are stable across such ranges provides an indication about the belief that can be put into them. Variations to a basic model can be constructed in different ways. The most obvious strategy is to consider various dependencies of the missing data process on the outcomes and/or on covariates. Alternatively, the distributional assumptions of the models can be changed. For example, it is natural to start from a primary model of MAR type, and then to consider variations of an MNAR nature.

Several authors have proposed the use of global and local influence tools (Verbeke et al., 2001, Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005). An important question is to what exactly are the sources causing an MNAR model to provide evidence for MNAR against MAR? There is evidence to believe that a multitude of outlying aspects, but not necessarily the (outlying) nature of the missingness mechanism in one or a few subjects, is responsible for an apparent MNAR mechanism (Jansen et al., 2006). The consequence of this is that local influence should be applied and interpreted with due caution. This methodology will be illustrated in Section 7.10.

Another route for sensitivity analysis is by making use of pattern-mixture models (Little, 1993, 1994; Thijs et al., 2002; Michiels et al., 2002). In a PMM, the joint distribution of $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$ is factored as the conditional distribution of $\boldsymbol{Y}_i$ given $\boldsymbol{R}_i$ and the marginal distribution of $\boldsymbol{R}_i$. Recently, this family has gained considerable interest, also due to the work of Carpenter, Roger, and Kenward (2013) and Carpenter and Kenward (2013). Some PMM-based strategies are implemented in the SAS procedure MI, through the MNAR statement. This family will be examined in detail in Section 7.11.

A further framework consists of so-called shared parameter models, where random effects are employed to describe the relationship between the measurement and dropout processes (Wu and Carroll, 1988; DeGruttola and Tu, 1994)

Robins, Rotnitzky, and Scharfstein (1998) discuss sensitivity analysis in a semi-parametric context.

Further, within the selection model framework, Baker, Rosenberger, and Der-Simonian (1992) proposed a model for multivariate and longitudinal binary data, subject to non-monotone missingness. Jansen et al. (2003) extended this model to allow for (possibly continuous) covariates, and developed a local influence strategy.

Finally, classical inference procedures account for the imprecision resulting from the stochastic component of the model. Less attention is devoted to the uncertainty arising from (unplanned) incompleteness in the data, even though the majority of clinical studies suffer from incomplete follow-up. Molenberghs et al. (2001) acknowledge both the status of imprecision, due to (finite) random sampling, as well as ignorance, due to incompleteness. Further, both can be combined into uncertainty (Kenward, Molenberghs, and Goetghebeur, 2001).

7.10 Sensitivity Analysis Using Local Influence

We first introduce the Diggle and Kenward (1994) model, combining a linear mixed model with a model for dropout based on logistic regression, in the spirit of (7.3.11). Thereafter, the use of local influence to examine sensitivity is described.

7.10.1 The Model of Diggle and Kenward (DK; 1994)

In agreement with notation introduced in Section 7.3, we assume that a vector of outcomes $\boldsymbol{Y}_i$ is designed to be measured. If dropout occurs, $\boldsymbol{Y}_i$ is only partially observed. We denote the occasion at which dropout occurs by $D_i > 1$, and $\boldsymbol{Y}_i$ is split into the $(D_i - 1)$-dimensional observed component $\boldsymbol{Y_i}^o$ and the $(n_i - D_i + 1)$-dimensional missing component $\boldsymbol{Y_i}^m$. In case of no dropout, we let $D_i = n_i + 1$, and $\boldsymbol{Y}_i$ equals $\boldsymbol{Y_i}^o$. The likelihood contribution of the ith subject, based on the observed data $(\boldsymbol{y_i}^o, d_i)$, is proportional to the marginal density function

$$\begin{aligned} f(\boldsymbol{y_i}^o, d_i|\boldsymbol{\theta}, \boldsymbol{\psi}) &= \int f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi})\ d\boldsymbol{y_i}^m \\ &= \int f(\boldsymbol{y}_i|\boldsymbol{\theta}) f(d_i|\boldsymbol{y}_i, \boldsymbol{\psi})\ d\boldsymbol{y_i}^m, \end{aligned} \tag{7.10.30}$$

in which a marginal model for $\boldsymbol{Y}_i$ is combined with a model for the dropout process, conditional on the response, and where $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are vectors of unknown parameters in the measurement model and dropout model, respectively.

Let $\boldsymbol{h}_{ij} = (y_{i1}, \ldots, y_{i;j-1})$ denote the observed history of subject i up to time $t_{i,j-1}$. The Diggle-Kenward model for the dropout process allows the conditional probability for dropout at occasion j, given that the subject was still observed at the previous occasion, to depend on the history $\boldsymbol{h}_{ij}$ and the possibly unobserved current outcome y_{ij}, but not on future outcomes y_{ik}, $k > j$. These conditional probabilities $P(D_i = j|D_i \geq j, \boldsymbol{h}_{ij}, y_{ij}, \boldsymbol{\psi})$ can now be used to calculate the probability of dropout at each occasion:

$$\begin{aligned} P(D_i = j|\boldsymbol{y}_i, \boldsymbol{\psi}) \ &= \ P(D_i = j|\boldsymbol{h}_{ij}, y_{ij}, \boldsymbol{\psi}) \\ &= \begin{cases} P(D_i = j|D_i \geq j, \boldsymbol{h}_{ij}, y_{ij}, \boldsymbol{\psi}) & j = 2, \\ P(D_i = j|D_i \geq j, \boldsymbol{h}_{ij}, y_{ij}, \boldsymbol{\psi}) & \\ \qquad \times \prod_{k=2}^{j-1} [1 - P(D_i = k|D_i \geq k, \boldsymbol{h}_{ik}, y_{ik}, \boldsymbol{\psi})] & j = 3, \ldots, n_i, \\ \prod_{k=2}^{n_i} [1 - P(D_i = k|D_i \geq k, \boldsymbol{h}_{ik}, y_{ik}, \boldsymbol{\psi})] & j = n_i + 1. \end{cases} \end{aligned}$$

Diggle and Kenward (1994) combine a multivariate normal model for the measurement process with a logistic regression model for the dropout process. More specifically, the measurement model assumes that the vector $\boldsymbol{Y}_i$ of repeated measurements for the ith subject satisfies the linear regression model $\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, V_i)$, $(i = 1, \ldots, N)$. The matrix V_i can be left unstructured or assumed to be of a specific form, e.g., resulting from a linear mixed model, a factor-analytic structure, or spatial covariance structure (Verbeke and Molenberghs, 2000).

In the particular case that a linear mixed model is assumed, we write

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \varepsilon_i, \tag{7.10.31}$$

(Verbeke and Molenberghs, 2000) where $\boldsymbol{Y}_i$ is the n dimensional response vector for subject i, $1 \leq i \leq N$; N is the number of subjects; X_i and Z_i are $(n \times p)$ and $(n \times q)$ known design matrices; $\boldsymbol{\beta}$ is the p dimensional vector containing the fixed effects; and $\boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$ is the q dimensional vector containing the random effects. The residual components $\boldsymbol{\varepsilon_i} \sim N(0, \Sigma_i)$.

The logistic dropout model can, for example, take the form:

$$\begin{aligned}&\text{logit}\,[P(D_i = j \mid D_i \geq j, \boldsymbol{h}_{ij}, y_{ij}, \boldsymbol{\psi})]\\ &= \psi_0 \;+\; \psi_1 y_{i,j-1} \;+\; \psi_2 y_{ij}.\end{aligned} \tag{7.10.32}$$

More general models can easily be constructed by including the complete history $\boldsymbol{h}_{ij} = (y_{i1}, \ldots, y_{i;j-1})$, as well as external covariates, in the above conditional dropout model. Note also that, strictly speaking, we could allow dropout at a specific occasion to be related to all future responses as well. However, this is rather counter-intuitive in many cases. Moreover, including future outcomes seriously complicates the calculations since computation of the likelihood (7.10.30) then requires evaluation of a possibly high-dimensional integral. Note also that special cases of model (7.10.32) are obtained from setting $\psi_2 = 0$ or $\psi_1 = \psi_2 = 0$, respectively. In the first case, dropout is no longer allowed to depend on the current measurement, implying MAR. In the second case, dropout is independent of the outcome, which corresponds to MCAR.

Diggle and Kenward (1994) obtained parameter and precision estimates by maximum likelihood. The likelihood involves marginalization over the unobserved outcomes $\boldsymbol{Y_i}^m$. Practically, this involves relatively tedious and computationally demanding forms of numerical integration. This, combined with likelihood surfaces tending to be rather flat, makes the model difficult to use. These issues are related to the problems to be discussed next.

7.10.2 Local Influence

The local influence approach, suggested by Cook (1986), can be used to investigate the effect of extending an MAR model for dropout in the direction of MNAR dropout (Verbeke et al., 2001).

We start from the DK model introduced in Section 7.10.1. Since no data would be observed otherwise, we assume that the first measurement Y_{i1} is obtained for every subject in the study. We denote the probability of dropout at occasion k, given that the subject was still in the study up to occasion k by $g(\boldsymbol{h}_{ik}, y_{ik})$. For the dropout process, we now consider an extension of model (7.10.32), which can be written as

$$\begin{aligned}\text{logit}\,[g(\boldsymbol{h}_{ik}, y_{ik})] &= \text{logit}\,[\mathrm{P}(D_i = k | D_i \geq k, \boldsymbol{y}_i)]\\ &= \boldsymbol{h}_{ik}\boldsymbol{\psi} + \omega y_{ij}.\end{aligned} \tag{7.10.33}$$

When ω equals zero and the model assumptions made are correct, the posited dropout model is MAR, and all parameters can be estimated using standard software since the measurement and dropout model can then be fitted separately. If $\omega \neq 0$, the dropout process is assumed to be MNAR. Now, a dropout model might be found to be MNAR solely because one or a few influential subjects have driven the analysis. To investigate sensitivity of estimation of quantities of interest, such as treatment effect, growth parameters, or the dropout model parameters, with respect to assumptions about the dropout model, we consider the following perturbed version of (7.10.33):

$$\begin{aligned}\text{logit}\,[g(\boldsymbol{h}_{ik}, y_{ik})] &= \text{logit}\,[\mathrm{P}(D_i = k | D_i \geq k, \boldsymbol{y}_i, W_i)]\\ &= \boldsymbol{h}_{ik}\boldsymbol{\psi} + \omega_i y_{ik} \qquad i = 1, \ldots, N.\end{aligned} \tag{7.10.34}$$

There is a fundamental difference with model (7.10.33) since the ω_i should not be viewed as parameters: They are local, individual-specific perturbations around a

null model. In our case, the null model will be the MAR model, corresponding to setting $\omega = 0$ in (7.10.33). Thus, the ω_i are perturbations that will be used only to derive influence measures (Cook, 1986).

This scheme enables studying the effect of how small perturbations in the MNAR direction can have a large impact on key features of the model. Practically, one way of doing this is to construct local influence measures (Cook, 1986). Clearly, not all possible forms of impact resulting from sensitivity to dropout model assumptions, will be found in this way, and the method proposed here should be viewed as one component of a sensitivity analysis (e.g., Molenberghs, Kenward, and Goetghebeur, 2001).

When small perturbations in a specific ω_i lead to relatively large differences in the model parameters, it suggests that the subject is likely to drive the conclusions.

Cook (1986) suggests that more confidence can be put in a model that is relatively stable under small modifications. The best known perturbation schemes are based on case deletion (Cook and Weisberg, 1982) in which the effect is studied of completely removing cases from the analysis. A quite different paradigm is the local influence approach where we investigate how the results of an analysis are changed under small perturbations of the model. In the framework of the linear mixed model, Beckman, Nachtsheim, and Cook (1987) used local influence to assess the effect of perturbing the error variances, the random-effects variances, and the response vector. In the same context, Lesaffre and Verbeke (1998) have shown that the local influence approach is also useful for the detection of influential subjects in a longitudinal data analysis. Moreover, since the resulting influence diagnostics can be expressed analytically, they often can be decomposed in interpretable components, which yield additional insights into the reasons why some subjects are more influential than others.

We are interested in the influence of MNAR dropout on the parameters of interest. This can be done in a meaningful way by considering (7.10.34) as the dropout model. Indeed, $\omega_i = 0$ for all i corresponds to an MAR process, which cannot influence the measurement model parameters. When small perturbations in a specific ω_i lead to relatively large differences in the model parameters, this suggests that these subjects might have a large impact on the final analysis. However, even though we might be tempted to conclude that such subjects drop out non-randomly, this conclusion is misguided because we are not aiming to detect (groups of) subjects that drop out non-randomly but rather subjects that have a considerable impact on the dropout and measurement model parameters. Indeed, a key observation is that a subject that drives the conclusions towards MNAR might be doing so, not only because its true data generating mechanism is of an MNAR type, but also for a wide variety of other reasons, such as an unusual mean profile or autocorrelation structure. Earlier analyses have shown that this might indeed be the case. Likewise, it is possible that subjects, deviating from the bulk of the data because they are generated under MNAR, go undetected by this technique.

Let us now introduce the key concepts of local influence. We denote the log-likelihood function corresponding to model (7.10.34) by

$$\ell(\boldsymbol{\gamma}|\boldsymbol{\omega}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\gamma}|\omega_i),$$

in which $\ell_i(\boldsymbol{\gamma}|\omega_i)$ is the contribution of the ith individual to the log-likelihood, and where $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\psi})$ is the s-dimensional vector, grouping the parameters of the measurement model and the dropout model, not including the $N \times 1$ vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)'$ of weights defining the perturbation of the MAR model. It is assumed that $\boldsymbol{\omega}$ belongs to an open subset Ω of $\mathbb{R}^N$. For $\boldsymbol{\omega}$ equal to $\boldsymbol{\omega_0} = (0, 0, \ldots, 0)'$, $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega_0})$ is the log-likelihood function that corresponds to a MAR dropout model.

Let $\widehat{\boldsymbol{\gamma}}$ be the maximum likelihood estimator for $\boldsymbol{\gamma}$, obtained by maximizing $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega_0})$, and let $\widehat{\boldsymbol{\gamma}}_\omega$ denote the maximum likelihood estimator for $\boldsymbol{\gamma}$ under $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega})$. The local influence approach now compares $\widehat{\boldsymbol{\gamma}}_\omega$ with $\widehat{\boldsymbol{\gamma}}$. Similar estimates indicate that the parameter estimates are robust with respect to perturbations of the MAR model in the direction of non-random dropout. Strongly different estimates suggest that the estimation procedure is highly sensitive to such perturbations, which, in turn, suggests that the choice between an MAR model and a non-random dropout model highly affects the results of the analysis. Cook (1986) proposed to measure the distance between $\widehat{\boldsymbol{\gamma}}_\omega$ and $\widehat{\boldsymbol{\gamma}}$ by the so-called likelihood displacement, defined by

$$LD(\boldsymbol{\omega}) = 2[\ell(\widehat{\boldsymbol{\gamma}}|\boldsymbol{\omega_0}) - \ell(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\omega}}|\boldsymbol{\omega_0})].$$

This takes into account the variability of $\widehat{\boldsymbol{\gamma}}$. Indeed, $LD(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega_0})$ is strongly curved at $\widehat{\boldsymbol{\gamma}}$, which means that $\boldsymbol{\gamma}$ is estimated with high precision, and small otherwise. Therefore, a graph of $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ contains essential information on the influence of perturbations. It is useful to view this graph as the geometric surface formed by the values of the $N+1$ dimensional vector $\boldsymbol{\xi}(\boldsymbol{\omega}) = (\boldsymbol{\omega}', LD(\boldsymbol{\omega}))'$ as $\boldsymbol{\omega}$ varies throughout Ω.

Since this influence graph can only be depicted when $N = 2$, Cook (1986) proposed to look at local influence, i.e., at the normal curvatures $C_{\boldsymbol{h}}$ of $\boldsymbol{\xi}(\boldsymbol{\omega})$ in $\boldsymbol{\omega}_0$, in the direction of some N dimensional vector $\boldsymbol{h}$ of unit length. Let $\boldsymbol{\Delta}_i$ be the s dimensional vector defined by

$$\boldsymbol{\Delta}_i = \left.\frac{\partial^2 \ell_i(\boldsymbol{\gamma}|\omega_i)}{\partial \omega_i \partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}},\omega_i=0}$$

and define Δ as the $(s \times N)$ matrix with $\boldsymbol{\Delta}_i$ as its ith column. Further, let $\ddot{L}$ denote the $(s \times s)$ matrix of second-order derivatives of $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega}_0)$ with respect to $\boldsymbol{\gamma}$, also evaluated at $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$. Cook (1986) has then shown that $C_{\boldsymbol{h}}$ can be easily calculated by

$$C_{\boldsymbol{h}} = 2|\boldsymbol{h}'\Delta'\ddot{L}^{-1}\Delta\boldsymbol{h}|.$$

Obviously, $C_{\boldsymbol{h}}$ can be calculated for any direction $\boldsymbol{h}$. One evident choice is the vector $\boldsymbol{h_i}$ containing one in the ith position and zero elsewhere, corresponding to the perturbation of the ith weight only. This reflects the influence of allowing the ith subject to drop out non-randomly, while the others can only drop out at random. The corresponding local influence measure, denoted by C_i, then becomes $C_i = 2|\boldsymbol{\Delta}_i'\ddot{L}^{-1}\boldsymbol{\Delta}_i|$. Another important direction is the direction $\boldsymbol{h}_{\max}$ of maximal normal curvature $C_{\max}$. It shows how to perturb the MAR model to obtain the largest local changes in the likelihood displacement. It is readily seen that $C_{\max}$ is the largest eigenvalue of $-2\ \Delta'\ddot{L}^{-1}\Delta$, and that $\boldsymbol{h}_{\max}$ is the corresponding eigenvector.

EXAMPLE: Age-related macular degeneration trial

In this section, in line with Beunckens et al. (2007) and Molenberghs and Kenward (2007), the visual acuity in the ARMD trial is first analyzed using the DK model. Apart from modeling the three missing data mechanisms MCAR, MAR, and MNAR, explicitly, an ignorable analysis is also conducted. For the measurement model, again, the linear mixed model was used, assuming different intercepts and treatment effects for each of the four time points, with an unstructured covariance matrix, as in (7.5.16). In the full selection models, the dropout is modeled as in (7.10.32). Parameter estimates and corresponding standard errors of the fixed effects of the measurement model and of the dropout model parameters are given in Table 7.10.

As expected, the parameter estimates and standard errors coincide for the ignorable likelihood analysis and the selection models under MCAR and MAR, except for some negligible numerical noise.

TABLE 7.10 **The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) assuming ignorability, as well as explicitly modeling the missing data mechanism under MCAR, MAR, and MNAR assumptions, for all data.**

Effect	Parameter	Ignorable	MCAR	MAR	MNAR
		Measurement model			
Int. 4	β_{11}	54.00 (1.47)	54.00 (1.46)	54.00 (1.47)	54.00 (1.47)
Int.12	β_{21}	53.01 (1.60)	53.01 (1.59)	53.01 (1.60)	52.98 (1.60)
Int.24	β_{31}	49.20 (1.74)	49.20 (1.73)	49.19 (1.74)	49.06 (1.74)
Int.52	β_{41}	43.99 (1.79)	43.99 (1.78)	43.99 (1.79)	43.52 (1.82)
Trt. 4	β_{12}	-3.11 (2.10)	-3.11 (2.07)	-3.11 (2.09)	-3.11 (2.10)
Trt. 12	β_{22}	-4.54 (2.29)	-4.54 (2.25)	-4.54 (2.29)	-4.67 (2.29)
Trt. 24	β_{32}	-3.60 (2.49)	-3.60 (2.46)	-3.60 (2.50)	-3.80 (2.50)
Trt. 52	β_{42}	-5.18 (2.59)	-5.18 (2.57)	-5.18 (2.62)	-5.71 (2.63)
		Dropout model			
Int.	ψ_0		-2.79 (0.17)	-1.86 (0.46)	-1.81 (0.47)
Previous	ψ_1			-0.020 (0.009)	0.016 (0.022)
Current	ψ_2				-0.042 (0.023)
		-2 log-likelihood			
		6488.7	6782.7	6778.4	6775.9
		Treatment effect at 1 year (p-value)			
		0.046	0.044	0.048	0.030

Given that the main interest lies in the treatment effect at one year, the corresponding p-values are displayed in Table 7.10. In all four cases, this treatment effect is significant.

Note that for the MNAR analysis, the estimates of the ψ_1 and ψ_2 parameters are more or less of the same magnitude, but with a different sign. This is in line with the argument of Molenberghs et al. (2001), stating that the dropout often depends on the increment $y_{ij} - y_{i,j-1}$. By rewriting the fitted dropout model in terms of the increment,

$$\text{logit}\,[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)] = -1.81 - 0.026 y_{i,j-1} - 0.042(y_{ij} - y_{i,j-1}),$$

we find that the probability of dropout increases with larger negative increments; that is, those patients who showed or would have shown a greater decrease in visual acuity from the previous visit are more likely to drop out.

Turning to local influence. Figure 7.2 displays overall C_i and influences for subvectors $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\psi}$. In addition, the direction $\boldsymbol{h}_{\max}$, corresponding to maximal local influence, is given. The main emphasis should be put on the relative magnitudes. We observe that patients #10, #27, #28, #114, #139, and #154 have larger C_i values compared to other patients, which means they can be considered influential. Virtually the same picture holds for $C_i(\boldsymbol{\psi})$.

Turning attention now to the influence on the measurement model, we see that for $C_i(\boldsymbol{\beta})$, there are no strikingly high peaks, whereas $C_i(\boldsymbol{\alpha})$ reveals considerable peaks for patients #68 and #185. Note that both patients fail to have a high peak for the overall C_i, owing to the fact that the scale for $C_i(\boldsymbol{\alpha})$ is relatively small compared to the overall C_i. Nevertheless, these patients can still be considered influential. Finally, the direction of maximum curvature reveals the same six influential patients as the overall C_i.

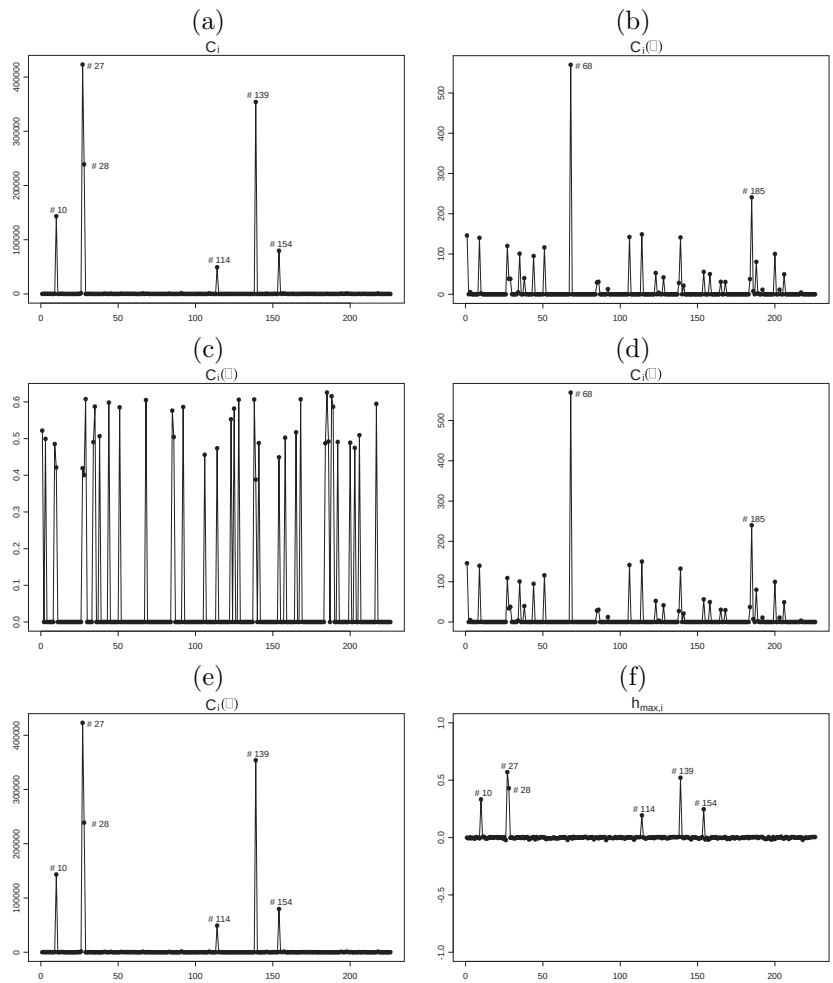


Figure 7.2 The Age-related Macular Degeneration Trial. Index plots of (a) C_i, (b) $C_i(\boldsymbol{\theta})$, (c) $C_i(\boldsymbol{\alpha})$, (d) $C_i(\boldsymbol{\beta})$, (e) $C_i(\boldsymbol{\psi})$, and (f) of the components of the direction $\boldsymbol{h}_{\max,i}$ of maximal curvature.

In Figure 7.3, the individual profiles of the influential observations are highlighted. Let us take a closer look at these cases. The six patients strongly influencing the dropout model parameters are those dropping out after the first measurement is taken at week 4. All of these patients are in the active treatment arm, except for #27. On the other hand, the two patients with strong influence on the measurement model parameters stay in the study up to week 24 and then have no observation for the last measurement occasion at 1 year. Patient #68 received the active treatment, and his/her visual acuity decreases substantially after week 4, thereafter staying more or less level. Conversely, patient #185 is enrolled in the placebo treatment arm and his/her visual acuity increases after week 4, then sloping downward a little after week 12.

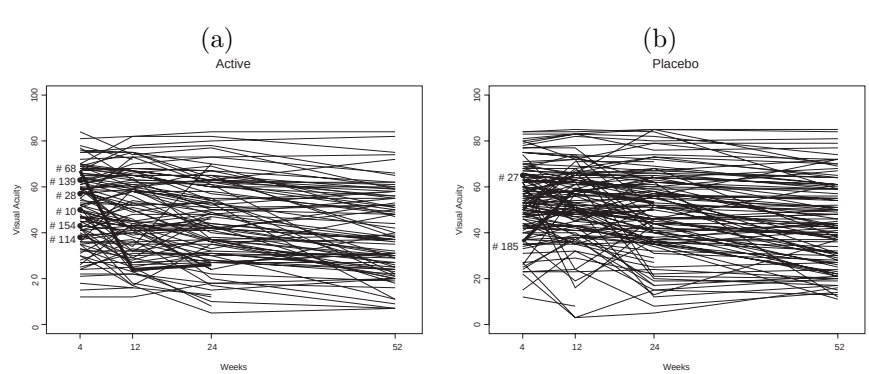


Figure 7.3 The Age-related Macular Degeneration Trial. Individual profiles for both treatment arms, with influential subjects highlighted.

It is of interest to consider an analysis without these influential observations. Therefore, we applied the selection model on three subsets of the data. The first subset obtains by removing all eight influential patients mentioned before. In the second subset of the data, patients #10, #27, #28, #114, #139, and #154 were removed, since these are overall the most influential ones. Finally, patients #68 and #185, which seemed to be influencing the measurement model the most, were removed, resulting in the third subset. Results of these analyses are shown in Tables 7.11 and 7.12. We compare the results of the MAR and MNAR analyses.

TABLE 7.11 **The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) explicitly modeling the missing data mechanism under MAR assumptions, after removing the following subsets of subjects Set 1:** $(10, 27, 28, 114, 139, 154, 68, 185)$**; Set 2:** $(10, 27, 28, 114, 139, 154)$**; and Set 3:** $(68, 185)$**.**

		Set 1	Set 2	Set 3
Effect	Parameter	MAR	MAR	MAR
		Measurement model		
Int. 4	β_{11}	54.14(1.51)	54.30(1.47)	53.84(1.48)
Int.12	β_{21}	53.09(1.64)	53.16(1.59)	52.94(1.60)
Int.24	β_{31}	49.56(1.77)	49.31(1.74)	49.44(1.73)
Int.52	β_{41}	44.40(1.82)	44.00(1.79)	44.38(1.78)
Trt. 4	β_{12}	-3.13(2.17)	-3.28(2.08)	-2.95(2.07)
Trt.12	β_{22}	-4.48(2.36)	-4.55(2.26)	-4.47(2.26)
Trt.24	β_{32}	-3.80(2.56)	-3.55(2.48)	-3.85(2.44)
Trt.52	β_{42}	-5.45(2.66)	-5.06(2.59)	-5.56(2.55)
		Dropout model		
Intercept	ψ_0	-1.90(0.47)	-1.90(0.47)	-1.85(0.46)
Previous	ψ_1	-0.019(0.010)	-0.019(0.010)	-0.020(0.009)
-2 log-likelihood		6535.3	6606.9	6706.4
Treatm. eff. at 1 year (p-value)		0.040	0.051	0.029

TABLE 7.12 **The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) explicitly modeling the missing data mechanism under MNAR assumptions, after removing the following subsets of subjects Set 1:** $(10, 27, 28, 114, 139, 154, 68, 185)$**; Set 2:** $(10, 27, 28, 114, 139, 154)$**; and Set 3:** $(68, 185)$**.**

		Set 1	Set 2	Set 3
Effect	Parameter	MNAR	MNAR	MNAR
		Measurement model		
Int. 4	β_{11}	54.15(1.49)	54.30(1.46)	53.84(1.47)
Int.12	β_{21}	53.06(1.62)	53.13(1.59)	52.91(1.59)
Int.24	β_{31}	49.46(1.75)	49.20(1.72)	49.31(1.72)
Int.52	β_{41}	43.97(1.84)	43.58(1.82)	43.90(1.82)
Trt. 4	β_{12}	-3.13(2.11)	-3.28(2.06)	-2.95(2.05)
Trt.12	β_{22}	-4.63(2.29)	-4.69(2.24)	-4.60(2.23)
Trt.24	β_{32}	-4.04(2.49)	-3.79(2.44)	-4.04(2.42)
Trt.52	β_{42}	-6.12(2.66)	-5.72(2.61)	-6.09(2.58)
		Dropout model		
Intercept	ψ_0	-1.85(0.49)	-1.85(0.49)	-1.81(0.47)
Previous	ψ_1	0.018(0.022)	0.017(0.022)	0.017(0.022)
Current	ψ_2	-0.044(0.024)	-0.043(0.024)	-0.043(0.024)
-2 log-likelihood		6532.7	6604.4	6703.8
Treatm. eff. at 1 year (p-value)		0.021	0.028	0.018

After removing the patients, who have large overall C_i and $C_i(\boldsymbol{\psi})$ values, the estimates of the dropout model parameters ψ_1 and ψ_2 are approximately the same, whereas the estimate of ψ_0 decreases from -1.86 to -1.90 under MAR, and from -1.81 to -1.85 under MNAR. The same can be seen after removing all patients. Considering the treatment effect at 1 year, its estimate under the MAR analysis increases from -5.18 to -5.06, yielding a slightly increased borderline p-value, whereas under the MNAR analysis it decreases with 0.01. Together with a decreased standard error this yields a small decrease in the p-value.

There is no impact on the likelihood ratio test for MAR against MNAR: After removing either patients #10, #27, #28, # 114, #139, and #154, or all influential patients, G^2 remains 2.5. If this likelihood ratio test would follow a standard χ^2_1-distribution, we would fail to reject the null hypothesis, which leads us to the MAR assumption. However, the test of MAR against MNAR is non-standard and the conventional chi-squared approximation cannot be used for its null distribution (Rotnitzky et al., 2000, Jansen et al., 2006).

Finally, we perform the same analyses on the third subset, with patients #68 and # 185 removed. Both for the MAR and MNAR analysis, the estimate of the treatment effect at 1 year decreases quite a lot, from -5.18 to -5.56 and from -5.71 to -6.09 respectively. Consequently, the p-value also drops down from 0.048 to 0.029 under MAR and from 0.030 to 0.018 under the MNAR analysis. The deviance for the likelihood ratio test for MAR against MNAR only changes slightly from 2.5 to 2.6.

7.11 Sensitivity Analysis Based on Multiple Imputation and Pattern-Mixture Models

In Section 7.11.1, the strategies to fit pattern-mixture models as described in Molenberghs and Kenward (2007, Ch. 17) are reviewed and applied to the ARMD data. In Section 7.11.2, sensitivity analyses methods, combining pattern-mixture models (PMM) and multiple imputation are described and their SAS implementation discussed.

7.11.1 Pattern-Mixture Strategies

PMM are inherently under-identified, because the data available for modeling within a given pattern are by definition confined to the observed components only. Verbeke and Molenberghs (2000) and Molenberghs and Kenward (2007) consider several identification strategies.

Strategy 1. Little (1993, 1994) addresses the under-identification through the use of identifying restrictions: Within a given pattern, the predictive distribtuion of the unobserved measurements, given the observed ones, is set equal to its counterpart from other patterns (e.g., the completers' pattern, termed CCMV; the neigboring pattern, termed NCMV; or a particular combination across all patterns from which the distribution is estimable, termed ACMV).

Strategy 2. As an alternative to identifying restrictions, model simplification can be undertaken to identify the parameters. The advantage is that the number of parameters decreases, which is desirable since the length of the parameter vector is a general issue with pattern-mixture models. Hogan and Laird (1997) noted that, to estimate the large number of parameters in general pattern-mixture

models, we have to make the awkward requirement that each dropout pattern occurs sufficiently often. Broadly, we distinguish between two interconnected types of simplifications.

- **Strategy 2a.** Trends can be restricted to functional forms supported by the information available within a pattern. For example, a linear or quadratic time trend is easily extrapolated beyond the last obtained measurement. We merely need to provide an ad hoc solution for the first or the first few patterns. To fit such models, a conventional model building exercise is conducted within each of the patterns separately.
- **Strategy 2b.** Alternatively, we can choose to let the model parameters vary across patterns in a controlled parametric way. Thus, rather than estimating a separate time trend within each pattern, we might for example assume that the time evolution within a pattern is unstructured, but parallel across patterns. This can be done by treating pattern as a covariate. The available data can be used to assess whether such simplifications are supported over the time ranges for which information is collected.

While the second strategy is computationally the simpler, there is a price to pay. Such simplified models, qualified as "assumption rich" by Sheiner, Beal, and Dunne (1997), are also making untestable assumptions, exactly as in the selection model case. Using the fitted profiles to predict their evolution beyond the time of dropout is nothing but extrapolation. It is possible only by making the models sufficiently simple. It is, for example, not possible to assume an unstructured time trend in incomplete patterns and then still extrapolate in an unambiguous fashion. In contrast, assuming a linear time trend allows estimation in all patterns containing at least two measurements. However, it is less obvious what the precise nature of the dropout mechanism is. Kenward, Molenberghs, and Thijs (2003) examined what restrictions need to be imposed, in the context of longitudinal data with dropout, to ensure that the dropout probability does not depend on future measurements, given past and current values. Strategy 2 is not compliant with this requirement, but the same is true for CCMV and NCMV.

A final observation, applying to both strategies, is that pattern-mixture models do not always automatically provide estimates and standard errors of marginal quantities of interest, such as overall treatment effect or overall time trend. Hogan and Laird (1997) provided a way to derive selection model quantities from the pattern-mixture model. This is a first instance in the PMM context where multiple imputation comes in handy. Several authors have followed this idea to formally compare the conclusions from a selection model with the selection model parameters derived from a pattern-mixture model (Verbeke, Lesaffre, and Spiessens, 1998; Michiels, Molenberghs, and Lipsitz, 1999).

To better see how this method works, we briefly sketch the sequence of steps to be followed.

1. Fit a model to the pattern t-specific identifiable densities: $f_t(y_1, \ldots, y_t)$. This results in a parameter estimate, $\widehat{\boldsymbol{\gamma}}_t$.
2. Select an identification method of choice.
3. Using this identification method, determine the conditional distributions of the unobserved outcomes, given the observed ones:
$$f_t(y_{t+1}, \ldots, y_T | y_1, \ldots, y_t). \tag{7.11.35}$$
4. Using standard MI methodology, draw multiple imputations for the unobserved components, given the observed outcomes and the correct pattern-specific density (7.11.35).

5. Analyze the multiply-imputed sets of data using the method of choice. This can be another pattern-mixture model, but also a selection model or any other desired model.
6. Inferences can be conducted using the standard combination rules.

EXAMPLE: Age-related macular degeneration trial

We now consider the use of pattern-mixture models for these data. Here, we will apply the first strategy making, use of CCMV, NCMV, and ACMV identifying restrictions.

The results for the three types of restrictions are shown in Table 7.13. After applying each one of the three restrictions, the same selection model as before is fitted. It can be seen from the estimates and associated standard errors that there is little difference in conclusions between the strategies.

TABLE 7.13 The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) and p-values resulting from the pattern-mixture model using identifying restrictions ACMV, CCMV, and NCMV.

Effect	Parameter	ACMV	CCMV	NCMV
		Parameter estimate (standard error)		
Intercept 4	β_{11}	54.00(1.47)	54.00(1.47)	54.00(1.47)
Intercept 12	β_{21}	52.87(1.68)	52.92(1.61)	52.86(1.63)
Intercept 24	β_{31}	48.65(2.00)	49.16(1.87)	48.77(1.78)
Intercept 52	β_{41}	44.19(2.14)	44.69(2.54)	44.00(1.80)
Treatment 4	β_{12}	-3.11(2.10)	-3.11(2.10)	-3.11(2.10)
Treatment 12	β_{22}	-4.18(2.48)	-4.07(2.30)	-4.40(2.42)
Treatment 24	β_{32}	-4.36(3.83)	-5.14(3.61)	-4.19(2.62)
Treatment 52	β_{42}	-5.04(3.86)	-2.33(4.93)	-4.89(2.70)
		p-values		
Intercept 4	β_{11}	---	---	---
Intercept 12	β_{21}	$< .0001$	$< .0001$	$< .0001$
Intercept 24	β_{31}	$< .0001$	$< .0001$	$< .0001$
Intercept 52	β_{41}	$< .0001$	$< .0001$	$< .0001$
Treatment 4	β_{12}	---	---	---
Treatment 12	β_{22}	0.092	0.077	0.069
Treatment 24	β_{32}	0.271	0.173	0.110
Treatment 52	β_{42}	0.211	0.647	0.071

In the pattern-mixture approach, we use information from different patterns to multiply impute new values whenever the observations are missing. Borrowing information from more distant patterns, such as the complete cases, can introduce extra variability, depending on the nature of the conditional distributions sampled from. It is not unexpected, therefore, for the variability to be smallest when applying NCMV, as seen in the standard errors.

It can be seen from these analyses that the treatment effect at week 52 is not statistically significant, in contrast to the conclusions based on Tables 7.10 and 7.11. The p-value is closest to significance with NCMV restrictions.

The fact that no significant treatment effect is found here, suggests caution concerning the conclusions obtained under the selection model formulation. This implies that a significant treatment effect is conditional upon the MAR assumption

holding. We would feel more comfortable about a significant treatment effect if it were holding across MAR and a number of MNAR scenarios. Thus, at best, it is fair to say that there is a weak evidence only for a treatment effect.

7.11.2 Pattern-Mixture Based Sensitivity Analysis

The PMM framework is very rich, and sensitivity analyses can be conducted from various perspectives. For example, we can use the PMM framework to examine the impact of certain departures from MAR (e.g., when patients would evolve differently after dropout, while still on the same treatment). A different perspective is the examination of potential outcomes that patients would have had, had they switched to alternative treatment strategies after dropping out (Little and Kang, 2015).

From a technical point of view, the algorithm described on page 372 is very general and allows for a variety of sensitivity analysis routes.

First, by simply varying the identifying restrictions (e.g., by juxtaposing ACMV, CCMV, and NCMV; but there are several others) a sensitivity analysis results. Note that ACMV corresponds to MAR in the PMM framework (Molenberghs et al., 1998); the other two can then be seen as deviations from it.

Second, it is of course possible to identify conditional densities of the form (7.11.35) in other ways than through setting them equal to other data-identified densities, or in ways that deviate from them in a controlled way. This is the route taken by Carpenter, Roger, and Kenward (2009), and discussed in detail in Carpenter and Kenward (2013, Ch. 10).

An advantage of using MI is that imputations can be generated in a PMM framework, with analysis conducted in the same or a different framework. For example, the models reported in Table 7.13 are of a selection model type, but imputations were obviously of a PMM signature.

Possible strategies for generating imputations, partially in line with Carpenter, Roger, and Kenward (2009) are as follows:

- Jump to reference. For example, patients receiving active treatment might be made to "jump" to the control group after dropout.
- After dropout in a given pattern (and perhaps in a given treatment group), subjects might be made to shift with a certain amount, relative to the MAR-based prediction. This amount in itself can be varied, from 0 (typically corresponding to MAR), to a prespecified maximal amount.
- Likewise, they might be made to change slope with a certain amount.

Carpenter and Kenward (2009) describe such strategies in the following generic terms:

1. *Separately for each treatment arm, take all patients' pre-deviation data and---assuming MAR---fit a multivariate normal distribution with unstructured mean (i.e., a separate mean for each of the* $1+p$ *baseline plus post-randomization observation times) and unstructured variance-covariance matrix (i.e., a* $(1+p)\times(1+p)$ *covariance matrix), (...).*
2. *Separately for each treatment arm, draw a mean vector and variance-covariance matrix from the posterior distribution.*
3. *For each patient who deviates before the end of the study, use the draws from step 2 to build the joint distribution of their pre- and post-deviation outcome data. Suggested options for constructing this are given below.*

4. *For each patient who deviates before the end, use their joint distribution in step 3 to construct their conditional distribution of post-deviation given pre-deviation outcome data. Sample their post-deviation data from this conditional distribution, to create a "completed" data set.*
5. *Repeated steps 2--4 M times, resulting in M imputed data sets.*
6. *Fit the substantive model to each imputed data set, and combine the resulting parameter estimates and standard errors using Rubin's rules for final inference.*

For precision estimation, we might also revert to resampling methods, as proposed by Lu (2014).

A special place is reserved for a so-called *tipping point analysis*. This can be undertaken whenever a continuous deviation from MAR is possible. For example, when in a given pattern for a given treatment group, subjects are systematically shifted by a certain amount, this amount can be changed continuously (or in small increments) until the point where significance of a key hypothesis test changes. If this point is unrealistically far away, then confidence in the primary analysis increases. Of course, there are a multitude of ways in which a given primary analysis can be subjected to a tipping point analysis. Exactly how it is conducted will likely depend on substantive considerations.

EXAMPLE: Age-related macular degeneration trial

To illustrate sensitivity analysis by way of MNAR adjustments in the multiple imputation process, we apply a shift to missing values in the treated arm, with magnitudes of 0, 10, 15, and 20, at 4, 12, 24, and 52 weeks, respectively. The results of both GEE and GLMM, without and with this adjustment, are shown in Table 7.14. We expect to see the same estimates for the standard (MAR) analyses

TABLE 7.14 The Age-related Macular Degeneration Trial. Parameter estimates (standard errors) for GEE and GLMM, comparing MAR versions with MNAR analyses based on shifts, identification using the placebo group only, and NCMV.

Effect	Par.	MAR	shift	placebo	NCMV
		Generalized estimating equations			
Int.4	β_{11}	-0.82(0.20)	-0.73(0.20)	-0.81(0.20)	-0.83(0.21)
Int.12	β_{21}	-0.97(0.22)	-0.71(0.19)	-0.98(0.22)	-1.06(0.21)
Int.24	β_{31}	-1.07(0.23)	-0.56(0.19)	-1.05(0.22)	-1.00(0.22)
Int.52	β_{41}	-1.66(0.27)	-0.82(0.20)	-1.58(0.29)	-1.59(0.27)
Trt.4	β_{12}	0.17(0.29)	0.07(0.28)	0.17(0.28)	0.17(0.29)
Trt.12	β_{22}	0.56(0.29)	0.29(0.27)	0.56(0.29)	0.67(0.28)
Trt.24	β_{32}	0.41(0.30)	-0.10(0.27)	0.39(0.29)	0.34(0.29)
Trt.52	β_{42}	0.41(0.35)	-0.43(0.30)	0.32(0.35)	0.32(0.35)
		Generalized linear mixed models			
Int.4	β_{11}	-1.46(0.36)	-1.32(0.36)	-1.39(0.35)	-1.42(0.35)
Int.12	β_{21}	-1.75(0.38)	-1.27(0.35)	-1.67(0.37)	-1.80(0.36)
Int.24	β_{31}	-1.83(0.38)	-1.01(0.34)	-1.78(0.38)	-1.70(0.38)
Int.52	β_{41}	-2.71(0.45)	-1.47(0.36)	-2.62(0.46)	-2.64(0.44)
Trt.4	β_{12}	0.32(0.48)	0.12(0.50)	0.25(0.48)	0.24(0.48)
Trt.12	β_{22}	0.99(0.49)	0.50(0.48)	0.91(0.48)	1.09(0.47)
Trt.24	β_{32}	0.67(0.51)	-0.19(0.48)	0.62(0.48)	0.53(0.49)
Trt.52	β_{42}	0.53(0.57)	-0.74(0.51)	0.45(0.56)	0.45(0.55)
R.I. s.d.	τ	2.21(0.26)	2.28(0.25)	2.17(0.25)	2.16(0.24)
R.I. var.	τ^2	4.90(1.14)	5.21(1.15)	4.72(1.09)	4.66(1.05)

as obtained in Table 7.9. However, there are slight differences because the results in Table 7.9 were based on the MCMC imputation method, whereas here FCS was used. The reason is that the MNAR statement requires either MONOTONE or FCS. Given that the data are slightly non-monotone, FCS is the obvious choice.

Turning to the difference between the MAR analyses and those after applying a shift, we observe reasonably large changes, including a sign change for the treatment effects at 24 and 52 weeks. Under MAR, there was only one marginally significant treatment effect (at 12 weeks), even though it is in favor of placebo. After applying the shift, nothing is nearly significant. It is noteworthy that a negative sign points to an effect in favor of the active treatment. This is not surprising, because by applying the shift, we progressively make the treatment more beneficial.

Two further MNAR-based analyses are conducted, both reported in Table 7.14. In the first case, imputation takes place based on the placebo group only. In other words, after dropout, the conditional distribution of the missing measurements given the observed ones is based on only the control arm. The final analysis, NCMV is applied, meaning that the distribution of a missing measurement given its predecessors is based on the adjacent pattern that contains all of these measurements. This imputation is done for each of the two treatment groups separately.

Unlike with the shift analysis, the SAS implementation for the latter two analyses requires data to be monotonically missing. Therefore, two multiple imputation calls are made. In the first one, the eight subjects with a non-monotone pattern are monotonized, under the assumption of MAR. Ten imputations are generated. These data sets are then used as input for one further MNAR imputation. The result is, evidently, ten fully imputed data sets. Details about the SAS code are presented in Section 7.11.3.

7.11.3 SAS and Sensitivity Analysis

The key statement to conduct sensitivity analyses of the type reported above is the MNAR statement. It is important to note that the statement requires either MONOTONE or FCS as imputation strategies. Hence, MCMC is not compatible with this tool.

There are two main strategies to apply the MNAR statement. The first one is that of adjustment, using the `adjust` option. It specifies a subset of the variables present in the VAR statement to which a certain adjustment should be applied. It is also possible to specify a subset of the observations to which the adjustment needs to be applied. For example, we can apply an adjustment at certain measurement occasions, for one of several treatment arms. An example is given in the following program (Program 7.29), which is needed to generate the imputations that lead to the results reported in Table 7.14.

PROGRAM 7.29 Sensitivity analysis using PROC MI, shift adjustment

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13as1
        nimpute=10 round=0.1;
   title 'Shift multiple imputation';
   class treat;
   var lesion diff4 diff12 diff24 diff52;
   fcs reg;
   mnar adjust (diff12 / shift=10 adjustobs=(treat='2'));
   mnar adjust (diff24 / shift=15 adjustobs=(treat='2'));
   mnar adjust (diff52 / shift=20 adjustobs=(treat='2'));
   by treat;
run;
```

Note that Program 7.29 replaced Program 7.18. The programs for the analysis and inference steps remain exactly the same. The `shift=` adjustment is but one of

several options. For example, rather than an additive shift, a multiplicative scale adjustment can be made using the `scale=` option.

The other main option is `model`. It can be used to specify for which variables what subgroup of the observations is to be used. Subgroups can be defined in a predefined way, using NCMV or CCMV. Alternatively, subgroups can be defined by way of levels of certain variables. An example of the latter is Program 7.30.

PROGRAM 7.30 Sensitivity analysis using PROC MI, subgroup adjustment

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13as2 nimpute=10;
title 'Model multiple imputation';
class treat;
var lesion diff4 diff12 diff24 diff52;
fcs reg;
mnar model (diff4  / modelobs= (treat='1'));
mnar model (diff12 / modelobs= (treat='1'));
mnar model (diff24 / modelobs= (treat='1'));
mnar model (diff52 / modelobs= (treat='1'));
run;
```

The method is particularly useful, and popular, when this group is defined as a control treatment group. Such a control-based imputation method is known as *copy reference*. The website `www.missingdata.org.uk` contains a suite of SAS macros, for various control-based imputation strategies, written by James Roger.

Note that NCMV and CCMV is available only for monotone data, whereas the third option is available with FCS as well.

Should we want to apply NCMV or CCMV, then we can first monotonize the data using standard imputation, and then apply the desired identifying restrictions, as in the following program:

PROGRAM 7.31 Sensitivity analysis using PROC MI, NCMV

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13as3 nimpute=10;
title 'Montone imputation';
var lesion diff4 diff12 diff24 diff52;
mcmc impute=monotone;
by treat;
run;

proc mi data=m.armd13as3 seed=486048 simple out=m.armd13as4 nimpute=1;
title 'Model multiple imputation';
var lesion diff4 diff12 diff24 diff52;
monotone reg;
mnar model (diff4 diff12 diff24 diff52 / modelobs=ncmv);
by treat;
run;
```

Output

```
Observations Used for Imputation Models Under MNAR Assumption

Imputed
Variable    Observations

diff4       Nonmissing lesion, diff4; Missing diff12, ..., diff52
diff12      Nonmissing lesion, ..., diff12; Missing diff24, diff52
diff24      Nonmissing lesion, ..., diff24; Missing diff52
diff52      Complete Cases
```

In the first MI call, 10 imputations are generated. The output data set of this call is used as input for the next one, where a single imputation is created. Evidently, this effectively creates $10 \times 1 = 10$ imputations. As a general rule, when multiple imputations are generated in a sequential fashion, the required number of imputations M should be generated the first time; in every subsequent call, there should then be a single imputation.

The patterns used in the imputation process (second call) are part of the printout:

7.12 Concluding Remarks

We have shown that analyzing incomplete (longitudinal) data, both of a Gaussian as well as of a non-Gaussian nature, can easily be done under the relatively relaxed assumption of missingness at random (MAR), using standard statistical software tools. Likelihood-based methods include the linear mixed model (e.g., implemented in the SAS procedure MIXED) and generalized linear mixed models (e.g., implemented in the SAS procedures GLIMMIX and NLMIXED). This is termed direct likelihood or ignorable likelihood. Under the same assumptions, ignorable Bayesian analyses can be conducted (e.g., using the SAS procedure MCMC).

In addition, weighted generalized estimating equations can be used under MAR. Its implementation is straightforward thanks to facilities of the SAS procedure GEE.

Finally, a versatile approach, valid under MAR, is to handle incompleteness by way of multiple imputation, after which standard, complete-data analysis methods can be used. SAS offers procedures MI and MIANALYZE to this effect.

All of this implies that traditionally popular but far more restricted modes of analysis, including complete case (CC) analysis, last observation carried forward (LOCF), or other simple imputation methods, ought to be abandoned, given the highly restrictive assumptions on which they are based.

Of course, general missingness not at random can never be entirely excluded, and we should therefore ideally supplement an MAR-based analysis with a suitable chosen set of sensitivity analyses. This area is still in full development, but thanks to the MNAR statement in PROC MI, an array of sensitivity analysis tools are now also provided within the context of standard SAS procedures.

7.13 References

Aerts, M., Geys, H., Molenberghs, G., and Ryan, L.M. (2002). Topics in Modelling of Clustered Data. London: Chapman and Hall.

Afifi, A. and Elashoff, R. (1966). Missing observations in multivariate statistics I: Review of the literature. *Journal of the American Statistical Association*, **61**, 595--604.

Baker, S.G., Rosenberger, W.F., and DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine*, **11**, 643--657.

Beckman, R.J., Nachtsheim, C.J., and Cook, R.D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413--426.

Beunckens, C., Molenberghs, G., Thijs, H., and Verbeke, G. (2007). Incomplete hierarchical data. *Statistical Methods in Medical Research*, **16**, 1--36.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9--25.

Buyse, M. and Molenberghs, G. (1998). The validation of surrogate endpoints in randomized experiments. *Biometrics*, **54**, 1014--1029.

Carey, V.C., Zeger, S.L., and Diggle, P.J. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, **80**, 517--526.

Carpenter, J.R. and Kenward, M.G. (2013). *Multiple Imputation and Its Applications.* Chichester: John Wiley & Sons.

Carpenter, J.R., Kenward, M.G., and Vansteelandt, S. (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A*, **169**, 571--584.

Carpenter, J.R., Roger, J.H., and Kenward, M.G. (2013). Analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics*, **23**, 1352--1371.

Cook, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, **48**, 133--169.

Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* London: Chapman & Hall.

DeGruttola, V. and Tu, X.M. (1994). Modelling progression of CD4 lymphocyte count and its relationship to survival time. *Biometrics*, **50**, 1003--1014.

Dempster, A.P., Laird, N.M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1--38.

Dempster, A.P. and Rubin, D.B. (1983). Overview. *Incomplete Data in Sample Surveys, Vol. II: Theory and Annotated Bibliography*, W.G. Madow, I. Olkin, and D.B. Rubin (Eds.). New York: Academic Press, pp. 3--10.

Diggle, P.J. and Kenward, M.G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49--93.

Fitzmaurice, G.M., Davidian, M., Verbeke, G., and Molenberghs, G. (2009) *Advances in Longitudinal Data Analysis.* London: CRC/Chapman Hall.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis*, Texts in Statistical Science. London: Chapman & Hall.

Glynn, R.J., Laird, N.M., and Rubin, D.B. (1986). Selection modelling versus mixture modelling with non-ignorable nonresponse. In: *Drawing Inferences from Self Selected Samples*, H. Wainer (Ed.). New York: Springer-Verlag, pp. 115--142.

Hartley, H.O. and Hocking, R. (1971). The analysis of incomplete data. *Biometrics*, **27**, 7783--808.

Heitjan, F. and Little, R.J.A. (1991). Multiple imputation for the fatal accident reporting system. *Applied Statistics*, **40**, 13-29.

Hogan, J.W. and Laird, N.M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, **16**, 239--258.

Jansen, I., Molenberghs, G., Aerts, M., Thijs, H., and Van Steen, K. (2003). A local influence approach applied to binary data from a psychiatric study. *Biometrics*, **59**, 409--418.

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G., and Kenward, M.G. (2006). The nature of sensitivity in missing not at random models. *Computational Statistics and Data Analysis*, **50**, 830--858.

Kenward, M.G. and Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, **12**, 236--247.

Kenward, M.G., Molenberghs, G., and Goetghebeur, E. (2001). Sensitivity analysis for incomplete categorical data. *Statistical Modelling*, **1**, 31--48.

Kenward, M.G., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika*, **90**, 53--71.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570--582.

Li, K.H., Raghunathan, T.E., and Rubin, D.B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distributions. *Journal of the American Statistical Association*, **86**, 1065--1073.

Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13--22.

Liang, K.--Y., Zeger, S.L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B*, **54**, 3--40.

Lipsitz, S.R., Laird, N.M., and Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, **78**, 153--160.

Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, **88**, 125--134.

Little, R.J.A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471--483.

Little, R.J.A., D'Agostino, R., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., Neaton, J.D., Rotnitzky, A., Scharfstein, D., Shih, W., Siegel, J.P., and Stern, H. National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials.* Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.

Little, R.J.A. and Kang, S. (2015). Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials. *Statistics in Medicine*, **34**, 2381--2390.

Little, R.J.A. and Rubin, D.B. (2014). *Statistical Analysis with Missing Data* (3rd ed.). New York: John Wiley & Sons. [The first edition appeared in 1987; the second edition in 2002.]

Lu, K. (2014). An analytic method for the placebo-based pattern-mixture model. *Statistics in Medicine*, **33**, 1134--1145.

Mallinckrodt, C.H. (2013). *Preventing and Treating Missing Data in Longitudinal Clinical Trials: A Practical Guide.* New York: Cambridge University Press.

Mallinckrodt, C.H., Clark, W.S., and Stacy R.D. (2001a). Type I error rates from mixed-effects model repeated measures versus fixed effects analysis of variance with missing values imputed via last observation carried forward. *Drug Information Journal*, **35**, 4, 1215--1225.

Mallinckrodt, C.H., Clark, W.S., and Stacy R.D. (2001b). Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceutical Statistics*, **11**, (1 & 2), 9--21.

Mallinckodt, C.H. and Lipkovich, I. (2016). *Analyzing Longitudinal Clinical Trial Data. A Practical Guide.* Boca Raton: Chapman & Hall/CRC.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* London: Chapman & Hall.

Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T., and Thijs, H. (2002). Selection models and pattern-mixture models to analyze longitudinal quality of life data subject to dropout. *Statistics in Medicine*, **21**, 1023--1041.

Michiels, B., Molenberghs, G., and Lipsitz, S.R. (1999). Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics*, **55**, 978--983.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M.G. (2008). Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, **70**, 371--388.

Molenberghs, G., Fitzmaurice, G., Kenward, M.G., Verbeke, G., and Tsiatis, A.A. (2015). *Handbook of Missing Data*. Boca Raton: Chapman & Hall/CRC.

Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies.* Chichester: John Wiley & Sons.

Molenberghs, G., Kenward, M.G., and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom dropout. *Biometrika* **84**, 33--44.

Molenberghs, G., Kenward, M.G., and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **50**, 15--29.

Molenberghs, G., Kenward, M.G., Verbeke, G., and Teshome Ayele, B. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica*, **21**, 187--206.

Molenberghs, G., Michiels, B., Kenward, M.G., and Diggle, P.J. (1998). Missing data mechanisms and pattern-mixture models. *Statistica Neerlandica*, **52**, 153--161.

Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M.G., Mallinckrodt, C., and Carroll, R.J. (2003). Analyzing incomplete longitudinal clinical trial data. *Submitted for publication.*

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* New York: Springer.

Molenberghs, G., Verbeke, G., Thijs, H., Lesaffre, E., and Kenward, M.G. (2001). Mastitis in dairy cattle: influence analysis to assess sensitivity of the dropout process. *Computational Statistics and Data Analysis*, **37**, 93--113.

Murray G.D. and Findlay J.G. (1988). Correcting for the bias caused by drop-outs in hypertension trials. *Statististics in Medicine*, **7**, 941--946.

O'Kelly, M. and Ratitch, B. (2014). *Clinical Trials with Missing Data: A Guide for Practitioners.* New York: John Wiley & Sons.

Pharmacological Therapy for Macular Degeneration Study Group (1997). Interferon α-IIA is ineffective for patients with choroidal neovascularization secondary to age-related macular degeneration. Results of a prospective randomized placebo-controlled clinical trial. *Archives of Ophthalmology*, **115**, 865--872.

Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, **4**, 12--35.

Pinheiro, J. and Bates, D. M. (2000). *Mixed-effects Models in S and S-plus*. New York: Springer-Verlag.

Prentice, R.L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033-1048.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C: the art of scientific computing,* Chapter 10. Cambridge University Press, New York, second edition.

Robins, J.M., Rotnitzky, A., and Scharfstein, D.O. (1998). Semiparametric regression for repeated outcomes with non-ignorable non-response. *Journal of the American Statistical Association*, **93**, 1321--1339.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**, 106--121.

Rosenbaum, P.R. and Rubin, D.B. (1983). The central role of the propensity score method in observational studies for causal effects. *Biometrika*, **70**, 41--55.

Rotnitzky, A., Cox, D.R., Bottai, M., and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernouilli*, **6**, 243--284.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, **63**, 581--592.

Rubin, D.B. (1978). Multiple imputations in sample surveys -- a phenomenological Bayesian approach to nonresponse. In: *Imputation and Editing of Faulty or Missing Survey Data*. Washington, DC: U.S. Department of Commerce, pp. 1--23.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Rubin, D.B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, **81**, 366--374.

Rubin, D.B., Stern H.S., and Vehovar V. (1995). Handling "don't know" survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association*, **90**, 822--828.

Schafer J.L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Scharfstein, D.O., Rotnitzky, A., and Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*, **94**, 1096--1120.

Sheiner, L.B., Beal, S.L., and Dunne, A. (1997). Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. *Journal of the American Statistical Association*, **92**, 1235--1244.

Skrondal, A. and Rabe-Hesketh, S. (2004). Generalized latent variable modeling: Multilevel, longitudinal and structural equation models. London: Chapman & Hall.

Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528--550.

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics*, **3**, 245--265.

Thijs, H., Molenberghs, G., and Verbeke, G. (2000). The milk protein trial: influence analysis of the dropout process. *Biometrical Journal*, **42**, 617--646.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82--86.

Tuerlinckx, F., Rijmen, F., Molenberghs, G., Verbeke, G., Briggs, D., Van den Noortgate, W., Meulders, M., and De Boeck, P. (2004). Estimation and software. In P. De Boeck & M. Wilson (Eds.), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach* (pp. 343--373). New York: Springer.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, **16**, 219--242.

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

van Buuren, S., Boshuizen, H.C., and Knook, D.L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, **18**, 681--694.

Van Steen, K., Molenberghs, G., Verbeke, G., and Thijs, H. (2001). A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling: An International Journal* **1**, 125--142.

Verbeke, G., Lesaffre, E., and Brant L.J. (1998). The detection of residual serial correlation in linear mixed models. *Statistics in Medicine*, **17**, 1391--1402.

Verbeke, G., Lesaffre, E., and Spiessens, B. (2001). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, **35**, 419--434.

Verbeke, G. and Molenberghs, G. (1997). *Linear Mixed Models in Practice: A SAS-Oriented Approach*. Lecture Notes in Statistics 126. New York: Springer-Verlag.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M.G. (2001). Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics*, **57**, 7--14.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439--447.

White, I.R., Daniel, R., and Royston, P. (2010). Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Computational Statistics and Data Analysis*, **54**, 2267--2275.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, **48**, 233--243.

Wu, M.C. and Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, 175--188.

About This Book

What Does This Book Cover?

The main goal of this book is to introduce popular statistical methods used in clinical trials and to discuss their implementation using SAS software. To help bridge the gap between modern statistical methodology and clinical trial applications, the book includes numerous case studies based on real trials at all stages of drug development. It also provides a detailed discussion of practical considerations and relevant regulatory issues as well as advice from clinical trial experts.

The book focuses on fundamental problems arising in the context of clinical trials such as the analysis of common types of clinical endpoints and statistical approaches most commonly used in early- and late-stage clinical trials. The book provides detailed coverage of approaches utilized in Phase I/Phase II trials, e.g., dose-escalation and dose-finding methods. Important trial designs and analysis strategies employed in Phase II/Phase III include multiplicity adjustment methods, data monitoring methods and methods for handling incomplete data.

Is This Book for You?

Although the book was written primarily for biostatisticians, the book includes high-level introductory material that will be useful for a broad group of pre-clinical and clinical trial researchers, e.g., drug discovery scientists, medical scientists and regulatory scientists working in the pharmaceutical and biotechnology industries.

What Are the Prerequisites for This Book?

General experience with clinical trials and drug development, as well as experience with SAS/STAT procedures, will be desirable.

What's New in This Edition?

The second edition of this book has been thoroughly revised based on the feedback provided by numerous readers of the first edition. The topics covered in the book have been grouped into three parts. The first part provides detailed coverage of general statistical methods used across the three stages of drug development. The second and third parts focus on the topics specific to early-phase and late-phase clinical trials, respectively.

The chapters from the first edition have been expanded to cover new approaches to addressing the statistical problems introduced in the original book. Numerous revisions have been made to improve the explanations of key concepts, add more examples and case studies. A detailed discussion of new features of SAS procedures has been provided and, in some cases, new procedures are introduced that were not available when the first edition was released.

What Should You Know about the Examples?

The individual chapters within this book include tutorial material along with multiple examples to help the reader gain hands-on experience with SAS/STAT procedures used in the analysis of clinical trials.

Software Used to Develop the Book's Content

The statistical methods introduced in this book are illustrated using numerous SAS/STAT procedures, including PROC GLM, PROC FREQ, PROC LOGISTIC, PROC GENMOD, PROC LIFETEST and PROC PHREG (used in the analysis of different types of clinical endpoints), PROC MIXED, PROC NLMIXED and PROC GENMOD (used in dose-finding trials), PROC MULTTEST (used in clinical trials with multiple objectives), PROC SEQDESIGN and PROC SEQTEST (used in group-sequential trials), PROC MIXED, PROC GLIMMIX, PROC GEE, PROC MI and PROC MIANALYZE (used in clinical trials with missing data). These procedures are complemented by multiple SAS macros written by the chapter authors to support advanced statistical methods.

Example Code and Data

You can access the example code, SAS macros and data sets used in this book by linking to its author page at http://support.sas.com/publishing/authors/dmitrienko.html.

SAS University Edition

This book is compatible with SAS University Edition. If you are using SAS University Edition, then begin here: https://support.sas.com/ue-data.

Output and Graphics

The second edition takes full advantage of new graphics procedures and features of SAS software, including PROC SGPLOT, PROC SGPANEL and ODS graphics options.

We Want to Hear from You

SAS Press books are written *by* SAS Users *for* SAS Users. We welcome your participation in their development and your feedback on SAS Press books that you are using. Please visit sas.com/books to do the following:

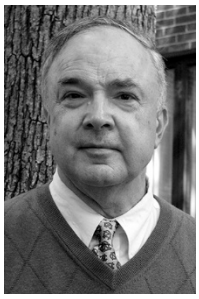
- Sign up to review a book
- Recommend a topic
- Request authoring information
- Provide feedback on a book

Do you have questions about a SAS Press book that you are reading? Contact the author through saspress@sas.com or https://support.sas.com/author_feedback.

SAS has many resources to help you find answers and expand your knowledge. If you need additional help, see our list of resources: sas.com/books.

About These Authors

Alex Dmitrienko, PhD, is Founder and President of Mediana Inc. He is actively involved in biostatistical research with an emphasis on multiplicity issues in clinical trials, subgroup analysis, innovative trial designs, and clinical trial optimization. Dr. Dmitrienko coauthored the first edition of *Analysis of Clinical Trials Using SAS®: A Practical Guide*, and he coedited *Pharmaceutical Statistics Using SAS®: A Practical Guide*.



Gary G. Koch, PhD, is Professor of Biostatistics and Director of the Biometrics Consulting Laboratory at the University of North Carolina at Chapel Hill. He has been active in the field of categorical data analysis for fifty years. Professor Koch teaches classes and seminars in categorical data analysis, consults in areas of statistical practice, conducts research, and trains many biostatistics students. He is coauthor of *Categorical Data Analysis Using SAS®, Third Edition*.

Learn more about these authors by visiting their author pages, where you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more:
http://support.sas.com/dmitrienko
http://support.sas.com/koch

Ready to take your SAS® and JMP® skills up a notch?

Be among the first to know about new books, special events, and exclusive discounts.
support.sas.com/newbooks

Share your expertise. Write a book with SAS.
support.sas.com/publish

sas.com/books
for additional books and resources.

sas
THE POWER TO KNOW.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies. © 2017 SAS Institute Inc. All rights reserved. M1588358 US.0217