Limburgs Universitair Centrum

Faculteit Wetenschappen

Center for Statistics

Flexible Model Strategies and Sensitivity Analysis Tools for Non-Monotone Incomplete Categorical Data

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen, richting Wiskunde
aan het Limburgs Universitair Centrum
te verdedigen door

Ivy Jansen

Promotors:
Prof. Dr. G. Molenberghs
Prof. Dr. M. Aerts

April 15, 2005

Contents

List of Tables

List of Figures

1

Introduction

1.1 The History of Missing Data

In applied sciences, data are more and more measured repeatedly over time, resulting in so-called longitudinal data. However, they typically suffer from incompleteness. Since incompleteness usually occurs for reasons outside of the control of the investigators and may be related to the outcome measurement of interest, it is generally necessary to address the process governing incompleteness. Only in special but important cases it is possible to ignore the missingness process. Over the last century, the focus, when formulating answers to the analysis of such incomplete data, has shifted. Indeed, early work on missing values was largely concerned with overcoming the lack of balance or deviations from the intended study design (Afifi and Elashoff, 1966; Hartley and Hocking, 1971). Later, general algorithms such as expectation-maximization (EM) (Dempster, Laird and Rubin, 1977), and data imputation and augmentation procedures (Rubin, 1987), combined with powerful computing resources have largely provided a solution to this aspect of the problem. During the last decade, a multitude of advanced models, allowing for potentially complicated ways in which missingness is influenced by observed and unobserved measurements, have been formulated.

In the meantime, practice has put a strong emphasis on methods such as complete case analysis (CC, Little and Rubin, 1987), restricting the analysis to subjects with all responses obtained, because it is simple to perform. In many cases, however,

researchers do not realize that there is a strong danger for bias. In addition, since data on incomplete records are deleted, the statistical efficiency is reduced, leading to larger standard errors. These and other dangers of a CC analysis have been reported in a number of statistical application areas (Little, 1992, 1995; Molenberghs *et al.*, 2004). Alternatively, filling-in of incomplete data with techniques such as last observation carried forward (LOCF), has been popular for a long time. Also here, there is no need for a full longitudinal model (e.g., when the scientific question is in terms of the last planned measurement occasion only). These so-called imputation methods are also described and discussed by Little and Rubin (1987) and carry even more problems than do CC methods. Especially when data are measured repeatedly, not only means and differences between groups, but also within-unit correlation and other modeling aspects tend to be distorted in non-trivial ways. CC and imputation methods require, at least, that missing data are missing completely at random (MCAR), a term coined by Rubin (1976) and indicating that the occurrence of missingness is independent of both observed and unobserved outcomes. This is in contrast to the much weaker condition of MAR (missing at random), where missingness is allowed to depend on observed outcomes but, given these, not on unobserved ones. A process that is neither MCAR nor MAR is termed non-random (MNAR). This taxonomy is valid in the selection model context (Little and Rubin, 1987), where the joint distribution of the outcomes and the non-response process is factorized into the marginal distribution of the outcomes and the conditional distribution of the non-response process given the outcomes. The reverse factorization is referred to as a pattern-mixture model (Little, 1993, 1994). When a common set of random-effects is thought to influence both the outcomes and non-response process, conditional upon which these processes are independent, then the so introduced model is referred to as a shared-parameter model. For reviews, see Little (1995) and Kenward and Molenberghs (1998).

A key result of Rubin (1976) is that likelihood-based and Bayesian inferential methods are valid, as soon as missingness is MAR, without the need for formulating an explicit missing data model. In other words, under this assumption, one only requires a method and a corresponding software tool, that allows to use incomplete records alongside the complete ones. For this reason, such an approach is often called an ignorable likelihood-based (or Bayesian) analysis.

Within the Gaussian setting, Molenberghs *et al.* (2004) suggest the use of a likelihood-based ignorable analysis, based on the linear mixed-effects model, as an alternative for CC and LOCF. They show that the incomplete sequences contribute to estimands of interests, even early dropouts when scientific interest is in the last planned measurement only. They also show that such an analysis is possible, without

the need of any additional data manipulation, using standard statistical software. Of course, a longitudinal model has to be specified for the entire vector of responses. In a clinical trial setting, with relatively short and balanced response sequences, full multivariate models, encompassing full treatment by group interactions, perhaps corrected for baseline covariates, and an unstructured variance-covariance matrix, are usually within reach. A model of this type is relatively mild in the restrictions made.

The non-Gaussian setting is different in the sense that there is no generally accepted counterpart to the linear mixed-effects model. Longitudinal models in this context include (1) marginal models (Bahadur, 1961; Ashford and Sowden, 1970; Molenberghs and Lesaffre, 1994, 1999), (2) random-effects (or subject-specific) approaches (Stiratelli, Laird and Ware, 1984; Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993), and (3) conditional models where parameters associated with a particular set of outcomes are interpreted relative to values for (a subset of) the other outcomes (Cox, 1972; Rosner, 1984; Liang and Zeger, 1989; Molenberghs and Ryan, 1999). Marginal and random-effects models both have their merit in the analysis of longitudinal clinical trial data. Two important representatives are the generalized estimating equations (GEE) approach within the marginal family and the generalized linear mixed-effects model (GLMM) within the random-effects family. There are important similarities and differences between these model families. While GLMM parameters can be fitted using maximum likelihood, the same is not true for the frequentist GEE method. Therefore, Robins, Rotnitzky and Zhao (1995) have devised so-called weighted generalized estimating equations (WGEE), valid under MAR but requiring the specification of a dropout model in terms of observed outcomes and/or covariates, in view of specifying the weights.

It is possible for the assumption of ignorability not to be true and then one might want to consider more general models. A lot of different modeling strategies have been developed, depending on the type of outcome and non-response process. A short overview will be given.

The non-response process can either be monotone, also called dropout, or non-monotone when there are intermittent missing values. Monotone missingness in continuous outcomes can be modeled, for example using a linear mixed model for the measurements, and a logistic regression for the dropout, depending on the previous, and possibly also the current, measurements (Diggle and Kenward, 1994). Monotone missingness in categorical outcomes is discussed in Molenberghs, Kenward and Lesaffre (1997) and Van Steen *et al.* (2001). They use a Dale model for the measurements and a logistic regression for dropout. Also for non-monotone missingness several modeling strategies exist. For continuous outcomes, Troxel, Harrington and

Lipsitz (1998) used a multivariate normal distribution for the measurements, together with the Markov Chain assumption, and a logistic regression, which only depends on the current measurement, for the missingness process. They encountered problems with the heavy computational load due to a multi-modal likelihood surface. For 2 binary outcomes Baker, Rosenberger and DerSimonian (1992) proposed a family of models for dealing with non-monotone missingness. It is based on log-linear models for the four-way classification of both outcomes, together with their respective missingness indicators. Baker (1995) proposed a model for three binary outcomes with non-monotone missingness, based on marginal and association models for the measurements, and a logistic regression for the missingness mechanism, depending on the last observed and last unobserved measurement.

Pattern-mixture models have gained interest during the last years. Several authors have contrasted selection models and pattern-mixture models. This is either done to compare their answer to the same scientific question, such as marginal treatment effect or time evolution, or to gain additional insight by supplementing the results form a selection model analysis with those from a pattern-mixture approach. Examples for continuous outcomes can be found in Verbeke, Lesaffre and Spiessens (2001a) and Michiels *et al.* (2002) and, while categorical outcomes have been treated by Michiels, Molenberghs and Lipsitz (1999a,b).

With the volume of literature on non-random missing data increasing, there has been growing concern as well (Glynn, Laird and Rubin, 1986). Conclusions based on such more complex models have often been questioned as unreliable because they depend on the specific form assumed for the MNAR process, which can, in principle, not be verified from the data. Also, formal tests for the null hypothesis of random missingness, while technically possible, should be approached with caution. As a result, several authors have advocated to investigate the sensitivity of the results with respect to model assumptions (Little, 1994; Rubin, 1994; Laird, 1994; Molenberghs *et al.*, 1999a). As a general rule, fitting an MNAR model should be subject to careful scrutiny. First, the impact of the assumed distributional form and the specific model choices on the conclusions, when an MNAR model is fitted, has been shown to be much higher than would be the case if data were complete (Kenward, 1998; Scharfstein, Rotnizky and Robins, 1999; Kenward, Goetghebeur and Molenberghs, 2001). Second, one may shift from the classical selection model framework to the pattern-mixture model framework. The use of pattern-mixture models for sensitivity analysis purposes has been explored by Thijs *et al.* (2002). Third, one may want to consider the impact one or a few influential subjects may have on the model parameters. It is natural, at first sight, to make use of the specific influence assessment methodology that has

been developed over the years (Cook, 1986; Chatterjee and Hadi, 1988). Applications of local influence analysis to the Diggle and Kenward (1994) model can be found in Verbeke *et al.* (2001b), Thijs, Molenberghs and Verbeke (2000), and Molenberghs *et al.* (2001b). Van Steen *et al.* (2001) adapted these ideas to the model of Molenberghs, Kenward and Lesaffre (1997), for monotone repeated ordinal data.

1.2 Overview of Subsequent Chapters

The main objective of this thesis is to develop non-random models to handle missing data within the non-Gaussian setting, since also categorical outcomes are very prominent in statistical practice, and, as was already mentioned in the previous section, techniques for this type of data are less standard, because of the lack of a simple analogue to the normal distribution. Also some other aspects of missing data will be discussed, such as the advantages of the use of likelihood-based ignorable analyses, the sensitivity of the developed models, and the behavior of local influence.

In **Chapter 2**, we will introduce the key example data that will be used throughout this work. Section 2.1 describes data on marital satisfaction, obtained from couples on two distinct moments in time. In Section 2.2 data from two clinical trials on the depression status of patients are considered. A multicenter, postmarketing study on the use of fluvoxamine is introduced in Section 2.3. Some details about the first Belgian Health Interview Survey in 1997 are presented in Section 2.4. Finally, Section 2.5 describes a data set, designed to study the effect of the inhibition of testosterone production in rats.

Chapter 3 considers some background on missing data. It starts with the introduction of the terminology and the definition of various missing data patterns in Section 3.1. Afterwards, some simple methods, such as a complete case (CC) analysis and imputation strategies (among which last observation carried forward, LOCF), and the method of direct likelihood for continuous outcomes will be reviewed in Section 3.2. In Section 3.3 an overview of the various modeling frameworks for the less straightforward non-Gaussian longitudinal data setting is provided. Subsequently, focus is on generalized linear mixed effects models on the one hand, of which the parameters can be estimated using full likelihood, and on generalized estimating equations on the other hand, which is a non-likelihood method and hence requires a modification to be valid under MAR. This modification will also be described in detail.

Molenberghs *et al.* (2004) argue that, in the context of clinical trials, a likelihood-

based analysis without modeling the dropout process, is more likely to be valid, and even easier to implement than CC and LOCF analyses, when the outcomes of interest are continuous. In **Chapter 4**, the same arguments will be used to strengthen similar issues in different settings. Section 4.1 shows that also in the context of sociology, when analyzing continuous data, there is no need to use CC above direct likelihood. The possible correlation between the parents' responses is taken into account, providing an extra dimension to the analyses. In Section 4.2, attention is devoted to the analysis of data from two clinical trials with binary outcomes. It is illustrated that also in this case more general models, such as the earlier introduced generalized linear mixed models or the (weighted) generalized estimating equations, are preferable to CC and LOCF.

Baker, Rosenberger and DerSimonian (1992) proposed a family of models for bivariate binary data subject to non-random non-response. In **Chapter 5**, Section 5.1, the original model family is sketched. In Section 5.2, these models are reformulated and extended to accommodate for, possibly continuous, covariates. This parameterization avoids the risk of invalid solutions. These BRD models are able to deal with non-monotone missingness but have some limitations as well, stemming from the conditional interpretation of the model parameters. Section 5.3 shows some insight in the derivatives of the log-likelihood function. Finally, this methodology is applied to the fluvoxamine data in Section 5.4.

In **Chapter 6**, the relatively unexplored domain of non-monotone missingness with multivariate ordinal responses will be broached. In this context, the multivariate Dale model (Molenberghs and Lesaffre, 1994) will be used and introduced in Section 6.1. Section 6.2 focuses on its bivariate version. To allow for multivariate categorical outcomes with non-monotone missingness, a multivariate Dale model for the measurements is combined with the same multivariate Dale model for the missingness mechanism, into a so-called Dale-Dale model in Section 6.3. Finally, this family of models is fitted to the Belgian Health Interview Survey in Section 6.4.

Besides the selection models, there is growing interest in pattern-mixture modeling. However, they are not yet available for non-monotone missing data. **Chapter 7** will focus on these type of models. In Section 7.1, the pattern-mixture context will be introduced. The strategy of identifying restrictions is the topic in Section 7.2, while Section 7.3 focuses on the special case of three measurements. Section 7.5 will discuss the assumptions needed when intermittent missingness is present. Attention is devoted to the determination of marginal effects across patterns in Section 7.6. In Section 7.7, the fluvoxamine data are reanalyzed using pattern-mixture models.

Since MNAR models are sensitive to the underlying model assumptions, gener-

ally performing a sensitivity analysis is strongly advisable. Many different routes might be followed, either at the level of the models, or at the level of the individuals (Draper, 1995; Glynn, Laird and Rubin, 1986; Molenberghs, Kenward and Goetghebeur, 2001a; Rubin, 1977, 1994; Scharfstein, Rotnizky and Robins, 1999). In **Chapter 8**, it is shown that also for multivariate and longitudinal binary data, subject to non-monotone missingness, methods to assess the influence can be developed. The method of global influence, also known as the case-deletion method (Cook and Weisberg, 1982), is reviewed in Section 8.1, and applied to the Dale-Dale model family in Section 8.2. A local influence strategy (Cook, 1986) for the BRD model family is developed in Section 8.3, and applied to the fluvoxamine trial in Section 8.4.

The original idea behind the use of local influence methods with an eye on sensitivity analysis was to detect observations that had a high impact on the conclusions *due to their aberrant missingness mechanism*. However, local influence tends to pick up a lot of different anomalies in the data at hand, not just deviations in the MNAR mechanism. In **Chapter 9**, the method of local influence is further studied, not only to better understand its behavior, but also to increase insight in the overall behavior and impact of MNAR mechanisms. Local influence in the context of the Diggle and Kenward (1994) model is described in Section 9.1. This methodology is applied to the rats case study in Section 9.2. Section 9.3 is dedicated to the behavior of local influence under standard conditions as well as under a number of anomalous scenarios. This is done using simulations and general modeling considerations.

2

Case Studies

In this chapter, we introduce several longitudinal data sets with missing observations. Not only clinical trials, but also two surveys are considered. These data sets will be used throughout the text as key examples to illustrate the techniques developed and other methods that are becoming more and more frequently used. In some studies, the continuous outcomes are retained, while other studies had binary outcomes recorded, or continuous ones were dichotomized.

2.1 Marriage Satisfaction Data

The research sample of the first study consists of married men and women participating in the longitudinal research project "Child-rearing and family in the Netherlands". In 1990 and 1995 the same family members (wife, husband and target child) provided information about essentially the same domains. Families were recruited using a multi-stage sampling method. In a first stage, a sample was taken from all Dutch municipalities, distinguished by regional zone and degree of urbanization. In a second stage, a sample of children aged 9 to 16 years was taken in the selected municipalities. The children were selected in such a way that in each city as many boys as girls and as many children aged 9 to 12 as in the age range 13–16 were chosen. In 1990, this procedure resulted in a sample of 1829 families. The response ratio was 43% ($N = 788$).

Despite the reduction of the number of families, the sample was representative regarding regional zone and degree of urbanization. Of the 656 families who agreed in 1990 to participate in the second wave, 627 could be traced five years later. Of these contacted families, 484 (77%) did actually participate in 1995. This sample proved to be still representative for regional zone but not for degree of urbanization. It appeared that primarily participants from the bigger cities refused to participate for the second time in the research project. More technical details on the database, and the reasons of refusal to participate, can be found in Gerris *et al.* (1992, 1993, 1998). The data were gathered by means of structured interviews and questionnaires, completed by both the child and its parents. To establish a homogeneous research group for a study on marital quality, only first marriages in which both men and women with the Dutch nationality were selected. This selection resulted in a research group of 646 couples in 1990 and 386 couples in 1995.

We consider three continuous responses to measure marital quality: marital satisfaction (satisfaction with the relationship and/or the partner), negative communication (to what degree certain forms of negative communication are characteristic to their marital relationship) and open communication (to what degree personal feelings and experiences are shared). All responses were measured on a 7-point Likert scale, ranging from 1 = "not at all applicable" to 7 = "very applicable". The uniqueness and stability of this concept was demonstrated in Van den Troost *et al.* (2001). The individual profiles for the three responses are shown in Figure 2.1, for males and females separately. The following covariates will be taken into account because of potential influence on marital quality: education (ranging from 1 = "elementary school" to 9 = "university education"), family income in euros (1 = "1100–1600", 2 = "1600–1800", 3 = "1800–2100", 4 = "2100–2500", 5 = "2500–3250", 6 = "3250–4500" and 7 = "more than 4500"), year of birth, year of marriage, number of children, marital status of the couple's parents (whether or not they were married while living with them, 1 = "yes", 2 = "no" and 3 = "not applicable").

2.2 Hamilton Depression Rating Scale Data

Secondly we consider two sets of data, coming from two clinical trials, enrolling 167 and 342 patients, respectively. The depression status of the patients is measured using the Hamilton Depression Rating Scale ($HAMD_{17}$). Analyses of this continuous $HAMD_{17}$ score are performed in Mallinckrodt *et al.* (2003a,b) and Molenberghs *et al.* (2004). Later, interest turned towards the dichotomized version of the $HAMD_{17}$ score

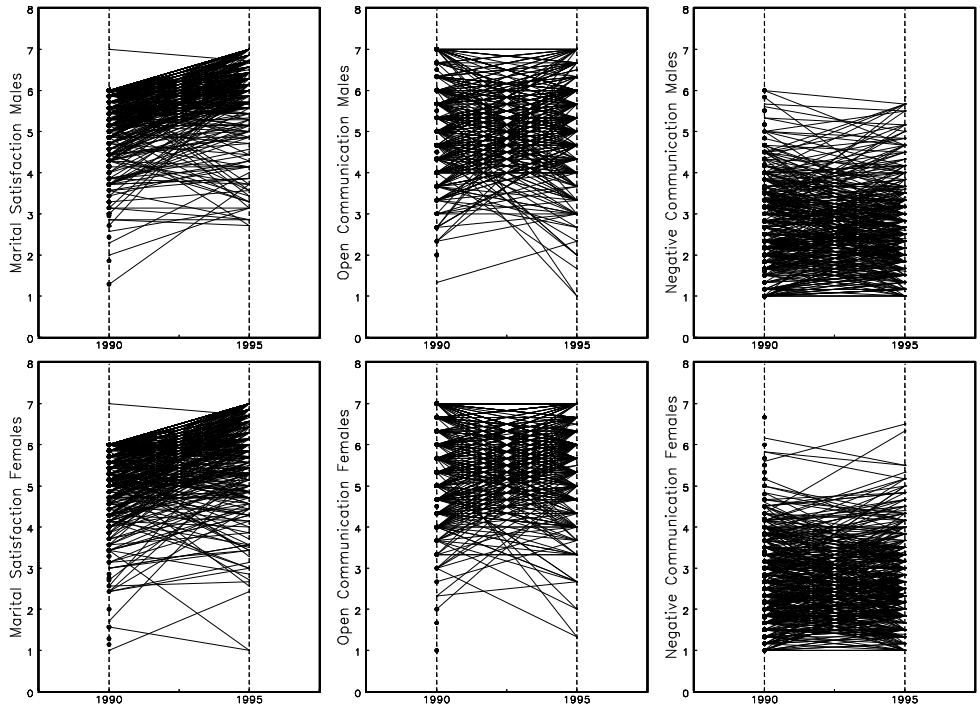


Figure 2.1: *Marriage Satisfaction Data. Individual profiles for the three outcomes and males/females separately.*

(1 if $HAMD_{17} > 7$, 0 otherwise). For each patient, a baseline assessment is available. Post-baseline visits differ between both studies (visits 4 to 11 for the first study, visits 4 to 8 for the second one).

For blinding purposes, therapies are recoded as A1 for primary dose of the experimental drug, A2 for secondary dose of experimental drug, and B and C for non-experimental drugs. The treatment arms across the two studies are as follows: A1, B, and C for Study 1; A1, A2, B, and C for Study 2. The primary contrast is between A1 and C. Emphasis is on the difference between arms at the end of the study. In both studies, the dropout at the end of the study lies between 30% and 40% per treatment arm. A graphical representation of the dropout, per study and per arm, is given in Figure 2.2.

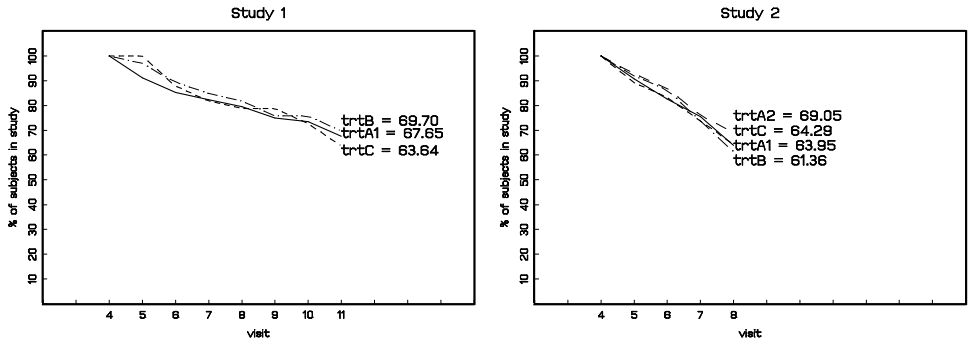


Figure 2.2: *Hamilton Depression Rating Scale Data. Evolution of dropout per study and per treatment arm. Treatment arms A1 and C, being the ones of primary interest, are shown in bolder typeface.*

2.3 Fluvoxamine Data

Another example comes from a multicenter, postmarketing study involving 315 patients that were treated by fluvoxamine for psychiatric symptoms described as possibly resulting from a dysregulation of serotonine in the brain. The data are discussed in Molenberghs and Lesaffre (1994), Kenward, Lesaffre and Molenberghs (1994), Molenberghs, Kenward and Lesaffre (1997), Michiels and Molenberghs (1997), Molenberghs *et al.* (1999b), and Jansen *et al.* (2003).

After enrollment into the study, a number of baseline characteristics was scored, and the patient was assessed at four follow-up visits. The therapeutic effect and the extent of worsening side effects were scored at each visit on an ordinal scale. A side effect occurs if new symptoms appear while there is therapeutic effect if old symptoms disappear. We will focus on a dichotomized version (present/absent) of side effects. In Chapter 5, focus is on the first and the last visit, while in Chapter 7, also the second visit will be considered.

Accumulated experience with fluvoxamine in controlled clinical trials has shown that it is effective as a conventional tricyclic antidepressant (Burton, 1991). However, many patients who suffer from depression have concomitant morbidity with conditions such as obsessive-compulsive disorder, anxiety disorders and, to some extent, panic disorders. In most trials, patients with comorbidity are excluded and therefore, it is of interest to gather evidence as to the importance of such factors, with a view on improved diagnosis and treatment. A useful, easy to obtain and quantitative covariate,

Table 2.1: *Fluvoxamine Data (315 subjects). 'Side effects' (yes/no) at the first and last visit.*

yes	89	13	26	2	0	14
no	57	65	49			

Table 2.2: *Fluvoxamine Data. 'Side effects' (yes/no) at the first and last visit. Information for 'prior duration' available.*

yes	88	13	24	2	0	13
no	56	65	49			

strongly related to the history of comorbidity, is duration of the mental illness, prior to inclusion in the trial (*prior duration*). Its effect on the clinical outcomes of the study is therefore of scientific importance (Lesaffre, Molenberghs and Dewulf, 1996) and will be studied.

The observed data are given in Table 2.1. A small subgroup has missing information on the duration covariate. Table 2.2 contains the data that will be used in the analyses (covariate information available). There are two patients with a non-monotone pattern of follow-up while 13 subjects have no follow-up data at all.

2.4 Belgian Health Interview Survey Data

The next data set is the first Belgian Health Interview Survey, which took place in 1997. The HIS1997 was conducted to evaluate the usefulness of a periodic health-related survey, with the idea of collecting information on the subjective health of the Belgian population, as well as on important predictor variables.

The main goal of the HIS1997 was to give a description of the health status of the overall population in Belgium as well as of the three regional subpopulations (Flemish, Walloon and Brussels region), and in addition of the German community. The idea was to obtain a reflection of how specific groups of people experience their health, to what extent they utilize health care facilities, and how they look after their own health by adopting a certain life-style or by relying on preventive and other health services.

Table 2.3: *Health Interview Survey Data (10786 subjects). 'Mental health' versus 'fixed general practitioner'.*

	Yes	No				
Good	5048	522	17	1737	266	571
Bad	2360	252	13			

Table 2.4: *Health Interview Survey Data. 'Mental health' versus 'fixed general practitioner'. Separate for males (5288) and females (5498), respectively.*

Males	2579	320	5	868	138	305
	928	140	5			
Females	2469	202	12	869	128	266
	1432	112	8			

The target population was defined as all people residing in Belgium at a particular point in time. The National Register was used as the sampling frame. The total number of successful interviews for the sample was set to 10,000 (0.1% of the Belgian population), 3500 in both the Flemish and Walloon regions, and 3000 in the Brussels region. Sampling was based on a combination of stratification, multistage sampling, and clustering. More details on the design of the study, together with a descriptive analysis of missingness, can be found in Burzykowski *et al.* (1999) and Renard *et al.* (1998).

We will focus on two specific aspects of the HIS1997, namely the status of mental health (0 = good, 1 = bad), and having a fixed general practitioner (0 = yes, 1 = no). Missingness occurred for several reasons: person is not interviewed, no valid information obtained, an error by the interviewer, question is not applicable, person does not know, no answer given. In total, information is available for 10786 subjects. Data are shown in Table 2.3. Covariates of interest are gender (male/female) and education (no/primary/lower secondary/higher secondary/higher). The distribution of the subjects over the different covariate levels of gender and education are shown in Tables 2.4 and 2.5, respectively.

Table 2.5: *Health Interview Survey Data. 'Mental health' versus 'fixed general practitioner'. Separate for no education (267), primary education (1486), lower secondary education (1844), higher secondary education (3265), higher education (3843), and education information missing (81), respectively.*

No	112 13 46 5	0 0	36 9	46
Primary	734 67 374 17	3 3	150 26	112
Lower secondary	899 74 446 47	2 1	255 36	84
Higher secondary	1611 131 699 51	4 3	581 62	123
Higher	1662 235 776 127	8 6	704 128	197
Missing	30 2 19 5	0 0	11 5	9

2.5 Rats Data

The last data set comes from a randomized experiment, designed to study the effect of the inhibition of testosterone production in rats. The experiment was conducted at the Department of Orthodontics of the Catholic University of Leuven (K.U.L.) in Belgium (Verdonck *et al.*, 1998). A total of 50 male Wistar rats has been randomized to either a control group or one of two treatment groups, consisting of a low and a high dose of the drug Decapeptyl, which is an inhibitor for the testosterone production in rats. The treatment started at the age of 45 days, and measurements were taken every

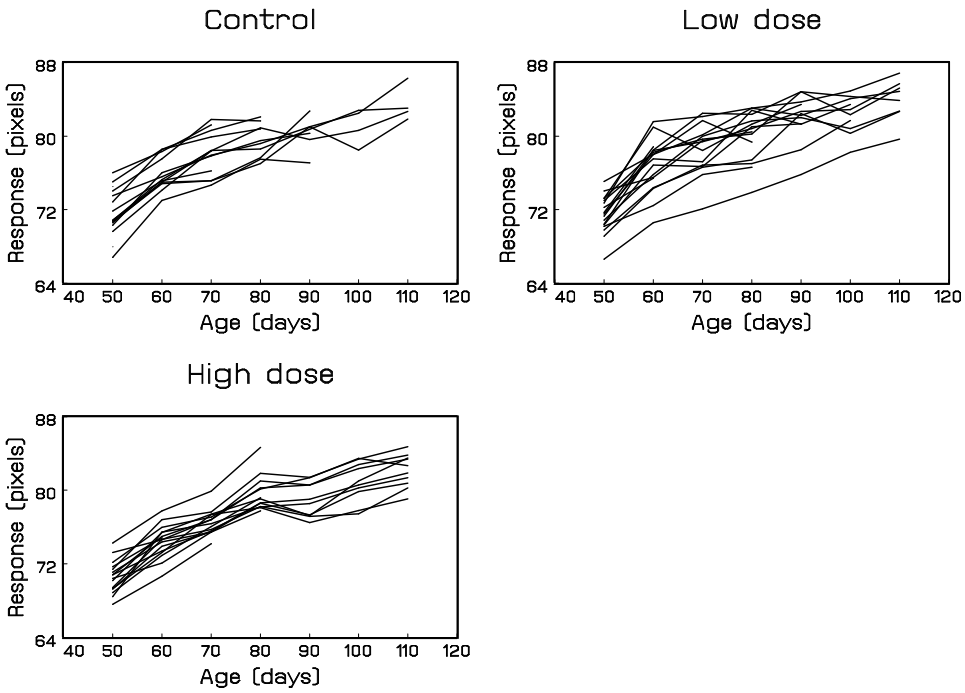


Figure 2.3: *Rats Data. Individual growth curves for the three treatment groups separately.*

10 days, with the first observation taken at the age of 50 days. The response of interest is a characterization of the height of the skull (in pixels), taken under anaesthesia. Unfortunately, many rats do not survive anaesthesia implying that for only 22 (44%) rats all 7 designed measurements could have been taken. The individual profiles are shown in Figure 2.3. The data have been analyzed extensively, see for example Verbeke and Lesaffre (1999), Verbeke *et al.* (2001b) or Verbeke and Molenberghs (2003).

3

Fundamental Concepts of Missing Data

Data from longitudinal studies are generally prone to incompleteness. Since incompleteness usually occurs for reasons outside of the control of the investigators and may be related to the outcome measurement of interest, it is generally necessary to address the process governing incompleteness. Only in special but important cases it is possible to ignore the missingness process.

Commonly used methods to analyze incomplete longitudinal data include *complete case analysis* (CC) and *last observation carried forward* (LOCF) or other simple forms of imputation. Claimed advantages include computational simplicity, no need for a full longitudinal model (e.g., when the interest is in terms of the last planned measurement occasion only) and, for LOCF, compatibility with the intention-to-treat (ITT) principle.

In this chapter, we first introduce some general concepts regarding incomplete data (Section 3.1). In Section 3.2 the use of a likelihood-based ignorable analysis, within the Gaussian setting, based on the linear mixed-effects model is suggested, and its advantage compared to CC and LOCF is discussed. In Section 3.3, we will focus on the less straightforward situation of non-Gaussian outcomes, such as binary, categorical, or count data. We therefore first sketch a general taxonomy for longitudinal models in

this context, including marginal, random-effects (or subject-specific), and conditional models. We then argue that marginal and random-effects models both have their merit in the analysis of longitudinal clinical trial data and focus on two important representatives, i.e., the generalized estimating equations (GEE, Liang and Zeger, 1986) approach within the marginal family and the generalized linear mixed-effects model (GLMM, Stiratelli, Laird and Ware, 1984; Wolfinger and O'Connell, 1993; Breslow and Clayton, 1993) within the random-effects family. We highlight important similarities and differences between these model families. While GLMM parameters can be fitted using maximum likelihood, the same is not true for the frequentist GEE method. Therefore, Robins, Rotnitzky and Zhao (1995) have devised so-called weighted generalized estimating equations (WGEE), valid under MAR but requiring the specification of a dropout model in terms of observed outcomes and/or covariates, in view of specifying the weights.

3.1 General Concepts of Modeling Incompleteness

To incorporate incompleteness into the modeling process, we need to reflect on the nature of the missing value mechanism and its implications for statistical inference. We will therefore introduce the necessary terminology and notation, together with various missing data patterns. Also the important case where the missing data mechanism can be excluded from the statistical analysis will be considered.

3.1.1 Terminology and Notation

Assume that for subject $i = 1, \ldots, N$ in the study a sequence of responses Y_{ij} is designed to be measured at a fixed set of occasions $j = 1, \ldots, n$. The outcomes are grouped into a vector $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in})'$. In addition, for each occasion j define

$$R_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$

The *missing data indicators* R_{ij} are grouped into a vector $\boldsymbol{R}_i$ which is of the same length as $\boldsymbol{Y}_i$. The underlying mechanism generating these $\boldsymbol{R}_i$ is denoted by the missingness process and it is possible to consider several missing data patterns within the study. In case the non-response process is restricted to dropout we have that all measurements for a subject from baseline onwards up to a certain time are recorded, after which all data are missing. The vector $\boldsymbol{R}_i$ can then be represented as a scalar D_i say, defined as $D_i = 1 + \sum_{j=1}^{n} R_{ij}$, representing the occasion at which dropout

occurs. In terms of missing data patterns we call this a *dropout* pattern and we can split this further into *monotone* dropout patterns where all measurements prior to the time of dropout are observed and *non-monotone* dropout patterns where for some reason it is possible that also measurements prior to the actual dropout are missing.

It is often necessary to split the vector $\boldsymbol{Y}_i$ into observed ($\boldsymbol{Y}_i^o$) and missing ($\boldsymbol{Y}_i^m$) components respectively. The following terminology is used:

Complete data $\boldsymbol{Y}_i$: the scheduled measurements. This is the hypothetical outcome vector that would have been recorded if there were no missing data.

Full data $(\boldsymbol{Y}_i, \boldsymbol{R}_i)$: the complete data, together with the missing data indicators. Note that one observes the measurements $\boldsymbol{Y}_i^o$ together with the dropout indicators $\boldsymbol{R}_i$.

Covariates $\boldsymbol{X}_i$: apart from the outcomes, additional information is measured. This information can be collected before or during the study. The covariate vector is allowed to change for different outcome components t and can include continuous as well as discrete variables. We assume no missing values appear in $\boldsymbol{X}_i$. Methods for the case of missing covariates have been explored by several authors (Little, 1992; Robins, Rotnitzky and Zhao, 1994; Zhao, Lipsitz and Lew, 1996).

3.1.2 Missing Data Mechanisms

In principle, one would like to consider a joint model for the measurement process together with the dropout process. In other words, interest is in the density of the full data

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}), \tag{3.1}$$

where the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describe the measurement and missingness processes, respectively. Covariates are assumed to be measured, but have been suppressed from notation for simplicity. We can factorize this density in several ways. In this and the next chapters we will discuss the factorization

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|\boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{\psi}). \tag{3.2}$$

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. This factorization forms the basis of *selection modeling* and can be explained intuitively by considering the second factor to correspond to the (self-)selection of individuals into 'observed' and 'missing' groups. In Chapter 7 we will focus on the reverse factorization

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|\boldsymbol{r}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{\psi}), \tag{3.3}$$

the so-called *pattern-mixture models*. This density can be seen as a mixture of different populations, characterized by the observed pattern of missingness. After initial mention of these models (Glynn, Laird and Rubin, 1986; Little and Rubin, 1987), they are receiving more attention lately (Little, 1993, 1994, 1995; Ekholm and Skinner, 1998; Hogan and Laird, 1997).

Selection models were used by Rubin (1976) and Little and Rubin (1987) to define their missing data terminology. This classical taxonomy is based on the second factor of (3.2)

$$f(\boldsymbol{r}_i|\boldsymbol{y}_i,\boldsymbol{\psi}) = f(\boldsymbol{r}_i|\boldsymbol{y}_i^o,\boldsymbol{y}_i^m,\boldsymbol{\psi}) \tag{3.4}$$

and can be described as follows:

- if (3.4) is independent of the measurements, i.e., when it assumes the form

  $$f(\boldsymbol{r}_i|\boldsymbol{\psi})$$

  then the process is termed *missing completely at random* (MCAR).

- if (3.4) is independent of the unobserved (missing) measurements $\boldsymbol{Y}_i^m$, but depends on the observed measurements $\boldsymbol{Y}_i^o$, thereby assuming the form

  $$f(\boldsymbol{r}_i|\boldsymbol{y}_i^o,\boldsymbol{\psi})$$

  then the process is referred to as *missing at random* (MAR).

- if (3.4) depends on the missing values $\boldsymbol{Y}_i^m$, the process is referred to as *informative* missingness or *missing not at random* (MNAR).

Focusing on the selection model (3.2) when analyzing the data, two choices have to be made.

Model for measurements. A choice has to be made regarding the modeling approach to the measurements. Several views are possible.

View 1. One can choose to analyze the entire profile of outcomes on a subject, irrespective of whether interest focuses on the entire profile or rather on a specific response at a specific time point. In the latter case, the motivation to model the entire profile is because, for example, earlier responses do provide statistical information on later ones. Inference is then based on the appropriate subset of parameters from the full longitudinal model.

View 2. One defines the scientific question and restrict the corresponding analysis to the *last planned occasion*. Of course, as soon as dropout occurs, such a measurement may not be available. In this case, one can either accept the dropout as it is or use one or other strategy (e.g. imputation) to incorporate the missing outcomes.

View 3. One can choose to define the question and the corresponding analysis in terms of the *last observed measurement*. While sometimes used as an alternative motivation for so-called *last observation carried forward* analyses (Siddiqui and Ali, 1998; Mallinckrodt *et al.*, 2003a,b), a common criticism is that the last observed measurement amalgamates measurements at real stopping times (for dropouts) and at a purely design-based time (for completers).

In all cases, we need to reflect on how to deal with missingness. Under View 1, both simple methods, like a complete case analysis, as well as the more advanced direct-likelihood method are possible. Also View 2 necessitates reflection on the missing data mechanism, although all other outcomes on the sample couple are ignored, so one is not making use of such additional information. View 3 completely ignores the missing data problem because the question is couched completely in terms of observed measurements. Under View 3, an LOCF analysis might be acceptable, provided it matched the scientific goals, but is then better described as a Last Observation analysis because nothing is carried forward. Such an analysis should properly be combined with an analysis of time to dropout, perhaps in a survival analysis framework. Of course, an investigator should reflect very carefully on whether View 3 represents a relevant and meaningful scientific question (see also Shih and Quan, 1997).

Method for handling missingness. A choice has to be made regarding the modeling approach for the missingness process. Under certain assumptions this process can be ignored (e.g., a likelihood-based ignorable analysis or direct-likelihood analysis). Some simple methods, such as a CC analysis or LOCF, do not explicitly address the missingness process.

The measurement model will depend on whether or not a full longitudinal analysis is done. When the focus is on the last observed measurement or on the last measurement occasion only, one typically opts for classical two- or multi-group comparisons (t test, Wilcoxon, etc.). When a longitudinal analysis is deemed necessary, the choice depends on the nature of the outcome. For continuous outcomes, one often assumes

a multivariate normal, a special case of the linear mixed-effects model (Verbeke and Molenberghs, 2000):

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \tag{3.5}$$

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 is a $(n \times p)$ known design matrix, $\boldsymbol{\beta}$ is the p dimensional vector containing the fixed effects, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, V_i)$, with V_i a general or structured variance covariance matrix. If necessary, a fully general linear mixed effects model can be considered, without any problem.

Assume that incompleteness is due to dropout only, and that the first measurement Y_{i1} is obtained for everyone. A possible model for the dropout process is a logistic regression for the probability of dropout at occasion j, given that 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

$$\begin{aligned} \text{logit}[g(\boldsymbol{h}_{ij}, y_{ij})] &= \text{logit}\left[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)\right] \\ &= \boldsymbol{h}'_{ij}\boldsymbol{\psi} + \omega y_{ij}, \qquad\qquad i = 1, \ldots, N \end{aligned} \tag{3.6}$$

(Diggle and Kenward, 1994). When ω equals zero, the dropout model is MAR, and all parameters can be estimated using standard software 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 (missing not at random), where missingness is allowed to depend on both observed and unobserved outcomes. Model (3.6) provides the building blocks for the dropout process $f(\boldsymbol{r}_i | \boldsymbol{y}_i, \boldsymbol{\psi})$.

3.1.3 Ignorability

Let us decide to use likelihood based estimation. The full data likelihood contribution for subject i assumes the form

$$L^*(\boldsymbol{\theta}, \boldsymbol{\psi} | \boldsymbol{y}_i, \boldsymbol{r}_i) \propto f(\boldsymbol{y}_i, \boldsymbol{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}).$$

Since inference has to be based on what is observed, the full data likelihood L^* has to be replaced by the observed data likelihood L:

$$L(\boldsymbol{\theta}, \boldsymbol{\psi} | \boldsymbol{y}_i^o, \boldsymbol{r}_i) \propto f(\boldsymbol{y}_i^o, \boldsymbol{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) \tag{3.7}$$

with

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

Under an MAR process, we obtain

$$\begin{aligned} f(\boldsymbol{y}_i^o, \boldsymbol{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) &= \int f(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m|\boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i^o, \boldsymbol{\psi}) d\boldsymbol{y}_i^m \\ &= f(\boldsymbol{y}_i^o|\boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i^o, \boldsymbol{\psi}), \end{aligned} \tag{3.8}$$

i.e., the likelihood factorizes into two components of the same functional form as the general factorization (3.2) of the complete data. If further $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are disjoint in the sense that the parameter space of the full vector $(\boldsymbol{\theta}', \boldsymbol{\psi}')'$ is the product of the individual parameter spaces (separability condition) then inference can be based on the marginal observed data density only.

In conclusion, when the separability condition is satisfied, *within the likelihood framework*, ignorability is equivalent to the union of MAR and MCAR. Hence, non-ignorability and MNAR are synonyms in this context. A formal derivation is given in Rubin (1976) (where it is also shown that the same requirements hold for Bayesian inference, but that frequentist inference is ignorable only under MCAR) and Little and Rubin (1987). 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, standard errors if based on the observed information matrix, and likelihood ratio values (Kenward and Molenberghs, 1998). Note that the estimands are the parameters of model (3.5), which is a model for complete data, corresponding to what one would expect to see in the absence of dropouts. A similar method is the full information maximum likelihood (FIML) method (Wothke, 2000), which casewise maximizes the likelihood of the observed data. This method has already been used within the sociological research.

A few cautionary remarks are warranted. First, when at least part of the scientific interest is directed towards the nonresponse process (e.g., to study reasons for non-response), obviously both processes need to be considered. 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 measures of precision for 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. Thirdly, it may be hard to rule out the operation of an MNAR mechanism, but this will not yet be discussed in this chapter. Fourthly, such an analysis can proceed only under View 1, i.e., a full longitudinal analysis is necessary, even when interest lies in a question concerning one specific occasion. In the latter case, the fitted model can be used as the basis for inference at that occasion. A common criticism is that a model needs to be considered, with the risk of model misspecification. However, it should be noted that in many sociological 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., a full group by time interaction model for the fixed effects and an unstructured variance-covariance matrix). A model of this type is relatively mild in the restrictions made.

These arguments, supplemented with the availability of software tools within which such multivariate models can be fitted to incomplete data (such as the MIXED, NLMIXED, GLIMMIX and GENMOD procedures in SAS), cast down regarding the usefulness of such simple methods as CC and LOCF. Apart from biases as soon as the missing data mechanism is not MCAR (Molenberghs *et al.*, 2004), CC can suffer from severe efficiency losses. Especially since tools have become available to include incomplete sequences along with complete ones into the analysis, one should do everything possible to avoid wasting patient data.

LOCF, a so-called imputation strategy (Dempster and Rubin, 1983; Little and Rubin, 2002), shares with other imputation methods that precision can be inflated artificially. Further, as Molenberghs *et al.* (2004) have shown, the method can produce severely biased treatment comparisons and, perhaps contrary to some common belief, such biases can be conservative but also liberal. The method rests on the strong assumption that a patient's outcome profile remains flat, at the level of the last observed measurement, throughout the remainder of follow up.

The complete case and LOCF methods, together with the direct-likelihood method, will be described in the next section and studied further in subsequent sections.

3.2 Simple Methods and Direct Likelihood

We will briefly review a number of relatively simple methods that still are commonly used. For the validity of many of these methods, MCAR is required. For others, such as LOCF, MCAR is necessary but not sufficient. The focus will be on the complete

case method, for which data are removed, and on imputation strategies, where data are filled in. Regarding imputation, one distinguishes 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 acknowledges the uncertainty stemming from filling in missing values rather than observing them (Rubin, 1987; Schafer, 1997). LOCF will be discussed within the context of imputation strategies, although LOCF can be placed in other frameworks as well.

A **complete case analysis** includes only those cases for which all measurements were recorded. This method has obvious advantages. It is simple to describe and almost any software can be used since there are no missing data. Unfortunately, the method suffers from severe drawbacks. Firstly, there is nearly always a substantial loss of information. Small amounts of missingness on each of the measurement occasions can result in an overal proportion of complete cases which is unacceptably low. Even though the reduction of the number of complete cases will be less severe in settings where the missingness indicators are correlated, this loss of information will usually militate against a complete case analysis. Secondly, 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. A CC analysis can be conducted when Views 1 and 2 of Section 3.1.2 are adopted. It obviously is not a reasonable choice with View 3.

An alternative way to obtain a data set on which complete data methods can be used is to fill in rather than delete (Little and Rubin, 1987). Concern has been raised regarding **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 (1987) show that the application of imputation could be considered acceptable in a linear model with one fixed effect and one error term, but that it is generally not acceptable for hierarchical and multivariate models, repeated measures with a complicated error structure, random-effects, and mixed-effects models.

A method that has received considerable attention (Siddiqui and Ali, 1998; Mallinckrodt *et al.*, 2003a,b) is **last observation carried forward** (LOCF). In the LOCF method, whenever a value is missing, the last observed value is substituted.

The technique can be applied to both monotone and non-monotone missing data. It is typically applied in settings where incompleteness is due to attrition.

LOCF can, but not necessarily, be regarded as an imputation strategy, depending on which of the views of Section 3.1.2 is taken. The choice of viewpoint has a number of consequences. First, when the problem is approached from a missing data standpoint, one has to think it plausible that subjects' measurements do not change from the moment of dropout onwards (or during the period they are unobserved in the case of intermittent missingness). In a clinical trial setting, one 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. Secondly, LOCF shares with other single imputation methods that it artificially increases the amount of information in the data, by treating imputed and actually observed values on an equal footing. This is especially true if a longitudinal view is taken. Verbeke and Molenberghs (1997, Ch. 5) have shown that all features of a linear mixed model (group difference, evolution over time, variance structure, correlation structure, random effects structure, ...) can be affected.

Thus, scientific questions with which LOCF is compatible will be those that are phrased in terms of the last obtained measurement (View 3). Whether or not such questions are sensible should be the subject of scientific debate, which is quite different from a *post hoc* rationale behind the use of LOCF. Likewise, it can be of interest to model the complete cases separately and to make inferences about them. In such cases, a CC analysis is of course the only reasonable way forward. This is fundamentally different from treating a CC analysis as one that can answer questions about the randomized population as a whole.

The user of imputation strategies faces several dangers. First, the imputation model could be wrong and, hence, the point estimates biased. Second, even for a correct imputation model, the uncertainty resulting from missingness is ignored. Indeed, even when one is reasonably sure about the mean value the unknown observation *would have had*, the actual stochastic realization, depending on both the mean and error structures, is still unknown. In addition, most methods require the MCAR assumption to hold while some even require additional and often unrealistically strong assumptions.

On the other hand, a **likelihood-based ignorable analysis** or **direct-likelihood method** produces expectations for the missing observations but no explicit imputation takes place, hence the amount of information in the data is not overestimated and important model elements, such as mean structure and variance components, are not distorted.

Historically, an important motivation behind the simpler methods was their very simplicity. Currently, with the availability of commercial software tools such as, for example, the SAS procedure MIXED, this motivation no longer applies. Arguably, an MAR analysis is the preferred choice. Of course, the correctness of an MAR analysis rests upon the truth of the MAR assumption, which is, in turn, never completely verifiable. Note that purely resorting to MNAR analyses is not satisfactory either, since such models are known to be sensitive to unverifiable model assumptions that necessarily have to be made. These and related issues will be discussed in subsequent chapters.

3.3 Discrete Repeated Measures

Whereas the linear mixed model, and its special cases, is seen as a unifying parametric framework for Gaussian repeated measures (Verbeke and Molenberghs, 2000), there are many more options available in the non-Gaussian setting. In a *marginal model*, marginal distributions are used to describe the outcome vector $\boldsymbol{Y}$, given a set $\boldsymbol{X}$ of predictor variables. The correlation among the components of $\boldsymbol{Y}$ can then be captured either by adopting a fully parametric approach or by means of working assumptions, such as in the semiparametric approach of Liang and Zeger (1986). Alternatively, in a *random-effects model*, the predictor variables $\boldsymbol{X}$ are supplemented with a vector $\boldsymbol{\theta}$ of random effects, conditional upon which the components of $\boldsymbol{Y}$ are usually assumed to be independent. This does not preclude that more elaborate models are possible if residual dependence is detected (Longford, 1993). Finally, a *conditional model* describes the distribution of the components of $\boldsymbol{Y}$, conditional on $\boldsymbol{X}$ but also conditional on (a subset of) the other components of $\boldsymbol{Y}$. Well-known members of this class of models are log-linear models (Gilula and Haberman, 1994). Let us give a simple example of each for the case of Gaussian outcomes. A marginal model starts from specifying:

$$E(Y_{ij}|\boldsymbol{x}_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}, \tag{3.9}$$

whereas in a random-effects model we focus on the expectation, conditional upon the random-effects vector:

$$E(Y_{ij}|\boldsymbol{b}_i, \boldsymbol{x}_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i. \tag{3.10}$$

The conditional model uses expectations of the form

$$E(Y_{ij}|Y_{i,j-1}, \ldots, Y_{i1}, \boldsymbol{x}_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha Y_{i,j-1}. \tag{3.11}$$

In the linear mixed model case, random-effects models imply a simple marginal model. This is due to the elegant properties of the multivariate normal distribution. In particular, the expectation (3.9) follows from (3.10) by either (a) marginalizing over the random effects or by (b) conditioning upon the random-effects vector $\boldsymbol{b_i} = \mathbf{0}$. Hence, the fixed-effects parameters $\boldsymbol{\beta}$ have both a marginal as well as a hierarchical model interpretation.

Since marginal and random-effects models are the most useful ones in many different contexts, and given this connection between them, it is clear why the linear mixed model provides a unified framework in the Gaussian setting. Such a close connection between the model families does not exist when outcomes are of a non-normal type, such as binary, categorical, or discrete. We will consider the marginal and random-effects model families in turn and then point to some particular issues arising within them or when comparisons are made between them.

3.3.1 Marginal Models

Thorough discussions on marginal modeling can be found in Diggle *et al.* (2002), and Fahrmeir and Tutz (2001). The specific context of clustered binary data has received attention in Aerts *et al.* (2002). Apart from full likelihood approaches, non-likelihood approaches, such as *generalized estimating equations* (GEE, Liang and Zeger, 1986) or *pseudo-likelihood* (Geys, Molenberghs and Lipsitz, 1998; le Cessie and van Houwelingen, 1994) have been considered.

Bahadur (1961) proposed a marginal model, accounting for the association via marginal correlations. Ekholm (1991) proposed a so-called success probabilities approach. George and Bowman (1995) proposed a model for the particular case of exchangeable binary data. Ashford and Sowden (1970) considered the multivariate probit model, for repeated ordinal data, thereby extending univariate probit regression. Molenberghs and Lesaffre (1994), and Lang and Agresti (1994) have proposed models which parameterize the association in terms of marginal odds ratios. Dale (1986) defined the bivariate global odds ratio model, based on a bivariate Plackett distribution (Plackett, 1965). Molenberghs and Lesaffre (1994, 1999), Lang and Agresti (1994), and Glonek and McCullagh (1995) extended this model to multivariate ordinal outcomes. They generalize the bivariate Plackett distribution in order to establish the multivariate cell probabilities.

While full likelihood methods are appealing because of their flexible ignorability properties (Section 3.1.3), their use for non-Gaussian outcomes can be problematic due to prohibitive computational requirements. Therefore, GEE is a viable alternative

within this family. Since GEE is frequentist in nature, it is ignorable only under MCAR and this motivates the proposal of so-called *weighted generalized estimating equations* (WGEE). We will discuss these in turn.

Generalized Estimating Equations

GEE, useful to circumvent the computational complexity of full likelihood, can be considered whenever interest is restricted to the mean parameters (treatment difference, time evolutions, effect of baseline covariates, etc.). It is rooted in the quasi-likelihood ideas expressed in McCullagh and Nelder (1989). Modeling is restricted to the correct specification of the marginal mean function, together with so-called *working assumptions* about the correlation structure of the vector of repeated measures.

Let us now introduce the classical form of GEE. Note that the score equations, to be solved when computing maximum likelihood estimates under a marginal normal model $\boldsymbol{y_i} \sim N(X_i\boldsymbol{\beta}, V_i)$, are given by

$$\sum_{i=1}^{N} X_i'(A_i^{1/2}C_iA_i^{1/2})^{-1}(\boldsymbol{y_i} - X_i\boldsymbol{\beta}) \quad = \quad \mathbf{0}, \tag{3.12}$$

in which the marginal covariance matrix V_i has been decomposed in the form $A_i^{1/2}C_iA_i^{1/2}$, with A_i the matrix with the marginal variances on the main diagonal and zeros elsewhere, and with C_i equal to the marginal correlation matrix. Switching to the non-Gaussian case, the score equations become

$$S(\boldsymbol{\beta}) \quad = \quad \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'}(A_i^{1/2}C_iA_i^{1/2})^{-1}(\boldsymbol{y_i} - \boldsymbol{\mu_i}) \;=\; \mathbf{0}, \tag{3.13}$$

which are less linear than (3.12) due to the presence of a link function (e.g., the logit link for binary data), and the mean-variance relationship. Typically the correlation matrix C_i contains a vector $\boldsymbol{\alpha}$ of unknown parameters 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 (3.13) is asymptotically normally distributed with mean $\boldsymbol{\beta}$ and with covariance matrix

$$I_0^{-1}I_1I_0^{-1}, \tag{3.14}$$

where

$$I_0 = \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu_i}'}{\partial \boldsymbol{\beta}} V_i^{-1} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'}\right),$$

$$I_1 = \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu_i}'}{\partial \boldsymbol{\beta}} V_i^{-1} \mathrm{Var}(\boldsymbol{y_i}) V_i^{-1} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'}\right).$$

In practice, $\mathrm{Var}(\boldsymbol{y_i})$ in (3.14) is replaced by $(\boldsymbol{y_i} - \boldsymbol{\mu_i})(\boldsymbol{y_i} - \boldsymbol{\mu_i})'$, which is unbiased on the sole condition of correct mean specification. One also needs estimates of the nuisance parameters $\boldsymbol{\alpha}$. Liang and Zeger (1986) proposed moment-based estimates for the working correlation. To this end, define deviations

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

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. An overdispersion parameter could be included as well, but we have suppressed it for ease of exposition. 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: $\boldsymbol{\mu_i}$; (3) compute Pearson residuals e_{ij}; (4) compute estimates for $\boldsymbol{\alpha}$; (5) compute $C_i(\boldsymbol{\alpha})$; (6) compute $V_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = A_i^{1/2}(\boldsymbol{\beta}) C_i(\boldsymbol{\alpha}) A_i^{1/2}(\boldsymbol{\beta})$; (7) 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)–(7) are iterated until convergence.

Weighted Generalized Estimating Equations

As Liang and Zeger (1986) pointed out, GEE-based inferences are valid only under MCAR. Robins, Rotnitzky and Zhao (1995) proposed a class of *weighted estimating equations* to allow for MAR, extending GEE.

The idea is to weight each subject's contribution in the GEEs by the inverse probability that a subject drops out at the time he dropped out. This can be calculated,

for example, as

$$\begin{aligned}\nu_{id_i} \equiv P[D_i = d_i] &= \prod_{k=2}^{d_i-1} (1 - P[R_{ik} = 0 | R_{i2} = \ldots = R_{i,k-1} = 1]) \\ &\quad \times P[R_{id_i} = 0 | R_{i2} = \ldots = R_{i,d_i-1} = 1]^{I\{d_i \leq T\}}. \quad (3.15)\end{aligned}$$

Recall that we partitioned $\boldsymbol{Y}_i$ into the unobserved components $\boldsymbol{Y}_i^m$ and the observed components $\boldsymbol{Y}_i^o$. Similarly, we can make the exact same partition of $\boldsymbol{\mu_i}$ into $\boldsymbol{\mu_i^m}$ and $\boldsymbol{\mu_i^o}$. In the weighted GEE approach, which is proposed to reduce possible bias of $\widehat{\boldsymbol{\beta}}$, the score equations to be solved when taking into account the correlation structure are:

$$\begin{aligned}S(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \frac{1}{\nu_{id_i}} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'} \left(A_i^{1/2} C_i A_i^{1/2}\right)^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \;=\; \mathbf{0} \\ &= \sum_{i=1}^{N} \sum_{d=2}^{n+1} \frac{I(D_i = d)}{\nu_{id}} \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'}(d) \left(A_i^{1/2} C_i A_i^{1/2}\right)^{-1}(d) \; (\boldsymbol{y_i}(d) - \boldsymbol{\mu_i}(d)) \;=\; \mathbf{0},\end{aligned}$$

where $\boldsymbol{y_i}(d)$ and $\boldsymbol{\mu_i}(d)$ are the first $d-1$ elements of $\boldsymbol{y_i}$ and $\boldsymbol{\mu_i}$, respectively. We define $\frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'}(d)$ and $\left(A_i^{1/2} C_i A_i^{1/2}\right)^{-1}(d)$ analogously.

3.3.2 Random-effects Models

Unlike for correlated Gaussian outcomes, the parameters of the random-effects and population-averaged models for correlated binary data describe different types of effects of the covariates on the response probabilities (Neuhaus, 1992). Therefore, the choice between population-averaged and random-effects strategies should heavily depend on the scientific goals. Population-averaged models evaluate the success probability as a function of covariates only. With a subject-specific approach, the response is modeled as a function of covariates and parameters, specific to the subject. In such models, interpretation of fixed-effects parameters is conditional on a constant level of the random-effects parameter. Population-averaged comparisons, on the other hand, make no use of within cluster comparisons for cluster varying covariates and are therefore not useful to assess within-subject effects (Neuhaus, Kalbfleisch and Hauck, 1991). While several non-equivalent random-effects models exist, one of the most popular ones is the *generalized linear mixed model* (GLMM, Breslow and Clayton, 1993), implemented in the SAS procedure NLMIXED. We will focus on this one.

Generalized Linear Mixed Models

A general formulation of mixed-effects models is as follows. Assume that $\boldsymbol{Y_i}$ (possibly appropriately transformed) satisfies

$$\boldsymbol{Y_i}|\boldsymbol{b_i} \quad \sim \quad F_i(\boldsymbol{\theta}, \boldsymbol{b_i}), \tag{3.16}$$

i.e., conditional on $\boldsymbol{b_i}$, $\boldsymbol{Y_i}$ follows a pre-specified distribution F_i, possibly depending on covariates, and parameterized through a vector $\boldsymbol{\theta}$ of unknown parameters, common to all subjects. Further, $\boldsymbol{b_i}$ is a q-dimensional vector of subject-specific parameters, called random effects, assumed to follow a so-called mixing distribution G which may depend on a vector $\boldsymbol{\psi}$ of unknown parameters, i.e., $\boldsymbol{b_i} \sim G(\boldsymbol{\psi})$. The $\boldsymbol{b_i}$ reflect the between-unit heterogeneity in the population with respect to the distribution of $\boldsymbol{Y_i}$. In the presence of random effects, conditional independence is often assumed, under which the components Y_{ij} in $\boldsymbol{Y_i}$ are independent, conditional on $\boldsymbol{b_i}$. The distribution function F_i in (3.16) then becomes a product over the n_i independent elements in $\boldsymbol{Y_i}$.

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 over the random effects distribution $G(\boldsymbol{\psi})$. Let $f_i(\boldsymbol{y_i}|\boldsymbol{b_i})$ and $g(\boldsymbol{b_i})$ denote the density functions corresponding to the distributions F_i and G, respectively. We then have that the marginal density function of $\boldsymbol{Y_i}$ equals

$$f_i(\boldsymbol{y_i}) = \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i}) g(\boldsymbol{b_i}) d\boldsymbol{b_i}, \tag{3.17}$$

which depends on the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. Assuming independence of the units, estimates of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\psi}}$ can be obtained from maximizing the likelihood function built from (3.17), and inferences immediately follow from classical maximum likelihood theory.

It is important to realize that the random-effects distribution G is crucial in the calculation of the marginal model (3.17). One often assumes G to be of a specific parametric form, such as a (multivariate) normal. Depending on F_i and G, the integration in (3.17) may or may not be possible analytically. Proposed solutions are based on Taylor series expansions of $f_i(\boldsymbol{y_i}|\boldsymbol{b_i})$, or on numerical approximations of the integral, such as (adaptive) Gaussian quadrature.

Note that there is an important difference with respect to the interpretation of the fixed effects $\boldsymbol{\beta}$. Under the classical linear mixed model (Verbeke and Molenberghs, 2000), we have that $E(\boldsymbol{Y_i})$ equals $X_i\boldsymbol{\beta}$, such that the fixed effects have a subject-specific as well as a population-averaged interpretation. Under non-linear mixed models, however, this does no longer hold in general. The fixed effects now

only reflect the conditional effect of covariates, and the marginal effect is not easily obtained anymore as $E(\boldsymbol{Y_i})$ is given by

$$E(\boldsymbol{Y_i}) = \int \boldsymbol{y_i} \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i}) g(\boldsymbol{b_i}) d\boldsymbol{b_i} d\boldsymbol{y_i}.$$

However, in a biopharmaceutical context, one is often primarily interested in hypothesis testing and the random-effects framework can be used to this effect.

A general formulation of GLMM is as follows. Conditionally on random effects $\boldsymbol{b_i}$, it assumes that the elements Y_{ij} of $\boldsymbol{Y_i}$ are independent, with density function usually based on a classical exponential family formulation, i.e., with mean $E(Y_{ij}|\boldsymbol{b_i}) = a'(\eta_{ij}) = \mu_{ij}(\boldsymbol{b_i})$ and variance $\mathrm{Var}(Y_{ij}|\boldsymbol{b_i}) = \phi a''(\eta_{ij})$, and where, apart from a link function h (e.g., the logit link for binary data or the log link for counts), a linear regression model with parameters $\boldsymbol{\beta}$ and $\boldsymbol{b_i}$ is used for the mean, i.e., $h(\boldsymbol{\mu_i}(\boldsymbol{b_i})) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}$. Note that the linear mixed model is a special case, with identity link function. The random effects $\boldsymbol{b_i}$ are again assumed to be sampled from a (multivariate) normal distribution with mean $\mathbf{0}$ and covariance matrix D. Usually, the canonical link function is used, i.e., $h = a'^{-1}$, such that $\boldsymbol{\eta}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}$. When the link function is chosen to be of the logit form and the random effects are assumed to be normally distributed, the familiar logistic-linear GLMM follows.

3.3.3 Marginal *versus* Random-effects Models

It is useful to underscore the difference between both model families, as well as the nature of this difference. To see the nature of the difference, consider a binary outcome variable and assume a random-intercept logistic model with linear predictor $\mathrm{logit}[P(Y_{ij} = 1|t_{ij}, b_i)] = \beta_0 + b_i + \beta_1 t_{ij}$, where t_{ij} is the time covariate. The conditional means $E(Y_{ij}|b_i)$, as functions of t_{ij}, are given by

$$E(Y_{ij}|b_i) = \frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})}, \tag{3.18}$$

whereas the marginal average evolution is obtained from averaging over the random effects:

$$\begin{aligned} E(Y_{ij}) = E[E(Y_{ij}|b_i)] &= E\left[\frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})}\right] \\ &\neq \frac{\exp(\beta_0 + \beta_1 t_{ij})}{1 + \exp(\beta_0 + \beta_1 t_{ij})}. \end{aligned} \tag{3.19}$$

A graphical representation of both (3.18) and (3.19) is given in Figure 3.1. This implies that the interpretation of the parameters in both types of model is completely different.

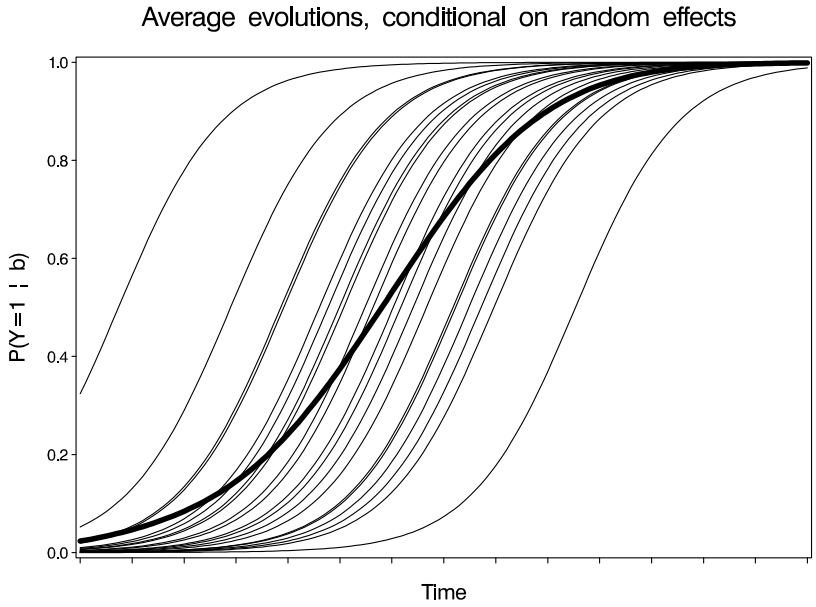


Figure 3.1: *Graphical representation of a random-intercept logistic curve, across a range of levels of the random intercept, together with the corresponding marginal curve.*

A schematic display is given in Figure 3.2. Depending on the model family (marginal or random-effects), one is led to either marginal or hierarchical inference. It is important to realize that in the general case the parameter $\boldsymbol{\beta}^M$ resulting from a marginal model are different from the parameter $\boldsymbol{\beta}^{RE}$ even when the latter is estimated using marginal inference. Some of the confusion surrounding this issue may result from the equality of these parameters in the very special linear mixed model case. When a random-effects model is considered, the marginal mean profile can be derived, but it will generally not produce a simple parametric form. In Figure 3.2 this is indicated by putting the corresponding parameter between quotes.

As an important example, consider our GLMM with logit link function, and where the only random effects are intercepts b_i. It can then be shown that the marginal mean $\boldsymbol{\mu_i} = E(Y_{ij})$ satisfies $h(\boldsymbol{\mu_i}) \approx X_i \boldsymbol{\beta}^M$ with

$$\frac{\boldsymbol{\beta}^{RE}}{\boldsymbol{\beta}^M} = \sqrt{c^2\sigma^2 \ + \ 1} \ > \ 1, \tag{3.20}$$

in which c equals $16\sqrt{3}/15\pi$. Hence, although the parameters $\boldsymbol{\beta}^{RE}$ in the generalized linear mixed model have no marginal interpretation, they do show a strong relation

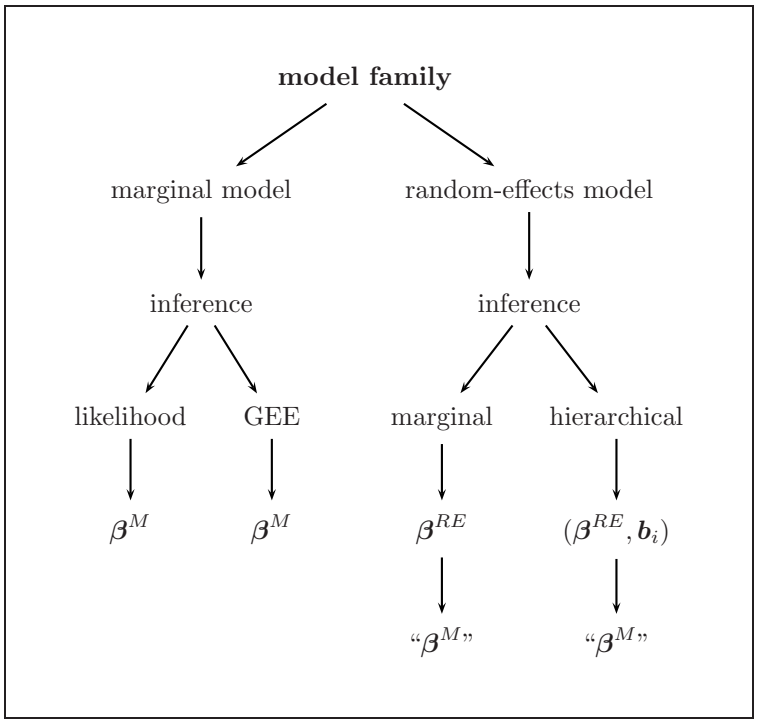


Figure 3.2: *Representation of model families and corresponding inference. A superscript 'M' stands for marginal, 'RE' for random effects. A parameter between quotes indicates that marginal functions but no direct marginal parameters are obtained.*

to their marginal counterparts. Note that, as a consequence of this relation, larger covariate effects are obtained under the random-effects model in comparison to the marginal model.

4

Likelihood-Based Ignorable Analyses in Sociology and Clinical Practice

In Chapter 3, methods to analyze incomplete (non-)Gaussian data were discussed. In this chapter we will apply the likelihood-based ignorable analyses on two data sets. Section 4.1 focuses on the analysis of data on marriage satisfaction, being bivariate longitudinal Gaussian. The results are contrasted with a complete case analysis. In Section 4.2, the non-Gaussian data on the Hamilton Depression Rating Scale are analyzed. Those results are compared with a complete case and with a last observation carried forward analysis. In both cases, the strength of the likelihood-based ignorable analysis (or direct-likelihood method) is shown. An account of these analyses can be found in Jansen *et al.* (2005c) and Jansen *et al.* (2005a).

4.1 Analysis of the Marriage Satisfaction Data

We analyze the data on marriage satisfaction, obtained from couples at two distinct moments in time (1990, 1995). The data are of a bivariate longitudinal type. Moreover, some couples provide incomplete records only, usually because the 1995

follow-up interview has not taken place. This incompleteness is a major challenge, since oftentimes, researchers opt for the complete case analysis. While simple to perform, there is a strong danger for bias, and additionally the statistical efficiency is reduced, leading to larger standard errors. We will compare the results of such a CC analysis (restricted to the couples with observations in 1990 as well as in 1995) with the direct-likelihood (DL) method (data as they are: for some couples both 1990 and 1995 are observed, for others only 1990), which uses all available data. Both concepts were introduced in Chapter 3. The analyses are performed using the SAS procedure MIXED, allowing the data to have repeated measures for each partner within a couple, and specifying a certain covariance structure between those repeated measures.

Only the couples with complete information for ALL covariates were used in the analyses, to overcome the problem of fitting submodels on different sets of data. First, we split the data into two separate sets of data, one only containing the information of the wives, the other with only the information of the husbands. Both data sets were analyzed separately, ignoring the possible correlation between husband's and wife's responses. Afterwards, the analyses were redone on the partners simultaneously, allowing us to also model the correlation between the partners' responses.

All analyses were repeated for the 3 outcomes of interest, namely marital satisfaction, open communication and negative communication.

4.1.1 Separate Analyses for Husband and Wife

To begin with, a separate analysis is performed for husbands and wives, to easily find gender specific covariates that might influence the outcomes of interest. Since our interest lies in a lot of covariate effects, and the model would become too extensive if all effects, together with their interactions with year of questionnaire, are included at once, we started the analyses with a single covariate selection, including the covariate itself, `year`, and the interaction effect, to allow the effect of this covariate to change over time. After selecting all significant effects from this single covariate approach, we combined them into a new model, from which we started removing the non-significant effects again. Finally, when all remaining covariates are (borderline) significant, the covariance matrix, which is kept unstructured until now, will be reduced. This matrix contains the covariances between the observations in 1990 and 1995. Therefore, it has to be a 2×2 matrix, for which only a few covariance structures are possible, namely unstructured (UN), banded main diagonal (UN(1)), compound symmetry (CS), and

simple (SIMPLE), respectively, as shown below:

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \quad \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \quad \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{pmatrix} \quad \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \tag{4.1}$$

Results from the analyses for marital satisfaction, open and negative communication are shown in Tables 4.1, 4.2 and 4.3, respectively. For the 3 outcomes, we see that all significant covariate effects in the CC model remain significant in the final model for the DL analysis, unless they were borderline significant in the CC case, then they turned out to be non-significant in the DL analysis. In the latter, sometimes even more covariate effects appeared to be significant, and the significance becomes stronger for almost all covariates.

Marital satisfaction increases between 1990 and 1995, both for males and females. For the husbands, education has a borderline significant decreasing effect on marital satisfaction when only considering the completers. For the wives, year of birth has a time dependent effect in the CC analysis, while a constant increasing effect in the DL analysis. There is a positive correlation between the 1990 and 1995 marital satisfaction of about 0.61, for males as well as for females.

Open communication increases between 1990 and 1995 for the males, but decreases for the females. For the husbands, in both CC and DL analysis, year of marriage has a borderline significant decreasing effect on open communication, and year of birth a time dependent effect over time. In the DL analysis, there is an additional interaction effect between income and time. For the wives, in both CC and DL analysis, income has a different effect on open communication in 1990 than in 1995, and year of birth a constant increasing effect. There is a positive correlation between the 1990 and 1995 open communication of about 0.41 for the males, and about 0.48 for the females.

Negative communication does not change between 1990 and 1995 for the males, and decreases for the females. For the husbands, there is only a borderline significant increasing effect of marital status on negative communication, in the CC analysis. For the wives, in both CC and DL analysis, the number of children and the year of marriage have a time dependent effect on negative communication. In the DL analysis, there is an additional interaction effect between income and time. There is a positive correlation between the 1990 and 1995 negative communication of about 0.61.

Table 4.1: *Marriage Satisfaction Data. Marital satisfaction, husband and wife separately. Remaining (borderline) significant effects. Year equals 1 for 1990, 0 for 1995. Marstat equals 1 if married, 0 if divorced.*

	HUSBAND						WIFE					
	CC (294 subj. / 588 obs.)			DL (603 subj. / 893 obs.)			CC (294 subj. / 588 obs.)			DL (608 subj. / 931 obs.)		
	est	s.e.	p-value	est	s.e.	p-value	est	s.e.	p-value	est	s.e.	p-value
intercept	6.1826	0.1060	<.0001	6.1982	0.0819	<.0001	6.6832	0.7317	<.0001	4.7177	0.4492	<.0001
year	-0.7785	0.0473	<.0001	-0.7870	0.0455	<.0001	-2.1303	0.6116	0.0006	-0.7350	0.0490	<.0001
education	-0.0400	0.0230	0.0831	-0.0587	0.0170	0.0006	—	—	—	—	—	—
birthyear	—	—	—	—	—	—	-0.0136	0.0146	0.3524	0.0238	0.0089	0.0078
birthyear × year	—	—	—	—	—	—	0.0270	0.0122	0.0273	—	—	—
covariance structure	UN			UN			UN			UN		
covariance matrix	$\begin{pmatrix} 0.6482 & 0.4652 \\ 0.4652 & 0.9405 \end{pmatrix}$			$\begin{pmatrix} 0.7204 & 0.5162 \\ 0.5162 & 0.9881 \end{pmatrix}$			$\begin{pmatrix} 0.8437 & 0.5873 \\ 0.5873 & 1.0980 \end{pmatrix}$			$\begin{pmatrix} 0.8739 & 0.6290 \\ 0.6290 & 1.2295 \end{pmatrix}$		
-2ℓ	1394.2			2195.8			1509.2			2478.4		

Table 4.2: *Marriage Satisfaction Data. Open communication, husband and wife separately. Remaining (borderline) significant effects. Year equals 1 for 1990, 0 for 1995. Marstat equals 1 if married, 0 if divorced.*

	HUSBAND						WIFE					
	CC (294 subj. / 588 obs.)			DL (603 subj. / 895 obs.)			CC (294 subj. / 588 obs.)			DL (608 subj. / 934 obs.)		
	est	s.e.	p-value	est	s.e.	p-value	est	s.e.	p-value	est	s.e.	p-value
intercept	8.0458	1.2288	<.0001	6.7478	1.0662	<.0001	3.1474	0.7571	<.0001	3.1285	0.5839	<.0001
year	-1.6311	0.6665	0.0150	-0.9555	0.7208	0.1860	1.0065	0.3522	0.0046	0.7704	0.2975	0.0100
maryear	-0.0371	0.0200	0.0640	-0.0327	0.0156	0.0364	—	—	—	—	—	—
income	—	—	—	0.0638	0.0470	0.1758	0.1172	0.0488	0.0170	0.1051	0.0412	0.0112
income × year	—	—	—	-0.1169	0.0514	0.0238	-0.1721	0.0524	0.0011	-0.1395	0.0444	0.0018
birthyear	-0.0035	0.0152	0.8184	0.0074	0.0136	0.5872	0.0347	0.0128	0.0069	0.0364	0.0097	0.0002
birthyear × year	0.0324	0.0140	0.0209	0.0340	0.0127	0.0078	—	—	—	—	—	—
covariance structure	CS			CS			CS			CS		
covariance matrix	$\begin{pmatrix} 1.2134 & 0.5000 \\ 0.5000 & 1.2134 \end{pmatrix}$			$\begin{pmatrix} 1.2334 & 0.4957 \\ 0.4957 & 1.2334 \end{pmatrix}$			$\begin{pmatrix} 1.1212 & 0.5271 \\ 0.5271 & 1.1212 \end{pmatrix}$			$\begin{pmatrix} 1.1591 & 0.5580 \\ 0.5580 & 1.1591 \end{pmatrix}$		
-2ℓ	1727.7			2674.3			1662.5			2478.4		

Table 4.3: *Marriage Satisfaction Data. Negative communication, husband and wife separately. Remaining (borderline) significant effects. Year equals 1 for 1990, 0 for 1995. Marstat equals 1 if married, 0 if divorced.*

	HUSBAND						WIFE					
	CC (294 subj. / 588 obs.)			DL (608 subj. / 932 obs.)			CC (294 subj. / 588 obs.)			DL (608 subj. / 932 obs.)		
	est	s.e.	p-value	est	s.e.	p-value	est	s.e.	p-value	est	s.e.	p-value
intercept	2.4711	0.1307	<.0001	2.7383	0.0401	<.0001	2.6771	1.2950	0.0396	2.4958	1.1744	0.0340
year	—	—	—	—	—	—	3.9312	1.1677	0.0009	3.6186	1.1047	0.0012
children	—	—	—	—	—	—	-0.0295	0.0467	0.5289	0.0202	0.0429	0.6385
children × year	—	—	—	—	—	—	-0.1113	0.0467	0.0179	-0.1080	0.0426	0.0116
maryear	—	—	—	—	—	—	0.0007	0.0176	0.9675	0.0019	0.0153	0.8990
maryear × year	—	—	—	—	—	—	-0.0502	0.0158	0.0016	-0.0402	0.0142	0.0051
marstat	0.2811	0.1443	0.0524	—	—	—	—	—	—	—	—	—
income	—	—	—	—	—	—	—	—	—	0.0051	0.0372	0.8911
income × year	—	—	—	—	—	—	—	—	—	-0.0651	0.0382	0.0897
covariance structure	CS			CS			CS			CS		
covariance matrix	$\begin{pmatrix} 1.1213 & 0.6889 \\ 0.6889 & 1.1213 \end{pmatrix}$			$\begin{pmatrix} 1.1032 & 0.6706 \\ 0.6706 & 1.1032 \end{pmatrix}$			$\begin{pmatrix} 1.0313 & 0.6180 \\ 0.6180 & 1.0313 \end{pmatrix}$			$\begin{pmatrix} 1.0870 & 0.6754 \\ 0.6754 & 1.0870 \end{pmatrix}$		
-2ℓ	1596.6			2585.6			1556.0			2702.0		

4.1.2 Joint Analysis for Both Partners

When analyzing both partners at the same time, in one model, we can still allow the covariates to have a different effect on the outcome for both partners. Therefore not only the interaction with `year`, but also with `partner` will be included into the model, and even higher order interactions of the form `covariate` × `year` × `partner` will be considered. Since this results in even more independent variables in the model, we will explore the variables of interest one by one, together with all relevant interaction terms. Afterwards, we will again remove the non-significant terms which remained after combining all significant effects from the single covariate selection, and reduce the covariance matrix. Since we are now not only interested in possible correlations between the years of questionnaire, but also in possible correlations between the partners, the covariance matrices in (4.1) should be extended to 4×4 matrices. These matrices are of the most general form possible and require in the unstructured case 10 parameters to be estimated. Therefore, we will attempt to reduce the covariance matrix to the Kronecker product of an unstructured matrix, modeling the covariance across the multivariate observations (partners) with an additional covariance matrix (unstructured, compound symmetry, or first-order autoregressive (AR(1))), modeling the covariance across time. The upper left value in the second matrix is constrained to be equal to 1, to identify the model. This Kronecker product also results in a 4×4 matrix, but with less parameters to estimate. Shown below are the UN@UN and UN@CS types, respectively (SAS uses @ for the Kronecker product). Since we only consider 2 measurement occasions, UN@AR(1) is identical to UN@CS.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \sigma_{34} \\ \sigma_{34} & \sigma_4^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1^2\sigma_{23} & \sigma_{12} & \sigma_{12}\sigma_{23} \\ \sigma_1^2\sigma_{23} & \sigma_1^2\sigma_4^2 & \sigma_{12}\sigma_{23} & \sigma_{12}\sigma_4^2 \\ \sigma_{12} & \sigma_{12}\sigma_{23} & \sigma_2^2 & \sigma_2^2\sigma_{23} \\ \sigma_{12}\sigma_{23} & \sigma_{12}\sigma_4^2 & \sigma_2^2\sigma_{23} & \sigma_2^2\sigma_4^2 \end{pmatrix}$$

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1^2\rho & \sigma_{12} & \sigma_{12}\rho \\ \sigma_1^2\rho & \sigma_1^2 & \sigma_{12}\rho & \sigma_{12} \\ \sigma_{12} & \sigma_{12}\rho & \sigma_2^2 & \sigma_2^2\rho \\ \sigma_{12}\rho & \sigma_{12} & \sigma_2^2\rho & \sigma_2^2 \end{pmatrix}$$

To use these structures in the SAS procedure MIXED, it is necessary to specify two distinct REPEATED effects, e.g. `partner` and `year`, both included in the CLASS statement, and in the same order as they appear in the Kronecker product.

Table 4.4: *Marriage Satisfaction Data. Marital satisfaction, both parents jointly. Remaining (borderline) significant effects. Parent equals 1 for husband, 0 for wife. Year equals 1 for 1990, 0 for 1995. Marstat equals 1 if married, 0 if divorced.*

	COUPLE					
	CC (294 subj. / 1176 obs.)			DL (647 subj. / 1963 obs.)		
	est	s.e.	p-value	est	s.e.	p-value
intercept	6.9913	0.6895	<.0001	5.2928	0.4264	<.0001
partner	-0.8399	0.7170	0.2424	1.0007	0.4472	0.0256
year	-2.2705	0.6202	0.0003	-0.7562	0.0401	<.0001
partner × year	1.4424	0.6449	0.0261	—	—	—
education	—	—	—	-0.0172	0.0204	0.4004
education × partner	—	—	—	-0.0461	0.0225	0.0400
birthyear	-0.0197	0.0137	0.1510	0.0138	0.0082	0.0922
birthyear × year	0.0298	0.0123	0.0159	—	—	—
birthyear × partner	0.0171	0.0145	0.2385	-0.0160	0.0087	0.0656
birthyear × partner × year	-0.0288	0.0130	0.0276	—	—	—
covariance structure	UN@UN			UN@UN		
covariance matrix	$\begin{pmatrix} 0.6423 & 0.2930 \\ 0.2930 & 0.7787 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.6308 \\ 0.6308 & 1.3692 \end{pmatrix}$			$\begin{pmatrix} 0.7025 & 0.3374 \\ 0.3374 & 0.8508 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.6368 \\ 0.6368 & 1.3952 \end{pmatrix}$		
-2ℓ	2802.1			4865.4		

Results from the analyses for marital satisfaction, open and negative communication are shown in Tables 4.4, 4.5 and 4.6, respectively.

In general, the same significant covariate effects appear as in the separate analyses for husband and wife. If a certain covariate was significant for both partners, then this effect is present in the same way as before, only the main effect, or also the interaction with `year`. If, on the other hand, the covariate was only significant for one of both partners, then now this effect, and a possible interaction with `year`, is present in the model, together with its interaction with `partner`. Some covariates disappear from the model, but this was restricted to those that were borderline significant when analyzing the partners separately. So we can say that both ways of analyses give comparable results. The simultaneous modeling of the partners, however, is easier to

Table 4.5: *Marriage Satisfaction Data. Open communication, both parents jointly. Remaining (borderline) significant effects. Parent equals 1 for husband, 0 for wife. Year equals 1 for 1990, 0 for 1995. Marstat equals 1 if married, 0 if divorced.*

	COUPLE					
	CC (294 subj. / 1176 obs.)			DL (619 subj. / 1862 obs.)		
	est	s.e.	p-value	est	s.e.	p-value
intercept	3.2290	0.6807	<.0001	4.4062	0.9370	<.0001
partner	1.5998	0.6856	0.0203	1.9449	1.0121	0.0551
year	0.7856	0.3054	0.0106	-0.4031	0.5614	0.4732
maryear	—	—	—	-0.0014	0.0136	0.9175
maryear × partner	—	—	—	-0.0321	0.0139	0.0211
income	0.0886	0.0406	0.0296	0.0698	0.0353	0.0481
income × year	-0.1360	0.0455	0.0029	-0.1145	0.0391	0.0034
birthyear	0.0368	0.0119	0.0021	0.0168	0.0107	0.1148
birthyear × year	—	—	—	0.0209	0.0096	0.0296
birthyear × partner	-0.0417	0.0139	0.0028	—	—	—
covariance structure	UN@CS			UN@CS		
covariance matrix	$\begin{pmatrix} 1.2199 & 0.3553 \\ 0.3553 & 1.0842 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.3983 \\ 0.3983 & 1 \end{pmatrix}$			$\begin{pmatrix} 1.2644 & 0.3659 \\ 0.3659 & 1.1468 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.4039 \\ 0.4039 & 1 \end{pmatrix}$		
-2ℓ	3341.2			5427.0		

conduct (no separation of the data for males and females necessary). It also allows us to have an idea about the baseline outcomes, i.e. the outcome without taking any covariate information into account, for husband and wife, both in 1990 and 1995, and to decide whether they significantly differ from each other. Finally, we get a handle on the association between them.

We can draw the following conclusions for the 3 outcomes of interest. The marital satisfaction for wives is lower than that of the husband in 1990, and both increase in 1995. In the DL analysis, this increase is similar for both partners, while in the CC analysis, the increase for wives is much higher than for husbands. The open communication is higher for the husband than for the wife. The evolution over time is the same for both partners, but is of a decreasing trend in the CC analysis, while of an increasing trend in the DL analysis. Finally, the conclusions for the last outcome of

Table 4.6: *Marriage Satisfaction Data. Negative communication, both parents jointly. Remaining (borderline) significant effects. Parent equals 1 for husband, 0 for wife. Year equals 1 for 1990, 0 for 1995. Marstat equals 1 if married, 0 if divorced.*

	COUPLE					
	CC (294 subj. / 1176 obs.)			DL (647 subj. / 1965 obs.)		
	est	s.e.	p-value	est	s.e.	p-value
intercept	2.8189	1.2762	0.0280	2.5777	1.0959	0.0190
partner	1.1281	1.4064	0.4231	0.9572	1.2173	0.4319
year	3.7888	1.2004	0.0018	2.6618	1.0481	0.0115
partner × year	-3.5734	1.2976	0.0063	-3.0844	1.1604	0.0082
children	-0.0101	0.0396	0.7988	0.0106	0.0369	0.7736
children × year	-0.0879	0.0413	0.0337	-0.0880	0.0376	0.0195
maryear	0.0004	0.0172	0.9825	0.0021	0.0149	0.8854
maryear × year	-0.0490	0.0163	0.0027	-0.0335	0.0142	0.0185
maryear × partner	-0.0202	0.0190	0.2884	-0.0161	0.0167	0.3351
maryear × partner × year	0.0495	0.0178	0.0056	0.0428	0.0159	0.0073
education	—	—	—	-0.0118	0.0220	0.5922
education × year	—	—	—	0.0567	0.0253	0.0252
marstat	-0.1951	0.1337	0.1492	—	—	—
marstat × partner	0.4431	0.1793	0.0140	—	—	—
covariance structure	UN@CS			UN@CS		
covariance matrix	$\begin{pmatrix} 1.0634 & 0.4269 \\ 0.4269 & 0.9943 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.5549 \\ 0.5549 & 1 \end{pmatrix}$			$\begin{pmatrix} 1.1741 & 0.4830 \\ 0.4830 & 1.0814 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.5705 \\ 0.5705 & 1 \end{pmatrix}$		
-2ℓ	3042.5			5337.7		

interest (negative communication) are much more complicated. In the CC analysis, negative communication decreases over time, but only marginally so for the husband, and much more for the wife, while in the DL analysis, negative communication increases a bit for males, and decreases for females. So in summary, we can say that negative communication approximately stays the same for the husband, and decreases considerably for the wife. In both ways of analysis, the wife has a higher value than the husband in 1990, but a lower in 1995.

4.1.3 Comparison With Literature Results

Importantly, we have to note that our conclusions differ somewhat from those often found in the literature. For example, a higher score for males on positive communication is somewhat counterintuitive (Canary, Stafford and Semic, 2002; Rubin, 1983; Wood, 1993; Weigel and Ballard-Reisch, 2001), as well as an increase in marital satisfaction between 1990 and 1995 (Vaillant and Vaillant, 1993; Van Laningham, Johnson and Amato, 2001). Also, in the literature there is evidence that individuals with a lower level of education or less material resources report lower marital satisfaction (Lewis and Spanier, 1979). Johnson, Amoloza and Booth (1992) suggest that marital satisfaction is lower in longer-term marriages than in those of short duration. Individuals who married at a young age report a lower marital quality (Holman, 2001). In a similar vein, individuals confronted with the divorce of their parents have a higher risk to develop less satisfying and unstable marriages themselves (Amato, 1996). Besides, the more children present, the higher the parental demands are and the more likely this results in negative marital outcomes (Lavee, Sharlin and Katz, 1996; Rogers and White, 1998).

Of course, this is at the heart of our proposed methodology, which not only takes the multivariate and longitudinal nature of the data into account, it also properly incorporates information from incomplete records. These results underscore that the traditional modes of analysis, in particular CC, may be too restrictive and even lead to false intuition, since for it to be valid the completers have to be a perfectly random subset of the entire set of data. In contrast, our analysis is valid under the much more flexible MAR assumption, which merely assumes that there is sufficient information on the missingness process in the (partially) observed data. Thus, it is better to use linear mixed models in combination with the assumption of MAR. Moreover, such analyses can be conducted routinely using standard statistical software such as the SAS procedure MIXED.

Since the covariates in our models are relatively highly correlated, one might be inclined to center these variables prior to including them into the models. However, duplication of our analyses with centered covariates led to the same results.

4.2 Analysis of the Hamilton Depression Rating Scale Data

Let us now analyze the two clinical trials on the Hamilton Depression Rating Scale, introduced in Section 2.2. The primary null hypothesis (zero difference between the

treatments and placebo in terms of proportion of the HAMD17 total score above the level of 7) will be tested using both marginal models (GEE and WGEE) and random-effects models (GLMM). According to the study protocol, the models will include the fixed categorical effects of treatment, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline score and baseline score-by-visit interaction. A random intercept will be included when considering the random-effects models. Analyses will be implemented using the SAS procedures GENMOD and NLMIXED.

Missing data will be handled in three different ways: (1) imputation using LOCF, (2) deletion of incomplete profiles, leading to a CC analysis, and (3) analyzing the data as they are, consistent with ignorability (for GLMM and WGEE). A fully longitudinal approach (View 1, see Section 3.1.2) is considered in Section 4.2.1. Section 4.2.2 compares the results of the marginal and random-effects models. Section 4.2.3 focuses on Views 2 (treatment effect at last planned occasion) and 3 (last measurement obtained), respectively.

4.2.1 View 1: Longitudinal Analysis

Marginal Models

First, let us consider the **GEE** approach. Within the SAS procedure GENMOD the exchangeable working correlation matrix is used. Parameter estimates, together with empirically-corrected as well as model-based standard errors and p-values, are given in Tables 4.7 and 4.8 for Studies 1 and 2, respectively.

In both studies, approximately the same conclusions are reached for LOCF, CC, and the analysis of the data as they are (subsequently abbreviated as MAR). There are no significant treatment main effects, while a placebo group change over time is observed in the first study only. Since there is no interaction between visit and treatment, the time evolution in the active groups is the same as in the placebo group. The main effect of baseline score is highly significant in the second study, as well as its interaction with visit from the sixth visit onwards. In contrast, the first study shows no effects of baseline at all. In line with expectation, the empirically-corrected standard errors are larger than the model-based ones, except for the effect of visit (in both studies), and some other effects in the second study.

On the other hand, **WGEE** is applied to perform an analysis that is correct under MAR, not only under MCAR as in ordinary GEE. This procedure is a bit more involved in terms of fitting the model to the data. We will outline the main steps. To compute the necessary weights, we first fit the dropout model, using a logistic regres-

Table 4.7: *Hamilton Depression Rating Scale Data. Study 1, GEE and WGEE: parameter estimates, standard errors (model-based, empirically-corrected) and p-values (model-based, empirically-corrected) for each approach.*

	CC (GEE)		LOCF (GEE)		MAR (GEE)		MAR (WGEE)	
	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value
intercept	2.11 (1.87; 1.09)	(0.260; 0.0528)	2.15 (1.90; 1.17)	(0.258; 0.0659)	2.04 (1.89; 1.10)	(0.281; 0.0638)	3.75 (1.33; 1.84)	(0.0048; 0.0413)
trt A1	-0.36 (1.19; 1.28)	(0.762; 0.779)	-0.39 (1.18; 1.32)	(0.743; 0.770)	-0.46 (1.19; 1.30)	(0.701; 0.727)	-0.12 (0.70; 1.31)	(0.870; 0.930)
trt B	-0.059 (1.26; 1.44)	(0.963; 0.967)	-0.16 (1.24; 1.43)	(0.899; 0.912)	-0.20 (1.25; 1.42)	(0.874; 0.889)	0.17 (0.75; 1.35)	(0.823; 0.901)
visit 5	-2.43 (1.86; 1.27)	(0.190; 0.0558)	-3.20 (1.82; 1.31)	(0.0791; 0.0145)	-3.16 (1.84; 1.30)	(0.0856; 0.0150)	-4.94 (1.29; 2.07)	(0.0001; 0.0168)
visit 6	-4.30 (1.84; 1.45)	(0.0194; 0.0031)	-4.80 (1.83; 1.48)	(0.0086; 0.0012)	-4.66 (1.85; 1.49)	(0.0119; 0.0018)	-5.98 (1.30; 2.20)	(<.0001; 0.0066)
visit 7	-3.84 (1.80; 1.29)	(0.0327; 0.0029)	-4.25 (1.78; 1.31)	(0.0171; 0.0012)	-4.05 (1.81; 1.31)	(0.0247; 0.0019)	-4.99 (1.28; 1.77)	(<.0001; 0.0049)
visit 8	-4.64 (1.80; 1.40)	(0.0100; 0.0009)	-5.15 (1.79; 1.42)	(0.0040; 0.0003)	-4.98 (1.82; 1.44)	(0.0060; 0.0005)	-5.90 (1.28; 2.13)	(<.0001; 0.0056)
visit 9	-4.28 (1.79; 1.44)	(0.0167; 0.0029)	-4.55 (1.78; 1.47)	(0.0103; 0.0019)	-4.33 (1.80; 1.50)	(0.0165; 0.0038)	-4.72 (1.28; 2.28)	(0.0002; 0.0389)
visit 10	-4.92 (1.79; 1.29)	(0.0059; 0.0001)	-4.71 (1.77; 1.31)	(0.0077; 0.0003)	-4.58 (1.80; 1.33)	(0.0108; 0.0006)	-4.87 (1.27; 1.99)	(0.0001; 0.0141)
visit 11	-4.74 (1.79; 1.41)	(0.0080; 0.0008)	-4.58 (1.77; 1.40)	(0.0095; 0.0011)	-4.75 (1.82; 1.46)	(0.0091; 0.0012)	-5.79 (1.28; 2.03)	(<.0001; 0.0043)
visit 5 * trt A1	-0.55 (1.19; 0.85)	(0.643; 0.517)	-0.11 (1.12; 0.96)	(0.921; 0.908)	-0.15 (1.14; 0.97)	(0.899; 0.881)	-0.84 (0.68; 1.00)	(0.218; 0.402)
visit 5 * trt B	-0.27 (1.26; 0.94)	(0.829; 0.773)	-0.012 (1.18; 1.04)	(0.992; 0.991)	-0.020 (1.20; 1.06)	(0.987; 0.985)	-1.37 (0.72; 1.19)	(0.0574; 0.248)
visit 6 * trt A1	0.29 (1.15; 1.31)	(0.801; 0.825)	0.51 (1.11; 1.31)	(0.644; 0.695)	0.54 (1.14; 1.36)	(0.634; 0.690)	0.27 (0.67; 1.40)	(0.691; 0.848)
visit 6 * trt B	0.15 (1.21; 1.40)	(0.902; 0.915)	0.26 (1.16; 1.36)	(0.821; 0.847)	0.38 (1.19; 1.42)	(0.751; 0.790)	-0.11 (0.72; 1.46)	(0.874; 0.938)
visit 7 * trt A1	0.47 (1.13; 1.15)	(0.680; 0.685)	0.48 (1.10; 1.17)	(0.662; 0.682)	0.52 (1.12; 1.22)	(0.646; 0.672)	-0.14 (0.67; 1.26)	(0.831; 0.910)
visit 7 * trt B	0.39 (1.19; 1.23)	(0.745; 0.753)	0.41 (1.15; 1.23)	(0.724; 0.742)	0.53 (1.18; 1.28)	(0.656; 0.681)	-0.14 (0.72; 1.29)	(0.844; 0.913)
visit 8 * trt A1	0.24 (1.13; 1.32)	(0.834; 0.857)	0.57 (1.10; 1.30)	(0.601; 0.660)	0.59 (1.13; 1.37)	(0.602; 0.670)	0.03 (0.67; 1.38)	(0.962; 0.982)
visit 8 * trt B	0.24 (1.19; 1.51)	(0.841; 0.874)	0.48 (1.15; 1.45)	(0.677; 0.741)	0.57 (1.18; 1.54)	(0.627; 0.709)	-0.28 (0.72; 1.49)	(0.692; 0.849)
visit 9 * trt A1	0.37 (1.13; 1.34)	(0.742; 0.783)	0.47 (1.09; 1.32)	(0.668; 0.722)	0.46 (1.12; 1.40)	(0.684; 0.744)	-0.12 (0.67; 1.42)	(0.859; 0.934)
visit 9 * trt B	0.21 (1.19; 1.48)	(0.861; 0.888)	0.23 (1.15; 1.42)	(0.842; 0.873)	0.26 (1.18; 1.51)	(0.825; 0.863)	-0.51 (0.72; 1.52)	(0.478; 0.737)
visit 10 * trt A1	0.11 (1.12; 1.19)	(0.922; 0.926)	0.30 (1.09; 1.21)	(0.783; 0.804)	0.31 (1.13; 1.29)	(0.783; 0.810)	-0.49 (0.67; 1.35)	(0.462; 0.715)
visit 10 * trt B	0.28 (1.18; 1.32)	(0.811; 0.830)	0.32 (1.15; 1.31)	(0.782; 0.810)	0.43 (1.18; 1.39)	(0.718; 0.759)	-0.50 (0.72; 1.42)	(0.486; 0.724)
visit 11 * trt A1	-0.70 (1.13; 1.17)	(0.535; 0.549)	-0.27 (1.09; 1.21)	(0.801; 0.820)	-0.53 (1.14; 1.27)	(0.639; 0.674)	-1.31 (0.67; 1.35)	(0.0528; 0.334)
visit 11 * trt B	-0.052 (1.18; 1.35)	(0.965; 0.969)	0.071 (1.14; 1.34)	(0.951; 0.958)	0.089 (1.19; 1.43)	(0.941; 0.950)	-0.76 (0.72; 1.43)	(0.291; 0.593)
baseline	0.054 (0.096; 0.10)	(0.575; 0.598)	0.077 (0.097; 0.11)	(0.427; 0.472)	0.085 (0.097; 0.10)	(0.380; 0.407)	0.039 (0.069; 0.15)	(0.579; 0.795)
baseline * visit 5	0.10 (0.096; 0.076)	(0.284; 0.176)	0.12 (0.095; 0.082)	(0.193; 0.133)	0.12 (0.096; 0.082)	(0.210; 0.142)	0.24 (0.068; 0.13)	(0.0006; 0.073)
baseline * visit 6	0.17 (0.096; 0.11)	(0.0830; 0.129)	0.18 (0.095; 0.11)	(0.0537; 0.104)	0.17 (0.097; 0.12)	(0.0815; 0.149)	0.24 (0.070; 0.17)	(0.0005; 0.145)
baseline * visit 7	0.10 (0.092; 0.089)	(0.268; 0.249)	0.12 (0.092; 0.093)	(0.183; 0.188)	0.10 (0.093; 0.095)	(0.281; 0.287)	0.16 (0.067; 0.13)	(0.0159; 0.221)
baseline * visit 8	0.15 (0.093; 0.11)	(0.105; 0.179)	0.16 (0.092; 0.12)	(0.0749; 0.154)	0.14 (0.094; 0.12)	(0.127; 0.234)	0.21 (0.068; 0.17)	(0.0022; 0.217)
baseline * visit 9	0.11 (0.092; 0.12)	(0.213; 0.325)	0.13 (0.091; 0.12)	(0.158; 0.275)	0.10 (0.093; 0.13)	(0.263; 0.404)	0.13 (0.067; 0.18)	(0.0575; 0.476)
baseline * visit 10	0.13 (0.091; 0.096)	(0.169; 0.189)	0.11 (0.090; 0.10)	(0.228; 0.280)	0.083 (0.093; 0.11)	(0.371; 0.434)	0.10 (0.067; 0.15)	(0.122; 0.504)
baseline * visit 11	0.13 (0.092; 0.099)	(0.163; 0.197)	0.11 (0.090; 0.10)	(0.222; 0.288)	0.11 (0.094; 0.11)	(0.253; 0.318)	0.16 (0.067; 0.15)	(0.0181; 0.293)
ℓ	-423.1		-533.0		-465.4		-2042.0	

Table 4.8: *Hamilton Depression Rating Scale Data. Study 2, GEE and WGEE: parameter estimates, standard errors (model-based, empirically-corrected) and p-values (model-based, empirically-corrected) for each approach.*

	CC (GEE)		LOCF (GEE)		MAR (GEE)		MAR (WGEE)	
	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value
intercept	-1.39 (0.84;0.85)	(0.0978;0.102)	-1.25 (0.73;0.74)	(0.0848;0.0905)	-1.36 (0.74;0.73)	(0.0641;0.0622)	-1.00 (0.41;0.76)	(0.0157;0.190)
trt A1	-0.20 (0.60;0.63)	(0.742;0.754)	-0.26 (0.53;0.59)	(0.628;0.661)	-0.33 (0.54;0.58)	(0.544;0.573)	-0.72 (0.33;0.66)	(0.0311;0.275)
trt A2	0.50 (0.67;0.70)	(0.459;0.476)	0.20 (0.57;0.61)	(0.722;0.737)	0.20 (0.58;0.62)	(0.730;0.744)	-0.46 (0.34;0.70)	(0.181;0.513)
trt B	0.28 (0.62;0.66)	(0.656;0.675)	0.44 (0.57;0.61)	(0.438;0.467)	0.41 (0.58;0.61)	(0.479;0.504)	0.19 (0.35;0.65)	(0.578;0.766)
visit 5	-1.04 (0.87;0.74)	(0.230;0.161)	-0.50 (0.70;0.67)	(0.475;0.456)	-0.53 (0.75;0.70)	(0.480;0.454)	-0.24 (0.41;0.92)	(0.555;0.791)
visit 6	-0.13 (0.82;0.75)	(0.871;0.859)	-0.080 (0.69;0.66)	(0.907;0.905)	-0.078 (0.73;0.68)	(0.915;0.909)	-0.037 (0.41;0.72)	(0.929;0.959)
visit 7	-0.46 (0.83;0.91)	(0.581;0.615)	-0.32 (0.69;0.78)	(0.644;0.684)	-0.41 (0.76;0.87)	(0.588;0.633)	-0.45 (0.44;0.88)	(0.306;0.608)
visit 8	-0.35 (0.83;0.93)	(0.673;0.707)	-0.20 (0.69;0.79)	(0.770;0.800)	-0.19 (0.77;0.88)	(0.807;0.830)	-0.26 (0.45;0.92)	(0.557;0.776)
visit 5 * trt A1	0.10 (0.59;0.65)	(0.866;0.880)	0.086 (0.50;0.60)	(0.862;0.885)	0.19 (0.53;0.62)	(0.715;0.755)	0.36 (0.32;0.78)	(0.260;0.638)
visit 5 * trt A2	-0.49 (0.65;0.64)	(0.454;0.450)	-0.11 (0.54;0.56)	(0.833;0.840)	-0.026 (0.57;0.59)	(0.964;0.965)	0.33 (0.34;0.75)	(0.325;0.662)
visit 5 * trt B	0.24 (0.61;0.55)	(0.689;0.659)	-0.044 (0.53;0.52)	(0.934;0.932)	0.047 (0.56;0.54)	(0.934;0.931)	0.012 (0.33;0.71)	(0.971;0.986)
visit 6 * trt A1	0.14 (0.57;0.62)	(0.803;0.819)	0.30 (0.49;0.58)	(0.536;0.602)	0.46 (0.52;0.60)	(0.373;0.440)	0.99 (0.32;0.71)	(0.0021;0.164)
visit 6 * trt A2	-0.40 (0.63;0.69)	(0.521;0.560)	0.040 (0.52;0.59)	(0.939;0.946)	0.10 (0.56;0.61)	(0.860;0.873)	0.72 (0.33;0.68)	(0.0318;0.295)
visit 6 * trt B	0.39 (0.59;0.66)	(0.505;0.555)	0.17 (0.52;0.60)	(0.744;0.775)	0.27 (0.55;0.62)	(0.630;0.670)	0.38 (0.33;0.70)	(0.251;0.590)
visit 7 * trt A1	-0.079 (0.57;0.66)	(0.889;0.905)	-0.13 (0.49;0.60)	(0.795;0.834)	-0.15 (0.53;0.64)	(0.778;0.816)	0.23 (0.33;0.72)	(0.482;0.747)
visit 7 * trt A2	-0.011 (0.63;0.70)	(0.986;0.987)	0.093 (0.52;0.62)	(0.858;0.879)	0.12 (0.57;0.66)	(0.839;0.860)	0.66 (0.35;0.72)	(0.0595;0.360)
visit 7 * trt B	0.63 (0.59;0.69)	(0.280;0.356)	0.19 (0.52;0.62)	(0.719;0.764)	0.33 (0.56;0.68)	(0.563;0.630)	0.50 (0.34;0.74)	(0.144;0.496)
visit 8 * trt A1	-0.54 (0.57;0.68)	(0.342;0.424)	-0.41 (0.49;0.61)	(0.394;0.497)	-0.43 (0.54;0.66)	(0.429;0.517)	-0.11 (0.34;0.75)	(0.737;0.880)
visit 8 * trt A2	-0.40 (0.63;0.71)	(0.529;0.578)	-0.20 (0.52;0.62)	(0.702;0.747)	-0.10 (0.58;0.66)	(0.858;0.876)	0.44 (0.35;0.73)	(0.211;0.549)
visit 8 * trt B	0.27 (0.58;0.65)	(0.644;0.677)	-0.073 (0.52;0.59)	(0.889;0.903)	0.095 (0.57;0.64)	(0.868;0.882)	0.34 (0.35;0.71)	(0.322;0.629)
baseline	0.21 (0.051;0.049)	(<.0001;<.0001)	0.23 (0.044;0.042)	(<.0001;<.0001)	0.23 (0.045;0.042)	(<.0001;<.0001)	0.26 (0.025;0.046)	(<.0001;<.0001)
baseline * visit 5	-0.0019 (0.052;0.046)	(0.971;0.967)	-0.043 (0.042;0.041)	(0.309;0.297)	-0.052 (0.045;0.042)	(0.244;0.219)	-0.10 (0.025;0.052)	(<.0001;0.0610)
baseline * visit 6	-0.11 (0.049;0.044)	(0.0256;0.0127)	-0.12 (0.041;0.039)	(0.0047;0.0025)	-0.13 (0.044;0.039)	(0.0026;0.0007)	-0.16 (0.025;0.042)	(<.0001;<.0001)
baseline * visit 7	-0.11 (0.050;0.049)	(0.0337;0.0327)	-0.10 (0.041;0.042)	(0.0132;0.0149)	-0.11 (0.045;0.045)	(0.0124;0.0124)	-0.14 (0.026;0.047)	(<.0001;0.0033)
baseline * visit 8	-0.11 (0.050;0.050)	(0.0246;0.0242)	-0.11 (0.041;0.042)	(0.0082;0.0094)	-0.14 (0.046;0.046)	(0.0031;0.0030)	-0.16 (0.027;0.049)	(<.0001;0.0008)
ℓ	-591.2		-848.2		-707.1		-2485.5	

sion for the probability that dropout occurs at a given time. The response value at the previous occasion and treatment are included as covariates. Next, the predicted probabilities of dropout are translated into weights, defined at the individual measurement level, as in (3.15). After these preparations, we merely need to include the weights by means of the `scwgt` statement within the GENMOD procedure. Together with the use of the `repeated` statement, WGEE follows. Also here, we use the exchangeable working correlation matrix. Parameter estimates, together with empirically-corrected as well as model-based standard errors and p-values, are given in the last columns of Tables 4.7 and 4.8 for Studies 1 and 2, respectively.

The results under WGEE are similar but not identical to the ones in the classical GEE setting. In some cases, the evidence, while not changing from significant to nonsignificant or vice versa, change in strength. This is the case, for example, for the baseline score-by-visit interaction for the second study.

Random-effects Models

To fit generalized linear mixed models, we use the SAS procedure NLMIXED, which allows fitting a wide class of linear, generalized linear, and non-linear mixed models. It relies on numerical integration. Not only different integral approximations are available, the principal ones being (non-)adaptive Gaussian quadrature, it also includes a number of optimization algorithms. The difference between non-adaptive and adaptive Gaussian quadrature is that for the first procedure the quadrature points are centered at zero for each of the random-effects and the current random-effects covariance matrix is used as the scale matrix, while for the latter the quadrature points will be appropriately centered and scaled, such that more quadrature points lie in the region of interest (see Figure 4.1). We will use both adaptive and non-adaptive quadrature, with several choices for the number of quadrature points, to check the stability of the results over a variety of choices for these numerical integrations.

Precisely, we initiate the model fitting using non-adaptive Gaussian quadrature, together with the quasi-Newton optimization algorithm. The number of quadrature points is left to be determined by the procedure, and all starting values are set equal to 0.5 (step 0). Using the resulting parameter estimates, we keep these choices but hold the number of quadrature points fixed at the values 2, 3, 5, 10, 20 and 50 (step 1). Subsequently, we switch to adaptive Gaussian quadrature (step 2). Finally, the quasi-Newton optimization is replaced by the Newton-Raphson optimization (step 3). Results for Study 1 are shown in Tables 4.9 to 4.11. While the differences between these choices are purely numerical, we do notice differences between the results, illustrating

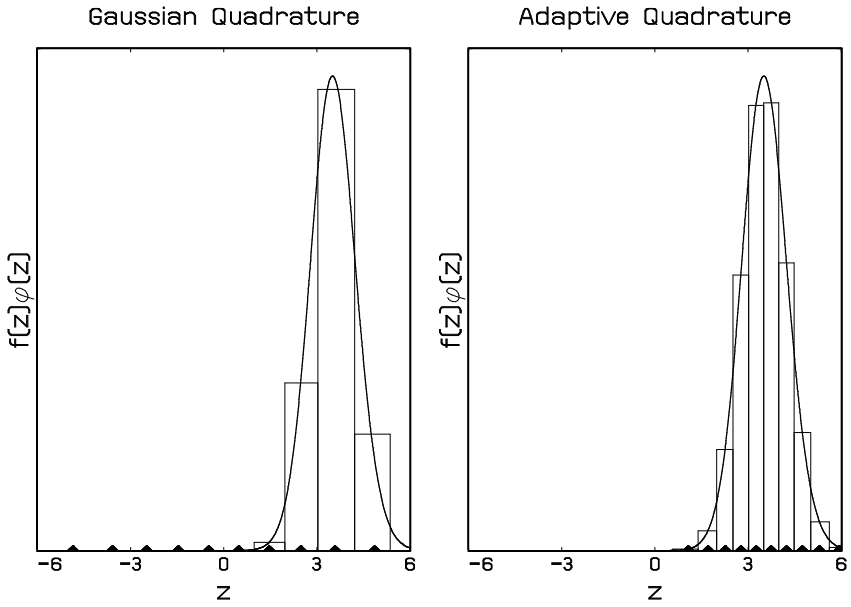


Figure 4.1: *Graphical illustration of non-adaptive Gaussian (left window) and adaptive Gaussian (right window) quadrature of order* $Q = 10$. *The black triangles indicate the position of the quadrature points, while the rectangles indicate the contribution of each point to the integral.*

that a numerical sensitivity analysis matters. The parameter estimates tend to stabilize with increasing number of quadrature points. However, non-adaptive Gaussian quadrature needs obviously more quadrature points than adaptive Gaussian quadrature.

Focusing on the results for 50 quadrature points, we observe that the parameter estimates for step 1 and step 2 are the same. On the other hand, parameter estimates for step 3 are different (order of 10^{-3}, visible in p-values). In spite of the differences in parameter estimates, is the noteworthy fact that the likelihood is the same in all steps, due to a flat likelihood. This was confirmed by running all steps again, but now using the parameter estimates of step 3 as starting values, at which point the parameter estimates all coincide. Thus, it may happen that the optimization routine has only seemingly converged.

The final conclusions reached here (see Tables 4.12 and 4.13), are very similar to the ones reached with the marginal models, at least in terms of hypothesis testing. An important exception is that there are now differential conclusions for the baseline

Table 4.9: *Hamilton Depression Rating Scale Data. Study 1, MAR: GLMM using proc NLMIXED with non-adaptive Gaussian quadrature, quasi-Newton optimization and different number of quadrature points.*

	$Q = 2$		$Q = 3$		$Q = 5$		$Q = 10$		$Q = 20$		$Q = 50$	
	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value
intercept	3.30 (1.94)	0.0904	3.50 (2.35)	0.137	3.03 (2.54)	0.234	3.10 (2.56)	0.227	3.37 (2.67)	0.209	3.39 (2.65)	0.203
trt A1	-0.90 (1.29)	0.485	-0.40 (1.27)	0.755	-0.59 (1.44)	0.683	-0.88 (1.41)	0.533	-0.56 (1.53)	0.713	-0.58 (1.51)	0.702
trt B	-0.39 (1.35)	0.775	0.28 (1.32)	0.835	-0.57 (1.47)	0.697	-0.17 (1.47)	0.910	-0.20 (1.60)	0.901	-0.16 (1.59)	0.920
visit 5	-3.89 (2.35)	0.0990	-4.80 (2.73)	0.0800	-4.90 (2.84)	0.0862	-4.72 (2.84)	0.0987	-4.65 (2.85)	0.104	-4.65 (2.85)	0.105
visit 6	-6.06 (2.46)	0.0147	-7.40 (2.75)	0.0077	-7.56 (2.83)	0.0083	-7.29 (2.97)	0.0150	-7.26 (2.94)	0.0145	-7.26 (2.94)	0.0146
visit 7	-5.77 (2.37)	0.0157	-6.64 (2.64)	0.0128	-6.66 (2.73)	0.0159	-6.53 (2.87)	0.0240	-6.52 (2.85)	0.0232	-6.52 (2.85)	0.0232
visit 8	-7.56 (2.42)	0.0022	-8.44 (2.72)	0.0023	-8.46 (2.82)	0.0031	-8.52 (2.95)	0.0044	-8.46 (2.93)	0.0044	-8.46 (2.93)	0.0044
visit 9	-6.58 (2.39)	0.0065	-7.30 (2.65)	0.0065	-7.30 (2.75)	0.0086	-7.32 (2.89)	0.0123	-7.29 (2.87)	0.0120	-7.29 (2.87)	0.0119
visit 10	-7.52 (2.39)	0.0020	-7.99 (2.65)	0.0030	-7.89 (2.79)	0.0052	-8.21 (2.89)	0.0050	-8.19 (2.89)	0.0052	-8.19 (2.89)	0.0052
visit 11	-7.43 (2.47)	0.0030	-8.35 (2.70)	0.0024	-8.13 (2.84)	0.0048	-8.47 (2.96)	0.0047	-8.48 (2.97)	0.0048	-8.48 (2.97)	0.0048
visit 5 * trt A1	0.43 (1.58)	0.787	0.26 (1.58)	0.871	0.35 (1.65)	0.833	0.30 (1.70)	0.858	0.26 (1.70)	0.879	0.27 (1.70)	0.874
visit 5 * trt B	-0.072 (1.60)	0.964	-0.23 (1.59)	0.887	-0.055 (1.67)	0.974	-0.20 (1.69)	0.908	-0.21 (1.71)	0.900	-0.21 (1.71)	0.902
visit 6 * trt A1	0.93 (1.55)	0.552	0.75 (1.54)	0.626	0.73 (1.64)	0.657	0.72 (1.68)	0.669	0.62 (1.67)	0.711	0.63 (1.68)	0.708
visit 6 * trt B	0.82 (1.54)	0.598	0.63 (1.53)	0.681	0.61 (1.63)	0.708	0.56 (1.66)	0.736	0.50 (1.66)	0.766	0.50 (1.67)	0.764
visit 7 * trt A1	0.89 (1.54)	0.564	0.53 (1.52)	0.727	0.64 (1.66)	0.698	0.50 (1.68)	0.768	0.50 (1.68)	0.766	0.50 (1.69)	0.767
visit 7 * trt B	0.48 (1.55)	0.757	0.17 (1.53)	0.912	0.18 (1.63)	0.912	0.12 (1.67)	0.942	0.065 (1.66)	0.969	0.070 (1.67)	0.967
visit 8 * trt A1	-0.42 (1.54)	0.788	-0.72 (1.51)	0.635	-0.63 (1.62)	0.700	-0.66 (1.61)	0.681	-0.63 (1.62)	0.699	-0.63 (1.62)	0.698
visit 8 * trt B	0.36 (1.58)	0.819	-0.10 (1.54)	0.947	0.024 (1.68)	0.989	-0.18 (1.70)	0.915	-0.15 (1.70)	0.929	-0.15 (1.71)	0.930
visit 9 * trt A1	0.49 (1.49)	0.742	0.35 (1.45)	0.808	0.13 (1.58)	0.933	0.23 (1.58)	0.886	0.30 (1.58)	0.849	0.30 (1.58)	0.850
visit 9 * trt B	0.44 (1.52)	0.774	0.27 (1.50)	0.855	0.29 (1.61)	0.855	0.33 (1.61)	0.840	0.39 (1.61)	0.810	0.38 (1.61)	0.814
visit 10 * trt A1	0.41 (1.49)	0.784	0.31 (1.45)	0.833	0.049 (1.59)	0.976	0.13 (1.58)	0.933	0.23 (1.58)	0.885	0.23 (1.59)	0.885
visit 10 * trt B	0.62 (1.49)	0.680	0.55 (1.46)	0.707	0.30 (1.60)	0.852	0.42 (1.58)	0.793	0.48 (1.59)	0.761	0.48 (1.59)	0.764
visit 11 * trt A1	-1.15 (1.52)	0.451	-1.26 (1.48)	0.396	-1.69 (1.66)	0.310	-1.60 (1.62)	0.324	-1.52 (1.64)	0.354	-1.53 (1.64)	0.354
visit 11 * trt B	0.25 (1.49)	0.868	0.24 (1.45)	0.871	-0.070 (1.62)	0.966	0.011 (1.59)	0.994	0.096 (1.60)	0.952	0.100 (1.61)	0.951
baseline	0.081 (0.10)	0.414	0.089 (0.12)	0.447	0.17 (0.12)	0.158	0.16 (0.13)	0.199	0.14 (0.13)	0.298	0.14 (0.13)	0.295
baseline * visit 5	0.16 (0.12)	0.210	0.21 (0.14)	0.143	0.20 (0.14)	0.160	0.19 (0.14)	0.180	0.19 (0.14)	0.192	0.19 (0.14)	0.193
baseline * visit 6	0.23 (0.13)	0.0810	0.30 (0.14)	0.0362	0.30 (0.14)	0.0344	0.29 (0.15)	0.0568	0.28 (0.15)	0.0588	0.28 (0.15)	0.0597
baseline * visit 7	0.16 (0.12)	0.190	0.20 (0.13)	0.136	0.20 (0.13)	0.138	0.19 (0.14)	0.183	0.19 (0.14)	0.189	0.19 (0.14)	0.190
baseline * visit 8	0.24 (0.13)	0.0548	0.28 (0.14)	0.0415	0.28 (0.14)	0.0405	0.28 (0.15)	0.0559	0.28 (0.14)	0.0584	0.28 (0.15)	0.0591
baseline * visit 9	0.19 (0.12)	0.132	0.22 (0.13)	0.108	0.21 (0.13)	0.110	0.21 (0.14)	0.139	0.21 (0.14)	0.144	0.21 (0.14)	0.145
baseline * visit 10	0.17 (0.12)	0.152	0.18 (0.13)	0.162	0.17 (0.13)	0.196	0.19 (0.14)	0.182	0.19 (0.14)	0.188	0.19 (0.14)	0.191
baseline * visit 11	0.19 (0.12)	0.121	0.23 (0.13)	0.0842	0.21 (0.14)	0.116	0.24 (0.14)	0.105	0.23 (0.14)	0.109	0.23 (0.15)	0.111
σ	1.83 (0.15)	$<.0001$	1.99 (0.17)	$<.0001$	2.35 (0.21)	$<.0001$	2.48 (0.26)	$<.0001$	2.49 (0.30)	$<.0001$	2.52 (0.30)	$<.0001$
-2ℓ	760.0		730.0		717.0		719.5		721.3		720.9	

Table 4.10: *Hamilton Depression Rating Scale Data. Study 1, MAR: GLMM using proc NLMIXED with adaptive Gaussian quadrature, quasi-Newton optimization and different number of quadrature points.*

	$Q = 2$		$Q = 3$		$Q = 5$		$Q = 10$		$Q = 20$		$Q = 50$	
	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value
intercept	3.38 (2.58)	0.192	3.40 (2.60)	0.193	3.39 (2.63)	0.199	3.39 (2.65)	0.203	3.39 (2.65)	0.203	3.39 (2.65)	0.203
trt A1	-0.59 (1.47)	0.688	-0.56 (1.48)	0.704	-0.58 (1.49)	0.698	-0.58 (1.51)	0.701	-0.58 (1.51)	0.702	-0.58 (1.51)	0.702
trt B	-0.15 (1.54)	0.921	-0.17 (1.56)	0.914	-0.16 (1.57)	0.920	-0.16 (1.59)	0.920	-0.16 (1.59)	0.920	-0.16 (1.59)	0.920
visit 5	-4.65 (2.80)	0.0987	-4.66 (2.81)	0.0994	-4.65 (2.84)	0.103	-4.65 (2.85)	0.105	-4.65 (2.85)	0.105	-4.65 (2.85)	0.105
visit 6	-7.26 (2.89)	0.0131	-7.26 (2.91)	0.0134	-7.26 (2.93)	0.0141	-7.26 (2.94)	0.0146	-7.26 (2.94)	0.0146	-7.26 (2.94)	0.0146
visit 7	-6.52 (2.79)	0.0208	-6.52 (2.81)	0.0215	-6.52 (2.83)	0.0225	-6.52 (2.85)	0.0232	-6.52 (2.85)	0.0232	-6.52 (2.85)	0.0232
visit 8	-8.46 (2.87)	0.0037	-8.46 (2.89)	0.0039	-8.46 (2.91)	0.0041	-8.46 (2.93)	0.0044	-8.46 (2.93)	0.0044	-8.46 (2.93)	0.0043
visit 9	-7.29 (2.81)	0.0104	-7.29 (2.83)	0.0109	-7.29 (2.85)	0.0114	-7.29 (2.87)	0.0119	-7.29 (2.87)	0.0119	-7.29 (2.87)	0.0119
visit 10	-8.19 (2.84)	0.0044	-8.18 (2.85)	0.0047	-8.19 (2.88)	0.0050	-8.19 (2.89)	0.0052	-8.19 (2.89)	0.0052	-8.19 (2.89)	0.0052
visit 11	-8.48 (2.91)	0.0041	-8.47 (2.93)	0.0043	-8.48 (2.95)	0.0046	-8.48 (2.97)	0.0048	-8.48 (2.97)	0.0048	-8.48 (2.97)	0.0048
visit 5 * trt A1	0.27 (1.67)	0.872	0.27 (1.68)	0.873	0.27 (1.69)	0.874	0.27 (1.70)	0.874	0.27 (1.70)	0.874	0.27 (1.70)	0.874
visit 5 * trt B	-0.21 (1.68)	0.900	-0.21 (1.69)	0.900	-0.21 (1.70)	0.902	-0.21 (1.71)	0.902	-0.21 (1.71)	0.902	-0.21 (1.71)	0.902
visit 6 * trt A1	0.63 (1.65)	0.702	0.63 (1.66)	0.704	0.63 (1.67)	0.706	0.63 (1.68)	0.708	0.63 (1.68)	0.708	0.63 (1.68)	0.708
visit 6 * trt B	0.50 (1.64)	0.759	0.50 (1.65)	0.760	0.50 (1.66)	0.762	0.50 (1.67)	0.764	0.50 (1.67)	0.764	0.50 (1.67)	0.764
visit 7 * trt A1	0.50 (1.65)	0.762	0.50 (1.67)	0.766	0.50 (1.68)	0.766	0.50 (1.69)	0.767	0.50 (1.69)	0.767	0.50 (1.69)	0.767
visit 7 * trt B	0.070 (1.64)	0.966	0.070 (1.65)	0.966	0.070 (1.66)	0.966	0.070 (1.67)	0.967	0.070 (1.67)	0.967	0.070 (1.67)	0.967
visit 8 * trt A1	-0.63 (1.60)	0.694	-0.63 (1.61)	0.695	-0.63 (1.61)	0.697	-0.63 (1.62)	0.698	-0.63 (1.62)	0.698	-0.63 (1.62)	0.698
visit 8 * trt B	-0.15 (1.68)	0.931	-0.16 (1.69)	0.927	-0.15 (1.70)	0.930	-0.15 (1.71)	0.930	-0.15 (1.71)	0.930	-0.15 (1.71)	0.930
visit 9 * trt A1	0.30 (1.55)	0.847	0.30 (1.57)	0.850	0.30 (1.57)	0.850	0.30 (1.58)	0.850	0.30 (1.58)	0.850	0.30 (1.58)	0.850
visit 9 * trt B	0.38 (1.58)	0.810	0.38 (1.60)	0.812	0.38 (1.60)	0.813	0.38 (1.61)	0.814	0.38 (1.61)	0.814	0.38 (1.61)	0.814
visit 10 * trt A1	0.23 (1.56)	0.883	0.23 (1.57)	0.884	0.23 (1.58)	0.885	0.23 (1.59)	0.885	0.23 (1.59)	0.885	0.23 (1.59)	0.885
visit 10 * trt B	0.48 (1.56)	0.759	0.48 (1.58)	0.760	0.48 (1.58)	0.762	0.48 (1.59)	0.763	0.48 (1.59)	0.764	0.48 (1.59)	0.764
visit 11 * trt A1	-1.54 (1.62)	0.342	-1.52 (1.63)	0.353	-1.53 (1.63)	0.350	-1.53 (1.64)	0.354	-1.53 (1.65)	0.354	-1.53 (1.65)	0.354
visit 11 * trt B	0.096 (1.58)	0.952	0.11 (1.59)	0.946	0.10 (1.60)	0.950	0.10 (1.61)	0.951	0.10 (1.61)	0.951	0.10 (1.61)	0.951
baseline	0.13 (0.13)	0.297	0.13 (0.13)	0.302	0.13 (0.13)	0.301	0.14 (0.13)	0.296	0.14 (0.13)	0.295	0.14 (0.13)	0.295
baseline * visit 5	0.19 (0.14)	0.180	0.19 (0.14)	0.184	0.19 (0.14)	0.189	0.19 (0.14)	0.193	0.19 (0.14)	0.193	0.19 (0.14)	0.193
baseline * visit 6	0.29 (0.15)	0.0533	0.28 (0.15)	0.0557	0.28 (0.15)	0.0577	0.28 (0.15)	0.0598	0.28 (0.15)	0.0597	0.28 (0.15)	0.0597
baseline * visit 7	0.19 (0.14)	0.173	0.19 (0.14)	0.181	0.19 (0.14)	0.184	0.19 (0.14)	0.190	0.19 (0.14)	0.190	0.19 (0.14)	0.190
baseline * visit 8	0.28 (0.14)	0.0515	0.28 (0.14)	0.0548	0.28 (0.14)	0.0567	0.28 (0.15)	0.0593	0.28 (0.15)	0.0592	0.28 (0.15)	0.0591
baseline * visit 9	0.21 (0.14)	0.131	0.21 (0.14)	0.138	0.21 (0.14)	0.141	0.21 (0.14)	0.145	0.21 (0.14)	0.146	0.21 (0.14)	0.145
baseline * visit 10	0.19 (0.14)	0.172	0.19 (0.14)	0.183	0.19 (0.14)	0.185	0.19 (0.14)	0.192	0.19 (0.14)	0.191	0.19 (0.14)	0.191
baseline * visit 11	0.24 (0.14)	0.098	0.23 (0.14)	0.104	0.23 (0.14)	0.106	0.23 (0.15)	0.111	0.23 (0.15)	0.111	0.23 (0.15)	0.111
σ	2.40 (0.27)	$<.0001$	2.39 (0.28)	$<.0001$	2.46 (0.28)	$<.0001$	2.52 (0.29)	$<.0001$	2.52 (0.30)	$<.0001$	2.52 (0.30)	$<.0001$
-2ℓ	725.0		725.8		721.8		721.0		720.9		720.9	

Table 4.11: *Hamilton Depression Rating Scale Data. Study 1, MAR: GLMM using proc NLMIXED with adaptive Gaussian quadrature, Newton-Raphson optimization and different number of quadrature points.*

	$Q = 2$		$Q = 3$		$Q = 5$		$Q = 10$		$Q = 20$		$Q = 50$	
	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value
intercept	3.31 (2.57)	0.200	3.24 (2.58)	0.211	3.36 (2.62)	0.202	3.39 (2.65)	0.203	3.39 (2.65)	0.203	3.39 (2.65)	0.203
trt A1	-0.56 (1.47)	0.703	-0.57 (1.48)	0.702	-0.58 (1.49)	0.700	-0.58 (1.51)	0.700	-0.59 (1.51)	0.699	-0.58 (1.51)	0.699
trt B	-0.12 (1.54)	0.936	-0.14 (1.55)	0.926	-0.15 (1.57)	0.922	-0.16 (1.59)	0.920	-0.16 (1.59)	0.920	-0.16 (1.59)	0.920
visit 5	-4.57 (2.79)	0.103	-4.53 (2.79)	0.107	-4.63 (2.83)	0.104	-4.64 (2.85)	0.105	-4.65 (2.85)	0.105	-4.65 (2.85)	0.105
visit 6	-7.20 (2.88)	0.0135	-7.09 (2.89)	0.0151	-7.22 (2.92)	0.0144	-7.26 (2.94)	0.0145	-7.26 (2.94)	0.0145	-7.26 (2.94)	0.0145
visit 7	-6.47 (2.79)	0.0213	-6.34 (2.79)	0.0242	-6.49 (2.83)	0.0229	-6.51 (2.84)	0.0232	-6.52 (2.84)	0.0232	-6.52 (2.84)	0.0232
visit 8	-8.38 (2.86)	0.0039	-8.25 (2.87)	0.0046	-8.41 (2.90)	0.0043	-8.45 (2.92)	0.0044	-8.46 (2.93)	0.0044	-8.46 (2.93)	0.0044
visit 9	-7.24 (2.80)	0.0107	-7.08 (2.81)	0.0126	-7.24 (2.85)	0.0118	-7.28 (2.86)	0.0119	-7.29 (2.87)	0.0119	-7.29 (2.87)	0.0119
visit 10	-8.13 (2.83)	0.0046	-7.92 (2.83)	0.0058	-8.12 (2.87)	0.0052	-8.17 (2.89)	0.0053	-8.18 (2.89)	0.0053	-8.18 (2.89)	0.0053
visit 11	-8.41 (2.90)	0.0043	-8.20 (2.90)	0.0053	-8.40 (2.94)	0.0049	-8.46 (2.96)	0.0049	-8.47 (2.97)	0.0049	-8.47 (2.97)	0.0049
visit 5 * trt A1	0.25 (1.67)	0.879	0.26 (1.68)	0.877	0.27 (1.69)	0.873	0.26 (1.70)	0.876	0.26 (1.70)	0.877	0.26 (1.70)	0.877
visit 5 * trt B	-0.24 (1.68)	0.885	-0.22 (1.69)	0.896	-0.21 (1.70)	0.904	-0.21 (1.71)	0.902	-0.21 (1.71)	0.902	-0.21 (1.71)	0.902
visit 6 * trt A1	0.62 (1.64)	0.706	0.63 (1.66)	0.706	0.64 (1.67)	0.702	0.63 (1.68)	0.707	0.63 (1.68)	0.707	0.63 (1.68)	0.707
visit 6 * trt B	0.50 (1.63)	0.760	0.50 (1.64)	0.762	0.51 (1.66)	0.757	0.51 (1.66)	0.762	0.50 (1.67)	0.763	0.50 (1.67)	0.763
visit 7 * trt A1	0.48 (1.65)	0.770	0.48 (1.66)	0.773	0.50 (1.67)	0.764	0.50 (1.68)	0.768	0.50 (1.69)	0.769	0.50 (1.69)	0.769
visit 7 * trt B	0.059 (1.63)	0.971	0.060 (1.64)	0.971	0.074 (1.66)	0.964	0.066 (1.67)	0.969	0.064 (1.67)	0.969	0.064 (1.67)	0.969
visit 8 * trt A1	-0.64 (1.59)	0.689	-0.63 (1.60)	0.696	-0.62 (1.61)	0.702	-0.63 (1.62)	0.699	-0.63 (1.62)	0.698	-0.63 (1.62)	0.698
visit 8 * trt B	-0.16 (1.67)	0.922	-0.16 (1.69)	0.923	-0.15 (1.70)	0.931	-0.15 (1.71)	0.929	-0.15 (1.71)	0.928	-0.15 (1.71)	0.928
visit 9 * trt A1	0.30 (1.55)	0.847	0.30 (1.56)	0.847	0.31 (1.57)	0.845	0.30 (1.58)	0.850	0.30 (1.58)	0.850	0.30 (1.58)	0.850
visit 9 * trt B	0.37 (1.58)	0.814	0.38 (1.59)	0.811	0.39 (1.60)	0.809	0.38 (1.61)	0.812	0.38 (1.61)	0.813	0.38 (1.61)	0.813
visit 10 * trt A1	0.22 (1.55)	0.889	0.23 (1.56)	0.883	0.23 (1.57)	0.882	0.23 (1.58)	0.887	0.23 (1.59)	0.887	0.23 (1.59)	0.887
visit 10 * trt B	0.47 (1.56)	0.763	0.49 (1.57)	0.757	0.49 (1.58)	0.756	0.48 (1.59)	0.761	0.48 (1.59)	0.762	0.48 (1.59)	0.762
visit 11 * trt A1	-1.56 (1.61)	0.336	-1.51 (1.62)	0.353	-1.51 (1.63)	0.355	-1.53 (1.64)	0.354	-1.53 (1.64)	0.354	-1.53 (1.64)	0.354
visit 11 * trt B	0.076 (1.58)	0.962	0.099 (1.59)	0.950	0.11 (1.60)	0.945	0.099 (1.61)	0.951	0.096 (1.61)	0.953	0.096 (1.61)	0.953
baseline	0.14 (0.13)	0.288	0.14 (0.13)	0.274	0.14 (0.13)	0.297	0.14 (0.13)	0.296	0.14 (0.13)	0.295	0.14 (0.13)	0.295
baseline * visit 5	0.19 (0.14)	0.187	0.18 (0.14)	0.199	0.19 (0.14)	0.192	0.19 (0.14)	0.193	0.19 (0.14)	0.194	0.19 (0.14)	0.194
baseline * visit 6	0.28 (0.15)	0.0555	0.27 (0.15)	0.0636	0.28 (0.15)	0.0593	0.28 (0.15)	0.0596	0.28 (0.15)	0.0596	0.28 (0.15)	0.0596
baseline * visit 7	0.19 (0.14)	0.179	0.18 (0.14)	0.203	0.19 (0.14)	0.189	0.19 (0.14)	0.190	0.19 (0.14)	0.190	0.19 (0.14)	0.190
baseline * visit 8	0.28 (0.14)	0.0546	0.27 (0.14)	0.0648	0.27 (0.14)	0.0592	0.28 (0.15)	0.0595	0.28 (0.15)	0.0595	0.28 (0.15)	0.0594
baseline * visit 9	0.21 (0.14)	0.135	0.20 (0.14)	0.158	0.21 (0.14)	0.145	0.21 (0.14)	0.145	0.21 (0.14)	0.145	0.21 (0.14)	0.145
baseline * visit 10	0.19 (0.14)	0.178	0.17 (0.14)	0.216	0.18 (0.14)	0.194	0.19 (0.14)	0.193	0.19 (0.14)	0.193	0.19 (0.14)	0.192
baseline * visit 11	0.23 (0.14)	0.102	0.22 (0.14)	0.126	0.23 (0.14)	0.113	0.23 (0.14)	0.112	0.23 (0.15)	0.112	0.23 (0.15)	0.112
σ	2.40 (0.27)	$<.0001$	2.38 (0.28)	$<.0001$	2.45 (0.28)	$<.0001$	2.51 (0.29)	$<.0001$	2.52 (0.30)	$<.0001$	2.52 (0.30)	$<.0001$
-2ℓ	725.0		725.8		721.8		721.0		720.9		720.9	

Table 4.12: *Hamilton Depression Rating Scale Data. Study 1, GLMM using proc NLMIXED with adaptive Gaussian quadrature, Newton-Raphson optimization and 50 quadrature points, for CC, LOCF and MAR.*

	CC		LOCF		MAR	
	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value
intercept	3.56 (2.77)	0.202	3.68 (2.90)	0.207	3.39 (2.65)	0.203
trt A1	-0.60 (1.61)	0.711	-0.71 (1.64)	0.664	-0.58 (1.51)	0.699
trt B	-0.13 (1.68)	0.937	-0.27 (1.72)	0.876	-0.16 (1.59)	0.920
visit 5	-3.69 (2.93)	0.211	-4.96 (2.97)	0.0975	-4.65 (2.85)	0.105
visit 6	-6.91 (3.00)	0.0232	-7.93 (3.06)	0.0104	-7.26 (2.94)	0.0145
visit 7	-6.30 (2.92)	0.0332	-7.26 (2.98)	0.0158	-6.52 (2.84)	0.0232
visit 8	-8.02 (3.00)	0.0086	-9.33 (3.06)	0.0027	-8.46 (2.93)	0.0044
visit 9	-7.42 (2.95)	0.0135	-8.19 (3.00)	0.0070	-7.29 (2.87)	0.0119
visit 10	-9.03 (3.01)	0.0034	-9.11 (3.02)	0.0030	-8.18 (2.89)	0.0053
visit 11	-8.61 (3.01)	0.0050	-8.84 (3.01)	0.0038	-8.47 (2.97)	0.0049
visit 5 * trt A1	0.052 (1.77)	0.977	0.099 (1.76)	0.955	0.26 (1.70)	0.877
visit 5 * trt B	-0.47 (1.82)	0.798	-0.19 (1.78)	0.914	-0.21 (1.71)	0.902
visit 6 * trt A1	0.23 (1.74)	0.897	0.54 (1.74)	0.758	0.63 (1.68)	0.707
visit 6 * trt B	0.46 (1.73)	0.791	0.35 (1.73)	0.838	0.50 (1.67)	0.763
visit 7 * trt A1	0.45 (1.74)	0.796	0.37 (1.74)	0.832	0.50 (1.69)	0.769
visit 7 * trt B	0.18 (1.73)	0.918	0.008 (1.73)	0.996	0.064 (1.67)	0.969
visit 8 * trt A1	-1.16 (1.74)	0.505	-0.62 (1.68)	0.712	-0.63 (1.62)	0.698
visit 8 * trt B	-0.24 (1.74)	0.893	-0.23 (1.73)	0.895	-0.15 (1.71)	0.928
visit 9 * trt A1	0.38 (1.65)	0.819	0.29 (1.64)	0.861	0.30 (1.58)	0.850
visit 9 * trt B	0.10 (1.69)	0.952	0.38 (1.67)	0.819	0.38 (1.61)	0.813
visit 10 * trt A1	0.25 (1.66)	0.882	0.30 (1.64)	0.854	0.23 (1.59)	0.887
visit 10 * trt B	0.006 (1.67)	0.997	0.54 (1.65)	0.742	0.48 (1.59)	0.762
visit 11 * trt A1	-1.63 (1.68)	0.333	-1.14 (1.65)	0.491	-1.53 (1.64)	0.354
visit 11 * trt B	-0.092 (1.67)	0.956	0.15 (1.65)	0.929	0.096 (1.61)	0.953
baseline	0.098 (0.13)	0.469	0.17 (0.14)	0.231	0.14 (0.13)	0.295
baseline * visit 5	0.16 (0.15)	0.269	0.20 (0.15)	0.180	0.19 (0.14)	0.194
baseline * visit 6	0.28 (0.15)	0.0628	0.32 (0.15)	0.0386	0.28 (0.15)	0.0596
baseline * visit 7	0.18 (0.14)	0.201	0.23 (0.15)	0.122	0.19 (0.14)	0.190
baseline * visit 8	0.28 (0.15)	0.0559	0.32 (0.15)	0.0329	0.28 (0.15)	0.0594
baseline * visit 9	0.22 (0.14)	0.127	0.26 (0.15)	0.0805	0.21 (0.14)	0.145
baseline * visit 10	0.26 (0.14)	0.0781	0.24 (0.15)	0.101	0.19 (0.14)	0.192
baseline * visit 11	0.26 (0.15)	0.0755	0.25 (0.15)	0.0912	0.23 (0.15)	0.112
σ	2.54 (0.32)	$<.0001$	3.12 (0.37)	$<.0001$	2.52 (0.30)	$<.0001$
-2ℓ	654.8		769.7		720.9	

Table 4.13: *Hamilton Depression Rating Scale Data. Study 2, GLMM using proc NLMIXED with adaptive Gaussian quadrature, Newton-Raphson optimization and 50 quadrature points, for CC, LOCF and MAR.*

	CC		LOCF		MAR	
	est.(s.e.)	p-value	est.(s.e.)	p-value	est.(s.e.)	p-value
intercept	-3.45 (1.78)	0.0536	-3.54 (1.55)	0.0233	-3.12 (1.40)	0.0268
trt A1	-0.44 (1.14)	0.700	-0.91 (1.04)	0.380	-0.78 (0.93)	0.400
trt A2	1.49 (1.31)	0.258	0.76 (1.17)	0.515	0.65 (1.04)	0.530
trt B	1.14 (1.19)	0.339	1.61 (1.13)	0.156	1.29 (1.01)	0.202
visit 5	-2.34 (1.74)	0.179	-1.25 (1.44)	0.388	-1.26 (1.40)	0.370
visit 6	-0.15 (1.70)	0.928	0.03 (1.45)	0.982	-0.05 (1.41)	0.969
visit 7	-0.82 (1.71)	0.631	-0.45 (1.46)	0.758	-0.37 (1.45)	0.799
visit 8	-0.55 (1.72)	0.748	-0.074 (1.46)	0.960	-0.49 (1.49)	0.744
visit 5 * trt A1	0.24 (1.10)	0.830	0.41 (0.97)	0.670	0.48 (0.93)	0.609
visit 5 * trt A2	-1.43 (1.32)	0.279	-0.58 (1.13)	0.611	-0.35 (1.07)	0.744
visit 5 * trt B	0.40 (1.16)	0.732	-0.22 (1.08)	0.838	-0.019 (1.03)	0.985
visit 6 * trt A1	0.59 (1.12)	0.602	1.13 (0.99)	0.254	1.21 (0.96)	0.208
visit 6 * trt A2	-1.33 (1.34)	0.322	-0.26 (1.15)	0.822	-0.051 (1.08)	0.962
visit 6 * trt B	0.71 (1.16)	0.541	0.19 (1.09)	0.860	0.40 (1.03)	0.701
visit 7 * trt A1	0.11 (1.15)	0.924	0.12 (1.00)	0.906	-0.014 (0.99)	0.989
visit 7 * trt A2	-0.39 (1.33)	0.772	-0.12 (1.15)	0.920	0.21 (1.10)	0.852
visit 7 * trt B	1.26 (1.17)	0.283	0.23 (1.10)	0.836	0.47 (1.06)	0.662
visit 8 * trt A1	-0.92 (1.17)	0.434	-0.59 (1.01)	0.557	-0.65 (1.04)	0.536
visit 8 * trt A2	-1.35 (1.35)	0.318	-0.91 (1.16)	0.433	-0.60 (1.13)	0.600
visit 8 * trt B	0.36 (1.18)	0.763	-0.51 (1.10)	0.645	-0.053 (1.10)	0.962
baseline	0.48 (0.11)	<.0001	0.56 (0.10)	<.0001	0.49 (0.090)	<.0001
baseline * visit 5	0.019 (0.11)	0.862	-0.073 (0.089)	0.411	-0.07 (0.086)	0.391
baseline * visit 6	-0.24 (0.11)	0.0246	-0.27 (0.090)	0.0025	-0.26 (0.087)	0.0028
baseline * visit 7	-0.23 (0.11)	0.0301	-0.24 (0.090)	0.0073	-0.25 (0.090)	0.0064
baseline * visit 8	-0.25 (0.11)	0.0214	-0.27 (0.091)	0.0036	-0.26 (0.091)	0.0046
σ	3.34 (0.36)	<.0001	3.75 (0.35)	<.0001	3.06 (0.30)	<.0001
-2ℓ	897		1226.7		1114.1	

score-by-visit interaction, in both studies. In the first study, we obtain a borderline significant interaction for some visits under LOCF, while the corresponding p-values under CC and MAR are borderline insignificant. The difference is more pronounced

in the second study in the sense that for several visits a highly significant effect is found under both LOCF and MAR, while the corresponding effects under CC are borderline significant only.

4.2.2 Marginal *versus* Random-effects Models

In all cases, the random-intercept variability (standard deviation parameter σ) is highly significant. This implies that the GEE parameters and the random-effects parameters cannot be compared directly. If the conversion factor (3.20) is computed, then one roughly finds a factor of about 2.5. We note that this factor is not reproduced when directly comparing the two sets of estimates. This is due to the fact that (3.20) operates at the true population parameter level, while we only have parameter estimates at our disposition. Since many of the estimates are not or only marginally significant, it is not unexpected, therefore, to observe deviations from this relationship, even though the general tendency is preserved in most cases.

4.2.3 Views 2 and 3: Single Time Point Analysis

When emphasis is on the last measurement occasion, LOCF and CC are straightforward to use. When the last observed measurement is of interest, while a different scientific question, the analysis is not different from the one obtained under LOCF but, of course, in this case CC is not an option.

Since the outcome is a dichotomous response, the data can be summarized in a $2\times k$ table, where k represents the number of treatments. The analysis essentially consists of comparing the proportions of succes or failure in all groups. For this purpose, both Pearson's chi-squared test (Agresti, 2002) and Fisher's Exact test (Freeman and Halton, 1951) will be used. Nevertheless, it is still possible to obtain inferences from a full longitudinal model in this context. We add these for the sake of reference, but it should be understood that the analysis using a simple model for the last time point only is more in line with practice.

When an ignorable analysis is considered, one has to explicitly consider all incomplete profiles, in order to correctly incorporate all information available. Thus, one has to consider a longitudinal model.

For both studies, non-experimental drug C is considered as reference treatment. Let α_i be the effect of treatment arm i at the last measurement occasion, where $i =$ A1, A2, B or C. We wish to test whether at the last measurement occasion all treatment effects are equal. For Study 1, this translates into $\alpha_{A1} = \alpha_B = \alpha_C$, or equivalently into $\alpha_{A1}-\alpha_C = \alpha_B-\alpha_C = 0$, whereas for Study 2 this means $\alpha_{A1} = \alpha_{A2} = \alpha_B = \alpha_C$,

Table 4.14: *Hamilton Depression Rating Scale Data. Views 2 and 3. p-values are reported ('mixed' refers to the assessment of treatment at the last visit based on a generalized linear mixed model).*

Method	Model	Study 1	Study 2
CC	mixed	0.0463	0.0614
	Pearson's Chi-squared Test	0.0357	0.0350
	Fisher's Exact Test	0.0336	0.0350
LOCF	mixed	0.1393	0.1067
	Pearson's Chi-squared Test	0.1553	0.0384
	Fisher's Exact Test	0.1553	0.0405
MAR	mixed	0.0500	0.0677

or similarly $\alpha_{A1} - \alpha_C = \alpha_{A2} - \alpha_C = \alpha_B - \alpha_C = 0$. Such contrasts can be obtained very easily using the SAS procedure NLMIXED. Table 4.14 shows a summary of the results in terms of p-values.

For Study 1, we have the following conclusions. The GLMMs lead to a small difference between CC and MAR: both are borderline. On the other hand, the GLMM for LOCF leads to a non-significant result. An endpoint analysis (i.e., using the last available measurement) shows the same result for LOCF (non-significant), whereas the result for CC becomes significant. For Study 2, the GLMM again lead to a small difference between CC and MAR, both with non-significant results. The GLMM for LOCF clearly gives a non-significant result. An endpoint analysis leads to a completely different picture, with results that are strongly different (significant) from the GLMM model. This illustrates that the choice between modeling technique is far from an academic question, but can have profound impact on the study conclusions, ranging from highly significant over borderline (non-)significant to highly non-significant.

4.3 Conclusions

In this chapter, we have indicated that a variety of approaches is possible, when analyzing incomplete longitudinal data. In the continuous case the linear mixed model (LMM) is the main mode of analysis. For binary outcomes, one has the choice

between a marginal model (generalized estimating equations, GEE) and a random-effects approach (generalized linear mixed models, GLMM). While GLMM and GEE may provide similar results in terms of hypothesis testing, things are different when the models are used for estimation purposes, because the parameters have quite different meanings. All of the methods LMM, GLMM and GEE can be used when data are incomplete. For both LMM and GLMM this holds under the fairly general assumption of an MAR mechanism to operate, while for GEE the stronger MCAR is required. However, GEE can be extended to weighted GEE, making it also valid under MAR.

Current statistical computing power has brought LMM, GLMM and WGEE within reach. This underscores that simple but potentially highly restrictive modes of analyses, such as CC or LOCF, should no longer be seen as the preferred mode of analysis, neither in sociology, nor in clinical practice or related fields.

While in the studies considered here there are no extreme differences between the various analyses conducted, one cannot be sure that such a conclusion would hold in all settings (Molenberghs *et al.*, 2004). Although in Section 4.1, the results from CC and direct-likelihood do not differ regarding the importance of the predictor variables, the significance as well as non-significance of the effects are much more pronounced in the DL analysis.

Whether our more generally valid analysis would consistently provide differences with those found in the literature (see Section 4.1.3) is an interesting subject of further study. However, the type of direct-likelihood analysis proposed here often still provides sensible assessments of important aspects of the data, even if the assumption of MAR is violated in favor of MNAR. Indeed, an ignorable analysis takes all information into account, not only from complete observations, but also from incomplete ones, through the conditional expectation of the missing measurements given the observed ones.

By definition, MNAR missingness cannot be fully ruled out based on the observed data. 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. A discussion of this phenomenon in the survey context has been given in Rubin, Stern and Vehovar (1995). These authors firstly 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. Secondly, and very important for confirmatory trials, an MAR analysis can be specified *a priori* without additional work relative to a situation with complete data. Thirdly, when there is residual doubt about the plausibility of MAR, a number of MNAR models, which are more general and explicitly incorporate

the dropout mechanism, can be fitted, provided one is prepared to approach formal aspects of model comparison with due caution, since the inferences they produce are typically highly dependent on the 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. Based on these considerations, we recommend, for primary analysis purposes, the use of ignorable likelihood-based methods or appropriately modified frequentist methods. To explore the impact of deviations from the MAR assumption on the conclusions, one should ideally conduct a sensitivity analysis (Verbeke and Molenberghs, 2000, Ch. 18–20).

In the next chapters, several MNAR models for non-continuous longitudinal data with non-monotone missingness are proposed. Afterwards, the MNAR analyses can be complemented with appropriate (global and/or local) influence analyses, as will be shown in Chapter 8.

5

An Extension and Reparameterization of the Baker, Rosenberger and DerSimonian (1992) Model

In previous chapters, focus was on the advantage of MAR models above simple methods such as CC and LOCF, and more specific for data with dropout. In this chapter, we switch attention to several MNAR models, since they are more general and explicitly incorporate the dropout mechanism. We will propose several models that can handle non-monotone missing data. We start with the model family proposed by Baker, Rosenberger and DerSimonian (1992) (further denoted as BRD). A number of contributions are made. First, the model is reformulated such that its membership of the selection model family is unambiguously clear. Second, the original model is extended to accommodate for, possibly continuous, covariates, turning the model into a regression tool for several categorical outcomes. Third, a parameterization is proposed that avoids the risk of invalid solutions. In other words, all combinations of the natural parameters produce probabilities between 0 and 1. As a consequence,

Table 5.1: *Theoretical distribution over complete and observed cells of a bivariate binary outcome. Tables correspond to completely observed subjects and subjects with the second, the first, and both measurements missing, respectively.*

$\pi_{11,11}$	$\pi_{11,12}$
$\pi_{11,21}$	$\pi_{11,22}$

$\pi_{10,11}$	$\pi_{10,12}$
$\pi_{10,21}$	$\pi_{10,22}$

$\pi_{01,11}$	$\pi_{01,12}$
$\pi_{01,21}$	$\pi_{01,22}$

$\pi_{00,11}$	$\pi_{00,12}$
$\pi_{00,21}$	$\pi_{00,22}$

$\pi_{11,11}$	$\pi_{11,12}$
$\pi_{11,21}$	$\pi_{11,22}$

$\pi_{10,1+}$
$\pi_{10,2+}$

$\pi_{01,+1}$	$\pi_{01,+2}$

$\pi_{00,++}$

the closed-form solutions of BRD no longer apply; given the focus on continuous covariates, the derivation of closed-form solutions should not be of primary concern.

In Section 5.1 we sketch the original BRD models, which are extended to incorporate covariate effects in Section 5.2. Section 5.3 shows detailed calculations of the derivatives of the log-likelihood function. An application to the fluvoxamine data is presented in Section 5.4.

5.1 The Original BRD Model Family

Baker, Rosenberger and DerSimonian (1992) considered a log-linear type of model for two possibly binary outcomes, subject to non-monotone missingness. They use a four-way classification of both outcomes, together with their respective missingness indicators. Denote the counts and corresponding probabilities by $Y_{r_1r_2,jk}$ and $\pi_{r_1r_2,jk}$ where $r_1, r_2 = 0, 1$ indicate whether a measurement is either missing or taken at occasions 1 and 2 respectively, and $j, k = 1, 2$ indicate the response categories for both outcomes. The complete data and observed data cell probabilities are presented in Table 5.1.

The models can be written as:

$$\begin{aligned} E(Y_{11,jk}) &= m_{jk}, & E(Y_{10,jk}) &= m_{jk}\tilde{\beta}_{jk}, \\ E(Y_{01,jk}) &= m_{jk}\tilde{\alpha}_{jk}, & E(Y_{00,jk}) &= m_{jk}\tilde{\alpha}_{jk}\tilde{\beta}_{jk}\tilde{\gamma}, \end{aligned}$$

with $m_{jk} = Y_{++++}\pi_{11,jk}$ and

$$\tilde{\alpha}_{jk} = \frac{q_{01|jk}}{q_{11|jk}}, \qquad \tilde{\beta}_{jk} = \frac{q_{10|jk}}{q_{11|jk}}, \qquad \tilde{\gamma} = \frac{q_{11|jk}q_{00|jk}}{q_{10|jk}q_{01|jk}},$$

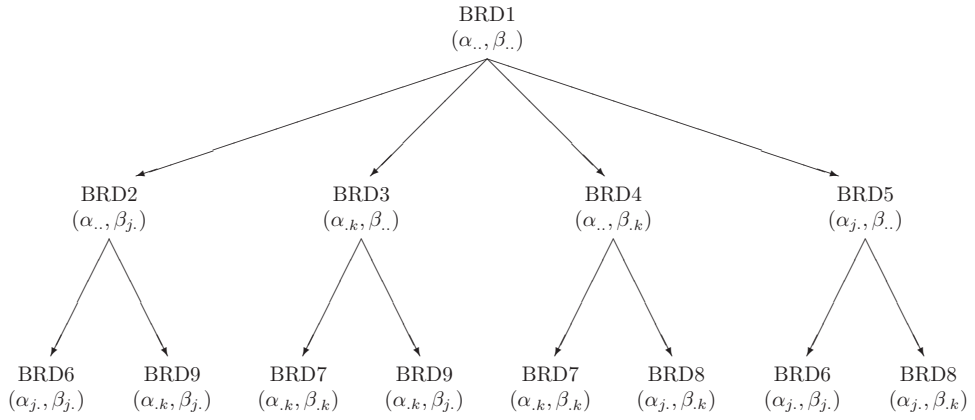


Figure 5.1: *Graphical representation of the BRD model nesting structure.*

such that $\tilde{\alpha}_{jk}$ models the non-response for the first variable, $\tilde{\beta}_{jk}$ the non-response for the second variable, and $\tilde{\gamma}$ the interaction between both non-response indicators. The subscripts are missing from $\tilde{\gamma}$ since Baker, Rosenberger and DerSimonian (1992) have shown that this quantity is independent of j and k in every identifiable model. From the expressions for $\tilde{\alpha}_{jk}$, $\tilde{\beta}_{jk}$ and $\tilde{\gamma}$, we see that selection model quantities are employed. However, $\pi_{11,jk}$ has a pattern-mixture flavor, and the typical selection model probabilities μ_{jk} are a combination of all other parameters:

$$\mu_{jk} = \pi_{11,jk}(1 + \tilde{\alpha}_{jk} + \tilde{\beta}_{jk} + \tilde{\alpha}_{jk}\tilde{\beta}_{jk}\tilde{\gamma}).$$

These authors consider nine identifiable models, based on setting $\tilde{\alpha}_{jk}$ and $\tilde{\beta}_{jk}$ constant in one or more indices:

$$\begin{array}{lllllllll} \text{BRD1} & : & (\tilde{\alpha}_{..}, \tilde{\beta}_{..}) & \text{BRD4} & : & (\tilde{\alpha}_{..}, \tilde{\beta}_{.k}) & \text{BRD7} & : & (\tilde{\alpha}_{.k}, \tilde{\beta}_{.k}) \\ \text{BRD2} & : & (\tilde{\alpha}_{..}, \tilde{\beta}_{j.}) & \text{BRD5} & : & (\tilde{\alpha}_{j.}, \tilde{\beta}_{..}) & \text{BRD8} & : & (\tilde{\alpha}_{j.}, \tilde{\beta}_{.k}) \\ \text{BRD3} & : & (\tilde{\alpha}_{.k}, \tilde{\beta}_{..}) & \text{BRD6} & : & (\tilde{\alpha}_{j.}, \tilde{\beta}_{j.}) & \text{BRD9} & : & (\tilde{\alpha}_{.k}, \tilde{\beta}_{j.}). \end{array} \tag{5.1}$$

The nesting structure of these models is schematically represented in Figure 5.1. Interpretation is straightforward. For example, BRD1 is MCAR, in BRD4 missingness in the first variable is constant, while missingness in the second variable depends on the, possibly unobserved, value of this variable.

5.2 The Extended BRD Model Family

Now we will extend the original BRD models to accommodate (possibly continuous) covariates. Let $i = 1, \ldots, n$ index distinct covariate levels, but this index i will be suppressed from notation in the rest of the section. We will use a selection model parameterization, differing from and extending the original one:

$$\pi_{r_1 r_2, jk} = p_{jk} q_{r_1 r_2 | jk}, \tag{5.2}$$

where p_{jk} parameterizes the measurement process and $q_{r_1 r_2|jk}$ describes the missingness mechanism, conditional on the measurements. In particular, we will assume

$$p_{jk} = \frac{\exp(\boldsymbol{\eta}_{jk})}{\sum_{j,k=1}^{2} \exp(\boldsymbol{\eta}_{jk})}, \tag{5.3}$$

$$q_{r_1 r_2|jk} = \frac{\exp[\alpha_{jk}(1-r_1) + \beta_{jk}(1-r_2) + \gamma(1-r_1)(1-r_2)]}{1 + \exp(\alpha_{jk}) + \exp(\beta_{jk}) + \exp(\alpha_{jk} + \beta_{jk} + \gamma)}, \tag{5.4}$$

where α_{jk}, β_{jk} and γ have the same interpretation as $\tilde{\alpha}_{jk}$, $\tilde{\beta}_{jk}$ and $\tilde{\gamma}$ in Section 5.1.

No a priori ordering is imposed on the outcomes. The advantage is that genuine multivariate settings (e.g., several questions in a survey) can be handled as well. When deemed necessary, the implications of ordering can be imposed by considering specific models and leaving out others. For example, one may want to avoid missingness on future observations. In the current bivariate case, the index k would have to be removed from α in the above model. To identify the model, we set $\boldsymbol{\eta}_{22} = \mathbf{0}$ and further $\boldsymbol{\eta}_{jk} = X_{jk}\boldsymbol{\theta}$. This allows inclusion of covariate effects which, together with (5.3), is related to the multigroup logistic model (Albert and Lesaffre, 1986). Even though the parameters $\boldsymbol{\theta}$ are conditional in nature and therefore somewhat difficult to directly interpret in case planned sequences are of unequal length, (5.3) allows easy calculation of the joint probabilities. Note that observed sequences could be of unequal length. The conditional interpretation of the parameters will be a major obstacle in the presence of time-varying covariates, as is often the case in longitudinal studies with longer measurement sequences. Arguably, in such a case, a different model may be more suitable. Generally, computational advantages become increasingly important as the length of the response vector grows. If necessary, specific functions of interest, such as a marginal treatment effect, can be derived. They will typically take the form of non-linear functions. Arguably, a model of the type here can be most useful as a component of a sensitivity analysis, in conjunction with the use of different (e.g., marginal) models.

In many examples, the design matrices X_{jk} will be equal. Stacking all parameters leads to $\boldsymbol{\eta} = X\boldsymbol{\theta}$ and similarly to $\boldsymbol{\delta} = Z\boldsymbol{\psi}$, where the vector $\boldsymbol{\delta}$ stacks the α_{jk}, β_{jk} and γ and Z is a design matrix. The vector $\boldsymbol{\psi}$ groups the parameters of interest. For example, if MCAR would be considered, the α and β parameters do not depend on neither j nor k and hence $\boldsymbol{\psi}' = (\alpha, \beta, \gamma)$. Both designs can be combined into one, using $\boldsymbol{\xi} = (\boldsymbol{\eta}', \boldsymbol{\delta}')'$,

$$W = \begin{pmatrix} X & 0 \\ 0 & Z \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\psi}')'. \tag{5.5}$$

The corresponding log-likelihood function can be written as:

$$\begin{aligned} \ell &= \sum_{j,k=1}^{2} Y_{11,jk} \ln \pi_{11,jk} + \sum_{j=1}^{2} Y_{10,j+} \ln(\pi_{10,j1} + \pi_{10,j2}) \\ &\quad + \sum_{k=1}^{2} Y_{01,+k} \ln(\pi_{01,1k} + \pi_{01,2k}) + Y_{00,++} \ln(\pi_{00,11} + \pi_{00,12} + \pi_{00,21} + \pi_{00,22}) \\ &= \sum_{j,k=1}^{2} \sum_{s=1}^{Y_{11,jk}} \ln \pi_{11,jk} + \sum_{j=1}^{2} \sum_{s=1}^{Y_{10,j+}} \ln \pi_{10,j+} \\ &\quad + \sum_{k=1}^{2} \sum_{s=1}^{Y_{01,+k}} \ln \pi_{01,+k} + \sum_{s=1}^{Y_{00,++}} \ln \pi_{00,++}. \end{aligned} \tag{5.6}$$

Computation of derivatives, needed for optimization and for the calculation of influence measures, will be described in Section 5.3. To include individual-specific covariates, a subscript i has to be introduced to the vector $\boldsymbol{\xi}_i$ and the matrix W_i and hence to their constituent components.

Similarly as in Section 5.1, nine identifiable models can be considered, by simply replacing $\tilde{\alpha}_{jk}$, $\tilde{\beta}_{jk}$ and $\tilde{\gamma}$ by α_{jk}, β_{jk} and γ in (5.1). The nesting structure in Figure 5.1 was considered by BRD using the original parameterization, but carry over to parameterization (5.4) is immediate.

5.3 Derivatives of the Log-Likelihood Function

We can write the joint probabilities as

$$\pi_{r_1 r_2, jk} = \frac{e^{\eta_{jk}}}{\sum_{v,w=1}^{2} e^{\eta_{vw}}} \frac{e^{\alpha_{jk}(1-r_1)+\beta_{jk}(1-r_2)+\gamma(1-r_1)(1-r_2)}}{\sum_{t,u=0}^{1} e^{\alpha_{jk}(1-t)+\beta_{jk}(1-u)+\gamma(1-t)(1-u)}}$$

$$\underbrace{p_{jk}}_{\eta\text{-parameters}} \cdot \underbrace{q_{(1-r_1)(1-r_2)|jk}}_{\delta\text{-parameters}}$$

From this we can calculate the log-likelihood function (5.6) and its derivatives. The first order derivative with respect to ω, the local influence parameter which will be introduced in Section 8.3 (only π_{****} contains ω), for the missingness patterns are

$$(11) \quad : \quad \frac{\partial \ell}{\partial \omega} = \sum_{j,k=1}^{2} \frac{1}{\pi_{11,jk}} \frac{\partial \pi_{11,jk}}{\partial \omega},$$

$$(10) \quad : \quad \frac{\partial \ell}{\partial \omega} = \sum_{j=1}^{2} \frac{1}{\pi_{10,j+}} \frac{\partial \pi_{10,j+}}{\partial \omega},$$

$$(01) \quad : \quad \frac{\partial \ell}{\partial \omega} = \sum_{k=1}^{2} \frac{1}{\pi_{01,+k}} \frac{\partial \pi_{01,+k}}{\partial \omega},$$

$$(00) \quad : \quad \frac{\partial \ell}{\partial \omega} = \frac{1}{\pi_{00,++}} \frac{\partial \pi_{00,++}}{\partial \omega},$$

respectively. The second order derivative to ω and an arbitrary parameter ζ from $\boldsymbol{\eta}$ or $\boldsymbol{\delta}$ $(\neq \omega)$, for a patient in category (11):

$$\frac{\partial^2 \ell}{\partial \omega \partial \zeta} = \sum_{j,k=1}^{2} \frac{\pi_{11jk} \dfrac{\partial^2 \pi_{11,jk}}{\partial \omega \partial \zeta} - \dfrac{\partial \pi_{11,jk}}{\partial \omega} \dfrac{\partial \pi_{11,jk}}{\partial \zeta}}{\pi_{11,jk}^2}$$

with similar expressions for the other categories.

We will use the following conventions. Since $t, u = 0, 1$ they obey $t^2 = t$ and $u^2 = u$. We will also, for the ease of notation, replace $e^{\alpha_{jk} t + \beta_{jk} u + \gamma t u}$ by $e^{\cdots}$. All summations that appear are for $t, u = 0, 1$.

It follows that

$$q_{11|jk} = \frac{1}{\sum e^{\cdots}}, \qquad q_{10|jk} = \frac{e^{\beta_{jk}}}{\sum e^{\cdots}}, \qquad q_{01|jk} = \frac{e^{\alpha_{jk}}}{\sum e^{\cdots}} \quad \text{and} \quad q_{00|jk} = \frac{e^{\alpha_{jk} + \beta_{jk} + \gamma}}{\sum e^{\cdots}}.$$

Further, it is easy to show that

$$\frac{\sum t e^{\cdots}}{\sum e^{\cdots}} = q_{0+|jk}, \qquad \frac{\sum u e^{\cdots}}{\sum e^{\cdots}} = q_{+0|jk}, \qquad \frac{\sum t u e^{\cdots}}{\sum e^{\cdots}} = q_{00|jk}, \qquad \frac{\sum t^2 e^{\cdots}}{\sum e^{\cdots}} = q_{0+|jk},$$

$$\frac{\sum u^2 e^{\cdots}}{\sum e^{\cdots}} = q_{+0|jk}, \qquad \frac{\sum t^2 u e^{\cdots}}{\sum e^{\cdots}} = q_{00|jk} \qquad \text{and} \qquad \frac{\sum t u^2 e^{\cdots}}{\sum e^{\cdots}} = q_{00|jk}.$$

Let us now study the pattern of completers $r_1 = r_2 = 1$. To calculate $(A) : \partial \pi_{11,jk} / \partial \zeta$ we use

$$(A1) \quad : \quad \frac{\partial \pi_{11,jk}}{\partial \eta} = \frac{\partial p_{jk}}{\partial \eta} . q_{11|jk},$$

$$(A2) \quad : \quad \frac{\partial \pi_{11,jk}}{\partial \delta} = p_{jk} . \frac{\partial q_{11|jk}}{\partial \delta},$$

$$(A3) \quad : \quad \frac{\partial \pi_{11,jk}}{\partial \theta} = \frac{\partial \pi_{11,jk}}{\partial \eta} . \frac{\partial \eta}{\partial \theta} = X . \frac{\partial \pi_{11,jk}}{\partial \eta}.$$

Calculation of $(B) : \partial\pi_{11,jk}/\partial\omega$ is straightforward:

$$(B) \quad : \quad \frac{\partial \pi_{11,jk}}{\partial \omega} = p_{jk} \cdot \left[\frac{\partial q_{11|jk}}{\partial \alpha_{jk}} \cdot \frac{\partial \alpha_{jk}}{\partial \omega} + \frac{\partial q_{11|jk}}{\partial \beta_{jk}} \cdot \frac{\partial \beta_{jk}}{\partial \omega} \right],$$

because only α_{jk} and β_{jk} can depend on ω (γ does not). The terms $\frac{\partial \alpha_{jk}}{\partial \omega}$ and $\frac{\partial \beta_{jk}}{\partial \omega}$ depend on the model, and can be equal to $j-1$, $k-1$, or 0. At least one of them will be zero.

To calculate $(C) : \partial^2\pi_{11,jk}/\partial\omega\partial\zeta$ we use

$$(C1) \quad : \quad \frac{\partial}{\partial \eta} \left(\frac{\partial \pi_{11,jk}}{\partial \omega} \right) = \frac{\partial p_{jk}}{\partial \eta} \cdot \left[\frac{\partial q_{11|jk}}{\partial \alpha_{jk}} \cdot \frac{\partial \alpha_{jk}}{\partial \omega} + \frac{\partial q_{11|jk}}{\partial \beta_{jk}} \cdot \frac{\partial \beta_{jk}}{\partial \omega} \right]$$

$$(C2) \quad : \quad \frac{\partial}{\partial \delta} \left(\frac{\partial \pi_{11,jk}}{\partial \omega} \right) = p_{jk} \cdot \left[\frac{\partial^2 q_{11|jk}}{\partial \delta \partial \alpha_{jk}} \cdot \frac{\partial \alpha_{jk}}{\partial \omega} + \frac{\partial^2 q_{11|jk}}{\partial \delta \partial \beta_{jk}} \cdot \frac{\partial \beta_{jk}}{\partial \omega} \right].$$

In addition, we need the following relations:

$$(a) \quad : \quad \frac{\partial p_{jk}}{\partial \eta_{jk}} = p_{jk}(1 - p_{jk}),$$

$$[(j',k') \neq (j,k)] \quad \frac{\partial p_{jk}}{\partial \eta_{j'k'}} = -p_{jk} p_{j'k'},$$

$$(b) \quad : \quad \frac{\partial q_{11|jk}}{\partial \alpha_{jk}} = -q_{11|jk} \cdot q_{0+|jk},$$

$$[(j',k') \neq (j,k)] \quad \frac{\partial q_{11|jk}}{\partial \alpha_{j'k'}} = 0,$$

$$(c) \quad : \quad \frac{\partial q_{11|jk}}{\partial \beta_{jk}} = -q_{11|jk} \cdot q_{+0|jk},$$

$$[(j',k') \neq (j,k)] \quad \frac{\partial q_{11|jk}}{\partial \beta_{j'k'}} = 0,$$

$$(d) \quad : \quad \frac{\partial q_{11|jk}}{\partial \gamma} = -q_{11|jk} \cdot q_{00|jk},$$

$$(e) \quad : \quad \frac{\partial^2 q_{11|jk}}{\partial \alpha_{jk}^2} = q_{11|jk} \cdot q_{0+|jk} \left(2q_{0+|jk} - 1 \right),$$

$$(f) \quad : \quad \frac{\partial^2 q_{11|jk}}{\partial \beta_{jk} \partial \alpha_{jk}} = q_{11|jk} \left(2q_{0+|jk} \cdot q_{+0|jk} - q_{00|jk} \right),$$

$$(g) \quad : \quad \frac{\partial^2 q_{11|jk}}{\partial \gamma \partial \alpha_{jk}} = q_{11|jk} \cdot q_{00|jk} \left(2q_{0+|jk} - 1 \right),$$

$$(h) \quad : \quad \frac{\partial^2 q_{11|jk}}{\partial \beta_{jk}^2} = q_{11|jk} \cdot q_{+0|jk} \left(2q_{+0|jk} - 1 \right),$$

$$(i) \quad : \quad \frac{\partial^2 q_{11|jk}}{\partial \gamma \partial \beta_{jk}} = q_{11|jk} \cdot q_{00|jk} \left(2q_{+0|jk} - 1 \right).$$

These calculations have to be redone for each of the three incomplete patterns as well. We will indicate here only how they are done when the first outcome is observed and the second one is not (i.e., $r_1 = 1$ and $r_2 = 0$).

Now,

$$(A) \quad : \quad \frac{\partial \pi_{10,j+}}{\partial \zeta} = \frac{\partial(\pi_{10,j1} + \pi_{10,j2})}{\partial \zeta} = \sum_{k=1}^{2} \frac{\partial \pi_{10,jk}}{\partial \zeta}$$

$$(A1) \quad : \quad \frac{\partial \pi_{10,j+}}{\partial \eta} = \sum_{k=1}^{2} \frac{\partial p_{jk}}{\partial \eta} . q_{10|jk},$$

$$(A2) \quad : \quad \frac{\partial \pi_{10,j+}}{\partial \delta} = \sum_{k=1}^{2} p_{jk} . \frac{\partial q_{10|jk}}{\partial \delta},$$

$$(A3) \quad : \quad \frac{\partial \pi_{10,j+}}{\partial \theta} = \frac{\partial \pi_{10,j+}}{\partial \eta} . \frac{\partial \eta}{\partial \theta} = X . \frac{\partial \pi_{10,j+}}{\partial \eta},$$

$$(B) \quad : \quad \frac{\partial \pi_{10,j+}}{\partial \omega} = \sum_{k=1}^{2} p_{jk} . \left[\frac{\partial q_{10|jk}}{\partial \alpha_{jk}} . \frac{\partial \alpha_{jk}}{\partial \omega} + \frac{\partial q_{10|jk}}{\partial \beta_{jk}} . \frac{\partial \beta_{jk}}{\partial \omega} \right],$$

$$(C) \quad : \quad \frac{\partial^2 \pi_{10,j+}}{\partial \omega \partial \zeta} = \frac{\partial}{\partial \zeta} \left(\frac{\partial \pi_{10,j+}}{\partial \omega} \right),$$

$$(C1) \quad : \quad \frac{\partial}{\partial \eta} \left(\frac{\partial \pi_{10,j+}}{\partial \omega} \right) = \sum_{k=1}^{2} \frac{\partial p_{jk}}{\partial \eta} . \left[\frac{\partial q_{10|jk}}{\partial \alpha_{jk}} . \frac{\partial \alpha_{jk}}{\partial \omega} + \frac{\partial q_{10|jk}}{\partial \beta_{jk}} . \frac{\partial \beta_{jk}}{\partial \omega} \right],$$

$$(C2) \quad : \quad \frac{\partial}{\partial \delta} \left(\frac{\partial \pi_{10,j+}}{\partial \omega} \right) = \sum_{k=1}^{2} p_{jk} . \left[\frac{\partial^2 q_{10|jk}}{\partial \delta \partial \alpha_{jk}} . \frac{\partial \alpha_{jk}}{\partial \omega} + \frac{\partial^2 q_{10|jk}}{\partial \delta \partial \beta_{jk}} . \frac{\partial \beta_{jk}}{\partial \omega} \right].$$

Further (we only show those that are different from the completers' expressions),

$$(b) \quad : \quad \frac{\partial q_{10|jk}}{\partial \alpha_{jk}} = -q_{10|jk} . q_{0+|jk},$$

$$(c) \quad : \quad \frac{\partial q_{10|jk}}{\partial \beta_{jk}} = q_{10|jk} \left(1 - q_{+0|jk}\right),$$

$$(d) \quad : \quad \frac{\partial q_{10|jk}}{\partial \gamma} = -q_{10|jk} . q_{00|jk},$$

$$(e) \quad : \quad \frac{\partial^2 q_{10|jk}}{\partial \alpha_{jk}^2} = q_{10|jk} . q_{0+|jk} \left(2q_{0+|jk} - 1\right),$$

$$(f) \quad : \quad \frac{\partial^2 q_{10|jk}}{\partial \beta_{jk} \partial \alpha_{jk}} = -q_{10|jk} \left(q_{0+|jk} - 2q_{+0|jk} . q_{0+|jk} + q_{00|jk}\right),$$

$$(g) \quad : \quad \frac{\partial^2 q_{10|jk}}{\partial \gamma \partial \alpha_{jk}} = q_{10|jk} . q_{00|jk} \left(2q_{0+|jk} - 1\right),$$

$$(h) \quad : \quad \frac{\partial^2 q_{10|jk}}{\partial \beta_{jk}^2} = q_{10|jk} \left(1 - 3q_{+0|jk} + 2q_{+0|jk}^2\right),$$

$$(i) \quad : \quad \frac{\partial^2 q_{10|jk}}{\partial\gamma\partial\beta_{jk}} = -2q_{10|jk}.q_{00|jk}\left(1 - q_{+0|jk}\right).$$

5.4 Models Fitted to the Fluvoxamine Data

In the analysis all patients with known duration level are considered, leaving a total of 310 out of 315 subjects in the study. In the measurement model, the effect of duration is held constant over both visits. Regarding the missingness model, an effect of duration is assumed in both the α and the β parameters. Each of the 9 models is represented by a specific choice for the design matrices and the corresponding parameter vector. For example, for BRD1, without any effect of duration, we obtain $\boldsymbol{\phi}^1$, X_i^1 and Z_i^1, while for BRD8, with a constant duration effect on the measurement model as well as on the missingness model, we obtain $\boldsymbol{\phi}^8$, X_i^8 and Z_i^8, which are constructed as follows:

$$\boldsymbol{\phi}^1 = (\theta_1, \theta_2, \theta_3, \alpha, \beta, \gamma)', \qquad \boldsymbol{\phi}^8 = (\theta_1, \theta_2, \theta_3, \theta_4, \alpha_{1.}, \alpha_{2.}, \alpha_{\text{cov}}, \beta_{.1}, \beta_{.2}, \beta_{\text{cov}}, \gamma)',$$

$$X_i^1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad X_i^8 = \begin{pmatrix} 1 & 0 & 0 & \text{cov}_i \\ 0 & 1 & 0 & \text{cov}_i \\ 0 & 0 & 1 & \text{cov}_i \end{pmatrix},$$

$$Z_i^1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Z_i^8 = \begin{pmatrix} 1 & 0 & \text{cov}_i & 0 & 0 & 0 & 0 \\ 1 & 0 & \text{cov}_i & 0 & 0 & 0 & 0 \\ 0 & 1 & \text{cov}_i & 0 & 0 & 0 & 0 \\ 0 & 1 & \text{cov}_i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \text{cov}_i & 0 \\ 0 & 0 & 0 & 0 & 1 & \text{cov}_i & 0 \\ 0 & 0 & 0 & 1 & 0 & \text{cov}_i & 0 \\ 0 & 0 & 0 & 0 & 1 & \text{cov}_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We will consider three sets of BRD models in detail. Table 5.2 presents models (estimates, s.e., negative loglikelihoods) without duration. In Table 5.3, duration is added as a covariate to the measurement model only, whereas in the final set (Table 5.4) the effect of duration is included in both measurement and missingness

Table 5.2: *Fluvoxamine Data. Maximum likelihood estimates and standard errors of BRD models. All observations included. No covariates.*

Effect	BRD1	BRD2	BRD3	BRD4	BRD5	BRD6	BRD7	BRD8	BRD9
Measurement model									
Int._{11}	0.22(0.15)	0.20(0.15)	0.28(0.15)	0.03(0.17)	0.32(0.15)	0.32(0.15)	0.14(0.16)	0.16(0.17)	0.27(0.15)
Int._{12}	-1.72(0.30)	-1.74(0.30)	-1.72(0.30)	-1.61(0.30)	-1.62(0.30)	-1.62(0.30)	-1.61(0.30)	-1.44(0.32)	-1.72(0.30)
Int._{21}	-0.12(0.18)	-0.12(0.18)	-0.05(0.18)	-0.42(0.23)	-0.13(0.18)	-0.13(0.18)	-0.31(0.21)	-0.39(0.22)	-0.04(0.17)
Dropout model									
α	-4.72(0.71)	-4.72(0.71)		-4.72(0.71)					
$\alpha_{1.}$					-3.87(0.71)	-3.93(0.71)		-3.93(0.71)	
$\alpha_{2.}$					$-\infty$	$-\infty$		$-\infty$	
$\alpha_{.1}$			-4.27(0.71)				-4.29(0.71)		-4.29(0.71)
$\alpha_{.2}$			$-\infty$				$-\infty$		$-\infty$
β	-1.09(0.13)		-1.09(0.13)		-1.09(0.13)				
$\beta_{1.}$		-1.37(0.22)				-1.37(0.22)			-1.37(0.22)
$\beta_{2.}$		-0.91(0.17)				-0.91(0.17)			-0.91(0.17)
$\beta_{.1}$				-1.57(0.38)			-1.57(0.38)	-1.56(0.37)	
$\beta_{.2}$				-0.55(0.29)			-0.56(0.29)	-0.56(0.29)	
γ	3.04(0.77)	3.04(0.77)	3.04(0.77)	3.04(0.77)	3.04(0.77)	3.31(0.79)	3.51(0.84)	3.31(0.79)	3.11(0.77)
- loglik	565.96	564.55	565.07	564.55	565.34	563.97	563.70	563.97	563.70

Table 5.3: *Fluvoxamine Data. Maximum likelihood estimates and standard errors of BRD models. All observations included. Duration as covariate in the measurement model.*

Effect	BRD1	BRD2	BRD3	BRD4	BRD5	BRD6	BRD7	BRD8	BRD9
Measurement model									
Int._{11}	0.46(0.17)	0.45(0.17)	0.53(0.17)	0.23(0.20)	0.57(0.17)	0.57(0.17)	0.35(0.18)	0.36(0.19)	0.52(0.18)
Int._{12}	-1.46(0.31)	-1.48(0.31)	-1.46(0.31)	-1.26(0.32)	-1.37(0.31)	-1.37(0.31)	-1.26(0.32)	-1.06(0.33)	-1.46(0.31)
Int._{21}	0.10(0.20)	0.10(0.19)	0.17(0.20)	-0.25(0.23)	0.09(0.21)	0.09(0.20)	-0.13(0.21)	-0.21(0.22)	0.18(0.20)
Duration	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)
Dropout model									
α	-4.71(0.71)	-4.71(0.71)		-4.71(0.71)					
$\alpha_{1.}$					-3.85(0.71)	-3.92(0.71)		-3.94(0.71)	
$\alpha_{2.}$					-∞	-∞		-∞	
$\alpha_{.1}$			-4.24(0.71)				-4.28(0.71)		-4.26(0.71)
$\alpha_{.2}$			-∞				-∞		-∞
β	-1.11(0.13)		-1.11(0.13)		-1.11(0.13)				
$\beta_{1.}$		-1.44(0.23)				-1.44(0.23)			-1.44(0.23)
$\beta_{2.}$		-0.90(0.17)				-0.90(0.17)			-0.90(0.17)
$\beta_{.1}$				-1.86(0.45)			-1.87(0.46)	-1.86(0.45)	
$\beta_{.2}$				-0.43(0.25)			-0.43(0.25)	-0.43(0.25)	
γ	2.98(0.77)	2.98(0.77)	2.98(0.77)	2.98(0.77)	2.98(0.77)	3.31(0.79)	3.74(0.89)	3.39(0.79)	3.07(0.77)
- loglik	550.15	548.31	549.12	546.60	549.39	547.57	545.55	545.84	547.30

Table 5.4: *Fluvoxamine Data. Maximum likelihood estimates and standard errors of BRD models. All observations included. Duration as covariate in both measurement and missingness model.*

Effect	BRD1	BRD2	BRD3	BRD4	BRD5	BRD6	BRD7	BRD8	BRD9
Measurement model									
Int.$_{11}$	0.46(0.18)	0.45(0.17)	0.53(0.18)	0.30 (0.20)	0.57(0.17)	0.57(0.17)	0.41(0.18)	0.43(0.19)	0.52(0.18)
Int.$_{12}$	-1.46(0.31)	-1.48(0.31)	-1.46(0.31)	-1.37 (0.31)	-1.37(0.31)	-1.37(0.31)	-1.37(0.31)	-1.22(0.33)	-1.46(0.31)
Int.$_{21}$	0.10(0.20)	0.10(0.20)	0.17(0.20)	-0.15 (0.24)	0.09(0.20)	0.09(0.21)	-0.04(0.22)	-0.13(0.23)	0.18(0.20)
Duration	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)	-0.02(0.01)
Dropout model									
$\alpha_{..}$	-4.57(0.72)	-4.57(0.72)		-4.57(0.72)					
$\alpha_{1.}$					-3.82(0.73)	-3.87(0.73)		-3.88(0.73)	
$\alpha_{2.}$					-∞	-∞		-∞	
$\alpha_{.1}$			-4.20(0.72)				-4.23(0.73)		-4.22(0.72)
$\alpha_{.2}$			-∞				-∞		-∞
α_{dur}	-0.02(0.02)	-0.02(0.02)	-0.01(0.02)	-0.02(0.02)	-0.01(0.02)	-0.01(0.02)	-0.01(0.02)	-0.00 (0.02)	-0.01(0.02)
$\beta_{..}$	-1.40(0.16)		-1.40(0.16)		-1.40(0.16)				
$\beta_{1.}$		-1.63(0.24)				-1.63(0.24)			-1.63(0.24)
$\beta_{2.}$		-1.22(0.20)				-1.22(0.20)			-1.22(0.20)
$\beta_{.1}$				-1.79(0.36)			-1.79(0.36)	-1.77(0.35)	
$\beta_{.2}$				-0.87(0.33)			-0.88(0.33)	-0.88(0.33)	
β_{dur}	0.02(0.01)	0.02(0.01)	0.02(0.01)	0.02 (0.01)	0.02(0.01)	0.02(0.01)	0.02(0.01)	0.02(0.01)	0.02(0.01)
γ	3.10(0.78)	3.10(0.78)	3.10(0.77)	3.10(0.78)	3.09 (0.78)	3.33(0.79)	3.50(0.84)	3.32(0.79)	3.16 (0.78)
- loglik	543.78	542.74	542.86	542.63	543.14	542.14	541.77	542.05	541.86
$p^{\dagger}$	0.0017	0.0038	0.0019	0.0189	0.0019	0.0044	0.0228	0.0226	0.0043

† p-value for the comparison with the corresponding BRD model in Table 5.3, to test the null hypothesis of no effect of duration in the missingness model.

parts. Sampling zeroes in some of the cells force certain parameters to lie on the boundary of their corresponding parameter space which, due to the parameterization, is equal to ∞. This should not be seen as a disadvantage of our model, since boundary solutions are a well known feature of MNAR models (Rubin, 1996). The advantage of our parameterization is that either an interior or a boundary solution is obtained, and never an invalid solution.

From Table 5.2, the likelihood ratio tests fails to reject BRD1 in favor of a more complex model, implying MCAR would be adequate. However, this conclusion changes when duration is included in the measurement model (Table 5.3). The effect of duration is highly significant, whichever of the BRD models is chosen to conduct a likelihood ratio test. Further, within Table 5.3, not BRD1 but rather BRD4 provides the most adequate description. The likelihood ratio test statistic for comparing BRD1–4 equals 7.10, while those for BRD4–7 and BRD4–8 are 2.10 and 1.52, respectively. Thus, from this set of models, one observes that duration improves the fit and apparently duration, included in the measurement model, has the effect of changing the nature of the missingness mechanism, by making it more complex, even though it is often believed that including explanatory variables may help explaining structure in the missingness mechanism. BRD4 states that missingness at the second occasion depends on the (possibly unobserved) value at that same occasion, a so-called type I model, in the typology of Baker (2000), in contrast to type II models, where missingness in a variable depends at least also on other, possibly incomplete, assessments. Obviously, such models are particularly vulnerable to assumptions. Up to this point, no covariate effects have been considered on the missingness parameters. When switching to Table 5.4, including duration in the missingness part, the conclusions change drastically. First, all evidence for non-MCAR missingness disappears and BRD1 comes out as the most adequate description. Second, comparing corresponding BRD models between Tables 5.3 and 5.4 (p-values in bottom line of Table 5.4), it is clear that the effect of duration on the missingness model cannot be neglected.

Important modeling and data analytic conclusions can be drawn. First, it clearly does not suffice to consider covariate effects on the measurement model, but one has to carefully contemplate such effects on the missingness model as well. Therefore, the models in Table 5.4, should be regarded as the ones of primary interest. Second, it is found that a longer duration implies a less favorable side effects outcome, as well as an increased change of missing visits. Obviously, duration acts as a confounding variable which, unless included in both parts of the model, may suggest a relationship between the measurement and missingness models and thus one may erroneously be led to

believe that the missing data are MNAR. Third, it should be noted that the parameter estimates of duration in the measurement part are remarkably stable. This implies that, in case one is primarily interested in the effect of duration on the occurrence of side effects all 18 models containing this effect (in Tables 5.3 and 5.4) provide very similar evidence. While this need not be the case in general, it is a comforting aspect of this particular data analysis. However, while we have reached plausible conclusions, one should still exercise caution, since non-random missingness models heavily rely on untestable assumptions (Verbeke and Molenberghs, 2000). Therefore, it is important to search for observations which may drive these conclusions (Verbeke *et al.*, 2001b). This naturally leads to sensitivity analysis, which will be undertaken in Chapter 8.

5.5 Conclusions

In this chapter, we have presented a set of analyses for incomplete binary data. Several plausible model strategies were considered, depending on the inclusion of duration as a covariate in the measurement and/or missingness models. To this end, a joint model for outcomes and non-response has been proposed in which (possibly continuous) covariates are allowed. The model is based on an extension of Baker, Rosenberger and DerSimonian (1992) towards the inclusion of covariates. While we focus on bivariate binary outcomes, the model can be extended to more than two assessments. To this end, extensions of (5.3) and (5.4) would have to be entertained. For both, log-linear type as well as marginal models can be considered. This model would be particularly attractive in the case of ordered missingness. This is an important feature since the particular model considered in this chapter suffers from interpretational problems when planned sequences are of unequal length and/or time-varying covariates are included.

In our case study, it turned out that the inclusion of a key covariate in both the measurement model and the missingness model, has the ability to substantially improve the fit of the model and to explain missingness in the sense that an otherwise seemingly MNAR mechanism is brought back to MCAR. The latter implies a number of methodological and interpretational advantages. The development of the extended BRD model family was published in Jansen *et al.* (2003).

6

A Dale-Dale Model for Categorical Outcomes with Non-Monotone Missingness

In Chapter 5, an extension of the Baker, Rosenberger and DerSimonian (1992) model for 2 binary outcomes with non-monotone missingness was proposed. Baker (1995) proposed a model for three binary outcomes with non-monotone missingness. Nevertheless, the domain of non-monotone missingness with multivariate ordinal responses is still relatively unexplored, and will be broached in this chapter as well as in Jansen and Molenberghs (2005). We propose a set of models based on the multivariate Dale model (Molenberghs and Lesaffre, 1994) for the measurements, as well as for the missingness mechanism. Since the data we will use only contain two responses of interest, the multivariate Dale model will be replaced by the bivariate Dale model in the application.

This chapter is organized as follows. In Section 6.1 we sketch the multivariate Dale model. This model will be reduced to the bivariate Dale model in Section 6.2, and in Section 6.3 it is shown how two copies of it can be combined, describing the measurement and missingness part of the model, respectively. Its application to the Health Interview Survey data is presented in Section 6.4.

6.1 The Multivariate Dale Model

The multivariate Dale model (Molenberghs and Lesaffre, 1994) extends the bivariate global cross-ratio model described by Dale (1986). This model accounts for the dependence between multiple ordinal responses, as well as for their dependence on covariate vector(s), which may be time-varying, continuous and/or discrete. The model arises from a decomposition of the joint probabilities into main effects (described by marginal probabilities) and interactions (described by cross-ratios of second and higher orders).

Let $i = 1, \ldots, N$ indicate the covariate level, containing n_i subjects. Every subject v at the ith level is evaluated at T distinct time points and at each visit the subject is scored using a categorical outcome variable. Hence, the outcome for subject v in the ith level is a series of measurements Y_{ivt} $(t = 1, \ldots, T)$, where Y_{ivt} can take on c_t distinct (possibly ordered) values j_t. Without loss of generality, we denote the category levels by $1, \ldots, c_t$. Along with the outcomes, a vector of covariates $\boldsymbol{x}$ is recorded, possibly time-dependent. For convenience, we assume that the first element of this covariate vector $\boldsymbol{x}$ equals 1, necessary for the intercept of the logistic regression. Both the marginal distributions and the cross-ratios can depend on these covariates.

Categorical data are typically presented in the form of frequency counts of observations. It is therefore convenient to summarize the categorical outcomes, measured for subjects with covariate vector $\boldsymbol{x}_i$, in a cross-classification of the outcomes Y_{ivt} into a $c_1 \times \ldots \times c_T$ dimensional contingency table with cell counts $Z_i^*(j_1, \ldots, j_T)$, denoting the number of subjects with outcome $(j_1, \ldots, j_T)$.

At every T-dimensional cutpoint $\boldsymbol{k} = (k_1, \ldots, k_T)$, the data table is collapsed into a $2 \times 2 \times \ldots \times 2$ table, each of which is assumed to arise as a discretization of a multivariate Plackett distribution (Plackett, 1965). In harmony with the desire to use cumulative measures, given the outcomes are ordinal, a data table of cumulative counts can be constructed:

$$Z_i(\boldsymbol{k}) = \sum_{\boldsymbol{\ell} \leq \boldsymbol{k}} Z_i^*(\boldsymbol{\ell}). \tag{6.1}$$

Thus, $Z_i(\boldsymbol{k})$ is just the number of individuals in group i whose observed response vector is $\boldsymbol{\ell}$, with $\boldsymbol{\ell} \leq \boldsymbol{k}$. The corresponding probabilities are

$$p_i(\boldsymbol{k}) = P(\boldsymbol{Y}_{iv} \leq \boldsymbol{k} | \boldsymbol{x}_i, \boldsymbol{\theta}) \tag{6.2}$$

and $p_i^*(\boldsymbol{k}) = P(\boldsymbol{Y}_{iv} = \boldsymbol{k} | \boldsymbol{x}_i, \boldsymbol{\theta})$, with $\boldsymbol{\theta}$ a vector of parameters of interest. Note that $Z_i(c_1, \ldots, c_T) = n_i$ and $p_i(c_1, \ldots, c_T) = 1$.

In addition, the marginal counts are given by all counts for which all but one index are equal to their maximal value: $Z_{itj_t} \equiv Z_i(c_1, \ldots, c_{t-1}, j_t, c_{t+1}, \ldots, c_T)$. Bivariate cell counts, i.e., cell counts of a cross-classification of a pair of outcomes, follow from setting all but two indices j_u equal to c_u, etc. Similarly, for example, bivariate probabilities pertaining to the tth and sth outcomes, are denoted by $p_{i,ts,j_tj_s} = p_i(c_1, \ldots, c_{t-1}, j_t, c_{t+1}, \ldots, c_{s-1}, j_s, c_{s+1}, \ldots, c_T)$. Generalizations to higher orders are straightforward. The order of the components is not important, but should be carried through the computations in a consistent fashion.

The multivariate Dale model involves describing T marginal distributions, $T(T-1)/2$ pairs of two-way interactions and three or higher order associations. The description is completed by specifying link functions and linear predictors for both the univariate margins and the association parameters. The latter are often assumed to be constant modeled on a log-odds ratio scale. For the univariate marginal links, a convenient choice is the logistic link function:

$$\eta_{itj_t} = \text{logit}(p_{itj_t}|\boldsymbol{x}_i) = \boldsymbol{\theta}'_{itj_t}\boldsymbol{x}_i, \quad (1 \leq t \leq T, 1 \leq j_t < c_t). \tag{6.3}$$

Full specification of the association is done in terms of marginal global odds ratios:

$$\varphi_{i,ts,j_tj_s} = \frac{(p_{i,ts,j_tj_s})(1 - p_{itj_t} - p_{isj_s} + p_{i,ts,j_tj_s})}{(p_{isj_s} - p_{i,ts,j_tj_s})(p_{itj_t} - p_{i,ts,j_tj_s})}. \tag{6.4}$$

They are usefully modeled on the log scale as

$$\eta_{i,ts,j_tj_s} = \ln \varphi_{i,ts,j_tj_s} = \boldsymbol{\theta}'_{i,ts,j_tj_s}\boldsymbol{x}_i, \quad (1 \leq t, s \leq T, 1 \leq j_t < c_t, 1 \leq j_s < c_s).$$

Higher order global odds ratios are easily introduced using ratios of conditional odds (ratios).

6.2 The Bivariate Dale Model

Since in the application we only have two responses of interest (mental health and fixed general practitioner), the multivariate Dale model can be replaced by its bivariate version (Dale, 1986). Both responses also only have two possible outcomes (good/bad and yes/no), such that it is not necessary to construct tables with cumulative counts and probabilities, and therefore the $*$ is omitted from the notation. The joint probabilities

$$p_{i,12,j_1j_2} = p_i(j_1, j_2) = P(Y_{iv1} = j_1, Y_{iv2} = j_2|\boldsymbol{x}_i), \quad (j_1, j_2 = 1, 2)$$

can be decomposed into two marginal distributions for the main effects, and one log cross-ratio for the association between both responses (indices i and $1, 2$ are omitted):

$$\begin{aligned} h_1(p_{1+}(\boldsymbol{x})) &= \boldsymbol{\theta}_1'\boldsymbol{x}, \\ h_2(p_{+1}(\boldsymbol{x})) &= \boldsymbol{\theta}_2'\boldsymbol{x}, \\ h_3\left(\frac{p_{11}(\boldsymbol{x})p_{22}(\boldsymbol{x})}{p_{12}(\boldsymbol{x})p_{21}(\boldsymbol{x})}\right) &= \boldsymbol{\theta}_3'\boldsymbol{x}, \end{aligned} \tag{6.5}$$

where h_1, h_2, and h_3 are link functions in the generalized linear model terminology, and $p_{1+}(\boldsymbol{x})$ and $p_{+1}(\boldsymbol{x})$ are the marginal probabilities for observing $Y_{v1} = 1$ and $Y_{v2} = 1$ respectively. The most popular choice for $h_1 \equiv h_2$ is the logit function, while for h_3 the natural logarithmic function is commonly used. This results in two marginal logistic regression models and the log cross-ratio

$$\ln \varphi = \ln \frac{p_{11}(\boldsymbol{x})p_{22}(\boldsymbol{x})}{p_{12}(\boldsymbol{x})p_{21}(\boldsymbol{x})}, \tag{6.6}$$

which is linear in the covariates. Solving equations (6.5) yields the probabilities $p_{j_1 j_2}(\boldsymbol{x})$ (the dependence on $\boldsymbol{x}$ is omitted for the ease of notation):

$$p_{11} = \begin{cases} \dfrac{1 + (p_{1+} + p_{+1})(\varphi - 1) - S(p_{1+}, p_{+1}, \varphi)}{2(\varphi - 1)} & \text{if } \varphi \neq 1, \\ p_{1+}p_{+1} & \text{if } \varphi = 1, \end{cases} \tag{6.7}$$

and

$$\begin{aligned} p_{12} &= p_{1+} - p_{11}, \\ p_{21} &= p_{+1} - p_{11}, \\ p_{22} &= 1 - p_{12} - p_{21} - p_{11}, \end{aligned} \tag{6.8}$$

with

$$S(\lambda_1, \lambda_2, \varphi) = \sqrt{[1 + (\lambda_1 + \lambda_2)(\varphi - 1)]^2 + 4\varphi(1 - \varphi)\lambda_1\lambda_2}. \tag{6.9}$$

Several extensions or variations to the model are possible, e.g., assume the associations to be constant, keep the intercepts and/or covariate parameters constant over time, include relations between the covariate parameters over time, etc.

6.3 Joint Model for the Measurement and Missingness Process

The bivariate Dale model, as introduced in Section 6.2, will be used for the measurement model and for the missingness model given the measurements, such that again

a selection model is obtained, as for the extended BRD model. Both discrete and continuous covariates can be included in both models.

Let $i = 1, \ldots, N$ index distinct covariate levels. In this section, the index i will be suppressed from notation. Let $j_1, j_2 = 1, 2$ correspond to the outcome categories of the first and second measurement, respectively and let $r_1, r_2 = 0, 1$ correspond to the missingness indicators (1 for an observed and 0 for a missing measurement). The complete data and observed data cell probabilities $\pi_{r_1 r_2, j_1 j_2}$ for this setting are identical to the ones for the BRD model in Table 5.1, and can be factorized as:

$$\pi_{r_1 r_2, j_1 j_2} = p_{j_1 j_2} q_{r_1 r_2 | j_1 j_2}, \tag{6.10}$$

where $p_{j_1 j_2}$ parameterizes the measurement process and $q_{r_1 r_2 | j_1 j_2}$ describes the missingness mechanism, conditional on the measurements, resulting in a selection model. In particular, we will assume

$$\begin{aligned} \eta_1 &= \text{logit } p_{1+} &&= X_1 \boldsymbol{\theta}, \\ \eta_2 &= \text{logit } p_{+1} &&= X_2 \boldsymbol{\theta}, \\ \eta_3 &= \ln \varphi_p &&= X_3 \boldsymbol{\theta}, \end{aligned} \tag{6.11}$$

$$\begin{aligned} \delta_{4|j_1 j_2} &= \text{logit } q_{1+|j_1 j_2} &&= Z_{1|j_1 j_2} \boldsymbol{\psi}, \\ \delta_{5|j_1 j_2} &= \text{logit } q_{+1|j_1 j_2} &&= Z_{2|j_1 j_2} \boldsymbol{\psi}, \\ \delta_{6|j_1 j_2} &= \ln \varphi_q &&= Z_{3|j_1 j_2} \boldsymbol{\psi}. \end{aligned} \tag{6.12}$$

As in Baker, Rosenberger and DerSimonian (1992) and Jansen *et al.* (2003), we will consider nine identifiable models:

Model 1 :	$(\psi^{(1)}_{\cdot\cdot}, \psi^{(2)}_{\cdot\cdot})$	Model 4 :	$(\psi^{(1)}_{\cdot\cdot}, \psi^{(2)}_{\cdot j_2})$	Model 7 :	$(\psi^{(1)}_{\cdot j_2}, \psi^{(2)}_{\cdot j_2})$
Model 2 :	$(\psi^{(1)}_{\cdot\cdot}, \psi^{(2)}_{j_1\cdot})$	Model 5 :	$(\psi^{(1)}_{j_1\cdot}, \psi^{(2)}_{\cdot\cdot})$	Model 8 :	$(\psi^{(1)}_{j_1\cdot}, \psi^{(2)}_{\cdot j_2})$
Model 3 :	$(\psi^{(1)}_{\cdot j_2}, \psi^{(2)}_{\cdot\cdot})$	Model 6 :	$(\psi^{(1)}_{j_1\cdot}, \psi^{(2)}_{j_1\cdot})$	Model 9 :	$(\psi^{(1)}_{\cdot j_2}, \psi^{(2)}_{j_1\cdot})$,

where $\psi^{(1)}_{j_1 j_2}$ models the non-response for the first outcome and $\psi^{(2)}_{j_1 j_2}$ models the non-response for the second outcome. Index j_1 indicates the dependence on the first outcome, index j_2 on the second outcome. The nesting structure of these models is schematically represented in Figure 6.1. Interpretation is similar to the BRD models. For example, Model 1 is MCAR, in Model 4 missingness in the first variable is constant, while missingness in the second variable depends on its value.

No a priori ordering is imposed on the outcomes. The advantage is that genuine multivariate settings (e.g., several questions in a survey) can be handled as well as

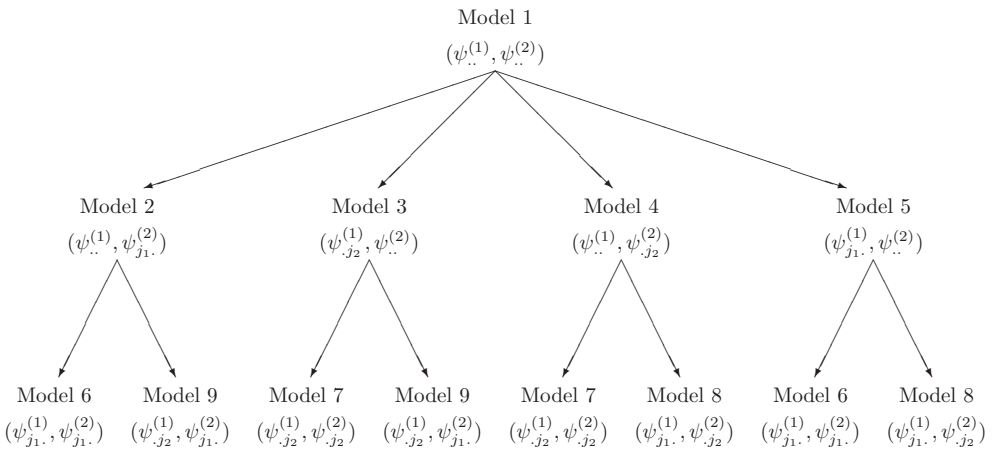


Figure 6.1: *Graphical representation of the Dale-Dale model nesting structure.*

longitudinal studies. When deemed necessary, the implications of ordering can be imposed by considering specific models and leaving out others. For example, one may want to avoid missingness on future observations. In the current bivariate case, the index j_2 would have to be removed from $\psi^{(1)}$ in the above model.

Models (6.11) and (6.12) can be combined into the model

$$\boldsymbol{\xi} = W\boldsymbol{\phi}, \tag{6.13}$$

where

$$\boldsymbol{\xi} = \left(\eta_1, \eta_2, \eta_3, \delta_{4|11}, \delta_{5|11}, \delta_{6|11}, \delta_{4|12}, \delta_{5|12}, \delta_{6|12}, \delta_{4|21}, \delta_{5|21}, \delta_{6|21}, \delta_{4|22}, \delta_{5|22}, \delta_{6|22}\right)',$$

$$W = \begin{pmatrix} X & 0 \\ 0 & Z \end{pmatrix},$$

and $\boldsymbol{\phi} = \left(\boldsymbol{\theta}', \boldsymbol{\psi}'\right)'$. This model will be referred to as the Dale-Dale model, a Dale model being used for both the measurement and missingness processes.

The design matrices X and Z, and the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ depend on the number of covariates that are taken into account, and whether or not their influence can be different on both outcomes. Let us assume that we include one covariate that is constant, and one that has a different influence on both outcomes. Then $\boldsymbol{\theta}$, $\boldsymbol{\psi}$, X and Z can be specified as follows:

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)',$$

$$\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8)',$$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \text{cov}_1 & \text{cov}_2 & 0 \\ 0 & 1 & 0 & \text{cov}_1 & 0 & \text{cov}_2 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

and

$$Z = \begin{pmatrix} Z_{1|11} \\ Z_{2|11} \\ Z_{3|11} \\ Z_{1|12} \\ Z_{2|12} \\ Z_{3|12} \\ Z_{1|21} \\ Z_{2|21} \\ Z_{3|21} \\ Z_{1|22} \\ Z_{2|22} \\ Z_{3|22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \text{cov}_1 & \text{cov}_2 & 0 & 0 & 0 \\ 0 & 1 & 0 & \text{cov}_1 & 0 & \text{cov}_2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \text{cov}_1 & \text{cov}_2 & 0 & I_1 & 0 \\ 0 & 1 & 0 & \text{cov}_1 & 0 & \text{cov}_2 & 0 & J_1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \text{cov}_1 & \text{cov}_2 & 0 & I_2 & 0 \\ 0 & 1 & 0 & \text{cov}_1 & 0 & \text{cov}_2 & 0 & J_2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \text{cov}_1 & \text{cov}_2 & 0 & I_3 & 0 \\ 0 & 1 & 0 & \text{cov}_1 & 0 & \text{cov}_2 & 0 & J_3 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $I_1 = 1$ in Models 3, 7 and 9, $I_2 = 1$ in Models 5, 6 and 8, $I_3 = 1$ if $I_1 = 1$ or $I_2 = 1$, $J_1 = 1$ in Models 4, 7 and 8, $J_2 = 1$ in Models 2, 6 and 9, $J_3 = 1$ if $J_1 = 1$ or $J_2 = 1$. These indicators are equal to 0 in all other models.

To include individual-specific covariates, a subscript i has to be introduced to the vector $\boldsymbol{\xi}$ and the matrix W and hence to their constituent components. $\boldsymbol{\theta}$, $\boldsymbol{\psi}$, X, and Z can easily be changed to increase or reduce the number of covariates.

6.4 Models Fitted to the Health Interview Survey Data

In the reported analysis all patients with known covariate information are considered. The effect of the covariate on the marginal probabilities can either be constant, or different for both probabilities. We assume that this choice is identical for the measurement model and the missingness model. So either θ_4 and ψ_4 in (6.13) are

included in the model, or θ_5, θ_6 and ψ_5, ψ_6, for parsimony. The association between both outcomes is assumed to be constant in all settings.

We will consider five different sets of Dale-Dale models in detail. First, models are considered without covariates. Results (estimates, standard errors, and negative loglikelihoods) are presented in Table 6.1. Second, a constant effect of gender is added (Table 6.2), with distribution of the data over males and females shown in Table 2.4. In Table 6.3 the effect of gender is allowed to differ for both marginal probabilities. Next, education is included. Table 6.4 shows the results of the models with constant effect of education, while in Table 6.5 this effect changes over the marginal probabilities. Table 2.5 contains the distribution of the data over the different categories of education. Several other ways to include covariate information are possible (only including a covariate effect in the measurement or missingness model, a constant covariate effect in one part of the model and a varying effect in the other part, more than one covariate effect, ...), but will not be presented here for conciseness.

Sampling zeroes or small counts in some of the cells forces certain parameters to lie on the boundary of their corresponding parameter space which, due to the parameterization, is equal to ∞. This should not be seen as a disadvantage of our model, since boundary solutions are a well known feature of MNAR models in general (Rubin, 1996). The advantage of our parameterization (see also Chapter 5) is that either an interior or a boundary solution is obtained, but never an invalid solution.

From Table 6.1, the likelihood ratio tests reject Model 1 in favor of Model 3 (likelihood ratio test = 24.14) and Model 4 ($LRT = 22.28$). Since these models make different assumptions about the missingness mechanism (in Model 3 a missing value for mental health depends on the (un)observed value for fixed general practitioner, in Model 4 a missing value for fixed general practitioner depends on the (un)observed value for fixed general practitioner), both have to be considered. The extension to Model 7, which combines the missingness mechanism assumptions from Model 3 and Model 4, is not significant (Models 3 and 7: $LRT = 0.00$, meaning they are equivalent at the observed data level, Models 4 and 7: $LRT = 1.86$). We also notice that in Model 4 the estimate for ψ_8, the MNAR parameter, lies on the boundary of the parameter space, which makes Model 4 less favorable than Model 3. Nevertheless, a more careful study is necessary to get insight into why two models with completely different assumptions about the missingness mechanism are significant. But first we will discuss the results of the analyses where covariates are included into the model.

The effect of gender, assumed to be equal for both marginal probabilities, on the measurement and missingness model simultaneously, is highly significant, whichever of the models is chosen to conduct a likelihood ratio test (with 2 degrees of freedom).

Table 6.1: *Health Interview Survey Data. Maximum likelihood estimates and standard errors of the Dale-Dale models. All observations included. No covariates.*

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Measurement model									
θ_1	0.76(0.02)	0.76(0.02)	0.76(0.02)	0.76(0.02)	0.87(0.08)	1.13(0.02)	0.76(0.02)	0.86(0.07)	0.76(0.02)
θ_2	2.17(0.03)	2.17(0.03)	2.17(0.03)	2.24(0.03)	2.17(0.03)	2.17(0.03)	2.16(0.03)	2.24(0.03)	2.17(0.03)
θ_3	0.03(0.08)	0.02(0.08)	0.03(0.08)	0.03(0.07)	-0.02(0.08)	-0.08(0.08)	0.03(0.08)	0.02(0.07)	0.03(0.08)
θ_4									
θ_5									
θ_6									
Missingness model									
ψ_1	1.16(0.02)	1.16(0.02)	1.20(0.02)	1.16(0.02)	1.02(0.08)	0.77(0.02)	1.20(0.02)	1.03(0.07)	1.20(0.02)
ψ_2	2.83(0.04)	2.97(0.13)	2.83(0.04)	2.71(0.04)	2.83(0.04)	2.56(0.04)	2.86(0.04)	2.71(0.04)	2.97(0.13)
ψ_3	4.35(0.17)	4.37(0.18)	4.36(0.17)	4.40(0.16)	4.43(0.20)	4.54(0.19)	4.35(0.17)	4.46(0.19)	4.38(0.18)
ψ_4									
ψ_5									
ψ_6									
ψ_7			-0.33(0.06)		0.55(0.38)	$+\infty$	-0.38(0.06)	0.50(0.31)	-0.33(0.06)
ψ_8		-0.40(0.30)		$+\infty$		2.75(0.30)	-0.29(0.02)	$+\infty$	-0.40(0.30)
-loglik	15991.13	15990.32	15979.06	15979.99	15990.33	15989.74	15979.06	15979.17	15978.21

Table 6.2: *Health Interview Survey Data. Maximum likelihood estimates and standard errors of the Dale-Dale models. All observations included. Constant effect of gender.*

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Measurement model									
θ_1	1.00(0.06)	1.00(0.06)	1.00(0.06)	1.00(0.06)	0.73(0.16)	0.20(0.05)	1.00(0.06)	1.03(0.14)	1.00(0.06)
θ_2	2.42(0.06)	2.42(0.06)	2.42(0.06)	2.49(0.06)	2.40(0.07)	2.30(0.05)	2.49(0.06)	2.49(0.06)	2.42(0.06)
θ_3	0.04(0.07)	0.03(0.07)	0.04(0.08)	0.04(0.07)	0.14(0.09)	0.21(0.07)	0.04(0.07)	0.04(0.07)	0.04(0.07)
θ_4	-0.16(0.04)	-0.16(0.04)	-0.16(0.04)	-0.16(0.04)	-0.14(0.04)	-0.09(0.03)	-0.16(0.04)	-0.16(0.04)	-0.16(0.04)
θ_5									
θ_6									
Missingness model									
ψ_1	1.00(0.07)	1.00(0.07)	1.06(0.07)	1.00(0.05)	1.30(0.24)	$+\infty$	1.02(0.07)	0.99(0.11)	1.05(0.07)
ψ_2	2.67(0.08)	2.79(0.14)	2.69(0.08)	2.56(0.07)	2.55(0.10)	5.25(0.26)	2.57(0.07)	2.57(0.09)	2.81(0.14)
ψ_3	4.35(0.18)	4.36(0.18)	4.36(0.18)	4.39(0.16)	4.56(0.30)	4.04(0.24)	4.39(0.17)	4.40(0.18)	4.37(0.18)
ψ_4	0.10(0.04)	0.10(0.04)	0.09(0.04)	0.10(0.04)	0.19(0.06)	0.38(0.05)	0.10(0.04)	0.10(0.06)	0.09(0.04)
ψ_5									
ψ_6									
ψ_7			-0.32(0.06)		-0.97(0.54)	$-\infty$	-0.09(0.07)	0.09(0.50)	-0.32(0.05)
ψ_8		-0.34(0.30)		$+\infty$		-3.75(0.26)	$+\infty$	$+\infty$	-0.35(0.31)
-loglik	15979.79	15979.17	15968.29	15968.30	15979.34	15957.23	15967.55	15968.29	15967.63

Table 6.3: *Health Interview Survey Data. Maximum likelihood estimates and standard errors of the Dale-Dale models. All observations included. Effect of gender different on both marginal probabilities.*

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Measurement model									
θ_1	1.44(0.07)	1.44(0.07)	1.44(0.07)	1.44(0.07)	1.53(0.12)	0.52(0.12)	1.44(0.07)	1.55(0.11)	1.44(0.07)
θ_2	1.60(0.10)	1.60(0.10)	1.60(0.09)	1.68(0.08)	1.60(0.10)	1.59(0.09)	1.47(0.16)	1.68(0.08)	1.60(0.10)
θ_3	0.09(0.08)	0.08(0.08)	0.09(0.08)	0.09(0.07)	0.05(0.09)	0.23(0.07)	0.09(0.08)	0.07(0.08)	0.09(0.08)
θ_4									
θ_5	-0.45(0.04)	-0.45(0.04)	-0.45(0.04)	-0.45(0.04)	-0.45(0.04)	-0.27(0.05)	-0.45(0.04)	-0.45(0.05)	-0.45(0.04)
θ_6	0.39(0.06)	0.39(0.06)	0.39(0.06)	0.38(0.05)	0.39(0.06)	0.39(0.06)	0.40(0.06)	0.38(0.05)	0.39(0.06)
Missingness model									
ψ_1	1.01(0.06)	1.01(0.06)	1.06(0.07)	1.01(0.06)	0.95(0.09)	3.29(1.32)	1.14(0.11)	0.95(0.08)	1.06(0.07)
ψ_2	2.62(0.13)	2.68(0.14)	2.63(0.12)	2.49(0.11)	2.62(0.13)	4.65(0.76)	2.95(0.45)	2.49(0.12)	2.70(0.14)
ψ_3	4.35(0.18)	4.36(0.18)	4.36(0.17)	4.39(0.16)	4.39(0.19)	4.09(0.25)	4.29(0.19)	4.44(0.19)	4.37(0.18)
ψ_4									
ψ_5	0.10(0.04)	0.10(0.04)	0.09(0.04)	0.10(0.04)	0.07(0.06)	0.39(0.05)	0.07(0.05)	0.06(0.05)	0.09(0.04)
ψ_6	0.14(0.08)	0.18(0.08)	0.14(0.08)	0.15(0.08)	0.14(0.08)	0.29(0.08)	0.10(0.10)	0.15(0.08)	0.17(0.09)
ψ_7			-0.32(0.06)		0.38(0.47)	-3.82(1.42)	-0.68(0.37)	0.44(0.38)	-0.32(0.06)
ψ_8		-0.33(0.30)		$+\infty$		-3.01(0.77)	-1.36(1.13)	$+\infty$	-0.36(0.30)
-loglik	15926.04	15925.49	15914.42	15915.71	15925.78	15920.05	15914.04	15915.14	15913.76

Further, within Table 6.2, Model 3 and Model 4 still provide the most adequate description. The likelihood ratio test statistics for comparing Models 1 and 3 and Models 1 and 4 equal 23.00 and 22.98, respectively, while those for extending Model 3 and Model 4 further along the edges of Figure 6.1, are all smaller than 1.50, and therefore not significant. Thus, from this set of models, one observes that gender improves the fit but the nature of the missingness mechanism does not change. In Chapter 5 a similar result was found. Without covariates a MCAR model was found to be the most adequate one. Including the effect of duration in the measurement model changed the nature of the missingness mechanism, by making it more complex. When further including duration in the missingness part, all evidence for non-MCAR missingness disappeared and again the same MCAR model was found to be best. It was also clear that the effect of duration on the missingness model could not be neglected.

Including a varying gender effect instead of a constant one, again improves the fit a lot (in all models $LRT > 100$ with $2df$). Here too, Models 3 and 4 are to be preferred, although the extension of Model 4 to Model 7 is borderline non-significant ($LRT = 3.34$), and in Model 7 the estimate of ψ_8 no longer lies on the boundary of the parameter space. Also worthwhile to mention is that in the measurement model, gender has an opposite effect on both marginal probabilities (negative for mental health, positive for fixed general practitioner), while in the missingness model there is a borderline significant positive effect. The negative effect of gender on mental health can be interpreted as men having a higher probability of bad mental health than women, while the positive effect of gender on fixed general practitioner means that men are having a higher probability of having a fixed general practitioner than women.

For the covariate education the same conclusions can be drawn concerning the model selection. Models 3 and 4, with a varying education effect, provide the most adequate description from all models in Tables 6.4–6.5, although the extension to Model 7 is again borderline non-significant, and will be considered as well. In the measurement model, the effect of education seems to be significant only for the marginal probability of fixed general practitioner, not for the marginal probability of mental health, with the interpretation that the higher the education, the higher the probability of having a fixed general practitioner. In the missingness model, the influence of education is opposite for both marginal probabilities (negative for missingness in mental health, positive for missingness in fixed general practitioner).

Important modeling and data analytic conclusions can be drawn. It is clear that covariate effects need not be the same on all marginal probabilities, in the measure-

Table 6.4: *Health Interview Survey Data. Maximum likelihood estimates and standard errors of the Dale-Dale models. 81 observations not included. Constant effect of education.*

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Measurement model									
θ_1	0.94(0.07)	0.94(0.07)	0.94(0.07)	0.96(0.06)	0.84(0.11)	0.34(0.06)	0.96(0.07)	0.90(0.09)	0.93(0.07)
θ_2	2.37(0.07)	2.36(0.07)	2.37(0.07)	2.45(0.06)	2.43(0.09)	2.44(0.07)	2.45(0.07)	2.48(0.08)	2.36(0.07)
θ_3	0.01(0.08)	0.01(0.08)	0.01(0.08)	0.01(0.07)	0.07(0.09)	0.17(0.08)	0.01(0.08)	0.01(0.08)	0.01(0.08)
θ_4	-0.05(0.02)	-0.05(0.02)	-0.05(0.02)	-0.05(0.01)	-0.06(0.02)	-0.07(0.02)	-0.05(0.02)	-0.06(0.02)	-0.05(0.02)
θ_5									
θ_6									
Missingness model									
ψ_1	1.43(0.08)	1.43(0.08)	1.44(0.08)	1.43(0.07)	1.67(0.26)	$+\infty$	1.43(0.07)	1.57(0.16)	1.44(0.08)
ψ_2	3.10(0.09)	3.15(0.15)	3.09(0.09)	2.99(0.08)	3.09(0.09)	5.91(0.24)	2.99(0.08)	2.99(0.09)	3.14(0.14)
ψ_3	4.36(0.18)	4.37(0.18)	4.37(0.18)	4.41(0.16)	4.46(0.24)	4.04(0.25)	4.41(0.18)	4.44(0.19)	4.37(0.18)
ψ_4	-0.07(0.02)	-0.07(0.02)	-0.06(0.02)	-0.07(0.02)	-0.07(0.02)	-0.03(0.02)	-0.07(0.02)	-0.07(0.02)	-0.06(0.02)
ψ_5									
ψ_6									
ψ_7			-0.31(0.06)		-0.65(0.56)	$-\infty$	-0.08(0.08)	-0.39(0.36)	-0.31(0.06)
ψ_8		0.15(0.33)		$+\infty$		-3.71(0.24)	$+\infty$	$+\infty$	-0.16(0.33)
-loglik	15839.06	15838.95	15828.54	15827.65	15838.57	15832.29	15827.09	15827.36	15828.42

Table 6.5: *Health Interview Survey Data. Maximum likelihood estimates and standard errors of the Dale-Dale models. 81 observations not included. Effect of education different on both marginal probabilities.*

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Measurement model									
θ_1	0.72(0.08)	0.72(0.08)	0.72(0.08)	0.72(0.07)	0.79(0.09)	0.99(0.07)	0.72(0.08)	0.78(0.09)	0.72(0.08)
θ_2	2.85(0.12)	2.85(0.12)	2.85(0.12)	2.95(0.11)	2.85(0.12)	2.85(0.12)	2.95(0.12)	2.95(0.12)	2.85(0.12)
θ_3	0.02(0.08)	0.02(0.08)	0.02(0.08)	0.02(0.07)	-0.02(0.08)	-0.08(0.07)	0.02(0.08)	0.01(0.08)	0.02(0.08)
θ_4									
θ_5	0.01(0.02)	0.01(0.02)	0.01(0.02)	0.01(0.02)	0.02(0.02)	0.04(0.02)	0.01(0.02)	0.02(0.02)	0.01(0.02)
θ_6	-0.17(0.03)	-0.17(0.03)	-0.17(0.03)	-0.18(0.03)	-0.17(0.03)	-0.17(0.03)	-0.18(0.03)	-0.18(0.03)	-0.17(0.03)
Missingness model									
ψ_1	1.53(0.07)	1.53(0.08)	1.54(0.08)	1.53(0.07)	1.40(0.11)	1.14(0.08)	1.53(0.08)	1.42(0.11)	1.54(0.08)
ψ_2	2.14(0.12)	2.27(0.16)	2.13(0.12)	2.05(0.11)	2.15(0.13)	1.83(0.12)	2.06(0.11)	2.05(0.11)	2.26(0.17)
ψ_3	4.42(0.18)	4.43(0.18)	4.43(0.18)	4.47(0.17)	4.47(0.20)	4.61(0.20)	4.47(0.16)	4.51(0.19)	4.44(0.18)
ψ_4									
ψ_5	-0.09(0.02)	-0.09(0.02)	-0.09(0.02)	-0.09(0.02)	-0.09(0.02)	-0.09(0.02)	-0.09(0.02)	-0.09(0.02)	-0.09(0.02)
ψ_6	0.19(0.03)	0.19(0.03)	0.19(0.03)	0.18(0.03)	0.19(0.03)	0.20(0.03)	0.18(0.03)	0.18(0.03)	0.19(0.03)
ψ_7			-0.30(0.06)		0.44(0.33)	$+\infty$	-0.07(0.07)	0.38(0.29)	-0.30(0.06)
ψ_8		-0.35(0.28)		$+\infty$		2.74(0.32)	$+\infty$	$+\infty$	-0.35(0.27)
-loglik	15789.46	15788.71	15779.33	15778.05	15788.78	15787.34	15777.57	15777.41	15778.58

ment model as well as in the missingness model. Thus, a varying covariate effect needs to be investigated. It should also be noted that the parameter estimates of the covariates in the measurement part are remarkably stable among the 9 Dale-Dale models. This implies that, in case one is primarily interested in a specific covariate effect on mental health and fixed general practitioner, all models containing this effect provide very similar evidence. This need not be the case in general. However, attention is more and more devoted to the nature of the missingness mechanism. Therefore, one should still exercise caution, since non-random missingness models heavily rely on untestable assumptions (Verbeke and Molenberghs, 2000). In our analyses, more than one model, with completely different assumptions, turned out to describe the data well. So it is important to search for observations which may drive these conclusions (Verbeke *et al.*, 2001b).

6.5 Conclusions

We have presented a set of models for multivariate ordinal data, based on the multivariate Dale model (Molenberghs and Lesaffre, 1994). A natural hierarchy between the models exist. We considered two binary outcomes, and therefore a bivariate Dale model for the outcomes is combined with the same bivariate Dale model for the non-response, resulting in the so-called Dale-Dale model. Several plausible model strategies (the inclusion of constant or varying covariate effects in both the measurement and missingness model) have been considered. While we focus on bivariate binary outcomes, the model formulation to incorporate more than two assessments and/or possibly ordinal outcomes, is straightforward, using the multivariate Dale model. The Dale-Dale model is particularly attractive when a varying covariate effect is deemed necessary. In our case study, the models allowing for this varying covariate effect in both the measurement and the missingness model, provided the best fit.

In our data analysis, the estimates of the measurement model parameters are remarkably stable, no matter which assumptions are made regarding the reasons for non-response. When interest lies only in the effect of a specific covariate on the measurement model, all models containing this effect provide very similar evidence. The model selection (independent of the included covariate effects) pointed out that a missing value for either mental health or fixed general practitioner depends on the value of fixed general practitioner. The model that combines both assumptions is (borderline) non-significant. To strengthen the data analytic findings, it is best to conduct a sensitivity analysis, which will be the topic of Chapter 8.

7

Pattern-Mixture Models for Categorical Outcomes with Non-Monotone Missingness

Whereas most models for incomplete longitudinal data are formulated within the selection model framework, pattern-mixture models have gained considerable interest in recent years (Little, 1993, 1994), since it is often argued that selection models, although identifiable, should be approached with caution, especially in the context of MNAR models (Glynn, Laird and Rubin, 1986).

In this chapter, focus is on several strategies to fit pattern-mixture models for non-monotone categorical outcomes. Also in this setting the issue of under-identification in pattern-mixture models is present. Little (1993, 1994) solves this problem through the use of identifying restrictions: inestimable parameters of the incomplete patterns are set equal to (functions of) the parameters describing the distribution of the completers. In this way, the conditional distribution of the unobserved measurements, given the observed ones in a specific pattern, is specified. While several authors perceive this under-identification as a drawback, we believe it is an asset since it forces one to reflect on the assumptions made.

Generally, interest is not only in the pattern-specific effect of included covariates

(e.g., treatment effect), but also in the overall effect of this covariate. However, this cannot be obtained by simply averaging the pattern-specific effects, as is the case when using linear mixed models. Therefore, attention will also be given to the derivation of the marginal covariate effect in pattern-mixture models for non-monotone categorical data.

In Section 7.1, the general context of pattern-mixture models will be sketched. The strategy of identifying restrictions is the topic in Section 7.2, while Section 7.3 focuses on the special case of three measurements. Section 7.4 gives attention to the use of the multivariate Dale model, introduced in Section 6.1, to fit pattern-mixture models for categorical outcomes, while Section 7.5 will discuss the assumptions needed when intermittent missingness is present. Section 7.6 focuses on the derivation of marginal effects in pattern-mixture models. Finally, in Section 7.7, the developed techniques will be used to reanalyze the fluvoxamine data.

7.1 General Form of Pattern-Mixture Models

The family of pattern-mixture models is based on the factorization

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|\boldsymbol{r}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{\psi}),$$

where dependence on covariates is suppressed from notation. When restricting attention to dropout only, we obtain

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

Thus, the conditional density of the measurements given the dropout pattern is combined with the marginal density describing the dropout mechanism. Note that the second factor can depend on covariates, but not on outcomes. It is, of course, possible to have different covariate dependencies in both components of the factorization.

The measurement model has to reflect dependence on dropout. We will illustrate this idea on a simple setting. Consider a continuous response at three measurement occasions, modeled using a trivariate Gaussian distribution. Assume that dropout may occur at time points 2 or 3, and let the dropout indicator t_i take value 1 or 2 indicating that the last observation occurred at this time point, and 3 indicating there is no dropout. Then, in a first instance, the model assumes a different distribution for each dropout pattern. We can write

$$\boldsymbol{y}_i \mid t_i \quad \sim \quad N(\boldsymbol{\mu}(t_i), \Sigma(t_i)), \tag{7.1}$$

where

$$\boldsymbol{\mu}(t) = \begin{pmatrix} \mu_1(t) \\ \mu_2(t) \\ \mu_3(t) \end{pmatrix} \quad \text{and} \quad \Sigma(t) = \begin{pmatrix} \sigma_{11}(t) & \sigma_{21}(t) & \sigma_{31}(t) \\ \sigma_{21}(t) & \sigma_{22}(t) & \sigma_{32}(t) \\ \sigma_{31}(t) & \sigma_{32}(t) & \sigma_{33}(t) \end{pmatrix},$$

for $t = 1, 2, 3$. Recall that t indicates the length of a sequence. Let $P(t) = \pi_t = f(t_i|\boldsymbol{\psi})$, then the marginal distribution of the response is a mixture of normals with, for example, mean

$$\boldsymbol{\mu} = \sum_{t=1}^{3} \pi_t \boldsymbol{\mu}(t).$$

Its variance can be derived by application of the delta method. Suppose that, for large n, T_n is normally distributed around θ with standard error $\frac{\sigma}{\sqrt{n}}$, and let g be a function that is at least twice differentiable at θ, then $g(T_n)$ will be approximately normal around $g(\theta)$ with variance $\dfrac{[g'(\theta)]^2\sigma^2}{n}$ (Agresti, 2002).

However, although the π_t can be simply estimated from the observed proportions in each dropout group, only 16 of the 27 response parameters can be identified from the data without making further assumptions. These 16 parameters comprise all those from the completers plus those from the following two submodels:

$$N\left(\begin{pmatrix} \mu_1(2) \\ \mu_2(2) \end{pmatrix}; \begin{pmatrix} \sigma_{11}(2) & \sigma_{21}(2) \\ \sigma_{21}(2) & \sigma_{22}(2) \end{pmatrix}\right) \quad \text{and} \quad N\left(\mu_1(1); \sigma_{11}(1)\right).$$

This is a *saturated* pattern-mixture model and the representation makes it clear which information is provided by each dropout group and, consequently, which assumptions are needed in order to predict the behavior of the unobserved responses, and to obtain marginal models for the response. However, this model contains under-identified members since it describes the full set of measurements in pattern t_i, even though there are no measurements after occasion t_i. At first sight, this leaves them open to the same criticism as selection models, but Little (1993) claims that the pattern-mixture approach is more honest, because parameters for which the data provide information are clearly distinguished from parameters for which there is no information at all. Several strategies can be followed to solve this problem of under-identification. A first one is based on Little (1993, 1994), who advocated the use of identifying restrictions which works well in relatively simple settings. Molenberghs *et al.* (1998) proposed a particular set of restrictions for the monotone case which corresponds to MAR and in Thijs *et al.* (2002) a formal way how to deal with these

kind of restrictions is introduced. Section 7.2 will focus on these identifying restrictions. Alternatively, several types of simplified (identified) models can be considered. The advantage is that the number of parameters decreases, which is generally an issue with pattern-mixture models. Hogan and Laird (1997) noted that in order to estimate the large number of parameters in general pattern-mixture models, one has to make the awkward requirement that each dropout pattern is sufficiently "filled", in other words, one has to require large numbers of dropouts. This problem is less prominent in simplified models. Note however that simplified models, qualified as "assumption rich" by Sheiner, Beal and Dunne (1997), are also making untestable assumptions and therefore illustrate that even pattern-mixture models do not provide a free lunch. A main advantage however is that the need of assumptions and their implications is more obvious.

7.2 Identifying Restrictions

For the time being, we restrict attention to monotone patterns. In general, let us assume that we have patterns t ($t = 1, \ldots, n$, but not necessary all of them are present), where the dropout indicator is $d = t + 1$. For pattern t, the complete data density is given by

$$f_t(y_1, \ldots, y_n) = f_t(y_1, \ldots, y_t) f_t(y_{t+1}, \ldots, y_n | y_1, \ldots, y_t). \tag{7.2}$$

The first factor is clearly identified from the observed data, while the second factor is not. It is assumed that the first factor is known or, more realistically, can be modeled using the observed data. Then, identifying restrictions are applied in order to identify the second component.

While, in principle, completely arbitrary restrictions can be used by means of any valid density function over the appropriate support, strategies which relate back to the observed data deserve privileged interest. One can base identification on all patterns for which a given component y_s is identified. A general expression for this is

$$f_t(y_s | y_1, \ldots y_{s-1}) = \sum_{j=s}^{n} \omega_{sj} f_j(y_s | y_1, \ldots y_{s-1}), \quad s = t+1, \ldots, n. \tag{7.3}$$

Let $\boldsymbol{\omega}_s = (\omega_{ss}, \ldots, \omega_{sn})'$. Every $\boldsymbol{\omega}_s$ with components summing to one, provides a valid identification scheme. Let us incorporate (7.3) into (7.2):

$$f_t(y_1, \ldots, y_n) = f_t(y_1, \ldots, y_t) \prod_{k=0}^{n-t-1} \left[\sum_{j=n-k}^{n} \omega_{n-k,j} f_j(y_{n-k} | y_1, ..., y_{n-k-1}) \right]. \tag{7.4}$$

Expression (7.4) clearly shows which information is used to complement the observed data density in pattern t in order to establish the complete data density.

Let us consider three special but important cases. Little (1993) proposed *complete case missing value* (CCMV), which uses the following identification:

$$f_t(y_s|y_1,\dots y_{s-1}) = f_n(y_s|y_1,\dots y_{s-1}), \quad s = t+1,\dots,n.$$

In other words, the conditional distribution beyond time t is always borrowed from the conditional distribution from the completers. This strategy can be defended in cases where the bulk of the subjects are complete and only small proportions are assigned to the various dropout patterns. Also, extension of this approach to non-monotone patterns is particularly easy.

Alternatively, the nearest identified pattern can be used:

$$f_t(y_s|y_1,\dots y_{s-1}) = f_s(y_s|y_1,\dots y_{s-1}), \qquad s = t+1,\dots,n.$$

We will refer to these restrictions as *neighboring case missing value* (NCMV).

The third special case of (7.3) will be *available case missing value* (ACMV). It has been shown in Molenberghs *et al.* (1998), that for monotone missing data ACMV in the pattern-mixture context is equivalent with MAR in the selection model framework. Let us derive the corresponding $\boldsymbol{\omega}_s$ vectors. Expression (7.3) can be restated as

$$f_t(y_s|y_1,\dots,y_{s-1}) = f_{(\geq s)}(y_s|y_1,\dots,y_{s-1}), \tag{7.5}$$

for $s = t+1,\dots,n$. Here, $f_{(\geq s)}(.|.) \equiv f(.|., d > s)$, with d an indicator for time of dropout, which is one more than the length of the observed sequence. Now, we can transform (7.5) as follows:

$$\begin{aligned} f_t(y_s|y_1,\dots,y_{s-1}) &= f_{(\geq s)}(y_s|y_1,\dots,y_{s-1}) \\ &= \frac{f_{(\geq s)}(y_1,\dots,y_s)}{f_{(\geq s)}(y_1,\dots,y_{s-1})} \\ &= \frac{\sum_{j=s}^n \pi_j f_j(y_1,\dots,y_s)}{\sum_{\ell=s}^n \pi_\ell f_\ell(y_1,\dots,y_{s-1})} \\ &= \sum_{j=s}^n \frac{\pi_j f_j(y_1,\dots,y_{s-1})}{\sum_{\ell=s}^n \pi_\ell f_\ell(y_1,\dots,y_{s-1})} f_j(y_s|y_1,\dots,y_{s-1}). \end{aligned} \tag{7.6}$$

Next, comparing (7.6) to (7.3) yields:

$$\omega_{sj} = \frac{\pi_j f_j(y_1,...,y_{s-1})}{\sum_{\ell=s}^n \pi_\ell f_\ell(y_1,...,y_{s-1})}, \tag{7.7}$$

where π_j is the fraction of observations in pattern j (Molenberghs *et al.*, 1998). Clearly, $\boldsymbol{\omega}_s$ defined by (7.7) consists of components which are nonnegative and sum to one. In other words, a valid density function is defined. Note that the number of parameters in $\boldsymbol{\omega}_s$ increases rapidly. There are several ways to deal with this. First, special but important restrictions such as CCMV, NCMV and ACMV do not suffer from this problem since each of the $\boldsymbol{\omega}$'s involved is then determined by the choice of restriction. Second, one may set all $\boldsymbol{\omega}$'s equal to the same constant, chosen from a small set, for example spanning the unit interval. Third, one could put prior distributions on the $\boldsymbol{\omega}$'s. The first solution is followed in the remainder of this chapter.

Since there exists a link between MAR in the selection and ACMV in the pattern-mixture families, it is also of interest to consider pattern-mixture alternatives for the MNAR family. The class of models where dropout may depend on the current, possibly unobserved, measurement, but not on future measurements, will be termed missing non-future dependent (MNFD). While they are natural and easy to consider in a selection model context, there exist important examples of mechanisms that do not satisfy MNFD, such as shared-parameter models (Wu and Bailey, 1989; Little, 1995). Kenward, Molenberghs and Thijs (2003) have shown there is a counterpart to MNFD in the pattern-mixture context. They defined non-future dependent missing value restrictions (NFMV), which exclude mechanisms such as CCMV and NCMV. Kenward, Molenberghs and Thijs (2003) have shown that, for longitudinal data with dropouts, MNFD and NFMV are equivalent.

Restrictions (7.3), with the CCMV, NCMV, and ACMV forms as special cases, can be incorporated in a comprehensive strategy to fit pattern-mixture models. We will briefly sketch the strategy to fit those models.

1. Fit a model to the pattern-specific identifiable densities: $f_t(y_1, \ldots, y_t)$. This results in a parameter estimate, $\boldsymbol{\phi}_t$.

2. Select an identification method of choice.

3. Using this identification method, determine the conditional distributions of the unobserved outcomes, given the observed ones:

$$f_t(y_{t+1}, \ldots, y_n | y_1, \ldots, y_t). \tag{7.8}$$

4. Draw multiple imputations for the unobserved components, given the observed outcomes and the correct pattern-specific density (7.8).

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 in the standard multiple imputation way (Rubin, 1987; Schafer, 1997; Verbeke and Molenberghs, 2000): the M within-imputation estimates for $\boldsymbol{\phi}$ are pooled to give the multiple imputation estimate

$$\hat{\boldsymbol{\phi}}^* = \frac{1}{M}\sum_{m=1}^{M}\hat{\boldsymbol{\phi}}^m.$$

The variance is obtained as a weighted sum of the within-imputation variance and the between-imputations variance:

$$\boldsymbol{V} = \boldsymbol{W} + \left(\frac{M+1}{M}\right)\boldsymbol{B} \tag{7.9}$$

where

$$\boldsymbol{W} = \frac{1}{M}\sum_{m=1}^{M}\mathrm{Var}(\hat{\boldsymbol{\phi}}^m)$$

is the average *within* imputation variance, and

$$\boldsymbol{B} = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\phi}}^m - \hat{\boldsymbol{\phi}}^*)(\hat{\boldsymbol{\phi}}^m - \hat{\boldsymbol{\phi}}^*)'$$

is the *between* imputation variance.

7.3 A Special Case: Three Measurements

In this section, we consider the special but insightful case of three measurements. Identification (7.4) then takes the following form:

$$\begin{aligned}
f_3(y_1,y_2,y_3) &= f_3(y_1,y_2,y_3), &(7.10)\\
f_2(y_1,y_2,y_3) &= f_2(y_1,y_2)f_3(y_3|y_1,y_2), &(7.11)\\
f_1(y_1,y_2,y_3) &= f_1(y_1)\left[\omega f_2(y_2|y_1) + (1-\omega)f_3(y_2|y_1)\right]f_3(y_3|y_1,y_2). &(7.12)
\end{aligned}$$

Since $f_3(y_1,y_2,y_3)$ is completely identifiable from the data, and for $f_2(y_1,y_2,y_3)$ there is only one possible identification, given (7.3), the only place where a choice has to be made is pattern 1. Setting $\omega = 1$ corresponds to NCMV, while $\omega = 0$ implies CCMV. Using (7.7) in this particular case, ACMV corresponds to

$$\omega = \frac{\pi_2 f_2(y_1)}{\pi_2 f_2(y_1) + \pi_3 f_3(y_1)}. \tag{7.13}$$

The conditional density $f_1(y_2|y_1)$ in (7.12) can be rewritten as

$$\begin{aligned} f_1(y_2|y_1) &= \omega f_2(y_2|y_1) + (1-\omega) f_3(y_2|y_1) \\ &= \frac{\pi_2 f_2(y_1, y_2) + \pi_3 f_3(y_1, y_2)}{\pi_2 f_2(y_1) + \pi_3 f_3(y_1)}. \end{aligned}$$

Let us now show which steps are required to draw from the conditional densities, without specifying any parametric form for these densities.

1. Estimate the parameters of the identifiable densities: $f_3(y_1, y_2, y_3)$ from pattern 3, $f_2(y_1, y_2)$ from pattern 2, and $f_1(y_1)$ from pattern 1.

2. To properly account for the uncertainty with which the parameters are estimated, we need to draw from them as is customarily done in multiple imputation. More precisely, we will draw a parameter vector of its distribution and assume that in all densities from which we draw, this parameter vector is used.

3. **For pattern 2.** Given an observation in this pattern, with observed values (y_1, y_2), calculate the conditional density $f_3(y_3|y_1, y_2)$ and draw from it.

4. **For pattern 1.** We now have to distinguish three substeps.

   (a) There is now only one ω involved: for pattern 1, in order to determine $f_1(y_2|y_1)$, as a combination of $f_2(y_2|y_1)$ and $f_3(y_2|y_1)$. Every ω in the unit interval is valid. Specific cases are:

   - $\omega = 0$: CCMV,
   - $\omega = 1$: NCMV,
   - ω calculated from (7.13), and identifies a linear combination across patterns: ACMV. Note that, given y_1, this is a constant, depending on π_2 and π_3.

   To pick one of the two components f_2 or f_3, we need to generate a random uniform variate, U say, except in the boundary NCMV and CCMV cases.

   (b) If $U \leq \omega$, calculate $f_2(y_2|y_1)$ and draw from it. Otherwise, do the same based on $f_3(y_2|y_1)$.

   (c) Given the observed y_1 and given y_2 which has just been drawn, calculate the conditional density $f_3(y_3|y_1, y_2)$ and draw from it.

All steps but the first one have to be repeated M times.

In case the observed densities are assumed to be normal, the corresponding conditional densities are particularly straightforward. However, in several cases, the

conditional density is a mixture of normal densities. Then an additional and straightforward draw from the components of the mixture is necessary. Similar developments are possible with categorical data, ensuring that draws from the proper conditional multinomial distributions are made. In the next sections, attention will be devoted to this last feature.

7.4 Pattern-Mixture Models for Categorical Outcomes

In the remainder of this chapter, we will restrict attention to the special case of three binary measurements. Extension to more than three outcomes, or to more than two outcome categories, is straightforward. The multivariate Dale model, introduced in Section 6.1, will be used to estimate the parameters of the identifiable densities. For the completers (pattern 3), a trivariate Dale model will be used, for pattern 2 a bivariate Dale model, and a univariate Dale model for pattern 1. We will term this the *minimal approach*. The multivariate Dale model combines logistic regression for each of the measurements with marginal global odds ratios to describe the association between outcomes. For three measurements, this results in the following logistic regressions and odds ratios (subject-specific indices i are removed for the ease of notation):

$$\begin{aligned}
\eta_1 &= \ln\left(\frac{p_{1++}}{1-p_{1++}}\right) = X_1\boldsymbol{\theta},\\
\eta_2 &= \ln\left(\frac{p_{+1+}}{1-p_{+1+}}\right) = X_2\boldsymbol{\theta},\\
\eta_3 &= \ln\left(\frac{p_{++1}}{1-p_{++1}}\right) = X_3\boldsymbol{\theta},\\
\eta_4 &= \ln\varphi_{12} = \ln\left(\frac{p_{11+}(1-p_{1++}-p_{+1+}+p_{11+})}{(p_{1++}-p_{11+})(p_{+1+}-p_{11+})}\right) = X_4\boldsymbol{\theta},\\
\eta_5 &= \ln\varphi_{13} = \ln\left(\frac{p_{1+1}(1-p_{1++}-p_{++1}+p_{1+1})}{(p_{1++}-p_{1+1})(p_{++1}-p_{1+1})}\right) = X_5\boldsymbol{\theta},\\
\eta_6 &= \ln\varphi_{23} = \ln\left(\frac{p_{+11}(1-p_{++1}-p_{+1+}+p_{+11})}{(p_{++1}-p_{+11})(p_{+1+}-p_{+11})}\right) = X_6\boldsymbol{\theta},\\
\eta_7 &= \ln\varphi_{123} = \ln\left(\frac{p_{111}p_{122}p_{212}p_{221}}{p_{112}p_{121}p_{211}p_{222}}\right) = X_7\boldsymbol{\theta}.
\end{aligned}$$

Therefore, the incomplete patterns provide information neither about the unobserved outcomes, nor about the associations involving those unobserved outcomes. Thus, for

pattern 2, only η_1, η_2 and η_4 can be obtained from the data, while for pattern 1 only η_1 will be available.

Also in this setting, one is interested in model parameters for the full set of repeated outcomes, and thus identifying restrictions are necessary to determine the unknown probabilities by equating them to functions of known probabilities. In the normal case, restrictions are very natural to apply, because marginal as well as conditional distributions can be expressed as simple functions of the mean vector and the covariance matrix components. For categorical data however, and the Dale model in particular, there is no easy transition from marginal to conditional distributions in terms of the model parameters.

First, the minimal approach is followed in the sense that a trivariate Dale model for the complete pattern is combined with a bivariate and univariate Dale model for the incomplete patterns. From this approach the underlying probabilities $p_{y_1y_2y_3|3}$, $p_{y_1y_2|2}$ and $p_{y_1|1}$ can be estimated. For pattern 2, there is only one possibility to impute the missing cell counts, since information on the third measurement can only be borrowed from pattern 3. So, the partial counts $Z_{y_1y_2|2}$ and the conditional probabilities $p_{y_3|y_1y_2,3}$ have to be used to identify $Z^*_{y_1y_2y_3|2}$ from $Z_{y_1y_2|2}p_{y_3|y_1y_2,3}$. For pattern 1, we have several possibilities to impute the missing cell counts, since information on the second measurement can be borrowed from pattern 2 as well as from pattern 3. The joint probability of y_1, y_2 and y_3 in pattern 1 can be written as

$$p_{y_1y_2y_3|1} \quad = \quad p_{y_1|1}\left[\omega p_{y_2|y_1,2} + (1-\omega)p_{y_2|y_1,3}\right] p_{y_3|y_1y_2,3},$$

where specific choices of ω lead to the previously defined identifying restrictions CCMV, NCMV and ACMV:

$$\begin{aligned}
\text{CCMV} \quad &: \quad p_{y_1|1}p_{y_2|y_1,3}p_{y_3|y_1y_2,3},\\
\text{NCMV} \quad &: \quad p_{y_1|1}p_{y_2|y_1,2}p_{y_3|y_1y_2,3},\\
\text{ACMV} \quad &: \quad \omega = \frac{\pi_2 p_{y_1|2}}{\pi_2 p_{y_1|2} + \pi_3 p_{y_1|3}},
\end{aligned}$$

such that the missing cell counts can be identified as follows:

$$\begin{aligned}
\text{CCMV} \quad &: \quad Z^*_{y_1y_2y_3|1} = Z_{y_1|1}p_{y_2|y_1,3}p_{y_3|y_1y_2,3},\\
\text{NCMV} \quad &: \quad Z^*_{y_1y_2y_3|1} = Z_{y_1|1}p_{y_2|y_1,2}p_{y_3|y_1y_2,3},\\
\text{ACMV} \quad &: \quad Z^*_{y_1y_2y_3|1} = Z_{y_1|1}\left[\frac{\pi_2 p_{y_1y_2|2} + \pi_3 p_{y_1y_2|3}}{\pi_2 p_{y_1|2} + \pi_3 p_{y_1|3}}\right] p_{y_3|y_1y_2,3}.
\end{aligned}$$

To perform the corresponding imputations, we use a uniform random number generator. Suppose the count Z is to be distributed over the cells Z_k, $k = 1, \ldots, c$.

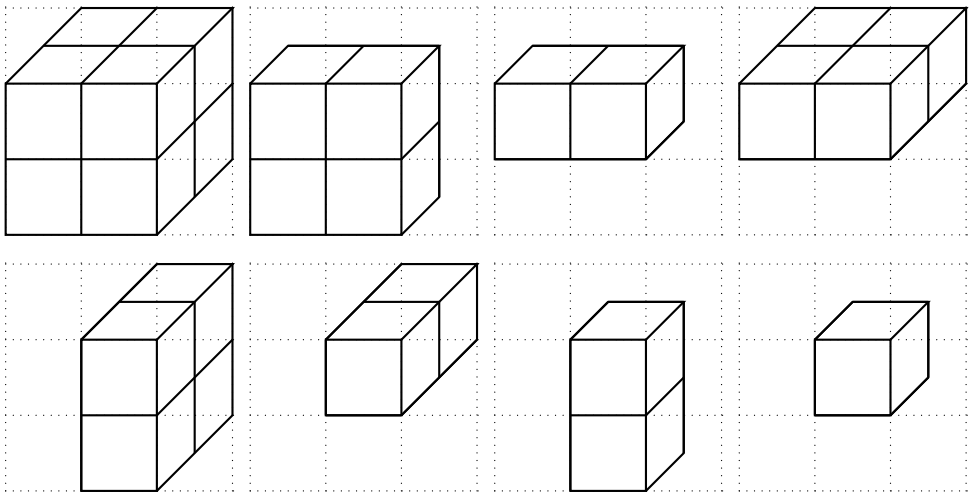


Figure 7.1: *Three-dimensional representation of all possible patterns for three binary outcomes with intermittent missingness. The horizontal axis displays the first measurement, the vertical axis corresponds to the second measurement, and the third axis with the last measurement.*

Then, the cumulative probabilities $\lambda_0, \ldots, \lambda_c$ are calculated and Z draws U_t from a uniform $U[0,1]$ distribution are made. Next, Z_k is set equal to $\sum_t (\lambda_{k-1} < U_t \leq \lambda_k)$.

From the completed counts $Z^*_{y_1 y_2 y_3|1}$ and $Z^*_{y_1 y_2 y_3|2}$, and from $Z_{y_1 y_2 y_3|3}$, one can estimate the parameters of interest, for example a trivariate Dale model for the three patterns separately, or a trivariate Dale model where pattern is included as a covariate. Also other possible models can be fitted to the completed counts.

Although parameter estimation is very elegant and computationally simple with the above two-step procedure, precision estimation is less straightforward. Indeed, treating the filled-in table as if it represented observed data fails to reflect random variability in the unobserved counts. Therefore, multiple imputation will be used to construct an asymptotic covariance matrix of the form (7.9).

7.5 Specific Assumptions for Intermittent Missingness

Since this thesis is mainly devoted to the analysis of non-monotone missing data, we will also consider this case in the setting of pattern-mixture models.

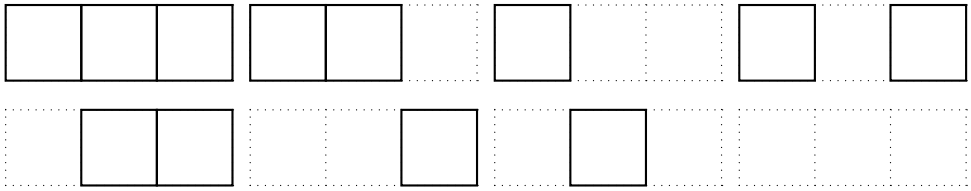


Figure 7.2: *Two-dimensional representation of all possible patterns for three outcomes with intermittent missingness, in the same order as in Figure 7.1. A solid square represents an observed measurement. From left to right, and from top to bottom, we have patterns 3, 2 and 1 as defined before, and further the non-monotone patterns 4, 5, 6, 7 and 8.*

In Figure 7.1, a three-dimensional graphical representation is given of all possible patterns for three binary outcomes when intermittent missingness is allowed. Figure 7.2 gives an equivalent two-dimensional representation. The first three patterns are the monotone patterns, which have already been discussed in Section 7.4. Pattern 3 is the fully observed pattern, and does not need any imputation. Patterns 1 and 2 will be considered again in this section, since many more possibilities will be available now to impute the unobserved data.

Let us first consider the patterns for which only one measurement is missing, namely patterns 2, 4 and 5, where the third, the second and the first outcome, respectively, are not observed. A bivariate Dale model can be used to fit the observed data densities $f_1(y_1, y_2)$, $f_4(y_1, y_3)$ and $f_5(y_2, y_3)$. Since it is recommended to use as much of the available data as possible to impute the conditional distributions of the unobserved outcomes, given the observed ones, we can only use information from pattern 3 to impute the unobserved data. This results in the following complete data densities:

$$\begin{aligned} f_2(y_1, y_2, y_3) &= f_2(y_1, y_2) f_3(y_3|y_1, y_2), \\ f_4(y_1, y_2, y_3) &= f_4(y_1, y_3) f_3(y_2|y_1, y_3), \\ f_5(y_1, y_2, y_3) &= f_5(y_2, y_3) f_3(y_1|y_2, y_3). \end{aligned}$$

Next, patterns 1, 6 and 7 will be discussed. Here, only one out of the three outcomes is measured, and a univariate Dale model can be used to obtain $f_1(y_1)$, $f_6(y_3)$ and $f_7(y_2)$. First, we have to decide which of the two unobserved outcomes will be imputed first. In the case of monotone missingness, the obvious choice for pattern 1 was to impute first y_2 and then y_3. In the case of non-monotone missingness, there is

no such obvious choice. Therefore, we will look at both possibilities. For pattern 6, for example, we can first consider the conditional density of y_1, given y_3. Information on this density can be borrowed from either the completers (pattern 3) or the neighbors (pattern 4), or a combination of both densities. Thus, we obtain the previously defined identifying restrictions CCMV, NCMV and ACMV. The conditional density of y_2, given y_1 and y_3, can only be borrowed from pattern 3. Similarly, the conditional density of y_2, given y_3, can be obtained first, using one of the available identifying restrictions, and afterwards the conditional density of y_1, given y_2 and y_3. Thus, the complete data densities for patterns 1, 6 and 7 can be written as:

$$
\begin{aligned}
f_1(y_1, y_2, y_3) &= \begin{cases} f_1(y_1)\left[\omega f_2(y_2|y_1) + (1-\omega) f_3(y_2|y_1)\right] f_3(y_3|y_1, y_2), \\ f_1(y_1)\left[\omega f_4(y_3|y_1) + (1-\omega) f_3(y_3|y_1)\right] f_3(y_2|y_1, y_3), \end{cases} \\
f_6(y_1, y_2, y_3) &= \begin{cases} f_6(y_3)\left[\omega f_4(y_1|y_3) + (1-\omega) f_3(y_1|y_3)\right] f_3(y_2|y_1, y_3), \\ f_6(y_3)\left[\omega f_5(y_2|y_3) + (1-\omega) f_3(y_2|y_3)\right] f_3(y_1|y_2, y_3), \end{cases} \\
f_7(y_1, y_2, y_3) &= \begin{cases} f_7(y_2)\left[\omega f_2(y_1|y_2) + (1-\omega) f_3(y_1|y_2)\right] f_3(y_3|y_1, y_2), \\ f_7(y_2)\left[\omega f_5(y_3|y_2) + (1-\omega) f_3(y_3|y_2)\right] f_3(y_1|y_2, y_3), \end{cases}
\end{aligned}
$$

where, in all cases, $\omega = 0$ corresponds to CCMV, $\omega = 1$ to NCMV, and ω as in (7.13), with the corresponding densities and pattern probabilities, corresponds to ACMV. So, we can either choose one of the two possibilities to determine the complete data density, or use a linear combination of both expressions. This combination is topic of further research.

Finally, pattern 8 does not contain any observed data, such that it is not possible to impute the unobserved data conditional on the observed data. This pattern will therefore be ignored.

7.6 Marginal Effects Across Patterns

We already mentioned that several strategies can be followed to analyze the imputed data sets. When a selection model is used, an overall effect of the covariates of interest (e.g., treatment effect) is obtained immediately from the model. When, however, a pattern-mixture model is used to analyze the multiply imputed sets of data, the overall covariate effect cannot be obtained directly, since the effect is modeled for each pattern separately. In the case of continuous data, where linear models are used, the overall effect is simply a weighted average of the pattern-specific effects. We will show that this is not true for categorical data. We therefore assume that the logistic

regression

$$P(Y_{ij} = 1|\text{pattern } k) = \frac{e^{\alpha_k+\beta_k T_i}}{1 + e^{\alpha_k+\beta_k T_i}}$$

is used to model the data from pattern k (as in the multivariate Dale model). α and β can depend on j, but we suppress this index from notation.

Assume interest is in one particular effect T, e.g., treatment effect at the last occasion, and assume π_k to be the pattern probability as defined before. The marginal success probability is then equal to

$$\sum_{k=1}^{K} \pi_k \frac{e^{\alpha_k+\beta_k T}}{1 + e^{\alpha_k+\beta_k T}}. \tag{7.14}$$

There are three ways to calculate from this the marginal treatment effect at the last occasion. First, the direct linear approach (Park and Lee, 1999) can be used, where

$$\beta \simeq \sum_k \pi_k \beta_k \tag{7.15}$$

but this is clearly wrong. Second, the marginal probability can be approximated via a logistic regression, a probit model or fully using the longitudinal nature, through a Dale model, a generalized linear mixed model (GLMM), ... And third, classical averaging can be performed. To this effect, keep function (7.14) as is and compute and graph, or sample. Note that averaging in this way will be similar to the marginalization of random effects models (e.g., GLMM to GEE). Here, the marginalization is over pattern, rather than over random effects. When a GLMM is used in each pattern, then there is a double marginalization, one over the random effects and one over the patterns. We will focus on the second approach, using a marginal logistic model.

Let us approximate (7.14) by a logistic regression:

$$f(T) = \sum_k \pi_k \frac{e^{\alpha_k+\beta_k T}}{1 + e^{\alpha_k+\beta_k T}} \cong \frac{e^{A+BT}}{1 + e^{A+BT}}. \tag{7.16}$$

Then, the logit of $f(T)$ can be approximated by

$$F(T) = \text{logit}\left(f(T)\right) \cong A + BT.$$

Using the first order Taylor expansion, results in

$$F(0) + \left.\frac{\partial F}{\partial T}\right|_{T=0} T \cong A + BT,$$

such that

$$A \simeq F(T=0) = \text{logit}\left(\sum_k \pi_k \frac{e^{\alpha_k}}{1+e^{\alpha_k}}\right).$$

It is easily shown that

$$\frac{\partial \text{logit}(x)}{\partial x} = \frac{1}{x(1-x)}$$

and

$$\frac{\partial f}{\partial T} = \sum_k \pi_k \frac{\left(e^{\alpha_k+\beta_k T}\right)\beta_k}{\left(1+e^{\alpha_k+\beta_k T}\right)^2},$$

such that

$$\left.\frac{\partial F}{\partial T}\right|_{T=0} = \frac{1}{\sum_k \pi_k \frac{e^{\alpha_k}}{1+e^{\alpha_k}}} \frac{1}{\sum_k \pi_k \frac{1}{1+e^{\alpha_k}}} \sum_k \beta_k \pi_k \frac{e^{\alpha_k}}{\left(1+e^{\alpha_k}\right)^2},$$

and equivalently

$$B \simeq \frac{\sum_k \beta_k \pi_k \frac{e^{\alpha_k}}{1+e^{\alpha_k}} \frac{1}{1+e^{\alpha_k}}}{\left(\sum_k \pi_k \frac{e^{\alpha_k}}{1+e^{\alpha_k}}\right)\left(\sum_k \pi_k \frac{1}{1+e^{\alpha_k}}\right)}.$$

Let $P_k = \frac{e^{\alpha_k}}{1+e^{\alpha_k}}$, then the approximate marginalized treatment effect can be estimated using

$$B \simeq \frac{\sum_k \beta_k \pi_k P_k(1-P_k)}{\left(\sum_k \pi_k P_k\right)\left[\sum_k \pi_k (1-P_k)\right]}. \tag{7.17}$$

Note that direct expansion of (7.16), without taking the logit first, leads to exactly the same expression.

Let us now consider the special case where the treatment effect is the same in each pattern ($\beta_k = \beta, \forall k$), then

$$B \simeq \beta \frac{\sum_k \pi_k P_k(1-P_k)}{\left(\sum_k \pi_k P_k\right)\left(\sum_k \pi_k (1-P_k)\right)},$$

such that

$$|B| \leq |\beta|. \tag{7.18}$$

This means that the marginal treatment effect at the last occasion, obtained through approximation (7.16), will not be larger in absolute value than the marginal treatment effect, obtained from the direct linear approach (7.15), when the treatment effects are equal across patterns.

Proof of Equation (7.18)

Let

$$g = \left(\sum_k \pi_k P_k\right)\left(\sum_k \pi_k(1-P_k)\right) - \sum_k \pi_k P_k(1-P_k)$$

and

$$H = \sum_k \pi_k P_k.$$

To find the extrema of g, we calculate

$$\begin{aligned}\frac{\partial g}{\partial P_\ell} &= \pi_\ell(1-H) - \pi_\ell H - \pi_\ell(1-2P_\ell)\\ &= 2\pi_\ell(P_\ell - H)\\ &= 2\pi_\ell\left(P_\ell - \sum_k \pi_k P_k\right).\end{aligned}$$

g then reaches an extremum if $\dfrac{\partial g}{\partial P_\ell} = 0$ for all ℓ. Thus,

$$\pi_\ell\left(P_\ell - \sum_k \pi_k P_k\right) = 0.$$

We can exclude $\pi_\ell = 0$, since such a pattern would vanish. Thus, g reaches an extremum if

$$P_\ell = \sum_k \pi_k P_k \;\; \forall \ell \quad \Leftrightarrow \quad P_1 = \ldots = P_K \equiv P$$

and hence

$$\sum_k \pi_k P_k = P\sum_k \pi_k = P.$$

At this extremum, $g = P(1-P) - P(1-P) = 0$. Now we still have to check whether this extremum is a minimum or a maximum. Therefore we calculate the second order derivatives of g.

$$\left.\begin{aligned}\frac{\partial^2 g}{\partial P_\ell^2} &= 2\pi_\ell(1-\pi_\ell)\\ \frac{\partial^2 g}{\partial P_\ell \partial P_m} &= 2\pi_\ell(-\pi_m) = -2\pi_\ell\pi_m\end{aligned}\right\} \Rightarrow \frac{\partial^2 g}{\partial \boldsymbol{P}\partial \boldsymbol{P}'} = 2\left[\mathrm{diag}\boldsymbol{\pi} - \boldsymbol{\pi}\boldsymbol{\pi}'\right]$$

which is a positive definite matrix. We can conclude that this extremum is a minimum, and thus $g \geq 0$, which means $0 \leq \dfrac{\sum_k \pi_k P_k(1-P_k)}{\left(\sum_k \pi_k P_k\right)\left(\sum_k \pi_k(1-P_k)\right)} \leq 1$, and $|B| \leq |\beta|$. ■

Marginalization when the β_k's are different, may both increase and decrease the effect, in absolute value. Let us consider the example of two patterns ($K = 2$). Set $\pi_1 = \pi$, $\pi_2 = 1 - \pi$, $\beta_1 = 1$ and $\beta_2 = \rho$. Expressions (7.15) and (7.17) then reduce to

$$\pi + (1-\pi)\rho \quad \text{and} \quad \frac{\pi P_1(1-P_1) + \rho(1-\pi)P_2(1-P_2)}{[\pi P_1 + (1-\pi)P_2]\,[\pi(1-P_1) + (1-\pi)(1-P_2)]}. \tag{7.19}$$

Let $N = [\pi P_1 + (1-\pi)P_2]\,[\pi(1-P_1) + (1-\pi)(1-P_2)]$. Choose ρ such that the equality between both expressions in (7.19) holds:

$$\begin{aligned} N\pi + N(1-\pi)\rho &= \pi P_1(1-P_1) + (1-\pi)P_2(1-P_2)\rho \\ (1-\pi)\,[P_2(1-P_2) - N]\,\rho &= \pi\,[N - P_1(1-P_1)] \\ \Rightarrow \rho &= \frac{\pi\,[N - P_1(1-P_1)]}{(1-\pi)\,[P_2(1-P_2) - N]}. \end{aligned} \tag{7.20}$$

Since $\rho \in \mathbb{R}$, setting ρ equal to this value is sufficient to have both equations equal. $\rho + \varepsilon$ and $\rho - \varepsilon$ will then make the inequality go both ways.

If $P_1 = P_2 = P$ then the right expression in (7.19) reduces to

$$\frac{\pi P(1-P) + \rho(1-\pi)P(1-P)}{[\pi P + (1-\pi)P]\,[\pi(1-P) + (1-\pi)(1-P)]} = \frac{\pi P(1-P) + \rho(1-\pi)P(1-P)}{P\,[\pi + (1-\pi)]\,(1-P)\,[\pi + (1-\pi)]},$$

which is equal to $\pi + (1-\pi)\rho$, and hence, for all ρ, both expressions in (7.19) are the same. Thus, the difference emerges from a difference in background success probability P_k. Note also that then in (7.20) the numerator and denominator are both equal to zero, confirming that the result applies to every ρ.

Now we will determine the sign of ρ for $P_1 \neq P_2$. Denote the coefficient of π in the numerator of ρ by f_1, and in the denominator by f_2. Then

$$\begin{aligned} f_1 &= N - P_1(1-P_1) \\ &= [\pi P_1 + (1-\pi)P_2]\,[\pi(1-P_1) + (1-\pi)(1-P_2)] - P_1(1-P_1) \end{aligned}$$

and

$$\begin{aligned} f_2 &= -N + P_2(1-P_2) \\ &= -[\pi P_1 + (1-\pi)P_2]\,[\pi(1-P_1) + (1-\pi)(1-P_2)] + P_2(1-P_2). \end{aligned}$$

Since for $\pi = 0$, $f_1 = P_2(1-P_2) - P_1(1-P_1) = Q$ and $f_2 = 0$, and for $\pi = 1$, $f_1 = 0$ and $f_2 = P_2(1-P_2) - P_1(1-P_1) = Q$, both functions evolve in the interval $[0, Q]$. To determine whether there are internal extrema in f_1 and f_2, we calculate

$$\frac{\partial f_1}{\partial \pi} = (P_1 - P_2)\,[\pi(1-2P_1) + (1-\pi)(1-2P_2)]\,,$$

which, since $P_1 \neq P_2$ by assumption, equals 0 for π equal to

$$\pi^* = \frac{2P_2 - 1}{2(P_2 - P_1)}.$$

$\frac{\partial f_2}{\partial \pi}$ equals zero at the same point π^*. By calculating the second order derivatives of f_1 and f_2 to π,

$$\begin{aligned} \frac{\partial^2 f_1}{\partial \pi^2} &= -2(P_1 - P_2)^2 < 0, \\ \frac{\partial^2 f_2}{\partial \pi^2} &= 2(P_1 - P_2)^2 > 0, \end{aligned}$$

we see that f_1 reaches a maximum in π^*, while f_2 is minimal in π^*. For example, suppose $P_1 = 0.2$ and $P_2 = 0.7$, then $\pi^* = 0.4$. At π^*, $f_1 = \frac{1}{2}.\frac{1}{2} - P_1(1 - P_1) \geq 0$ and $f_2 = -\frac{1}{2}.\frac{1}{2} + P_2(1 - P_2) \leq 0$. Note that π^* is a valid extremum in $[0, 1]$ if for $P_1 < P_2$, $P_1 \leq \frac{1}{2} \leq P_2$, and for $P_1 > P_2$, $P_2 \leq \frac{1}{2} \leq P_1$.

Let us consider the following situations. When $\pi^* \in [0, 1]$, and for example $P_1 \leq \frac{1}{2} \leq P_2$, or even $P_1 < \frac{1}{2} < P_2$ since $P_1 \neq P_2$ ($P_2 \leq \frac{1}{2} \leq P_1$ is similar), then $f_1 > 0$ and $f_2 < 0$, and hence $\rho < 0$ in a neighborhood of π^*. When $\pi^* \notin [0, 1]$, then f_1 and f_2 are monotonic and both of the same sign, such that ρ is nonnegative. Then there exist treatment effects $(1, \rho)$ such that there is no dilution of effect, but equality or inflation. Figure 7.3 shows the curves of f_1 and f_2 for several values of P_1 and P_2. We will study two of those examples in further detail.

Example 1

Assume $P_1 = 0.2$ and $P_2 = 0.7$. Set π equal to $\pi^* = 0.4$. Then $f_1 = 0.09$ and $f_2 = -0.04$, such that

$$\rho = \frac{\pi}{1 - \pi}.\frac{f_1}{f_2} = -\frac{0.4}{0.6}.\frac{0.09}{0.04} = -1.5.$$

In this case, the treatment effects, $(1; -1.5)$ are in the opposite direction. Since $\pi P_1(1 - P_1) = 0.064$, $(1 - \pi)P_2(1 - P_2) = 0.126$, $\pi P_1 + (1 - \pi)P_2 = 0.5$ and $\pi(1 - P_1) + (1 - \pi)(1 - P_2) = 0.5$, the marginal treatment effect, calculated by (7.15) and (7.17) can be summarized as follows, for several values of ρ:

	$\lvert B\rvert$	versus	$\lvert\beta\rvert$
	$\lvert 0.256 + 0.504\rho\rvert$		$\lvert 0.4 + 0.6\rho\rvert$
$\rho = -2$	$\lvert -0.752\rvert$	$<$	$\lvert -0.8\rvert$
$\rho = -1.5$	$\lvert -0.5\rvert$	$=$	$\lvert -0.5\rvert$
$\rho = -1$	$\lvert -0.248\rvert$	$>$	$\lvert -0.2\rvert$

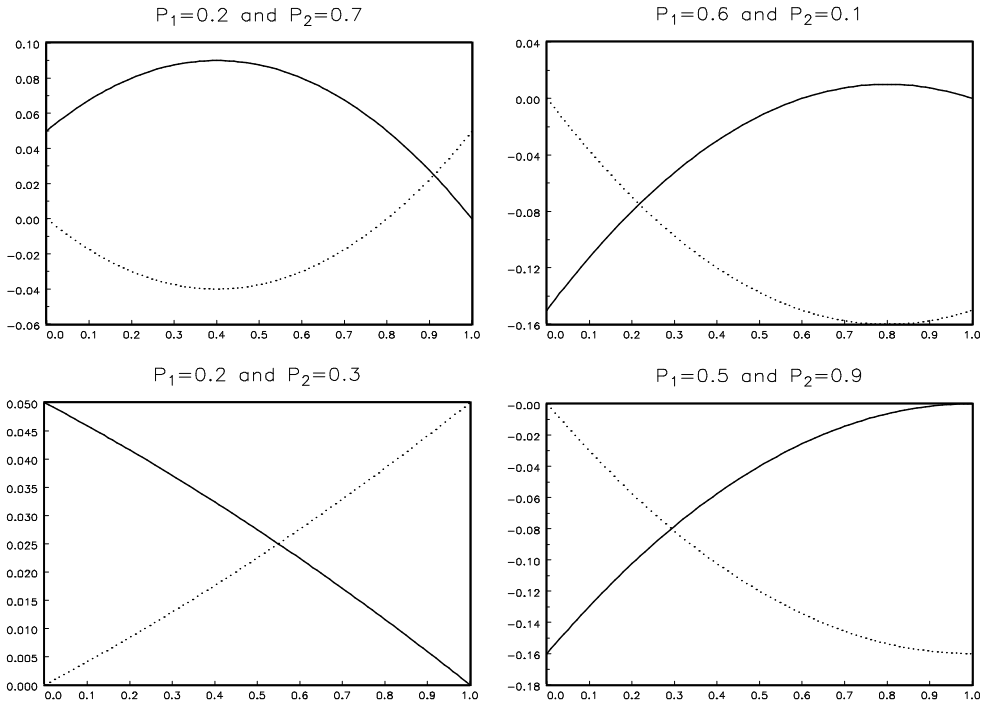


Figure 7.3: *Graphical representation of the f_1 (solid line) and f_2 (dotted line) curves for several values of P_1 and P_2. In the top panels, $\pi^* \in [0,1]$, in the bottom panels, $\pi^* \notin [0,1]$.*

Example 2

Assume now $P_1 = 0.2$ and $P_2 = 0.3$. Then $\pi^* = -2 \notin [0,1]$. We choose $\pi = 0.5$. Now $f_1 = 0.0275$ and $f_2 = 0.0225$, such that

$$\rho = \frac{\pi}{1-\pi}.\frac{f_1}{f_2} = \frac{0.5}{0.5}.\frac{0.0275}{0.0225} = \frac{11}{9}.$$

So, in this case, both treatment effects, $(1; 1.22)$ are quite close to each other. Since $\pi P_1(1-P_1) = 0.08$, $(1-\pi)P_2(1-P_2) = 0.105$, $\pi P_1 + (1-\pi)P_2 = 0.25$ and $\pi(1-P_1) + (1-\pi)(1-P_2) = 0.75$, the marginal treatment effect, calculated by (7.15) and (7.17) can be summarized as follows, for several values of ρ:

	$\lvert B\rvert$	versus	$\lvert\beta\rvert$
	$\lvert 128/300 + 0.56\rho\rvert$		$\lvert 0.5 + 0.5\rho\rvert$
$\rho = 10/9$	$\lvert 18.88/18\rvert$	$<$	$\lvert 19/18\rvert$
$\rho = 11/9$	$\lvert 10/9\rvert$	$=$	$\lvert 10/9\rvert$
$\rho = 12/9$	$\lvert 21.12/18\rvert$	$>$	$\lvert 21/18\rvert$

After making all of these considerations, it is clear that determining a marginal effect across patterns in the case of non-Gaussian data, is less straightforward than in the Gaussian case. One should bear in mind that the direct linear approach (Park and Lee, 1999) is not correct in the case of categorical data, and that this method can neither be considered to be conservative nor liberal, in the sense that there is not always a dilution of effect, but equality or inflation is also possible.

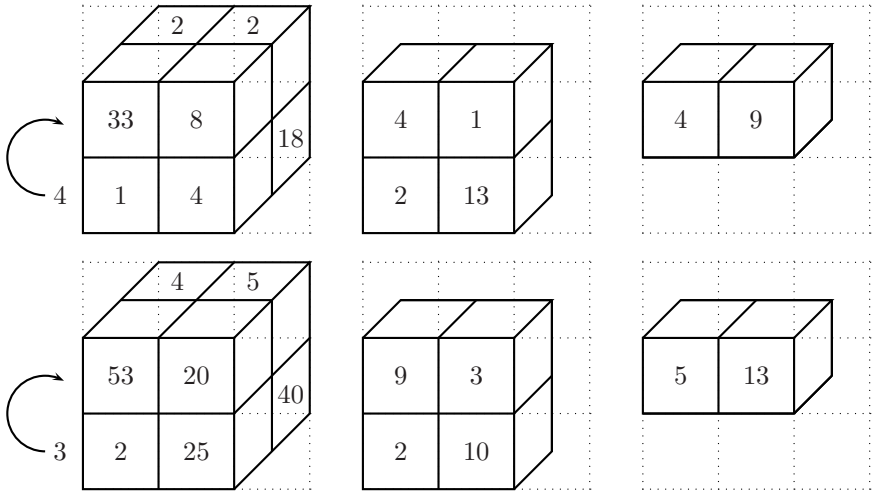
7.7 Models Fitted to the Fluvoxamine Data

In this section, we will reanalyze the fluvoxamine data. This time, we will consider the first, second and last side effects measurement, and take into account the gender of the patient (0 = males, 1 = females). For all patients, the gender is known, leaving 315 patients in the analyses. There are 224 completers (pattern 3), 44 patients missed the last visit (pattern 2), 31 only appeared at the first visit (pattern 1), 1 person belongs to pattern 5, 1 to pattern 6, and the remaining 14 patients do not have any observations at all (pattern 8). For those 14 patients, there is no solution to impute the missing outcomes, and therefore, they will not be considered in the analyses. Pattern 5 and pattern 6 both only contain 1 patient (0.33% of the total number of subjects in the study), so their effect on the results can be ignored. This leaves 299 patients in the study. The data are summarized in Table 7.1.

As described in Section 7.4, we start the analyses with fitting a trivariate Dale model to the completers, a bivariate Dale model to pattern 2, and a logistic regression to pattern 1. Then, an identifying restriction is chosen to define the conditional distributions of the unobserved outcomes, given the observed ones. Afterwards, we draw multiple imputations ($M = 10$). We thus obtain for each choice of identifying restrictions ten multiply-imputed sets of data, which then can be analyzed, using several possible models.

Let us first discuss the results reported in Table 7.2. A single trivariate Dale model is fitted, with a constant log odds ratio for each of the possible associations between outcomes, and a possible effect of gender on the marginal probabilities. We

Table 7.1: *Fluvoxamine Data. 'Side effects' (yes/no) at the first (horizontal), second (vertical) and last visit. Top table for males, bottom table for females.*



notice that the estimates for the association parameters are very close under the three possible identifying restrictions. The associations φ_{12}, φ_{13} and φ_{23} are highly significant ($p < 0.0001$), while φ_{123} is borderline significant ($p \approx 0.045$). Also, the estimates for the first marginal probability are almost equal under CCMV, NCMV and ACMV. This was to be expected, since the first outcome was observed for all subjects that were used in the analysis. The parameter estimates for the logistic regression of the third marginal probability are also quite similar. This is due to the fact that all identifying restrictions implied the same conditional density for the third outcome, given the first and second ones, namely borrow it from the completers. The small difference that is nevertheless observed, results from a difference in imputation for the second outcome, since the imputation of the third outcome is conditional on the second one. And as we can see, the estimates for the second marginal probability differ much between the three identifying restrictions. The CCMV and NCMV estimates, for the intercept as well as for gender, are lying furthest apart. ACMV estimates are closer to CCMV estimates, since many more completers are available than neighbors, thus ω will be smaller than 0.5. Finally, we will contemplate the effect of gender. We observe that the estimate is negative for the first marginal probability, approximately zero for the second one, and positive for the third one, meaning that the probability of no side effects is larger, equal or smaller for males than for females, for the first,

Table 7.2: *Fluvoxamine Data. Multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on gender, and constant associations.*

	CCMV	NCMV	ACMV
intercept_1	-0.1259(0.1949)	-0.1266(0.1951)	-0.1230(0.1949)
gender_1	-0.2528(0.2423)	-0.2516(0.2429)	-0.2574(0.2424)
intercept_2	0.1180(0.1995)	0.0385(0.1984)	0.1060(0.2005)
gender_2	-0.0022(0.2536)	0.0375(0.2435)	0.0020(0.2531)
intercept_3	0.3245(0.2134)	0.2901(0.2139)	0.3120(0.2166)
gender_3	0.2816(0.2675)	0.3159(0.2700)	0.2968(0.2703)
φ_{12}	3.1051(0.3433)	3.1218(0.3284)	3.1178(0.3386)
φ_{13}	2.0288(0.3072)	2.0047(0.3077)	2.0220(0.3121)
φ_{23}	2.8687(0.3583)	2.9588(0.3521)	2.8639(0.3548)
φ_{123}	1.8446(0.9272)	1.9283(0.9269)	1.8524(0.9386)

second and last measurement occasion, respectively. However, the effect of gender on the marginal probabilities is not significant.

Next, a more extended trivariate Dale model is presented in Table 7.3. Now, pattern-specific intercepts are allowed in the logistic regressions for the marginal probabilities. The gender effect is assumed to be the same for all patterns, and the associations between outcomes are still constant. The parameter $\texttt{intercept}_i$ is the intercept in the logistic regression for the ith marginal probability for pattern 3. $\texttt{pattern1}_i$ and $\texttt{pattern2}_i$ are dummy variables, such that they correspond to the difference in intercept between pattern 3 and pattern 1 or pattern 2, respectively. For the first marginal probability there is no significant difference between the pattern-specific intercepts. Only in the NCMV case, a borderline non-significant difference ($p \approx 0.077$) is observed between pattern 1 and pattern 3. We notice that the intercept for pattern 3 is higher than for the other patterns, resulting in a higher probability of no side effects at the first measurement occasion for the completers. Similar conclusions can be found for the second and last occasions. Taking a closer look at the results for the second marginal probability, we notice that the intercepts for pattern 2 and 3 are significantly different ($p \approx 0.035$), while patterns 1 and 3 are only borderline significantly different ($p \approx 0.05$) when NCMV is used, and not significant when

Table 7.3: *Fluvoxamine Data. Multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on a pattern-specific intercept and a fixed gender effect, and constant associations.*

	CCMV	NCMV	ACMV
intercept_1	0.0215(0.2134)	0.0266(0.2131)	0.0238(0.2133)
pattern1_1	-0.6731(0.4209)	-0.7339(0.4151)	-0.6736(0.4205)
pattern2_1	-0.3418(0.3379)	-0.3429(0.3376)	-0.3426(0.3379)
gender_1	-0.3027(0.2458)	-0.3013(0.2459)	-0.3060(0.2457)
intercept_2	0.3164(0.2250)	0.2935(0.2187)	0.3172(0.2240)
pattern1_2	-0.4485(0.4777)	-0.9597(0.4906)	-0.5451(0.4927)
pattern2_2	-0.6989(0.3324)	-0.6914(0.3323)	-0.7004(0.3325)
gender_2	-0.0709(0.2629)	-0.0424(0.2514)	-0.0725(0.2608)
intercept_3	0.4713(0.2326)	0.4503(0.2346)	0.4607(0.2321)
pattern1_3	-0.2846(0.5761)	-0.4108(0.5997)	-0.3162(0.5311)
pattern2_3	-0.5498(0.4615)	-0.5469(0.4639)	-0.5457(0.4620)
gender_3	0.2309(0.2778)	0.2654(0.2812)	0.2476(0.2779)
φ_{12}	3.1343(0.3469)	3.1410(0.3361)	3.1406(0.3444)
φ_{13}	2.0304(0.3084)	2.0168(0.3134)	2.0208(0.3112)
φ_{23}	2.8706(0.3589)	2.9654(0.3573)	2.8624(0.3561)
φ_{123}	1.7910(0.9649)	1.9351(0.9666)	1.8100(0.9778)

CCMV or ACMV is used. This can be explained by the fact that for NCMV, pattern 1 borrows all information from pattern 2 and thus takes distance from pattern 3, while under CCMV and ACMV all or most of the information is borrowed from pattern 3, and therefore there is only little distance between pattern 1 and 3. For the third marginal probability, there is no significant difference between the three patterns for all identifying restrictions, since the missing information is always identified from pattern 3. CCMV, NCMV and ACMV lead to almost the same estimates for all parameters concerning the third marginal probability. Finally, the effect of gender changes over the different measurement occasions as before, and again its effect on the marginal probabilities is not significant. The associations φ_{12}, φ_{13} and φ_{23} are highly significant ($p < 0.0001$), while φ_{123} is now borderline significant ($p \approx 0.045$)

Table 7.4: *Fluvoxamine Data. Multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on a pattern-specific intercept and a pattern-specific gender effect, and constant associations.*

	CCMV	NCMV	ACMV
intercept_1	0.1878(0.2364)	0.1933(0.2356)	0.1882(0.2361)
pattern1_1	-0.9785(0.6495)	-1.0467(0.6485)	-0.9689(0.6510)
pattern2_1	-1.0648(0.5433)	-1.0659(0.5424)	-1.0664(0.5431)
gender_1	-0.5432(0.2866)	-0.5438(0.2862)	-0.5434(0.2865)
$\text{pattern1}\times\text{gender}_1$	0.4447(0.8463)	0.4748(0.8506)	0.4224(0.8521)
$\text{pattern2}\times\text{gender}_1$	1.2278(0.6989)	1.2271(0.6981)	1.2297(0.6991)
intercept_2	0.5089(0.2448)	0.5067(0.2456)	0.5087(0.2448)
pattern1_2	-0.8513(0.7509)	-1.5285(0.7823)	-0.9168(0.7770)
pattern2_2	-1.4699(0.5386)	-1.4657(0.5382)	-1.4711(0.5390)
gender_2	-0.3519(0.2937)	-0.3517(0.2943)	-0.3519(0.2938)
$\text{pattern1}\times\text{gender}_2$	0.6298(1.1582)	0.8790(0.9177)	0.5870(1.1482)
$\text{pattern2}\times\text{gender}_2$	1.3098(0.6927)	1.3032(0.6929)	1.3095(0.6934)
intercept_3	0.5916(0.2445)	0.5942(0.2446)	0.5922(0.2446)
pattern1_3	-0.5736(0.7847)	-0.8706(0.8602)	-0.6826(0.8516)
pattern2_3	-0.9877(0.6146)	-0.9937(0.6158)	-0.9868(0.6134)
gender_3	0.0561(0.2979)	0.0542(0.2978)	0.0559(0.2979)
$\text{pattern1}\times\text{gender}_3$	0.4706(0.9388)	0.7796(1.0740)	0.5907(0.9538)
$\text{pattern2}\times\text{gender}_3$	0.7610(0.8722)	0.7683(0.8693)	0.7612(0.8700)
φ_{12}	3.1328(0.3456)	3.1271(0.3359)	3.1412(0.3424)
φ_{13}	2.0235(0.3102)	2.0092(0.3140)	2.0143(0.3139)
φ_{23}	2.9035(0.3702)	2.9732(0.3564)	2.8943(0.3669)
φ_{123}	1.7912(0.9537)	1.9162(0.9506)	1.8026(0.9504)

only for NCMV, and borderline non-significant ($p \approx 0.064$) for CCMV and ACMV.

Third, Table 7.4 contains parameter estimates of a trivariate Dale model where now not only the intercept, but also the gender effect is allowed to be different in the three patterns. Also here $\texttt{intercept}_i$ corresponds to the effect of pattern 3, while the

dummy variables $\texttt{pattern1}_i$ and $\texttt{pattern2}_i$ model the difference in success probability between pattern 3 and pattern 1 or 2, respectively. $\texttt{gender}_i$ represents the gender effect in pattern 3, while the interactions between the dummies and gender refer to the difference in gender effect between pattern 3 and pattern 1 or 2, respectively. The parameter estimates for the logistic regression of p_{1++} reveal the following results. The probability of no side effects is borderline significantly different ($p \approx 0.05$) between pattern 2 and 3, but not significantly different between pattern 1 and 3. Gender is borderline non-significant ($p \approx 0.058$) in pattern 3, and a borderline non-significant different gender effect occurred between pattern 2 and 3. For p_{+1+}, similar conclusions are reached, but now the difference in probability of no side effects is highly significant ($p \approx 0.006$) between pattern 2 and 3, and under NCMV borderline significant ($p \approx 0.05$) between pattern 1 and 3. The gender effect in pattern 3 is not significant anymore. Finally, the success probability p_{++1} is not different in the three patterns, and the gender effect is not significant. The associations φ_{12}, φ_{13} and φ_{23} are again highly significant ($p < 0.0001$), while φ_{123} is borderline significant ($p \approx 0.044$) only for NCMV, and borderline non-significant ($p \approx 0.059$) for CCMV and ACMV.

Finally, a trivariate Dale model is fitted to each of the patterns separately, with marginal probabilities depending on gender, and constant associations between outcomes. These results are summarized in Table 7.5. If the previous model was further extended, with, for the three patterns, different associations between outcomes, the same estimates would have been obtained, as in Table 7.5. We will now discuss the estimates that were obtained by fitting a separate trivariate Dale model to each pattern. For pattern 3, of course, there is no difference between the initial estimates and the multiple imputation estimates, since no imputation was necessary in this pattern. For patterns 1 and 2, several estimates are tending to infinity, since a lot of sparse or empty cells were present in the multiply-imputed sets of data, because the 13 males and 18 females in pattern 1, and the 20 males and 24 females in pattern 2, had to be distributed over 8 cells, with one more likely to be filled than the other. Especially the association parameters suffer from those empty cells. Therefore, it is hard to draw conclusions for patterns 1 and 2. Also, it leads to no avail to try to find the marginal effects of the covariate gender, using the technique of Section 7.6. If, however, the proportion of subjects was equal in each pattern, then the marginal gender effects, obtained by using those techniques, would correspond to the gender effects that resulted from the first model that was fitted.

From all the analyses that are performed here, we can conclude that the first model is too simple, since all patterns are treated equally, and from more complex models, we could conclude that there exists some difference in success probability between the

Table 7.5: *Fluvoxamine Data. Estimates from the initial Dale models for the incomplete data, together with multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on gender, and constant associations, fitted for each pattern separately.*

	initial	CCMV	NCMV	ACMV
Pattern 1				
intercept_1	-0.8109 (0.6009)	-0.8329 (0.4665)	-1.1210 (14.110)	-0.7455 (0.5671)
gender_1	-0.1446 (0.7988)	-0.0794 (0.6389)	6.0180 (21.981)	-0.3621 (0.9958)
intercept_2		-0.4042 (0.5910)	-13.062 (40.395)	-0.5381 (0.7385)
gender_2		0.2961 (0.9318)	0.5644 (0.8026)	0.3274 (0.9198)
intercept_3		0.0394 (0.7034)	-4.3260 (13.075)	-0.0734 (0.7340)
gender_3		0.5328 (0.8321)	12.117 (37.712)	0.6421 (0.8248)
φ_{12}		6.7429 (979.38)	-4010.0 (13044)	3.9641 (49.283)
φ_{13}		20.529 (2.86E6)	13.581 (2462i)	4.8504 (115.91)
φ_{23}		2.6890 (2.0828)	62.209 (205.04)	3.3699 (217.70)
φ_{123}		-4.5287 (7.69E43)	226.80 (45980)	-1.3710 (707.5i)
Pattern 2				
intercept_1	-0.8473 (0.4880)	1.4118 (7.2243)	1.4118 (7.2202)	1.4118 (7.2269)
gender_1	0.6802 (0.6371)	-1.9439 (8.2479)	-1.9439 (8.2396)	-1.9439 (8.2498)
intercept_2	-1.0986 (0.5164)	-4.4028 (12.501)	-4.4028 (12.497)	-4.4028 (12.500)
gender_2	1.0986 (0.6583)	4.1350 (12.071)	4.1350 (12.064)	4.1350 (12.070)
intercept_3		13.651 (46.566)	13.651 (46.555)	13.651 (46.566)
gender_3		-13.221 (46.112)	-13.221 (46.107)	-13.221 (46.111)
φ_{12}	2.9199 (0.8145)	3.9217 (28.152)	3.9217 (62.726)	3.9217 (48.873)
φ_{13}		2596.9 (2.86E6)	2596.9 (8249.1)	2596.9 (8609.7)
φ_{23}		-9.8258 (54.313)	-9.8258 (73.867)	-9.8258 (224.30)
φ_{123}		2581.9 (7.69E43)	2581.9 (2.62E22)	2581.9 (2.62E22)
Pattern 3				
intercept_1	0.1956 (0.2376)	0.1956 (0.2376)	0.1956 (0.2376)	0.1956 (0.2376)
gender_1	-0.5525 (0.2886)	-0.5525 (0.2886)	-0.5525 (0.2886)	-0.5525 (0.2886)
intercept_2	0.5107 (0.2437)	0.5107 (0.2437)	0.5107 (0.2437)	0.5107 (0.2437)
gender_2	-0.3522 (0.2929)	-0.3522 (0.2929)	-0.3522 (0.2929)	-0.3522 (0.2929)
intercept_3	0.5824 (0.2447)	0.5824 (0.2447)	0.5824 (0.2447)	0.5824 (0.2447)
gender_3	0.0679 (0.2987)	0.0679 (0.2987)	0.0679 (0.2987)	0.0679 (0.2987)
φ_{12}	3.1325 (0.3889)	3.1325 (0.3889)	3.1325 (0.3889)	3.1325 (0.3889)
φ_{13}	2.1026 (0.3533)	2.1026 (0.3533)	2.1026 (0.3533)	2.1026 (0.3533)
φ_{23}	2.9471 (0.3726)	2.9471 (0.3726)	2.9471 (0.3726)	2.9471 (0.3726)
φ_{123}	1.2110 (0.9510)	1.2110 (0.9510)	1.2110 (0.9510)	1.2110 (0.9510)

patterns. Thus, this should at least be taken into account. The last model, however, is too complex, and it was hard to reach convergence, due to a lot of sparse or even empty cells for the originally incomplete patterns. A golden mean has to be chosen between the simplest and most complex model. Also non-significant covariate effects should be removed from the model.

7.8 Conclusions

In this chapter, we reviewed the general concepts of pattern-mixture models and the technique of identifying restrictions to specify the conditional distribution of the unobserved measurements, given the observed ones. Then, these concepts were extended to categorical outcomes. Also, a solution is suggested to handle the intermittent missing data. They can be identified using the same identifying restrictions as were used for monotone missingness.

Since interest is often in an overall covariate effect, and not in the pattern-specific effects only, and since this overall effect cannot be obtained as simple as in the case of Gaussian data by averaging the pattern-specific effects, a complete section was devoted to the derivation of a marginal effect of interest. It was also shown that the method of averaging does not lead to a diluted effect, but likewise can lead to an equal or increased effect, and thus should not be used as a conservative estimate.

The fluvoxamine data were reanalyzed, using the method of pattern-mixture models, including identifying restrictions. Several models were fitted to the multiply-imputed sets of data. Some were too simple, others too complex, leading to sparse or even empty cells for the originally incomplete patterns, and resulted in convergence problems. Nevertheless, the different ways in which the data were analyzed, can be seen as a sensitivity analysis. Especially the use of different identifying restrictions is a first step in assessing the sensitivity of the assumptions made.

Further research can be devoted to the analysis of other data sets with more intermittent missingness, such that the suggestions made in this chapter to identify those missing values, can be explored.

8

Sensitivity Analysis Tools for Categorical Data

We already indicated in previous chapters that models for incomplete longitudinal data, especially the parametric MNAR models, are vulnerable to model-assumption related sensitivity, and therefore it is imperative to study this phenomenon. It turns out in practice that there are numerous subtle issues not encountered with complete tables. Some of these issues are purely technical (nonunique, invalid, or boundary estimates), others are of a more interpretational and philosophical nature (e.g., models that yield the same or similar fits to the observed data can produce qualitatively different predictions for the unobserved data; Molenberghs *et al.*, 1999a). With the growing volume of MNAR based selection models the need for a careful understanding of such sensitivities, and the development of tools to discern their impact, has been growing as well (Glynn, Laird and Rubin, 1986). Early, important contributions to sensitivity analysis have been made by Draper (1995) and Copas and Li (1997).

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 useful approaches. The simplest procedure is to fit a selected number of (non-random) models which 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.

However, a sensitivity analysis can also be performed on the level of individual observations instead of on the level of the models. In that case, interest is directed towards finding those individuals who drive the conclusions towards one or more MNAR models. Therefore, the influence of every individual separately will be explored. Two techniques exist, i.e., global influence and local influence (Cook, 1986). The global influence methodology, also known as the case-deletion method (Cook and Weisberg, 1982) is introduced by Cook (1979, 1986) in linear regression, and by Molenberghs *et al.* (2003) and Thijs, Molenberghs and Verbeke (2000) in linear mixed models. Verbeke *et al.* (2001b) and Thijs, Molenberghs and Verbeke (2000) already used local influence on the Diggle and Kenward (1994), which is based on a selection model, integrating a linear mixed model for continuous outcomes with logistic regression for dropout. Later, Van Steen *et al.* (2001) adapted these ideas to the model of Molenberghs, Kenward and Lesaffre (1997), for monotone repeated ordinal data.

In Section 8.1 we focus on the global influence for the Dale-Dale models from Chapter 6, and an application on the Health Interview Survey data is considered in Section 8.2. In Section 8.3 the idea of local influence is developed for the extension of the Baker, Rosenberger and DerSimonian (1992) model (Chapter 5), and applied to the fluvoxamine data in Section 8.4.

8.1 Global Influence

One of the tools to perform a sensitivity analysis is global influence, starting from case-deletion. This methodology is based on the difference in log-likelihood between the model fitted to the data set as a whole on the one hand, and the data set minus one subject on the other hand. We will apply these global influence ideas to the Dale-Dale models from Section 6.3. Denote the log-likelihood function, corresponding to model (6.13), as

$$\ell(\boldsymbol{\phi}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\phi}), \tag{8.1}$$

in which $\ell_i(\boldsymbol{\phi})$ is the contribution of the ith individual to the log-likelihood, and where $\boldsymbol{\phi}$ is the s-dimensional vector as defined in Section 6.3. Further, we denote by

$$\ell_{(-i)}(\boldsymbol{\phi}), \tag{8.2}$$

the log-likelihood function, where the contribution of the ith subject has been removed. Cook's distances (CD) are based on measuring the discrepancy between either the maximized log-likelihoods (8.1) and (8.2) or (subsets of) the estimated parameter vectors $\widehat{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\phi}}_{(-i)}$, with obvious notation. Precisely, we will consider both

$$CD_{1i} \;=\; 2(\widehat{\ell}(\boldsymbol{\phi}) - \widehat{\ell}_{(-i)}(\boldsymbol{\phi})), \tag{8.3}$$

as well as

$$CD_{2i}(\boldsymbol{\phi}) \;=\; 2\,(\widehat{\boldsymbol{\phi}} - \widehat{\boldsymbol{\phi}}_{(-i)})'\ddot{L}^{-1}\,(\widehat{\boldsymbol{\phi}} - \widehat{\boldsymbol{\phi}}_{(-i)}). \tag{8.4}$$

Formulation (8.4) easily allows to consider the global influence in a subvector of $\boldsymbol{\phi}$, such as the measurement model parameters $\boldsymbol{\theta}$, or missingness model parameters $\boldsymbol{\psi}$. This will be indicated using notation of the form $CD_{2i}(\boldsymbol{\theta})$ and $CD_{2i}(\boldsymbol{\psi})$.

Performing a global influence analysis on data with categorical outcomes is less time consuming than on data with continuous outcomes, since the data can be summarized in cells such as in Table 2.3. Thus, instead of removing every subject one by one, we only need to remove one subject per cell and per covariate level.

8.2 Global Influence Analysis of the Health Interview Survey Data

We will apply the global influence ideas to the HIS data, which were analyzed using the Dale-Dale models, in order to investigate the conclusions that were reached in Section 6.4. Whereas all 9 models, we focus on Model 3 and Model 4, since those were the most adequate ones when, on the one hand, a varying gender effect and, on the other hand, a varying education effect is included in both parts of the model. In addition, we will consider Model 7, a supermodel of both Model 3 and 4, which was, in both settings, borderline non-significant.

It was already mentioned in Section 8.1 that for categorical outcomes a global influence analysis is reduced to removing one subject per respons combination and per covariate level. In the HIS data, this means that, when including gender, only 2×9 subjects need to be removed, and when including education, 5×9 subjects. Results of the global influence analyses are shown in Figures 8.1–8.6.

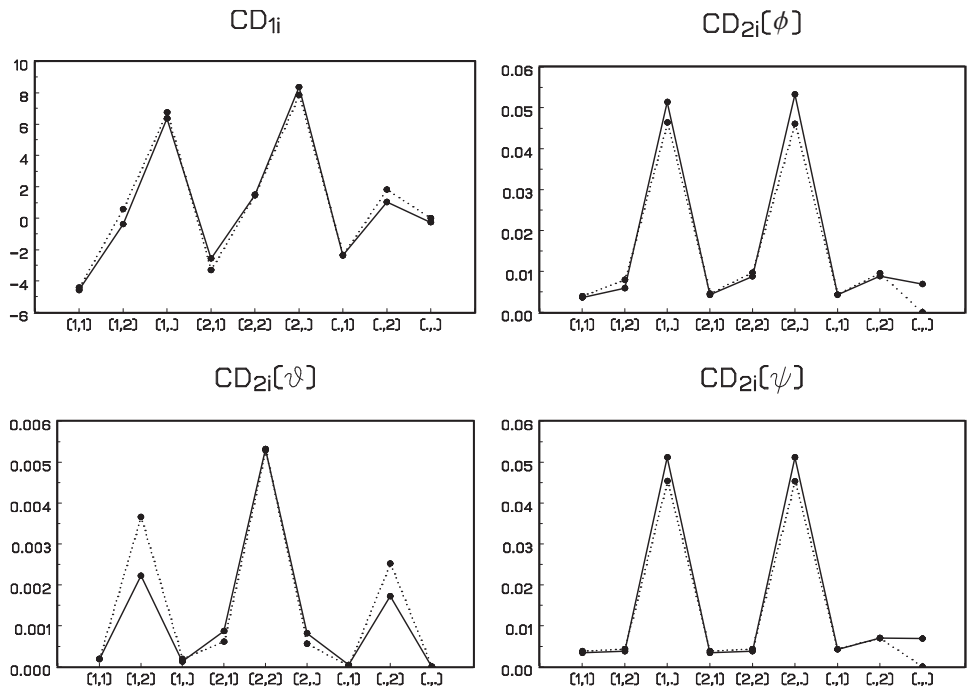


Figure 8.1: *Health Interview Survey Data. Global influence for Model 3 with varying gender effect. Solid line for males, dotted for females.*

When gender is included in the model, the influence graphs for the Cook's distances on the log-likelihood scale (CD_{1i}), for the complete parameter vector ($CD_{2i}(\boldsymbol{\phi})$), and for the parameter vector of the missingness model ($CD_{2i}(\boldsymbol{\psi})$) look similar in all models considered. The influence graphs for the Cook's distances for the parameter vector of the measurement model ($CD_{2i}(\boldsymbol{\theta})$) are similar for Models 3 and 4, but different for Model 7. We will discuss these in turn.

The largest CD_{1i}, $CD_{2i}(\boldsymbol{\phi})$, and $CD_{2i}(\boldsymbol{\psi})$ were measured for subjects with a missing value for fixed general practitioner. This can be explained as follows. Models 3, 4, and 7 all assume that a missing value for a specific outcome depends on the value of fixed general practitioner. Since the value of fixed general practitioner is not available, these assumptions heavily rely on the value such an individual *would have had* had the measurements been made, thus a strong sensitivity. Males having no measurements at all (for mental health and fixed general practitioner) have a non-negligible effect on the missingness model parameters, while females do not. In all other cases, the influence for males and females is comparable. Varying parameterization invariably

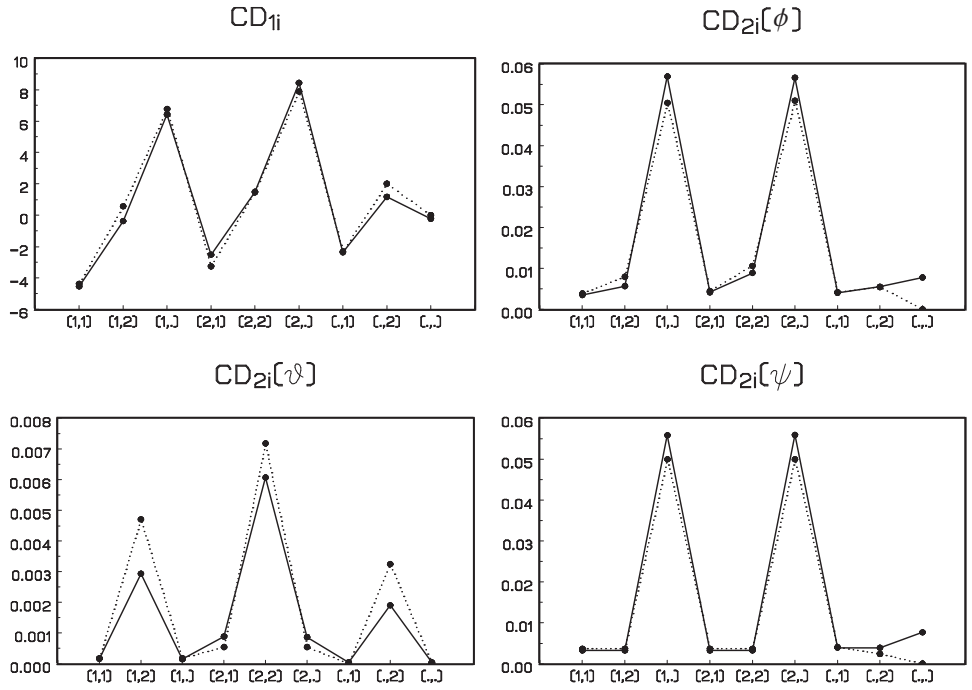


Figure 8.2: *Health Interview Survey Data. Global influence for Model 4 with varying gender effect. Solid line for males, dotted for females.*

produces the same result.

In Models 3 and 4, $CD_{2i}(\boldsymbol{\theta})$ is largest for subjects without fixed general practitioner, but this influence is negligible. In Model 7 subjects without fixed general practitioner are also influential, but they are not the only ones. Males with no measurement for fixed general practitioner have a relatively high influence, compared with females, while in all other cases their influence on the measurement model parameters is comparable, or higher for females than for males. Nevertheless, the Cook's distances for the measurement model parameters are much smaller than those for the missingness model parameters, which shows that the measurement model parameters are remarkably stable. This conclusion has also been drawn in Section 6.4.

When education is included in the model, the influence graphs look similar for all models considered. For CD_{1i}, $CD_{2i}(\boldsymbol{\phi})$ and $CD_{2i}(\boldsymbol{\psi})$ the largest Cook's distances were measured for subjects with a missing value for fixed general practitioner. This result is similar to the one for gender, and can be explained in the same way. The Cook's distances $CD_{2i}(\boldsymbol{\theta})$ are largest for subjects without fixed general practitioner,

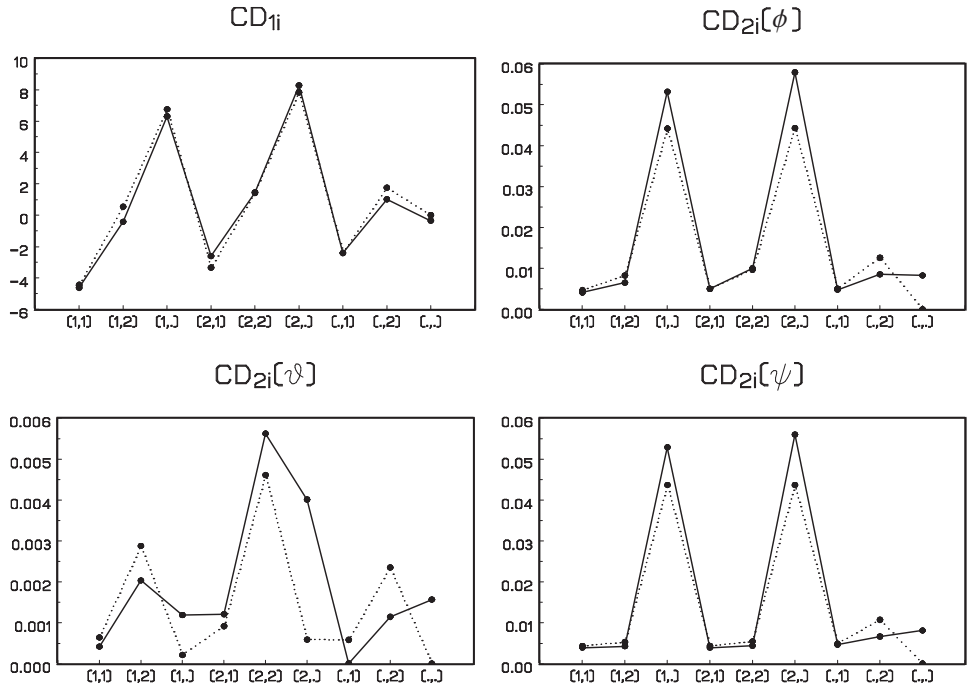


Figure 8.3: *Health Interview Survey Data. Global influence for Model 7 with varying gender effect. Solid line for males, dotted for females.*

for all levels of education, but this influence is negligible. Again, the Cook's distances for the measurement model parameters are much smaller than those for the missingness model parameters. Some more conclusions have to be drawn. Subjects having no measurements at all (for mental health and fixed general practitioner) have no influence on the measurement model parameters, but a non-negligible effect on the missingness model parameters, since the "empty" observations are explicitly modeled in the Dale-Dale models. Subjects without education (solid line) have the highest influence on all parameters of the model, possibly due to their low presence (267 out of 10705) in the data.

8.3 Local Influence

A drawback of global influence is that the specific cause of the influence cannot be retrieved, since by deleting a subject all types of influence stemming from it are lumped together. Local influence however, studies the effect of infinitesimally small

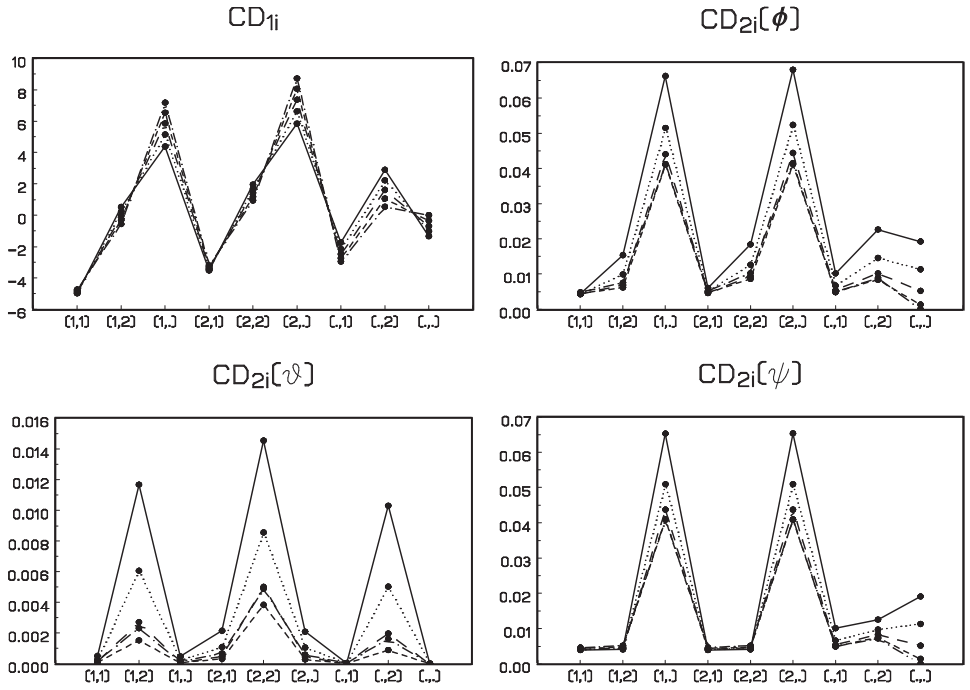


Figure 8.4: *Health Interview Survey Data. Global influence for Model 3 with varying education effect. Solid line for no education, dotted for primary education, dashed for low secondary education, short dashes for high secondary education, dots and dashes for higher education.*

model perturbations around a given null model.

Verbeke *et al.* (2001b), Thijs, Molenberghs and Verbeke (2000), and Molenberghs *et al.* (2001b) studied local influence in the context of the Diggle and Kenward (1994) model where a linear mixed measurement model is combined with logistic models for dropout. To this end, they considered the following perturbed version of dropout model (3.6):

$$\begin{aligned} \text{logit}(g(\boldsymbol{h}_{ij}, y_{ij})) &= \text{logit}\left[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)\right] \\ &= \boldsymbol{h}'_{ij}\boldsymbol{\psi} + \omega_i y_{ij} \qquad\qquad i = 1, \ldots, N, \end{aligned} \tag{8.5}$$

where the ω_i are local, individual-specific perturbations around a null model. They should not be confused with subject-specific parameters. Their null model was the MAR model, corresponding to setting $\omega = 0$ in (3.6). When small perturbations in a specific ω_i lead to relatively large differences in the model parameters, then this

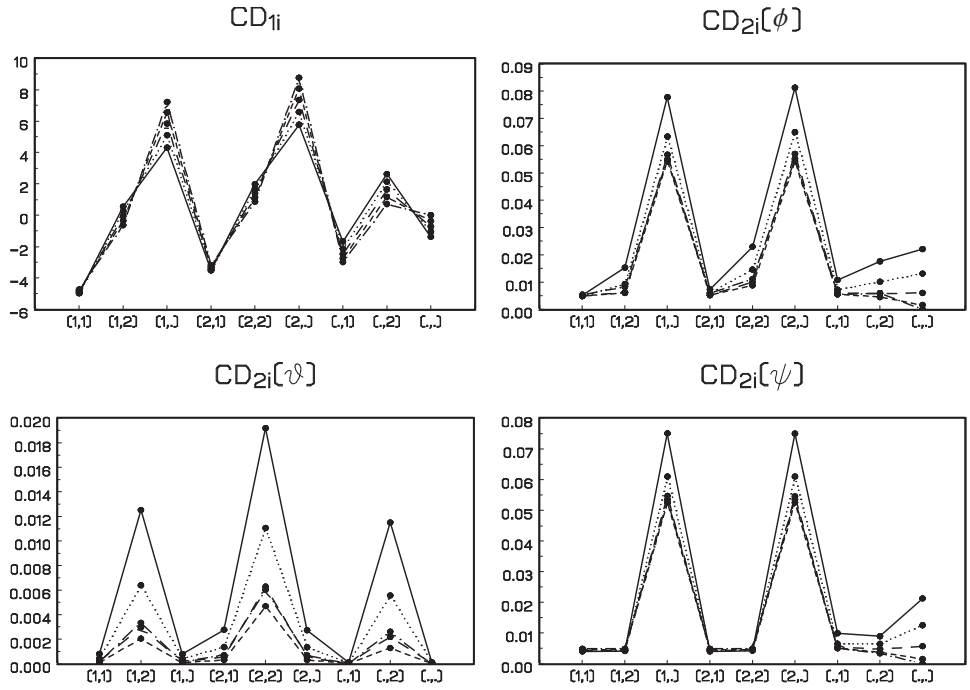


Figure 8.5: *Health Interview Survey Data. Global influence for Model 4 with varying education effect. Solid line for no education, dotted for primary education, dashed for low secondary education, short dashes for high secondary education, dots and dashes for higher education.*

suggests that these subjects may have a large impact on the final analysis.

We will consider perturbations of a given BRD model in the direction of a model with one more parameter in which the original model is nested, implying that perturbations lie along the edges of Figure 5.1: for each of the nested pairs in Figure 5.1, the simpler of the two models equates two parameters from the more complex one. For example, BRD4 includes $\beta_{.k}$, $(k = 1, 2)$, whereas in BRD1 only $\beta_{..}$ is included. For the influence analysis, ω_i is then included as a contrast between two such parameters; for the perturbation of BRD1 in the direction of BRD4, one considers $\beta_{..}$ and $\beta_{..} + \omega_i$. The vector of all ω_i's defines the direction in which such a perturbation is considered. The BRD family provides a versatile environment for sensitivity analysis, as opposed to the Diggle and Kenward model where, in its basic form, only a few missingness parameters are present. This is due in part to the ability to handle non-monotone missingness. Note that the influence analysis focuses on the missingness model, rather

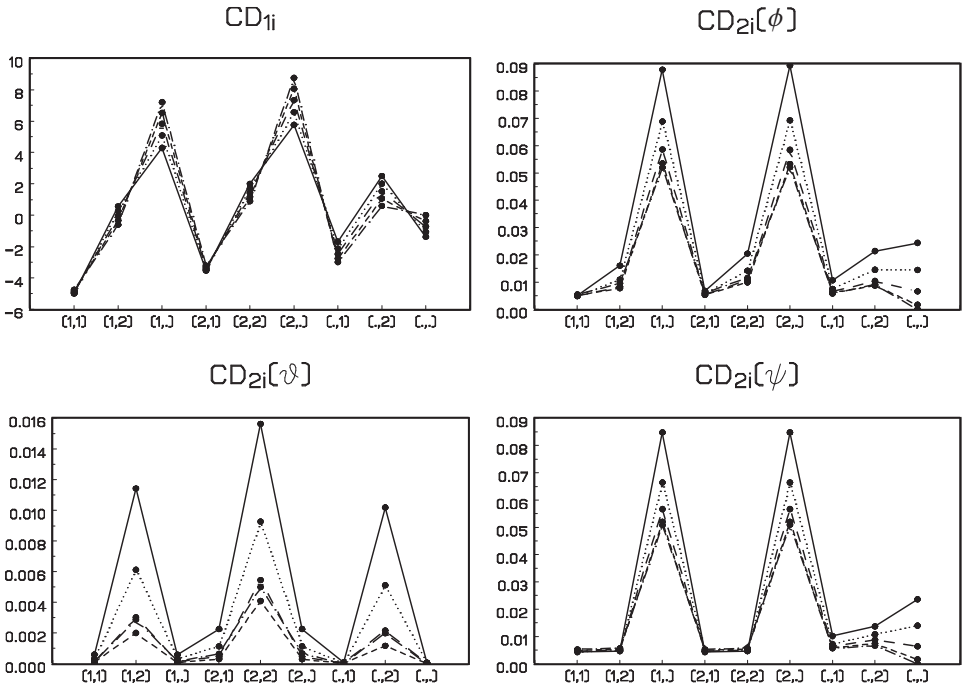


Figure 8.6: *Health Interview Survey Data. Global influence for Model 7 with varying education effect. Solid line for no education, dotted for primary education, dashed for low secondary education, short dashes for high secondary education, dots and dashes for higher education.*

than on the measurement model parameters. This may be seen as slightly odd since often scientific interest focuses on the measurement model parameters. However, it has been documented (Rubin, 1994; Kenward, 1998; Verbeke *et al.*, 2001b) that the missingness model parameters are often the most sensitive ones to take up all kinds of misspecification and influential features. These may then, in turn, impact conclusions coming from the measurement model parameters (e.g., time evolution) or combinations from both (e.g., covariate effects for certain groups of responders).

Let us now introduce the key concepts of local influence (Cook, 1986; Verbeke *et al.*, 2001b). Since the resulting influence diagnostics can in many cases be expressed analytically, they often can be decomposed in interpretable components, which yields additional insight. We denote the log-likelihood corresponding to model (5.3)–(5.4) by $\ell(\boldsymbol{\phi}|\boldsymbol{\omega}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\phi}|\omega_i)$, in which $\ell_i(\boldsymbol{\phi}|\omega_i)$ is the contribution of the ith individual, and where $\boldsymbol{\phi} = (\boldsymbol{\theta}, \boldsymbol{\psi})$ is the s-dimensional vector, grouping the parameters of the

measurement model and the dropout model as in (5.5), not including the $N \times 1$ vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)'$ of weights defining the perturbation. Assume that $\boldsymbol{\omega}$ belongs to an open subset Ω of $\mathbb{R}^N$. For $\boldsymbol{\omega}$ equal to $\boldsymbol{\omega_0} = (0, 0, \ldots, 0)'$, $\ell(\boldsymbol{\phi}|\boldsymbol{\omega_0})$ is the log-likelihood corresponding to the simpler of the two BRD models.

Let $\widehat{\boldsymbol{\phi}}$ be the maximum likelihood estimator for $\boldsymbol{\phi}$, obtained by maximizing $\ell(\boldsymbol{\phi}|\boldsymbol{\omega_0})$, and let $\widehat{\boldsymbol{\phi}}_\omega$ denote the maximum likelihood estimator for $\boldsymbol{\phi}$ under $\ell(\boldsymbol{\phi}|\boldsymbol{\omega})$. The local influence approach compares $\widehat{\boldsymbol{\phi}}_\omega$ with $\widehat{\boldsymbol{\phi}}$. Similar estimates indicate that the parameter estimates are robust w.r.t. perturbations in the direction of the extended model. Cook (1986) proposed to measure the distance between $\widehat{\boldsymbol{\phi}}_\omega$ and $\widehat{\boldsymbol{\phi}}$ by the likelihood displacement, defined by $LD(\boldsymbol{\omega}) = 2[\ell(\widehat{\boldsymbol{\phi}}|\boldsymbol{\omega_0}) - \ell(\widehat{\boldsymbol{\phi}}_{\boldsymbol{\omega}}|\boldsymbol{\omega_0})]$. This takes into account the variability of $\widehat{\boldsymbol{\phi}}$. Indeed, $LD(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\phi}|\boldsymbol{\omega_0})$ is strongly curved at $\widehat{\boldsymbol{\phi}}$, which means that $\boldsymbol{\phi}$ 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 $\boldsymbol{\zeta}(\boldsymbol{\omega}) = (\boldsymbol{\omega}', LD(\boldsymbol{\omega}))'$ as $\boldsymbol{\omega}$ varies throughout Ω. Since this so-called *influence graph*, as shown in Figure 8.7 (Lesaffre and Verbeke, 1998), can only be depicted when $N = 2$, Cook (1986) proposed to consider local influence, i.e., at the normal curvatures $C_{\boldsymbol{h}}$ of $\boldsymbol{\zeta}(\boldsymbol{\omega})$ 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

$$\boldsymbol{\Delta}_i = \left. \frac{\partial^2 \ell_i(\boldsymbol{\phi}|\omega_i)}{\partial \omega_i \partial \boldsymbol{\phi}} \right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}},\, \omega_i = 0} \tag{8.6}$$

and define Δ as the $(s \times N)$ matrix with $\boldsymbol{\Delta}_i$ as its ith column. Let $\ddot{L}$ denote the $(s \times s)$ matrix of second order derivatives of $\ell(\boldsymbol{\phi}|\boldsymbol{\omega}_0)$ with respect to $\boldsymbol{\phi}$, also evaluated at $\boldsymbol{\phi} = \widehat{\boldsymbol{\phi}}$. Cook (1986) has then shown that $C_{\boldsymbol{h}}$ can be easily calculated by $C_{\boldsymbol{h}} = 2|\boldsymbol{h}'\Delta'\ddot{L}^{-1}\Delta\boldsymbol{h}|$.

$C_{\boldsymbol{h}}$ can be calculated for any direction $\boldsymbol{h}$. One choice is the vector $\boldsymbol{h_i}$ containing one in the ith position and zero elsewhere, corresponding to the perturbation of the ith subject only, reflecting the influence of allowing the ith subject to drop out in a more general fashion than the others. 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 model to obtain the largest local changes in the likelihood displacement. It is readily seen that $C_{\max}$ is the largest eigenvalue of $-2\ \Delta'\ \ddot{L}^{-1}\ \Delta$, with $\boldsymbol{h}_{\max}$ the corresponding eigenvector. Calculation of local influence measures reduces to evaluation of Δ and $\ddot{L}$ and a convenient computational scheme can be used whenever a program is available

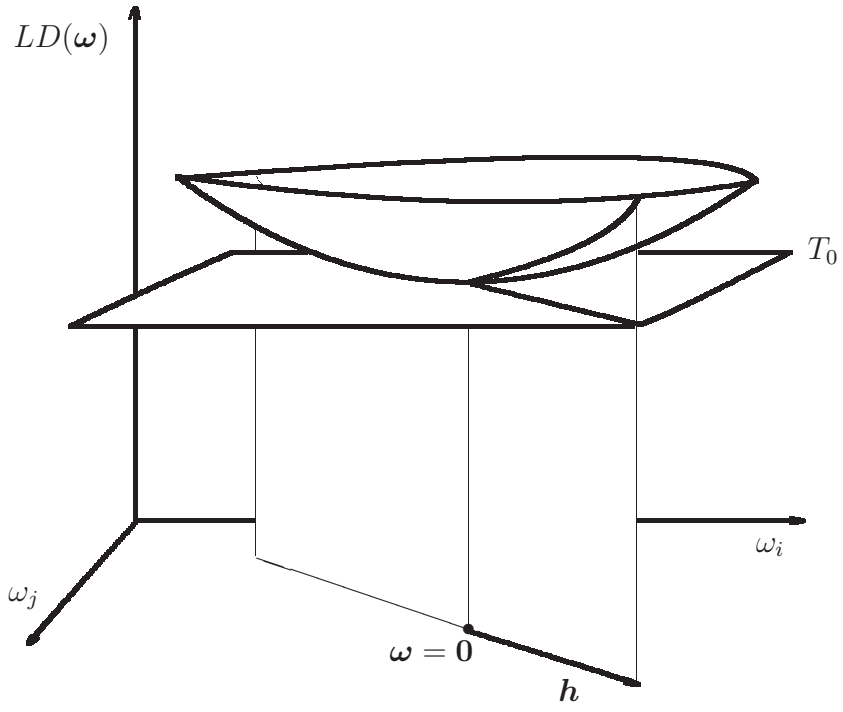


Figure 8.7: *Graphical representation of the influence graph and the local influence approach.*

to fit the full alternative model, i.e., the model at the end of edge in Figure 5.1 since it then suffices to compute the second derivative at $(\widehat{\boldsymbol{\phi}}, \omega_i = 0)$, for each observation separately, from which the $\boldsymbol{\Delta}_i = (\boldsymbol{\phi}, \omega)$ subvector is selected.

8.4 Local Influence Analysis of the Fluvoxamine Data

We will apply the local influence ideas to the fluvoxamine data, which were analyzed using the BRD models, in order to contradict or strengthen the conclusions of Section 5.4. Whereas all comparisons along the edges of Figure 5.1 are possible, we focus on the comparison BRD1–4 (Figure 8.8), since the first one was the most adequate model when no duration effect is included and when duration is included in both

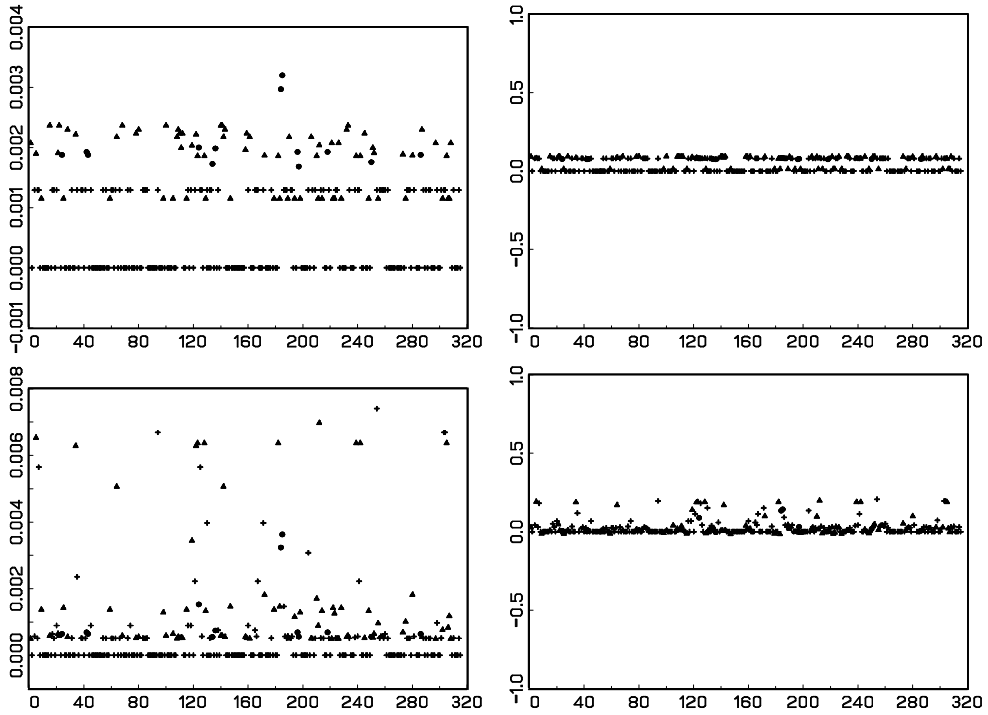


Figure 8.8: *Fluvoxamine Data. Index plots of C_i (left panels) and the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature (right panels) for comparison BRD1–4, without (top panels) or with (bottom panels) duration as a covariate in the missingness models.*

parts of the model, while the second one was the model of choice when duration is included in the measurement model only. In addition, we will consider the comparisons BRD4–7 (Figure 8.9) and BRD4–8 (Figure 8.10), the supermodels of BRD4. The symbols used in these figures are: +: both observations are available, (1,1) type; ▲: only the first observation is available, (1,0) type; ■: only the second observation is available, (0,1) type; ●: both measurements are missing, (0,0) type.

We consider C_i and $\boldsymbol{h}_{\max}$. The top right panel in Figure 8.8 essentially shows no structure, while in the top left there are two important observations. First, a layering effect is present. This is not surprising, since there are quite a number of discrete features in the model: the responses and the missingness patterns. On the other hand, the continuous covariate duration is included in the measurement model. In this case, mainly the missingness patterns are noticeable, although the top layer shows a good deal of variability.

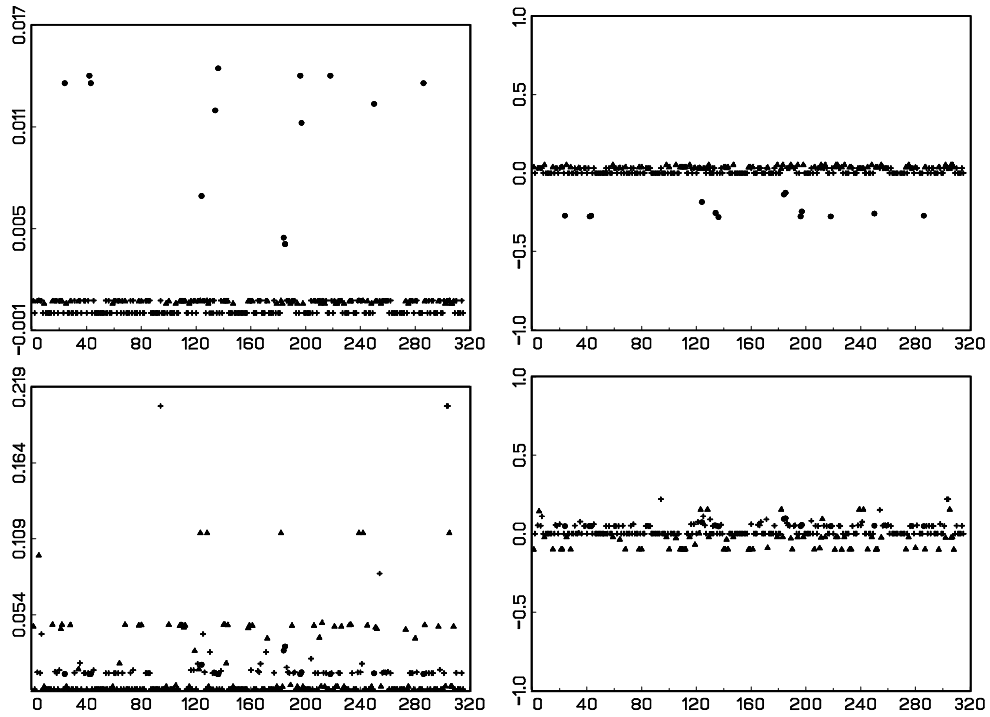


Figure 8.9: *Fluvoxamine Data. Index plots of C_i (left panels) and the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature (right panels) for comparison BRD4–7, without (top panels) or with (bottom panels) duration as a covariate in the missingness models.*

Two views can be taken. Either, focus is on two observations, #184 and #185, that stand out. These subjects have no measurements at all for side effects. Alternatively, the entire pattern without follow up measurements can be studied. We will return to this issue later in this section. This phenomenon is in contrast to the analyses made by Verbeke *et al.* (2001b) and Molenberghs, Kenward and Goetghebeur (2001a) who found that the influential observations are invariably completers. In this case, the situation is different since the "empty" observations are explicitly modeled in the BRD models. An equivalent conclusion was reached in Section 8.2, where the subjects with the largest global influence value for the missingness model parameters, were also empty observations. Therefore, assumptions about the perturbations in the direction of such observations have an impact on the values such an individual *would have had* had the measurements been made; hence a strong sensitivity. This illustrates that studying influence by means of perturbations in the missingness model

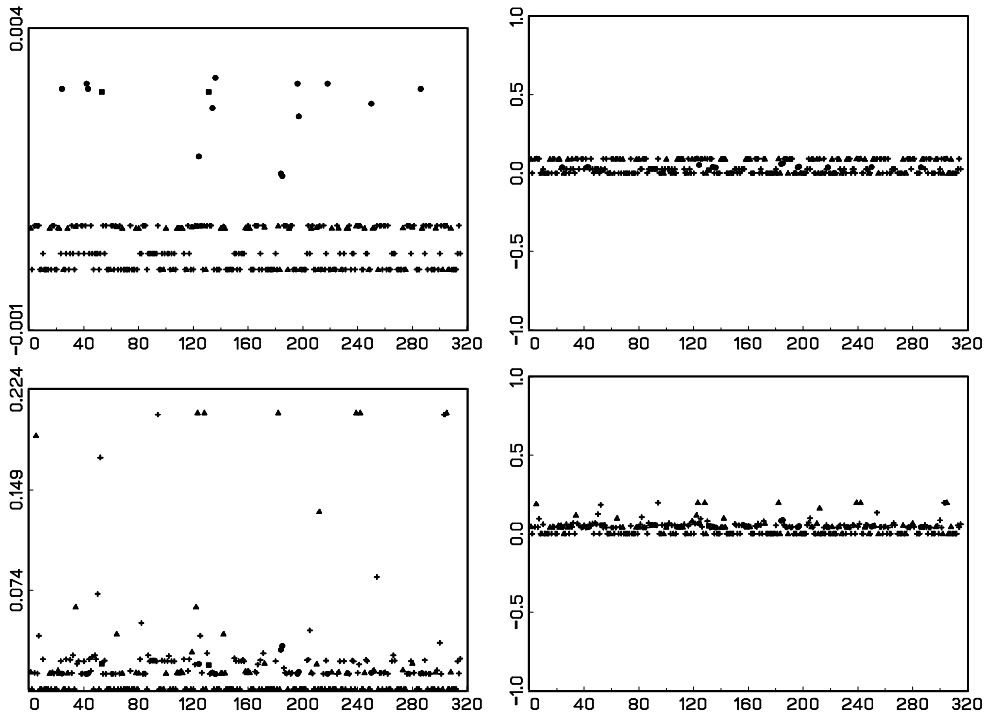


Figure 8.10: *Fluvoxamine Data. Index plots of C_i (left panels) and the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature (right panels) for comparison BRD4–8, without (top panels) or with (bottom panels) duration as a covariate in the missingness models.*

may lead to important conclusions regarding the measurement model parameters. Indeed, the measurement model conclusions depend, not only on the observations actually made, but also on the expectation of the missing measurements. In an MNAR model, such expectations depend on the missingness model as well, since they are made *conditional on an observation being missing*. A high level of sensitivity means that the expectations of the missing outcomes and the resulting measurement model parameters strongly depend on the missingness model. Based on this consideration, Verbeke *et al.* (2001b) showed that, in spite of the fact that completers cannot have a direct influence on the measurement model parameters, they still can do so implicitly. Given the strong level of dependence of missingness models on assumptions, it is crucial to investigate the sensitivity of the measurement model conclusions, using local influence that targets the missingness model. As stated earlier, the only continuous characteristics of the observations are the levels for duration. These are 38 and 41 for

observations #184 and #185, respectively, the largest values within the group without observations and the 91st and 92nd percentile values within the entire sample. Thus, the conclusions are driven by a very high value of duration.

Consider the bottom panels of Figure 8.8. The right hand panel still shows little or no structure. On the left hand side, the layering has been blurred due to the occurrence of duration as a continuous feature into the missingness model. The fact that no sets of observations stand out as such, confirms the impression that a good fit has been obtained by including duration in both parts of the model. Consider Figure 8.9. A qualitative difference with Figure 8.8 (top left panels) is that the entire group with no follow-up measurements shows more influential than all other subjects. In this case, $\boldsymbol{h}_{\max}$ displays the same group of subjects with no follow-up. In the top left panel of Figure 8.10 also the entire group with no follow-up measurements shows more influential, but now also the two non-monotone subjects have a large value for the C_i measure. The top right panel ($\boldsymbol{h}_{\max}$) shows no structure. However, all of this disappears when one turns to the bottom panels of Figures 8.9 and 8.10, again underscoring the importance of duration in the missingness model. The consequence of these findings is that, as soon as duration is included in the missingness model, reasonable confidence can be put into the conclusions. Nevertheless, based on the comparison BRD1–4, it seems wise to further study the effect of subjects #184 and #185, as well as from the group without follow up. To this effect, three additional analyses are considered: two sets pertain to removal of subjects #184 and #185: without (I) and with (II) duration as a covariate in the measurement model. We do not consider removal in case duration is included in the missingness model since, in this case, these two subjects did not show up as locally influential. Finally, removing all subjects without follow-up measurements and using duration as covariate in the measurement model is reported as family III. Results are presented in Table 8.1.

Analysis I prefers BRD1 and analysis II prefers BRD4, although slightly less extreme than before: likelihood ratio test statistics for BRD1–4, BRD4–7, and BRD4–8 are 6.60, 3.64, and 3.08, respectively, compared with 7.10, 2.10, and 1.52 obtained initially. However, while the two subjects deleted in I and II cannot explain the apparent non-random missingness, the same conclusions are reached when all subject in pattern (0,0) are deleted (analysis III), since then a few likelihood ratios are significant (7.17, attained for BRD3–7 and for BRD5–8; and 7.32 for BRD1–4). Thus, removing these subjects does not change the conclusions about the non-random nature of the data. This is useful supplemental information: it confirms that the largest impact on the conclusion regarding the nature of missingness is coming from the inclusion of duration, and neither from isolated individuals, nor from a specific missingness

Table 8.1: *Fluvoxamine Data. Negative log-likelihood values for three additional sets of analysis. I: #184 and #185 removed, no covariates; II: #184 and #185 removed, duration as covariate in the measurement model; III: all observations in the (0,0) group removed, duration as covariate in the measurement model.*

Set	BRD1	BRD2	BRD3	BRD4	BRD5	BRD6	BRD7	BRD8	BRD9
I	559.59	558.18	558.70	558.18	558.97	557.59	557.32	557.59	557.32
II	543.65	541.87	542.16	540.35	542.43	540.61	538.53	538.81	540.34
III	496.19	494.33	495.26	492.53	495.53	493.71	491.67	491.95	493.43

pattern. It is pleasing that the final analysis encompasses all subjects and therefore avoids the need of subject deletion.

Subjects in an influence graph are displayed without a particular order. Several alternatives are possible. For example, one could order the subjects by covariate level, but this method cannot be considered when there are several covariates. Alternatively, the subjects could be ordered by C_i or h_i level, but then different orderings would exist on different plots.

8.5 Conclusions

A number of sensitivity tools have been proposed and used to strengthen the data analytic findings from Chapters 5 and 6. The already existing methods of global and local influence have been adjusted to assess the influence in the case of selection models for bivariate binary outcomes subject to non-monotone missingness. The results are also discussed in Jansen *et al.* (2003) and Jansen and Molenberghs (2005).

Global influence has the advantage that it is very easy to conduct (remove the subjects one by one) and that the influence of removing a subject can be studied on several aspects separately (e.g., likelihood, measurement and missingness model parameters). Local influence on the other hand has the advantage that no subjects have to be removed, and therefore it is still possible to investigate and interpret the nature of the influence.

In our influence analyses, it turned out that different subsets of patients may be influential for different sets of parameters or for different model extensions or, as was the case here, several analyses may point to the same pair of influential obser-

vations. Second, especially in the local influence graphs, some subgroups of patients almost lay on a straight line, while others formed a cloud. This is due to the combination of categorical aspects (outcomes, non-response patterns) with continuous aspects (covariates). These can lead to different actions, ranging from design and protocol changes in future studies to removal of observations or groups of observations from the analysis of the current study. However, the latter seemed unnecessary, since apparently very stable conclusions were reached.

Clearly, the influence analyses performed here are not the only ones possible. For example, in local influence other perturbation schemes are possible as well, or one could consider a different route of sensitivity analysis altogether. Ideally, several could be considered within an integrated sensitivity analysis. However, the ones considered here already give a lot of insight in the data. The remaining question stays when a subject or a group of subjects can be considered as influential, or what reasons might be related to high influence values. This topic will be studied in the next chapter.

9

The Behavior of Local Influence

The original idea behind the use of local influence methods with an eye on sensitivity analysis was to detect observations that had a high impact on the conclusions *due to their aberrant missingness mechanism*. For example, most missing measurements might be MAR, while a few could be MNAR following one or a few deviating mechanisms. However, in most successful applications, where a seemingly MNAR mechanism turned out to be MAR or even MCAR after removing the influential subjects identified upon the use of local influence, the situation turned out to be more complex than anticipated. Indeed, the influential subjects often are influential for other than missingness related features. So, local influence tends to pick up a lot of different anomalies in the data at hand, not just deviations in the MNAR mechanism.

In this chapter, we aim to further study the method of local influence, not only to better understand its behavior, but also to increase insight in the overall behavior and impact of MNAR mechanisms. This is done using simulations and general modeling considerations.

In Section 9.1, we first repeat some general aspects of local influence and we work out some details for its use with the Diggle and Kenward (1994) model. Then, the rats data are reanalyzed (see also Verbeke *et al.*, 2001b) and a local influence analysis

is performed in Section 9.2. Section 9.3 is dedicated to the behavior of local influence under standard conditions as well as under a number of anomalous scenarios.

9.1 Local Influence Method Applied to the Model of Diggle and Kenward (1994)

Local influence was already introduced in Section 8.3 and applied to a data set of bivariate binary outcomes with non-monotone missingness in Section 8.4. In this chapter, we will focus on the local influence method applied to the Diggle and Kenward (1994) model, which is suitable for continuous outcomes subject to dropout only. Details about this model were already described in Section 3.1.2, more specifically in (3.5) and (3.6).

Verbeke *et al.* (2001b), Thijs, Molenberghs and Verbeke (2000), and Molenberghs *et al.* (2001b) investigated sensitivity of estimation of quantities of interest, w.r.t. assumptions regarding the dropout model. To this end, they considered the following perturbed version of dropout model (3.6):

$$\begin{aligned} \text{logit}(g(\boldsymbol{h}_{ij}, y_{ij})) &= \text{logit}\left[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)\right] \\ &= \boldsymbol{h}'_{ij}\boldsymbol{\psi} + \omega_i y_{ij} \qquad\qquad i = 1, \ldots, N, \end{aligned} \tag{9.1}$$

where the ω_i are local, individual-specific perturbations around a null model. They should not be confused with subject-specific parameters. Our null model will be the MAR model, corresponding to setting $\omega = 0$ in (3.6). Thus, the ω_i are perturbations that will be used only to derive influence measures (Cook, 1986).

Using this proposal, one can study the impact on key model features, induced by small perturbations in the direction, or seemingly so, of MNAR. We are interested in the influence exerted by the dropout model on the parameters of interest. This can be done, for example, by considering (9.1) as the dropout model. When small perturbations in a specific ω_i lead to relatively large differences in the model parameters, then this suggests that these subjects may have a large impact on the final analysis. However, even though we may be tempted to conclude that such subjects drop out non-randomly, this conclusion is misguided since 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 may 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. Likewise, it is possible that subjects, deviating from the bulk of the data because they are generated under MNAR, go undetected by this technique. This begs the question that one needs to reflect carefully upon which anomalous features are typically detected and which ones typically go unnoticed. This will be investigated in Section 9.3. But first, we will consider some more details about how to apply local influence to the model of Diggle and Kenward (1994). The key concepts of local influence were already introduced in Section 8.3, and will not be reviewed anymore.

First note that the dropout mechanism is described by

$$f(d_i|\boldsymbol{y}_i,\boldsymbol{\psi}) = \begin{cases} \displaystyle\prod_{j=2}^{n_i}[1-g(\boldsymbol{h}_{ij},y_{ij})] & \text{for a completer } (d_i = n_i+1), \\ \displaystyle\prod_{j=2}^{d-1}[1-g(\boldsymbol{h}_{ij},y_{ij})]g(\boldsymbol{h}_{id},y_{id}) & \text{for a dropout } (d_i = d \leq n_i), \end{cases}$$

where the g-factors follow from (9.1). The log-likelihood contribution for a complete sequence then is

$$\ell_{i\omega} = \ln f(\boldsymbol{y}_i) + \ln f(d_i|\boldsymbol{y}_i,\boldsymbol{\psi}),$$

where the parameter dependencies are suppressed for notational ease. The density $f(\boldsymbol{y}_i)$ is multivariate normal, following from the linear mixed model (3.5). The contribution from an incomplete sequence is more complicated. Its log-likelihood term is

$$\begin{aligned} \ell_{i\omega} &= \ln f(y_{i1},\ldots,y_{i,d-1}) + \sum_{j=2}^{d-1}\ln[1-g(\boldsymbol{h}_{ij},y_{ij})] \\ &\quad + \ln\int f(y_{id}|y_{i1},\ldots,y_{i,d-1})g(\boldsymbol{h}_{id},y_{id})dy_{id}. \end{aligned}$$

Further details can be found in Verbeke *et al.* (2001b). We need expressions for Δ and $\ddot{L}$. Straightforward derivation shows that the columns $\boldsymbol{\Delta}_i$ of Δ are given by

$$\left.\frac{\partial^2\ell_{i\omega}}{\partial\boldsymbol{\theta}\partial\omega_i}\right|_{\omega_i=0} = \boldsymbol{0}, \tag{9.2}$$

$$\left.\frac{\partial^2\ell_{i\omega}}{\partial\boldsymbol{\psi}\partial\omega_i}\right|_{\omega_i=0} = -\sum_{j=2}^{n_i}\boldsymbol{h}_{ij}y_{ij}g(\boldsymbol{h}_{ij})[1-g(\boldsymbol{h}_{ij})], \tag{9.3}$$

for complete sequences (no drop out) and by

$$\left.\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i}\right|_{\omega_i=0} = [1 - g(\boldsymbol{h}_{id})] \frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\theta}}, \tag{9.4}$$

$$\begin{aligned}\left.\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i}\right|_{\omega_i=0} &= -\sum_{j=2}^{d-1} \boldsymbol{h}_{ij} y_{ij} g(\boldsymbol{h}_{ij})[1 - g(\boldsymbol{h}_{ij})] \\ &\quad -\boldsymbol{h}_{id} \lambda(y_{id}|\boldsymbol{h}_{id}) g(\boldsymbol{h}_{id})[1 - g(\boldsymbol{h}_{id})],\end{aligned} \tag{9.5}$$

for incomplete sequences. All above expressions are evaluated at $\widehat{\boldsymbol{\gamma}}$, and $g(\boldsymbol{h}_{ij}) = g(\boldsymbol{h}_{ij}, y_{ij})|_{\omega_i=0}$, is the MAR version of the dropout model. In (9.4), we make use of the conditional mean

$$\lambda(y_{id}|\boldsymbol{h}_{id}) = \lambda(y_{id}) + V_{i,21} V_{i,11}^{-1} [\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]. \tag{9.6}$$

The variance matrices follow from partitioning the responses as $(y_{i1}, \ldots, y_{i,d-1}|y_{id})'$.

The derivatives of (9.6) w.r.t. the measurement model parameters are

$$\begin{aligned}\frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\beta}} &= \boldsymbol{x}_{id} - V_{i,21} V_{i,11}^{-1} X_{i,(d-1)}, \\ \frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\alpha}} &= \left[\frac{\partial V_{i,21}}{\partial \boldsymbol{\alpha}} - V_{i,21} V_{i,11}^{-1} \frac{\partial V_{i,11}}{\partial \boldsymbol{\alpha}}\right] V_{i,11}^{-1} [\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]\end{aligned}$$

where $\boldsymbol{x}'_{id}$ is the dth row of X_i, and where $X_{i,(d-1)}$ indicates the first $(d-1)$ rows of X_i. Further, $\boldsymbol{\alpha}$ indicates the subvector of covariance parameters within the vector $\boldsymbol{\theta}$.

In practice, the parameter $\boldsymbol{\theta}$ in the measurement model is often of primary interest. Since $\ddot{L}$ is block-diagonal with blocks $\ddot{L}(\boldsymbol{\theta})$ and $\ddot{L}(\boldsymbol{\psi})$, we have that for any unit vector $\boldsymbol{h}$, $C_{\boldsymbol{h}}$ equals $C_{\boldsymbol{h}}(\boldsymbol{\theta}) + C_{\boldsymbol{h}}(\boldsymbol{\psi})$, with

$$C_{\boldsymbol{h}}(\boldsymbol{\theta}) = -2\boldsymbol{h}' \left[\left.\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i}\right|_{\omega_i=0}\right]' \ddot{L}^{-1}(\boldsymbol{\theta}) \left[\left.\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i}\right|_{\omega_i=0}\right] \boldsymbol{h} \tag{9.7}$$

$$C_{\boldsymbol{h}}(\boldsymbol{\psi}) = -2\boldsymbol{h}' \left[\left.\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i}\right|_{\omega_i=0}\right]' \ddot{L}^{-1}(\boldsymbol{\psi}) \left[\left.\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i}\right|_{\omega_i=0}\right] \boldsymbol{h}, \tag{9.8}$$

evaluated at $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$. It now immediately follows from (9.2) and (9.4) that *direct* influence on $\boldsymbol{\theta}$ only arises from those measurement occasions at which dropout occurs. In particular, from (9.4) it is clear that the corresponding contribution is large only if (1) the dropout probability was small but the subject disappeared nevertheless and (2) the conditional mean 'strongly depends' on the parameter of interest. This implies that complete sequences cannot be influential in the strict sense ($C_i(\boldsymbol{\theta}) = 0$) and that incomplete sequences only contribute, in a direct fashion, at the actual dropout time.

However, we make an important distinction between direct and indirect influence. It was shown that complete sequences can have an impact by changing the conditional expectation of the unobserved measurements given the observed ones *and given the dropout mechanism*. Thus, a complete observation which has a strong impact on the *dropout model parameters*, can still drastically change the measurement model parameters and functions thereof.

Expressions (9.7)–(9.8) can be simplified further in specific cases. For example, Verbeke *et al.* (2001b) considered the compound-symmetric situation. Precisely, they were able to split the overal influence in the approximate sum of three components, describing the mean model parameter $\boldsymbol{\beta}$, the variance components σ^2 and τ^2, and the dropout model parameters $\boldsymbol{\psi}$, respectively:

$$\begin{aligned} C_i^{\mathrm{ap}}(\boldsymbol{\beta}) &= 2[1-g(\boldsymbol{h}_{id})]^2(\xi_{id}\boldsymbol{x}_{id}+(1-\xi_{id})\boldsymbol{\rho}_{id})' \\ &\quad \times \sigma^2\left[\sum_{i=1}^{N}\left(\xi_{id}X'_{i(d-1)}X_{i(d-1)}+(1-\xi_{id})R'_{i(d-1)}R_{i(d-1)}\right)\right]^{-1} \\ &\quad \times (\xi_{id}\boldsymbol{x}_{id}+(1-\xi_{id})\boldsymbol{\rho}_{id}), \qquad (9.9) \\ C_i^{\mathrm{ap}}(\sigma^2,\tau^2) &= 2[1-g(\boldsymbol{h}_{id})]^2\xi_{id}^2(1-\xi_{id})^2\overline{[\boldsymbol{h}_{id}-\lambda(\boldsymbol{h}_{id})]}^2 \\ &\quad \times \left(-1,\frac{1}{\tau^2}\right)\ddot{L}^{-1}(\sigma^2,\tau^2)\begin{pmatrix}-1\\ \frac{1}{\tau^2}\end{pmatrix}, \qquad (9.10) \end{aligned}$$

where $R_{i,d-1}=X_{i(d-1)}-\mathbf{1}_{d-1}\overline{X_{i(d-1)}}$, $\overline{X_{i(d-1)}}=\frac{1}{d-1}\mathbf{1}'_{d-1}X_{i(d-1)}$,

$$\overline{\boldsymbol{h}_{id}-\lambda(\boldsymbol{h}_{id})}=\frac{1}{d-1}\mathbf{1}'_{d-1}[\boldsymbol{h}_{id}-\lambda(\boldsymbol{h}_{id})],$$

and

$$\begin{aligned} \ddot{L}(\sigma^2,\tau^2) &= \sum_{i=1}^{N}\frac{d-1}{2(\sigma^2+(d-1)\tau^2)^2} \\ &\quad \times\begin{pmatrix}[\sigma^2+(d-1)\tau^2]^2-\tau^2[2\sigma^2+(d-1)\tau^2] & 1\\ 1 & (d-1)\end{pmatrix}, \end{aligned}$$

For the dropout model parameters, there are no approximations involved, and we have that

$$C_i(\boldsymbol{\psi}) = 2\left(\sum_{j=2}^{d}\boldsymbol{h}_{ij}y_{ij}v_{ij}\right)'\left(\sum_{i=1}^{N}\sum_{j=2}^{d}v_{ij}\boldsymbol{h}_{ij}\boldsymbol{h}'_{ij}\right)^{-1}\left(\sum_{j=2}^{d}\boldsymbol{h}_{ij}y_{ij}v_{ij}\right), \quad (9.11)$$

in which $d=n_i$ for a complete case and where y_{id} needs to be replaced with

$$\lambda(y_{id}|\boldsymbol{h}_{id})=\lambda(y_{id})+(1-\xi_{id})\overline{[\boldsymbol{h}_{id}-\lambda(\boldsymbol{h}_{id})]}$$

for incomplete sequences. Further, v_{ij} equals $g(h_{ij})[1 - g(h_{ij})]$ which is the variance of the estimated dropout probability under MAR.

9.2 Local Influence Analysis of the Rats Data

The rats data, which are introduced in Section 2.5, are first linearized, using the logarithmic transformation $t = \ln(1 + (\text{age} - 45)/10)$ for the time scale, as proposed by Verbeke and Lesaffre (1999). The transformation was chosen such that $t = 0$ corresponds to the start of the treatment. Let y_{ij} denote the jth measurement for the ith rat, taken at $t = t_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, N$. A simple statistical model, as considered by Verbeke *et al.* (2001b), then assumes that y_{ij} satisfies a model of the form (3.5) with common average intercept β_0 for all three groups, average slopes β_1, β_2 and β_3 for the three treatment groups, respectively, and assuming a so-called compound symmetry covariance structure, i.e., with common variance $\sigma^2 + \tau^2$ and common covariance τ^2. τ^2 denotes the random intercept, while σ^2 is the measurement error. The following specific version of dropout model (3.6) will be assumed:

$$\text{logit}\left[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)\right] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij}. \tag{9.12}$$

Parameter estimates are shown in Table 9.1. More details about these estimates and the performance of a local influence analysis can be found in Verbeke *et al.* (2001b). This section will focus on specific details of this local influence analysis.

Figure 9.1 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. Apart from the last one of these graphs, the scales are not unitless and therefore it would be hard to use a common one for all of the panels. This implies that the main emphasis should be on *relative* magnitudes.

The largest C_i are observed for rats #10, #16, #35, and #41, all belonging to the low dose group, and virtually the same picture holds for $C_i(\boldsymbol{\psi})$. They are highlighted in the second panel of Figure 9.2. All four belong to the low dose group. Arguably, their relatively large influence is caused by an interplay of three facts. First, the profiles are relatively high, and hence y_{ij} and h_{ij} in (9.11) are large. Second, since all four profiles are complete, the first factor in (9.11) contains a maximal number of large terms. Third, the computed v_{ij} are relatively large.

Turning attention to $C_i(\boldsymbol{\alpha})$ reveals peaks for rats #5 and #23. Both belong to the control group and drop out after a single measurement occasion. They are highlighted in the first panel of Figure 9.2. To explain this, observe that the relative magnitude

Table 9.1: *Rats Data. Maximum likelihood estimates (standard errors) of completely random, random and non-random dropout models, fitted to the rats data set, with and without modification.*

	Original Data			Modified Data		
Effect	MCAR	MAR	MNAR	MCAR	MAR	MNAR
Measurement model						
β_0	68.61 (0.33)	68.61 (0.33)	68.60 (0.33)	70.20 (0.92)	70.20 (0.92)	70.25 (0.92)
β_1	7.51 (0.22)	7.51 (0.22)	7.53 (0.24)	7.52 (0.25)	7.52 (0.25)	7.42 (0.26)
β_2	6.87 (0.23)	6.87 (0.23)	6.89 (0.23)	6.97 (0.25)	6.97 (0.25)	6.90 (0.25)
β_3	7.31 (0.28)	7.31 (0.28)	7.35 (0.30)	7.21 (0.31)	7.21 (0.31)	7.04 (0.33)
τ^2	3.44 (0.77)	3.44 (0.77)	3.43 (0.77)	40.38 (0.18)	40.38 (0.18)	40.71 (8.25)
σ^2	1.43 (0.14)	1.43 (0.14)	1.43 (0.14)	1.42 (0.14)	1.42 (0.14)	1.44 (0.15)
Dropout model						
ψ_0	-1.98 (0.20)	-8.48 (4.00)	-10.30 (6.88)	-1.98 (0.20)	-0.79 (1.99)	2.08 (3.08)
ψ_1		0.08 (0.05)	0.03 (0.16)		-0.015 (0.03)	0.23 (0.15)
ψ_2			0.07 (0.22)			-0.28 (0.17)
-loglik	550.20	548.80	548.75	609.00	608.85	607.40

of $C_i(\boldsymbol{\alpha})$, approximately given by (9.10), is determined by $1-g(h_{id})$ and $h_{id}-\lambda(h_{id})$. The first term is large when the probability of dropout is small. Now, when dropout occurs early in the sequence, the measurements are still relatively low, implying that the dropout probability is rather small (cf. Table 9.1). This feature is built into the model by writing the dropout probability in terms of the raw measurements with time-independent coefficients rather than, for example, in terms of residuals. Further, the residual $h_{id}-\lambda(h_{id})$ is large since these two rats are somewhat distant from the group by time mean. A practical implication of this is that the time-constant nature of the dropout model may be unlikely to hold.

Since all deviations are rather moderate, we further explore our approach by considering a second analysis where all responses for rats #10, #16, #35, and #41 have been increased with 20 units. The effect of this distortion will primarily be seen in the variance structure. Precisely, such a change is likely to inflate the random intercept variance, at the expense of the other variance components. In doing so, we will illustrate that (1) such a change is likely to show up in the assessment of the dropout

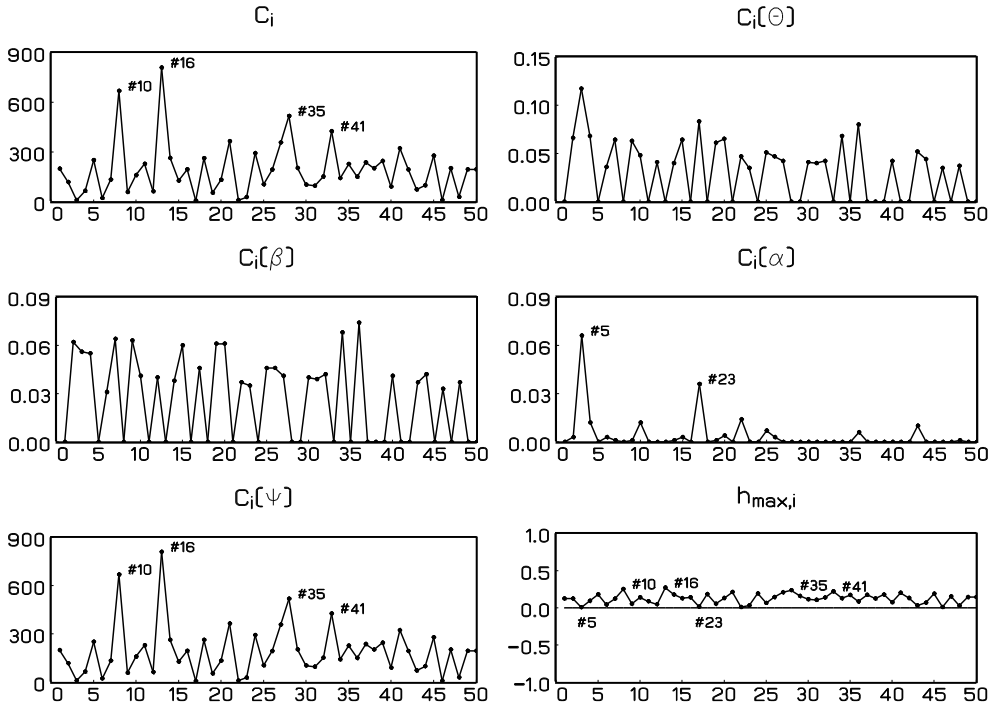


Figure 9.1: *Rats Data. Index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature.*

model, underscoring the sensitivity and that (2) the local influence approach is able to detect such an effect. The parameter estimates for all three models are also shown in Table 9.1. Clearly, while the fixed-effect parameters remain virtually unchanged, the random intercept parameter has, of course, drastically increased. Likewise, the dropout parameters are affected. In addition, the likelihood ratio statistic for MAR versus MCAR changes from 2.8 to 0.3 and for MNAR versus MAR changes from 0.1 to 2.9. Thus, the evidence has shifted from the first to the second test. While all of these statistics seem to be non-significant, there is an important qualitative effect. Moreover, as discussed in Jansen *et al.* (2005b), the use of the classical χ^2-distribution is very questionable for testing MNAR.

In order to check whether these findings are recovered by the local influence approach, let us study Figure 9.3. In line with the changes in parameter estimates, $C_i(\boldsymbol{\beta})$ shows no peaks in these observations but peaks in C_i and $C_i(\boldsymbol{\psi})$ indicate a relatively strong influence from the four extreme profiles.

It will be clear from the above that subjects may turn out to be influential, for

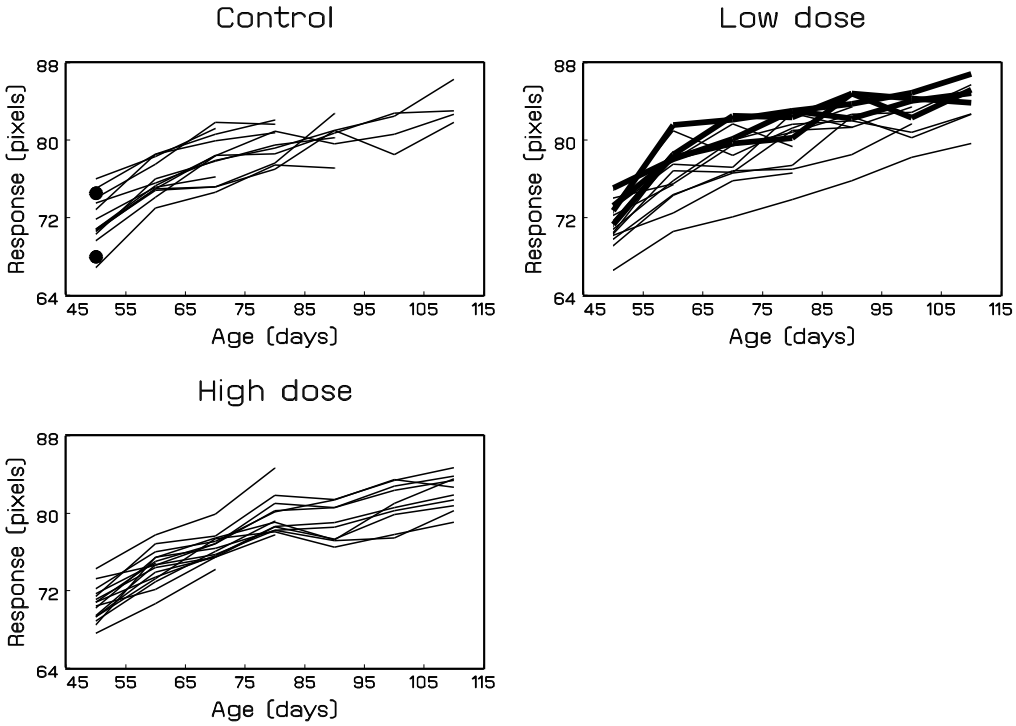


Figure 9.2: *Rats Data. Individual growth curves for the three treatment groups separately. Influential subjects are highlighted.*

reasons different from the nature of the dropout model. Indeed, increasing the profile by 20 units primarily changes the level of the random intercept and ultimately changes the form of the random-effects distribution. Nevertheless, this feature shows in our local influence analysis, where the perturbation is put into the dropout model and not, for example, in the measurement model. This feature requires careful study and will be addressed in the next section.

9.3 The Behavior of Local Influence Methods

A number of concerns have been raised, not only about sensitivity, but also about the tools used to assess sensitivity themselves. For example, Verbeke *et al.* (2001b) noted, based on a case study, that the local influence tool, as defined and implemented in Section 9.2, is able to pick up anomalous features of study subjects that are not necessarily related to the missingness mechanism. In particular, they found that subjects with an unusually high profile, or a somewhat atypical serial correlation

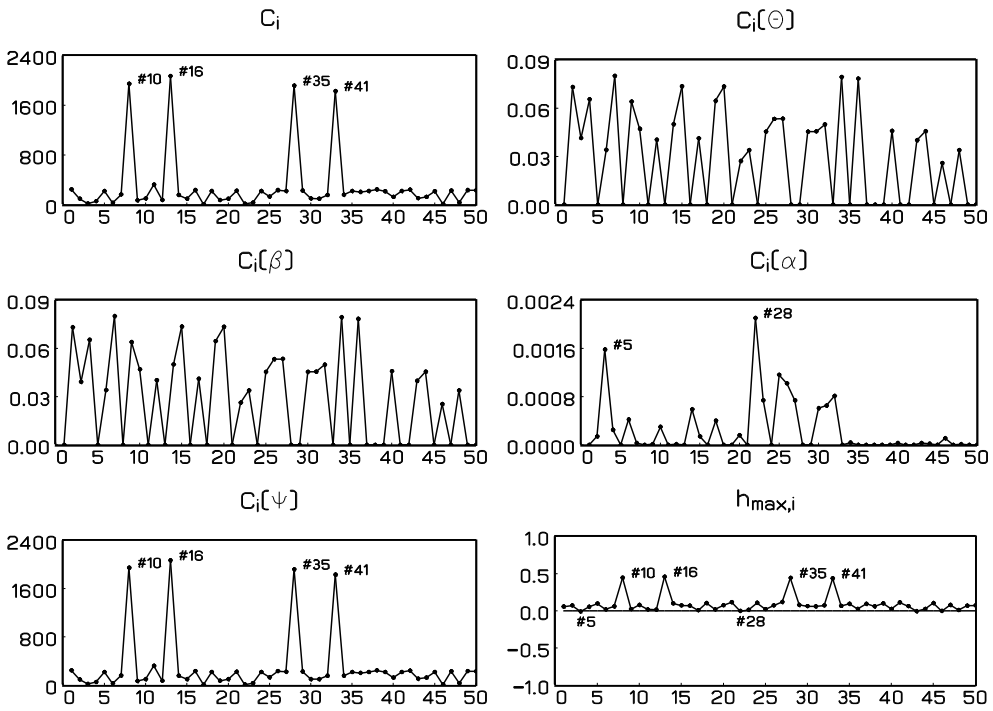


Figure 9.3: *Rats Data. Index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, where 4 profiles have been shifted upward.*

behavior, are detected with the local influence tool. At first sight, this is a little disconcerting, since the ω_i parameter in (9.1) is placed in the dropout model and not in the measurement model, necessitating further investigation regarding which effects are easy or difficult to detect with these local influence methods.

We aim to gain more insight into this phenomenon in a number of ways. To this effect, we undertake a targeted simulation study to explore various sources of influence. First, we took interest in the relative magnitudes of the influence measures to assess how feasible it is to separate influence values that are in line with regular behavior from those that are unduly large. This can be done by proposing a rule of thumb as well as by constructing sampling-based confidence limits and bounds. Second, the impact of one or a few subjects with an anomalous dropout mechanism was explored. Such anomalies are of the type one would intuitively expect to be picked up by the proposed tool. We illustrate that great care is needed. Third, impact due to anomalies in the measurement model was studied. We will show that precisely

such anomalies are relatively easily picked up by the tool, in spite of its conception for anomalies in the missingness mechanism. We offer an explanation for why such behavior is seen.

9.3.1 The Effect of Sample Size

Lesaffre and Verbeke (1998) applied local influence methods to the classical linear mixed-effects model. They introduced ω_i parameters as follows: $\ell = \sum_i \omega_i \ell_i$ where ℓ_i is the log-likelihood contribution for subject i. They were able to show that the sum of the influences is approximately equal to $2s$ with s the total number of parameters in $\boldsymbol{\beta}$ and V_i, when the sample size N is large enough. Their result is based on the fact that, in their local influence contributions, Δ_i in (8.6) becomes

$$\Delta_i = \frac{\partial \ell_i}{\partial \boldsymbol{\theta}},$$

so that the entire expression has the flavor of a contribution to the score test. In our case, as can be seen from (9.4) and (9.5), Δ_i is a second order rather than a first order derivative of the log-likelihood contributions, implying that a, perhaps linear, dependence on the sample size could be envisaged, instead of a constant sum of influence measures. Such a calibration would be beneficial since it would allow to determine critical values, or at least rules of thumb, to determine what is large enough for a subject's influence to undergo further scrutiny.

To this end, we generated a number of data sets, all under the assumption of MAR and with parameters equal to the ones from the rats example. The only difference between these simulations was the sample size. Selected quantiles for sample sizes 50, 500, and 1000 are shown in Table 9.2. Studying even larger sample sizes would be faced with increasing computation times. This also is the reason for considering a single run at size 1000. While the relationship is less clear, as is to be expected, for the maximum value, an obvious trend is seen in the median values and in the 95th percentile. We indeed notice that the influence for a subject decreases linearly with sample size, and hence the total influence for a data set is roughly constant. This is confirmed by a simple multiplicative regression model, which yields that the product of the median and the sample size is constant and equal to 7500. Similarly, the product of the 95th percentile and the sample size to the power 0.96 equals 15,367. To ensure calibration at the individual level, one could then multiply all influences by the sample size. This calibration result implies that the rescaled local influence can be used as a rough measure to determine whether large values are present. For example, one could investigate subjects for which the influence exceeds $1/N$ of the

Table 9.2: *Simulation Study. Selected quantiles of the local influence measures for data sets of different sample sizes, as obtained from simulations and after fitting a simple empirical model.*

	Simulated results						Empirical model		
sample size	50	50	50	500	500	1000	50	500	1000
median	176.7	138.7	146.8	16.5	15.6	7.6	150.0	15.0	7.5
95 percentile	384.5	359.9	317.3	40.6	39.5	18.7	359.4	39.4	20.3
maximum	683.7	674.1	950.7	138.0	137.8	53.6	—	—	—

calibrated total with a certain amount. However, while useful in its own right, we still do not learn anything about the actual distribution of a local influence profile under the null hypothesis. To gain further insight into this problem, we will derive confidence limits and simultaneous confidence bounds in the next section.

9.3.2 Pointwise Confidence Limits and Simultaneous Confidence Bounds for the Local Influence Measure

Since for practical purposes only high values of the influence measures are of interest, we will focus on one-sided (upper) limits and bounds. To this end, we simulated 1000 data sets of 50 rats, using the parameters of the MAR model in Table 9.1. To have a consistent ordering of the C_i values, not based on the arbitrary order of the rats within the set of data, we sorted them from large to small.

This can be seen as, say, 1000 repetitions of a bootstrap experiment with 50 grid points. At each grid point, the 95% pointwise upper confidence limit then simply is the 95% quantile of the $C_{i,j}$ values at that particular grid point. Construction of the simultaneous confidence bounds is based on Besag *et al.* (1995). For each grid point j, order the $C_{i,j}$ values to obtain order statistics $C_{i,j}^{[t]}$ and their corresponding ranks $r_j^{(t)}, t = 1, ..., 1000$. Next, for fixed k, define t_k as the k-th order statistic of the set

$$\left\{ \max \left(\max_{1 \le j \le 50} r_j^{(t)}; 1001 - \min_{1 \le j \le 50} r_j^{(t)} \right); t = 1, ..., 1000 \right\}.$$

Then, by construction, the intervals

$$\left\{ \left[C_{i,j}^{[1001-t_k]}; C_{i,j}^{[t_k]} \right]; j = 1, ..., 50 \right\}$$

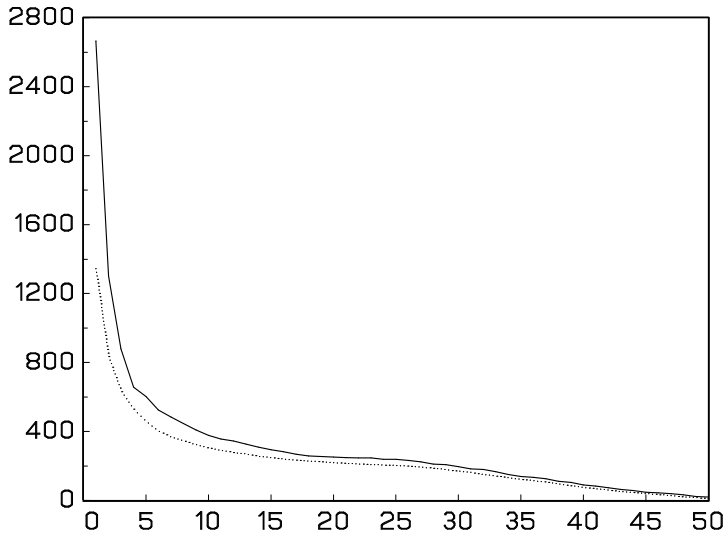


Figure 9.4: *Simulation Study.* 95% *pointwise upper confidence limit (dotted) and* 95% *simultaneous upper confidence bound (solid).*

have a global confidence level of at least $100(k/1000)\%$. To obtain the 95% simultaneous upper confidence bound, simply take $k = 900$, and restrict consideration to the upper bound $C_{i,j}^{[t_k]}$. A graphical representation of this result is given in Figure 9.4.

9.3.3 The Effect of Anomalies in the Missingness Mechanism

To get an idea of the effect of anomalies in the dropout mechanism, a general procedure was followed, as described next. Generate an MNAR data set, fit these data assuming an MAR mechanism in model (9.12), and use the estimates of those model parameters to generate 1000 data sets, which are then used to construct the pointwise confidence limits and simultaneous confidence bounds as outlined in Section 9.3.2. Afterwards, add the profile of ordered C_i values from the original MNAR data set on the graph with the pointwise confidence limits and simultaneous confidence bounds. Several different settings to generate the MNAR data set were explored, and will be discussed in the remainder of this section.

First, attention is paid to the creation of MNAR based on the model parameters. This was done in the following ways: (1) a set of 50 rats were generated using the MNAR parameters from the original rats data, as presented in the upper part of Table 9.1; (2) the same parameters were used, except for ψ_2, which was increased to 0.5. (3) only 10% of the data (equivalent to 5 rats) were generated taking $\psi_2 = 0.2$,

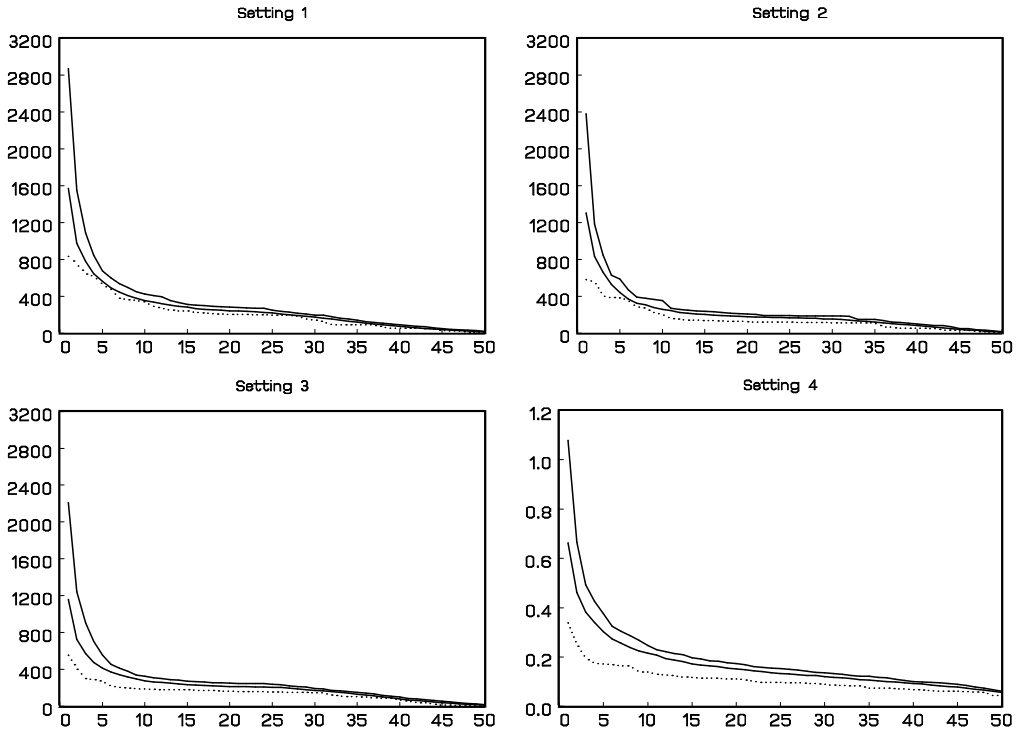


Figure 9.5: *Simulation Study. Graphical representation of the profiles of different parameter-based MNAR settings (dotted), compared with the* 95% *pointwise upper confidence limit and* 95% *simultaneous upper confidence bound (solid).*

while for the remaining 90% of the data (45 rats) $\psi_2 = 0$; (4) using the incremental parameterization introduced in Thijs, Molenberghs and Verbeke (2000), 10% of the rats were generated with $\lambda_2 = 0.2$, and the other 90% with $\lambda_2 = 0$.

All different settings of these simulations were repeated several times, but, since they all gave similar results, for each setting only one result is discussed and presented in Figure 9.5.

A general trend is observed in all settings. The C_i profile of the MNAR data set crosses neither the 95% pointwise upper confidence limit nor the simultaneous upper confidence bound for large values of C_i. On the other hand, in some settings they cross the 95% pointwise upper confidence limit for small values of C_i (near the end of the profile), but since we are only interested in highly influential subjects, this result is irrelevant for our purposes. Taking a closer look at the rats for whom $\psi_2 \neq 0$ (settings 3 and 4) we note that their C_i values are very small (all within the 10 lowest values). We can therefore conclude that this type of MNAR is not detectable using

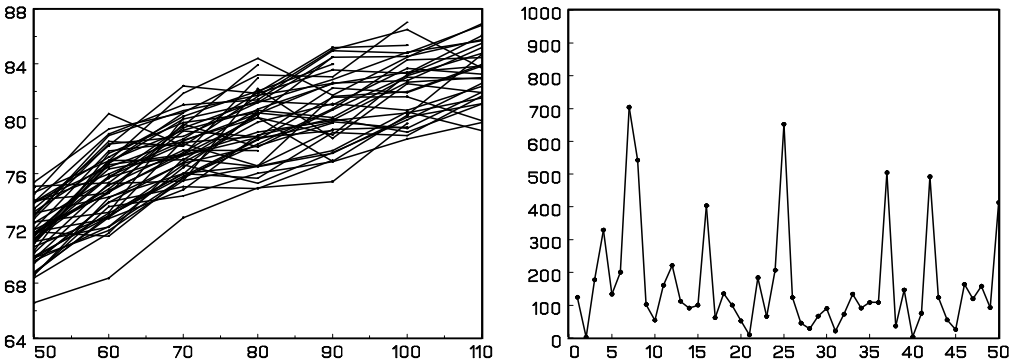


Figure 9.6: *Simulation Study. Individual growth curves (left panel) and C_i profile (right panel) of the generated MAR data set.*

local influence. Note that the limit and bound for setting (2) is more ragged than for the others. The reason is that convergence is quite difficult to obtain for this setting, in line with general convergence problems for situations where ψ_2 is substantially different from zero. The scale for setting (4) is completely different from the scale in the other settings, which is due to the fact that setting (4) considers the effect of the difference between the current and the previous measurement on the dropout process, rather than the raw effect of the current measurement in the other three settings.

Alternatively, in a second round of settings, MNAR was created in a deterministic way. Therefore a data set is generated using the MAR parameters of the original rats data in Table 9.1 (the C_i profile of this data set is shown in Figure 9.6). Afterwards, the MNAR part was created by manually deleting values from some profiles as follows: (5) all values of skull height from the moment that one of them exceeded 86 mm, resulting in shortened profiles for rats 7, 8, 25, 42 and 50; (6) all values of skull height from the moment that one of them exceeded 85 mm, giving shortened profiles for rats 4, 7, 8, 16, 24, 25, 37, 39, 42 and 50; (7) second to last values of skull height if the value at age 60 days (2nd value) exceeded 78.83 mm (95th percentile), rats 7, 8 and 25 are shortened; (8) third to last values of skull height if the value at age 70 days (3rd value) exceeded 80.82 mm (95th percentile), values deleted for rats 8, 42 and 50. Results of the local influence analyses are shown in Figure 9.7.

Our interest is now in seeing how such sets of data give qualitatively different influence graphs than under the original rats data set. Therefore, we have to proceed somewhat differently from the simulation study done for settings (1)–(4). We now rather directly compare the influence graphs from the four settings (5)–(8) with the original one. Under setting (5), we see that the peaks for rats #7 and #25, seen

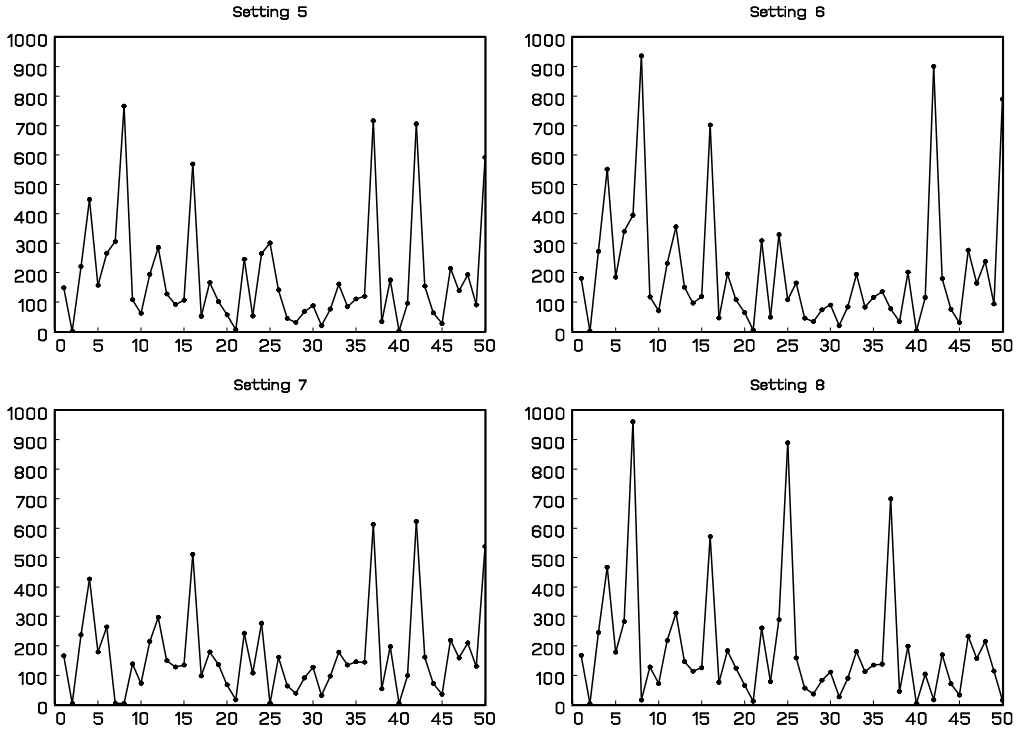


Figure 9.7: *Simulation Study. Graphical representation of the unordered C_i profiles of different settings with manually created anomalies in the missingness model.*

in the original analysis, are removed, while all other C_i values become larger (some moderate peaks now become the largest ones). This result can be ascribed to the result in Section 9.3.1, where it was shown that the total influence of a data set is roughly constant. Thus, the reduction in C_i for rats #7 and #25 of about 400 units each, will be counterbalanced by an increased C_i for all other rats. It is also noteworthy that the peaks for the other rats with shortened profiles (#8, #42 and #50) are still present. A similar phenomenon has been observed in Section 8.4 for categorical outcomes, where incomplete observations turned out to be the most influential ones. Settings (6)–(8) are similar in qualitative terms, even though the phenomena are tiny bit more extreme in setting (6) than in setting (5). Also, in setting (7) and (8), the influence of **all** rats with shortened profiles drop to almost 0.

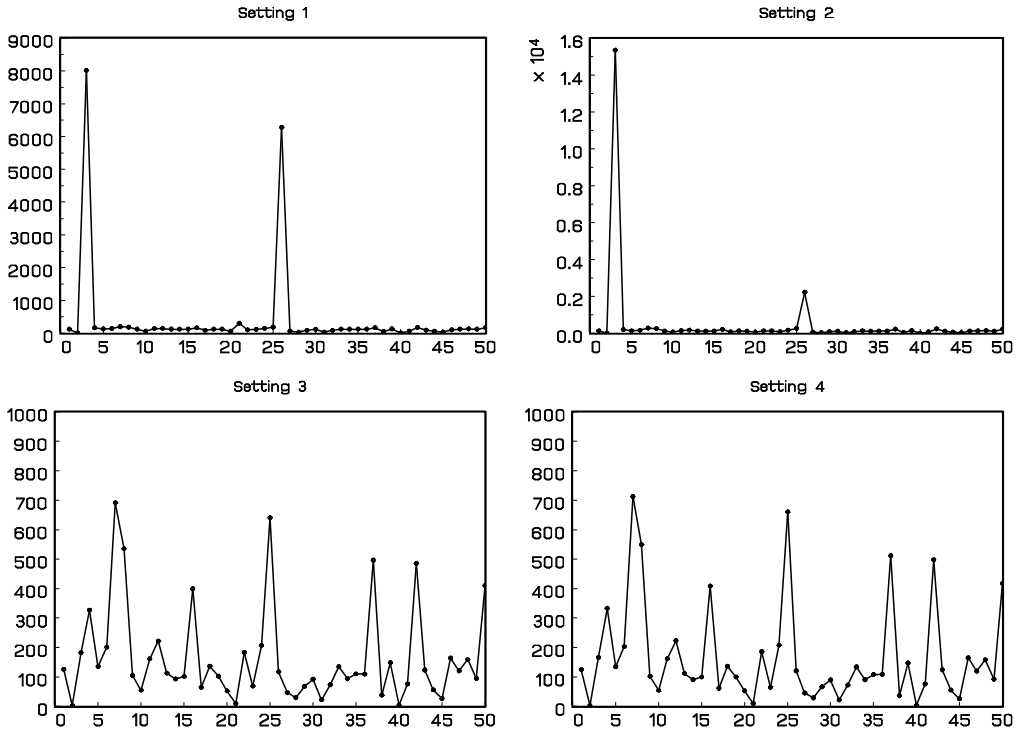


Figure 9.8: *Simulation Study. Graphical representation of the unordered C_i profiles of different settings with anomalies in the measurement model.*

9.3.4 The Effect of Anomalies in the Measurement Model

In this section, we shift attention to anomalies in the measurement model. We generated 4 MAR data sets, each of them with its specific changes to the measurement model for 3 randomly selected rats (#3, #21 and #26), namely (1) an increased mean profile by 20 units *after* the dropout probability was calculated, (2) an increased mean profile by 20 units *before* the dropout probability was calculated, (3) an increased variance component by 20 units, and (4) an increased τ^2 (covariance for the compound symmetry) by 20 units. The starting data set without any changes to the measurement model is the same as was used in Section 9.3.3, shown in Figure 9.6. Results of the local influence analyses after manipulating the original data set in several ways, are shown in Figure 9.8.

While settings (3) and (4), focusing on the variance-covariance structure, show virtually no impact, settings (1) and (2) exhibit a dramatic effect. The impact is larger in setting (2) because there also the dropout model is affected. In both settings,

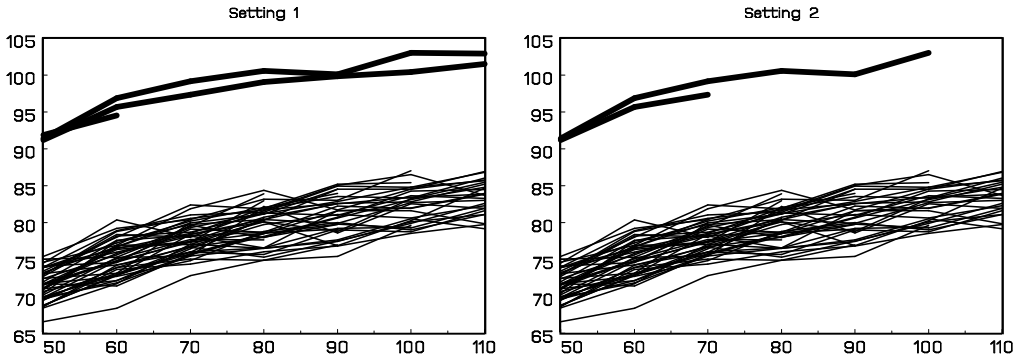


Figure 9.9: *Simulation Study. Individual growth curves for settings 1 and 2, where the mean profile is increased, before or after the dropout probability was calculated. Rats #3, #21 and #26 are highlighted.*

rats #3 and #26 clearly stick out, while with differing relative magnitudes. The effect of rat #21 is negligible. These results can be explained by taking a closer look at the individual profiles of those rats. Figure 9.9 shows that in setting (1) rat #21 has only 2 observations, while rats #3 and #26 have complete profiles. In setting (2), the profile of rat #21 reduces to only one observation, which explains the negligible influence, and the profiles of rats #3 and #26 reduce to 6 and 3 measurements, respectively. Previous conclusions again indicate that shortened profiles tend to give smaller influence values.

9.4 Conclusions

Over the last couple of decades, models for MNAR missingness have gained in popularity. However, as already noted in the discussion to Diggle and Kenward (1994), it has been made clear at various occasions that caution should be used when interpreting such models, due to the great sensitivity the results exhibit with respect to the model assumptions made. This has led to quite a bit of work on sensitivity analysis. One such tool is local influence, but this particular tool itself tends to behave in an, at first sight, non-intuitive fashion.

The results in this chapter indicate that there is little or no local influence stemming from having a few subjects that drop out in a non-random way, by setting ψ_2 for these equal to a nonzero value, while there is considerable influence in a number of settings where the measurement model is changed in the sense that a few profiles follow a deviating mean-model structure. This indicates that the non-random pa-

rameter ψ_2, rather than capturing true MNAR missingness, has a strong tendency to pick up other deviations, primarily in the measurement model. Many authors have noted that there is very little information in many sets of data for the parameter ψ_2, in addition to the information available for all other parameters. If this were to be true, this ought to show in the behavior of the likelihood ratio test statistic for ψ_2, as well as in the structure of the information matrix for the vector of model parameters. This particular behavior of the likelihood ratio test statistic for MAR missingness versus MNAR missingness, together with the analyses of this chapter, are published in Jansen *et al.* (2005b).

The bottom line of our simulation studies is that local influence tools in the incomplete data context are useful, not to detect individuals that drop out non-randomly, but rather to detect anomalous subjects that lead to a seemingly MNAR mechanism. A careful study of such subjects, combined with appropriate treatment (e.g., correction of errors, removal, ...), can lead to a final MAR model, in which more confidence can be put by the researchers, which ultimately is the goal of every sensitivity analysis.

Based on the simulation studies, it was also demonstrated that the local influence measure for a subject decreases linearly with the sample size of the data set, and hence the total influence for a data set is roughly constant. This might also explain a result that was obtained in Section 8.2, namely when including education as a covariate into the model, the group with the fewest observations led to the largest global influence measures. Thus, when the influence is measured for a group of subjects with identical characteristics, this influence has to be rescaled according to the number of subjects in that group. However, further investigation of this feature within the global influence context will be necessary.

10

General Conclusions and Future Research

Throughout this thesis, it has become evident that a variety of approaches is possible, when analyzing incomplete longitudinal data. First and foremost, an alternative was given for the frequently used but highly restrictive complete case analysis and last observation carried forward analysis. While the latter assume the data to be missing completely at random, the assumption of missing at random is sufficient for the linear mixed models, when the data are continuous, and for the generalized linear mixed models in the random-effects approach, or the weighted generalized estimating equations in the marginal model context, when the data are of the binary type. All methods can be easily performed with the currently available statistical analysis software. With the analyses of data in different scientific fields, we hope these methods will achieve more popularity in the near future. Within the fast growing field of statistical genetics, where missing data are very common, much simpler methods are used more frequently (Jansen *et al.*, 2002, 2005d). However, although these data are not necessarily of the longitudinal type, the GEE method has already been used (Van Steen *et al.*, 2005).

While in current statistical practice, still al lot of effort is needed before people come to recognize the necessity of those missing at random modeling approaches as

the main mode of analysis, research is, already for many years, directed towards the development of missing not at random models for all types of data and missingness mechanisms, since not at random missingness cannot be fully ruled out based on the observed data. In this thesis, several missing not at random models for non-continuous longitudinal data with non-monotone missingness were proposed.

First, a family of joint models for outcomes and non-response, based on an extension of Baker, Rosenberger and DerSimonian (1992), has been proposed in which (possibly continuous) covariates are allowed, in the measurement model as well as in the missingness model. Focus was on bivariate binary outcomes. Second, a set of models for multivariate ordinal data, based on the multivariate Dale model (Molenberghs and Lesaffre, 1994) was developed. A (bivariate or multivariate) Dale model for the outcomes was combined with the same Dale model for the non-response. In this hierarchy of models, the inclusion of constant or varying covariate effects was also possible in both the measurement and missingness model. In the data analyses, using either the extended BRD models or the Dale-Dale models, the estimates of the measurement model parameters were remarkably stable, no matter which assumptions were made regarding the reasons for non-response.

Based on the proposed models, and the already existing models by Diggle and Kenward (1994) for monotone continuous outcomes, and by Molenberghs, Kenward and Lesaffre (1997) for monotone discrete outcomes, it is possible to develop many more flexible models. For example, when combining a linear mixed model for the measurements with a Dale model for the non-response process, non-monotone continuous outcomes can be modeled. Or combining a Dale model for the measurements with the multinomial logit model from our extended BRD model family for the non-respons, is another possibility to model non-monotone discrete data.

All previous models were based on the selection model factorization of the joint model for the outcomes and non-response process. Also within the pattern-mixture models, a distinction can be made between continuous or discrete data, and monotone or non-monotone missingness. In this thesis, we developed a pattern-mixture model framework for non-monotone discrete outcomes, using the multivariate Dale model to analyze the data per pattern. Identifying restrictions were used to specify the conditional distribution of the unobserved measurements, given the observed ones in a specific pattern. Similar models can be developed, using for example our extended BRD models, or the generalized linear mixed models, instead of the multivariate Dale model. Insight was also given into the non-trivial way of determining a marginal effect across patterns in the case of non-Gaussian data.

Contrasting several models within one of the families proposed here (extended

BRD, Dale-Dale or pattern-mixture), is in itself a way to perform a sensitivity analysis. Also comparing the results from selection models and pattern-mixture models can be fruitful to assess the sensitivity of such models. On the other hand, a sensitivity analysis can also be conducted at the level of the individuals. Therefore, in this thesis, the already existing methods of global and local influence have been adjusted to assess the influence in the case of selection models for bivariate binary outcomes subject to non-monotone missingness. Clearly, it is advisable to combine several of those methods within an integrated sensitivity analysis.

Until now, local influence was mainly used to detect observations that had a high impact on the conclusions due to their aberrant missingness mechanism. However, in many applications, the situation turned out to be more complex than anticipated. Indeed, the influential subjects often are influential for other than missingness related features. For example, in Molenberghs *et al.* (2001b), the three influential cows were identified by an extreme increase between the measurements at two subsequent years. Thijs, Molenberghs and Verbeke (2000) observed a similar behavior. The questions that arose from those findings were (1) when can a subject or a group of subjects be considered as influential, and (2) what reasons might be related to high influence values. In the last chapter, an attempt was made to answer those questions. A similar feature may be present within the global influence context, so further investigation will be necessary.

Bibliography

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

Agresti, A. (2002) *Categorical Data Analysis.* Hoboken, NJ: Wiley.

Albert, A. and Lesaffre, E. (1986) Multiple group logistic discrimination. *Computers and Mathematics, with Applications*, **2**, 209–224.

Amato, P. R. (1996) Explaining the intergenerational transmission of divorce. *Journal of Marriage and the Family*, **58**, 628–640.

Ashford, J. R. and Sowden, R. R. (1970) Multivariate probit analysis. *Biometrics*, **26**, 535–546.

Bahadur, R. R. (1961) A representation of the joint distribution of responses to n dichotomous items. In: *Studies in Item Analysis and Prediction* (Ed. H. Solomon), Stanford Mathematical Studies in the Social Sciences VI. Stanford, CA: Stanford University Press.

Baker, S. G. (1995) Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics*, **51**, 1042–1052.

Baker, S. G. (2000) Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-none compliance. *Journal of the American Statistical Association*, **95**, 43–50.

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.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems. *Statistical Science*, **10**, 3–66.

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

Burton, S. W. (1991) Review of fluvoxamine and its uses in depression. *International Clinical Psychopharmacology*, **6**, 1–17.

Burzykowski, T., Molenberghs, G., Tafforeau, J., Van Oyen, H. and Demarest, S. (1999) Missing data in the health interview survey 1997 in belgium. *Archives of Public Health*, **57**, 107–129.

Canary, D. J., Stafford, L. and Semic, B. A. (2002) A panel study of the associations between maintenance strategies and relational characteristics. *Journal of Marriage and the Family*, **64**, 395–406.

le Cessie, S. and van Houwelingen, J. C. (1994) Logistic regression for correlated binary data. *Applied Statistics*, **43**, 95–108.

Chatterjee, S. and Hadi, A. S. (1988) *Sensitivity Analysis in Linear Regression.* New York: John Wiley & Sons.

Cook, R. D. (1979) Influential observations in linear regression. *Journal of the American Statistical Association*, **74**, 169–174.

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 and Hall.

Copas, J. B. and Li, H. G. (1997) Inference from non-random samples (with discussion). *Journal of the Royal Statistical Society, Series B*, **59**, 55–96.

Cox, D. R. (1972) The analysis of multivariate binary data. *Applied Statistics*, **21**, 113–120.

Dale, J. R. (1986) Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.

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

Diggle, P. J., Heagerty, P. J., Liang, K.-Y. and Zeger, S. L. (2002) *Analysis of Longitudinal Data (2nd ed.).* Oxford Science Publications. Oxford: Clarendon Press.

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

Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B*, **57**, 45–97.

Ekholm, A. (1991) Algorithms versus models for analyzing data that contain misclassification errors. *Biometrics*, **47**, 1171–1182.

Ekholm, A. and Skinner, C. (1998) The muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics*, **47**, 251–263.

Fahrmeir, L. and Tutz, G. (2001) *Multivariate Statistical Modelling Based on Generalized Linear Models.* Heidelberg: Springer-Verlag.

Freeman, G. H. and Halton, J. H. (1951) Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika*, **38**, 141–149.

Gerris, J. R. M., Houtmans, M. J. M., Kwaaitaal-Roosen, E. M. G., Schipper, J. C., Vermulst, A. A. and Janssens, J. M. A. M. (1998) *Parents, adolescents and young adults in Dutch families. A longitudinal study.* Nijmegen: University of Nijmegen, Institute of Family Studies.

Gerris, J. R. M., Van Boxtel, D. A. A. M., Vermulst, A. A., Janssens, J. M. A. M., Van Zutphen, R. A. H. and Felling, A. J. A. (1992) *Child-rearing, family relations and family processes in 1990.* Nijmegen: University of Nijmegen, Institute of Family Studies.

Gerris, J. R. M., Vermulst, A. A., Van Boxtel, D. A. A. M., Janssens, J. M. A. M., Van Zutphen, R. A. H. and Felling, A. J. A. (1993) *Parenting in Dutch families.* Nijmegen: University of Nijmegen, Institute of Family Studies.

Geys, H., Molenberghs, G. and Lipsitz, S. R. (1998) A note on the comparison of pseudo-likelihood and generalized estimating equations for marginal odds ratio models. *Journal of the Statistical Computation and Simulation*, **62**, 45–72.

Gilula, Z. and Haberman, S. J. (1994) Conditional log-linear models for analyzing categorical panel data. *Journal of the American Statistical Association*, **89**, 645–656.

Glonek, G. F. V. and McCullagh, P. (1995) Multivariate logistic models. *Journal of the Royal Statistical Society, Series B*, **81**, 477–482.

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

Hartley, H. O. and Hocking, R. (1971) The analysis of incomplete data. *Biometrics*, **27**, 783–808.

Hogan, J. W. and Laird, N. M. (1997) Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, **16**, 239–257.

Holman, T. B. (2001) *Premarital prediction of marital quality or breakup. Research, theory and practice.* New York: Kluwer Academic/Plenum Publishers.

Jansen, I., Beunckens, C., Molenberghs, G., Verbeke, G. and Mallinckrodt, C. (2005a) Analyzing incomplete binary longitudinal clinical trial data. *Submitted to Statistical Science.*

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G. and Kenward, M. G. (2005b) The nature of sensitivity in monotone missing not at random models. *Computational Statistics and Data Analysis*, **0**, 00–00.

Jansen, I. and Molenberghs, G. (2005) A flexible marginal modeling strategy for non-monotone missing data. *Submitted.*

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., Van den Troost, A., Molenberghs, G., Vermulst, A. and Gerris, J. (2005c) Modeling partially incomplete marital satisfaction data. *Submitted to Sociological Methods & Research*.

Jansen, I., Van Steen, K., Molenberghs, G., De Wit, M. and Peeters, M. (2002) Using word frequencies for testing the equivalence between two dna sequences. *Genetic epidemiology*, **23**, 287.

Jansen, I., Van Steen, K., Molenberghs, G., De Wit, M. and Peeters, M. (2005d) A similarity measure and test between two dna sequences based on a mahalanobis distance between word frequencies. *Submitted*.

Johnson, D. R., Amoloza, T. O. and Booth, A. (1992) Stability and developmental change in marital quality: A three-wave panel analysis. *Journal of Marriage and the Family*, **54**, 582–594.

Kenward, M. G. (1998) Selection models for repeated measurements with nonrandom dropout: an illustration of sensitivity. *Statistics in Medicine*, **17**, 2723–2732.

Kenward, M. G., Goetghebeur, E. J. T. and Molenberghs, G. (2001) Sensitivity analysis of incomplete categorical data. *Statistical Modelling*, **1**, 31–48.

Kenward, M. G., Lesaffre, E. and Molenberghs, G. (1994) An application of maximum likelihood and estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics*, **50**, 945–953.

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 Thijs, H. (2003) Pattern-mixture models with proper time dependence. *Biometrika*, **90**, 53–71.

Laird, N. M. (1994) Discussion to Diggle, P. J. and Kenward, M. G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 84.

Lang, J. B. and Agresti, A. (1994) Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625–632.

Lavee, Y., Sharlin, S. and Katz, R. (1996) The effect of parenting stress on marital quality: An integrated mother-father model. *Journal of Family Issues*, **17**, 114–135.

Lesaffre, E., Molenberghs, G. and Dewulf, L. (1996) Effect of dropouts in a longitudinal study: an application of a repeated ordinal model. *Statistics in Medicine*, **15**, 1123–1141.

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

Lewis, R. A. and Spanier, G. B. (1979) Theorizing about the quality and stability of marriage. In: *Contemporary theories about the family: research-base theories.* (Eds. W. R. Burr, R. Hill, F. I. Nye and I. L. Reiss), pp. 49–65. New York: Free Press.

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

Liang, K.-Y. and Zeger, S. L. (1989) A class of logistic regression models for multivariate binary time series. *Journal of the American Statistical Association*, **84**, 447–451.

Little, R. J. A. (1992) Regression with missing X's: a review. *Journal of the American Statistical Association*, **87**, 1227–1237.

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. (1995) Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.

Little, R. J. A. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data.* New York: Wiley.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data.* New York: Wiley.

Longford, N. (1993) Inference about variation in clustered binary data. Paper presented at the Multilevel Conference, Oct 1 and 2, 1993, Los Angeles, Rand Corporation.

Mallinckrodt, C. H., Clark, W. S., Carroll, R. J. and Molenberghs, G. (2003a) Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. *Journal of Biopharmaceutical Statistics*, **13**, 179–190.

Mallinckrodt, C. H., Sanger, T. M., Dube, S., Debrota, D. J., Molenberghs, G., Carroll, R. J., Zeigler Potter, W. M. and Tollefson, G. D. (2003b) Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry*, **53**, 754–760.

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

Michiels, B. and Molenberghs, G. (1997) Protective estimation of longitudinal categorical data with nonrandom dropout. *Communications in Statistics: Theory and Methods*, **26**, 65–94.

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**, 1023–1042.

Michiels, B., Molenberghs, G. and Lipsitz, S. R. (1999a) A pattern-mixture odds ratio model for incomplete categorical data. *Communications in Statistics: Theory and Methods*, **28**, 2843–2869.

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

Molenberghs, G., Goetghebeur, E., Lipsitz, S. R. and Kenward, M. G. (1999a) Nonrandom missingness in categorical data: strengths and limitations. *The American Statistician*, **53**, 110–118.

Molenberghs, G., Goetghebeur, E., Lipsitz, S. R., Kenward, M. G., Lesaffre, E. and Michiels, B. (1999b) Missing data perspectives of the fluvoxamine data set: a review. *Statistics in Medicine*, **18**, 2449–2464.

Molenberghs, G., Kenward, M. G. and Goetghebeur, E. (2001a) Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics*, **50**, 15–29.

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

Molenberghs, G. and Lesaffre, E. (1994) Marginal modelling of correlated ordinal data using a multivariate plackett distribution. *Journal of the American Statistical Association*, **89**, 633–644.

Molenberghs, G. and Lesaffre, E. (1999) Marginal modelling of multivariate categorical data. *Statistics in Medicine*, **18**, 2237–2255.

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

Molenberghs, G. and Ryan, L. M. (1999) Likelihood inference for clustered multivariate binary data. *Environmetrics*, **10**, 279–300.

Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C. and Carroll, R. J. (2004) Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, **5**, 445–464.

Molenberghs, G., Thijs, H., Kenward, M. G. and Verbeke, G. (2003) Sensitivity analysis for continuous incomplete longitudinal outcomes. *Statistica Neerlandica*, **57**, 112–135.

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

Neuhaus, J. M. (1992) Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*, **1**, 249–273.

Neuhaus, J. M., Kalbfleisch, J. D. and Hauck, W. W. (1991) A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, **59**, 25–35.

Park, T. and Lee, S.-Y. (1999) Simple pattern-mixture models for longitudinal data with missing observations: Analysis of urinary incontinence data. *Statistics in Medicine*, **18**, 2933–2941.

Plackett, R. L. (1965) A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516–522.

Renard, D., Molenberghs, G., Van Oyen, H. and Tafforeau, J. (1998) Investigation of the clustering effect in the belgian health interview survey. *Archives of Public Health*, **56**, 345–361.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994) Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846–866.

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.

Rogers, S. J. and White, L. K. (1998) Satisfaction with parenting: The role of marital happiness, family structure, and parents' gender. *Journal of Marriage and the Family*, **60**, 293–308.

Rosner, B. (1984) Multivariate methods in ophtalmology with applications to other paired-data situations. *Biometrics*, **40**, 1025–1035.

Rubin, D. B. (1976) Inference and missing data. *Biometrika*, **63**, 581–592. With comments by R. J. A. Little and a reply by the author.

Rubin, D. B. (1977) Formalizing subjective notions about the effect of nonresponse in sample surveys. *Journal of the American Statistical Association*, **72**, 538–543.

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

Rubin, D. B. (1994) Discussion to Diggle, P. J. and Kenward, M. G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 80–82.

Rubin, D. B. (1996) Multiple imputation after 18+ years. *Journal of the American Statistical Association*, **91**, 473–489.

Rubin, D. B., Stern, H. S. and Vehovar, V. (1995) Handling "dont know" survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association*, **90**, 822–828.

Rubin, L. (1983) *Intimate Strangers: Men and Women Together*. New York: Harper & Row.

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

Scharfstein, D. O., Rotnizky, A. and Robins, J. M. (1999) Adjusting for non-ignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*, **94**, 1096–1146.

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.

Shih, W. J. and Quan, H. (1997) Testing for treatment differences with dropouts present in clinical trials – a compositie approach. *Statistics in Medicine*, **16**, 1225–1239.

Siddiqui, O. and Ali, M. W. (1998) A comparison of the random-effects pattern mixture model with last observation carried forward (locf) analysis in longitudinal clinical trials with dropouts. *Journal of Biopharmaceutical Statistics*, **8**, 545–563.

Stiratelli, R., Laird, N. and Ware, J. (1984) Random effects models for serial observations with dichotomous response. *Biometrics*, **40**, 961–972.

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.

Van den Troost, A., Vermulst, A. A., Gerris, J. R. M. and Matthijs, K. (2001) Meetinvariantie van huwelijkskwaliteit en satisfactie. *Leuven/Nijmegen. Onderzoeksverslag van het Departement Sociologie - afdeling Gezin, Bevolking & Gezondheidszorg en Orthopedagogiek. - Gezin en Gedrag*, **GB/2001-13**.

Troxel, A. B., Harrington, D. P. and Lipsitz, S. R. (1998) Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics*, **47**, 425–438.

Vaillant, C. O. and Vaillant, G. E. (1993) Is the U-curve of marital satisfaction an illusion? a 40-year study of marriage. *Journal of Marriage and the Family*, **55**, 230–239.

Van Laningham, J., Johnson, D. R. and Amato, P. (2001) Marital happiness, marital duration, and the U-shaped curve: Evidence from a five-wave panel study. *Social Forces*, **79**, 1313–1341.

Van Steen, K., Molenberghs, G., De Wit, M. and Peeters, M. (2005) Comparing dna sequences using generalized estimating equations and pseudo-likelihood. *Submitted*.

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. and Lesaffre, E. (1999) The effect of drop-out on the efficiency of longitudinal experiments. *Applied Statistics*, **48**, 363–375.

Verbeke, G., Lesaffre, E. and Spiessens, B. (2001a) The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, **35**, 419–439.

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. and Molenberghs, G. (2003) The use of score tests for inference on variance components. *Biometrics*, **59**, 254–262.

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

Verdonck, A., De Ridder, L., Verbeke, G., Bourguignon, J. P., Carels, C., Kuhn, E. R., Darras, V. and de Zegher, F. (1998) Comparative effects of neonatal and prepubertal castration on craniofacial growth in rats. *Archives of Oral Biology*, **43**, 861–871.

Weigel, D. J. and Ballard-Reisch, D. S. (2001) The impact of relational maintenance behaviors on marital satisfaction: A longitudinal analysis. *Journal of Family Communication*, **1**, 265–279.

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

Wood, J. T. (1993) Engendered relations: Interaction, caring, power and responsibility in intimacy. In: *Social context and social relationships*. (Ed. S. Duck), pp. 26–54. Newbury Park/London: Sage Publications.

Wothke, W. (2000) Longitudinal and multi-group modeling with missing data. In: *Modeling longitudinal and multilevel data: practical issues, applied approaches and specific examples* (Eds. T. D. Little, K. U. Schnabel and J. Baumert), pp. 219–240. Mahwah, N.J.: Lawrence Erlbaum.

Wu, M. C. and Bailey, K. R. (1989) Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, **45**, 939–955.

Zhao, L. P., Lipsitz, S. and Lew, D. (1996) Regression analysis with missing covariate data using estimating equations. *Biometrics*, **52**, 1165–1182.

Samenvatting