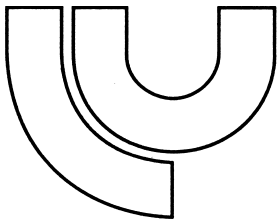

Limburgs Universitair Centrum

Faculteit Wetenschappen

Sensitivity Analysis for Incomplete Data

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

HERBERT THIJS

Promotor:
Prof. dr. G. Molenberghs

2002

Voor Mijn Ouders

Dankwoord

Heel graag zou ik hier even willen stilstaan en mijn dank betuigen aan iedereen die op de een of andere manier heeft bijgedragen tot de verwezenlijking van deze thesis.

Als eerste kom ik dan als vanzelfsprekend bij mijn promotor Prof. Dr. Geert Molenberghs die om te beginnen met zijn overtuigingskracht mij na een kort gesprek wist te overhalen om ook de stap te wagen vanuit een algebraïsche vertrouwde werkomgeving naar een toch wel (zo blijkt nu) zeker zo boeiende wereld van (bio-)statistiek. In de vele daaropvolgende gesprekken zorgde een combinatie van wetenschappelijk advies en een gezonde dosis humor voor een blijvende motivatie. Daarom Geert, Hartelijk bedankt voor je blijvende steun en vertrouwen en jouw schitterende begeleiding en late het geen geheim weze dat ik verdere samenwerking zeker niet uit de weg zal gaan.

Natuurlijk zijn er nog de andere collega's op het LUC die ik niet wil vergeten. Het zijn trouwens zij die er elke dag weer voor zorgden dat het wetenschappelijk onderzoek op tijd en stond werd ingeruild voor een lach of een traan. Gedurende die vier jaar kwamen er regelmatig mensen bij maar gingen er ook mensen weg, ieder met zijn eigen karakter maar beste collega's mag ik hier en nu effe mijn excuses aanbieden voor die momenten dat ik niet helemaal was die ik verwacht was te zijn?

En als laatste wil ik zeker ook mijn familie en vrienden bedanken en wel heel speciaal mijn ouders. Waar zij in deze laatste jaren niet altijd meer konden volgen wat ik uitspookte waren het wel zij die reeds 26 jaar achter mij staan en mij (met de nodige moeite) blijvend wisten te motiveren. Mama en Papa, bedankt! Ook mijn broer, mijn vrienden en nog vele andere mensen ben ik enorm dankbaar voor alles (hoe miniem soms) wat jullie betekend hebben zonder dat jullie het zelf waarschijnlijk beseffen.

Herbert Thijs
Diepenbeek, oktober 2001

Contents

List of Tables

List of Figures

Chapter 1

Introduction

During the last few decades, many experiments such as clinical trials but also studies outside the medical world, are designed to collect information about the subjects included in the study, repeatedly over a specific period of time. In dealing with this type of information existing simple univariate techniques are not sufficient to model the data. Hence the use of so called *longitudinal techniques* has grown and nowadays longitudinal data analysis is a widely spread topic of research to which many approaches have been found useful. Currently, a considerable amount of literature can be consulted, describing methods to handle longitudinal data (Lindsey 1993, Longford 1993, Diggle, Liang and Zeger 1994, Hand and Crowder 1995, Verbeke and Molenberghs 1997 and 2000, Vonesh and Chinchilli 1997). Among these methods, the linear mixed model for normally distributed endpoints (Laird and Ware 1982) can be considered as being mostly used due to a large extent, to the fact that for this model one can use simple and flexible software tools such as: the SAS procedure MIXED (Littell *et al.* 1996), the SPlus function LME, etc. This model is nowadays extended in such a way that one can describe serial association (Diggle 1988 and Molenberghs and Verbeke 2000) although one has to take care because its application is not always straightforward. On the other hand, considering categorical outcomes such as binary data or counts, less tools are available but also in this framework several proposals have been made including generalized estimating equations (Liang and Zeger 1986) and generalized linear mixed models (Wolfinger and O'Connell 1993, Breslow and Clayton 1993).

In dealing with longitudinal data as introduced above it is not unusual for some measurements to be unobserved. The problem of missing data is rather common and omnipresent in clinical trials (Piantadosi 1997, Green, Benedetti and Crowley 1997 and Friedman, Furberg and DeMets 1998), epidemiological studies (Kahn and Sempos 1989, Clayton and Hills 1993, Lilienfeld and Stoley 1994 and Selvin 1996) and maybe even most severely in sample surveys (Fowler 1988, Schefer, Khare and Ezatti-Rice 1993, Rubin 1987 and Rubin, Stern and Vehovar 1995). On the one hand there is the possibility of missing responses where one single measurement failed to be observed or even a sequence of measurements can terminate early, both for reasons outside the control of the investigator (recovery, lack of improvement, unwanted signs or symptoms that may be related to the investigational treatment,...). In the latter case of early termination of measurement sequences we denote the subjects so affected as *dropouts* where single missing observations are denoted as *intermittent missingness.* On the other hand missingness might appear in the covariates due to reasons of different type. The main interest of research until now has been the problem of dropout where only recently more effort is put in treating missing covariates.

Many authors, in a first attempt to handle missing data, concentrated on computational issues imposed due to unbalancedness of the data (Afifi and Elashoff 1966 and Hartley and Hocking 1971) while more recently Dempster, Laird and Rubin (1977) developed the Expectation-Maximization algorithm whereas Rubin (1987) and Tanner and Wong (1987) discussed imputation and augmentation methods. Both approaches will be briefly introduced in Chapter 3 and it is argued that together with an improvement of software these methods solve largely the computational problems related to missing data. However we still must consider the problems concerning statistical inference and its validity. Again several simple but ad hoc methods can be used to deal with missing data and we can mention complete case analysis or simple imputation as examples but the major problems with incomplete longitudinal data related to these methods, are efficiency loss and the introduction of bias but also the implementation of existing methods is not always straightforward (Laird 1988). Some approaches next to complete case analysis, are univariate analyzes with adjustments for variance estimates, two-step analyzes, and likelihood based methods where for the latter Laird (1988) further distinguishes between likelihood based methods that include an explicit model for dropout and methods that only model the measurement process. Also several other authors recently have indicated that it might be necessary to accommodate dropout in the modeling process. In other words one must model the measurement process jointly with a model for dropout which itself is sometimes

of scientific interest.

Rubin (1976) and Little and Rubin (1987, Ch. 6) make important distinctions between different missing values processes. A dropout process is said to be *completely random* (MCAR) if the dropout is independent of both unobserved and observed data and *random* (MAR) if, conditional on the observed data, the dropout is independent of the unobserved measurements; otherwise the dropout process is termed *non-random* (MNAR). If a dropout process is random then a valid analysis, can be obtained through a likelihood-based analysis that ignores the dropout mechanism, provided the parameter describing the measurement process is functionally independent of the parameter describing the dropout process, the so-called parameter distinctness condition. This situation is termed *ignorable* by Rubin (1976) and Little and Rubin (1987). This leads to considerable simplification in the analysis. Furthermore, there are situations where the MAR assumption gives better results than MNAR (Rubin, Stern and Vehovar 1995). In many examples, however, the reasons for dropout are many and varied and it is therefore difficult to justify on a priori grounds the assumption of random dropout. Arguably, in the presence of non-random dropout, a wholly satisfactory analysis of the data is not feasible and where the treatment of missing data that are missing at random requires some caution, one needs to be even more careful with non-randomly missing data.

One approach is to estimate from the available data the parameters of a model representing a non-random dropout mechanism. It may be difficult to justify the particular choice of dropout model, and it does not necessarily follow that the data contain information on the parameters of the particular model chosen, but where such information exists the fitted model may provide some insight into the nature of the dropout process and of the sensitivity of the analysis to assumptions about this process. This is the route taken by Diggle and Kenward (1994) in the context of continuous longitudinal data. More precisely, these authors perform an MNAR analyzes by defining a model for dropout dependent on the unobserved data and combining this model with a linear mixed model for the measurements (see also Diggle, Liang and Zeger 1994, Ch. 11). Further approaches are proposed by Schluchter (1988), Laird, Lange and Stram (1987), Wu and Bailey (1988, 1989), Wu and Carroll (1988). These last authors use random-effects models to describe the censoring or non-response process. Greenlees, Reece and Zieschang (1982), combine the probit of the dropout probability with the general linear measurement model. Little (1995) gives a careful review of the different modeling approaches.

Reconsidering the case of categorical outcomes one notices that also this field has received considerable attention. A general framework is provided by Fay (1986) and an overview of methods for missing data in longitudinal data is given in Laird (1988) who distinguishes between ignorable and non-ignorable missingness, in the context of both normally distributed and categorical data. Baker and Laird (1988) extend the work of Fay (1986) and give a thorough account of the modeling of contingency tables in which there is one response dimension and an additional dimension indicating whether the response is absent. They pay particular attention to the circumstances in which no solution exists for the non-random dropout models. Molenberghs, Kenward and Lesaffre (1997) introduce a method for the analysis of longitudinal ordinal data with non-random dropout. Their approach is based on Diggle and Kenward (1994), who treat non-random dropout in continuous longitudinal data. Molenberghs and Goetghebeur (1997) have introduced a simple method to construct and maximize the observed data likelihood, whilst still formulating their models at the complete data level. Other proposals have been made as well: Lehnen and Koch (1974) present a saturated likelihood approach. They have to assume that the missingness is completely random. Dempster, Laird and Rubin (1977) use the EM algorithm to maximize the likelihood in case of incomplete categorical data. Fuchs (1982) uses the EM algorithm to fit Log-linear models for ignorable incomplete data. Conaway *et al.* (1992) use loglinear models and perform fitting within GLIM, with the aid of the EM algorithm.

With the volume of literature on non-random missing data increasing, there has been growing concern about the fact that models often rest on strong assumptions and relatively little evidence from the data themselves. Here fore there is a clear need for methods that investigate the sensitivity of the results with respect to the model assumptions. Some progress can be made by examining the effect of dropout mechanism parameters for which there is no information in the data on the analyzes of the measurement parameters and on the remaining dropout parameters; Nordheim (1984) gives an example of such a sensitivity analysis with cross-sectional binomial outcomes. A similar principle is advocated by Little (1994). Fitzmaurice, Molenberghs and Lipsitz (1995) assess the effect of the non-response mechanism on the estimation of marginal regression parameters for repeated binary outcomes. Molenberghs, Goetghebeur, Lipsitz and Kenward (1999) illustrate the need for sensitivity analysis by reviewing some of the issues that arise with models for non-random missing data.

1. Models with the same or similar fit at the level of the observed data, can dif-

fer considerably in terms of prediction and interpretation of the (hypothetical) complete data.

2. Care has to be taken with boundary solutions or even solutions that violate parameter space restrictions.

3. Some models are overspecified in the sense that they lead to a whole family of solutions, rather than to a point estimate.

Despite this considerable amount of literature referring to the need for sensitivity analysis, only few actual proposals have been made. Moreover, many of these are to be considered as useful but ad hoc approaches. On the other hand, most methods are formulated within the selection modeling frame (Little and Rubin 1987) as opposed to pattern-mixture modeling (Little 1993). A selection model factors the joint distribution of the measurement and missingness processes into the marginal distribution of the measurement process and the conditional distribution of the missingness process given the measurements and this is intuitively appealing since the marginal measurement distribution would be of interest also with complete data. Also the terminology introduced by Little and Rubin's concerning MCAR, MAR and MNAR is most easily developed in the selection setting. However, it is often argued that, especially in the context of non-random missingness models, selection models, although identifiable, should be approached with caution. This point was already raised by Rubin 1994, Laird 1994 as discussants to Diggle and Kenward (1994), and Glynn, Laird and Rubin (1986) who indicate that this is typical for so-called selection models, while it is much less so for a pattern-mixture model (Little 1993, 1994, Hogan and Laird 1997), where the reverse factorization is used. In our view, a more formal approach to sensitivity analyzes should be fruitful as well.

Concerning the selection model of Diggle and Kenward (1994) it is fair to say that it raised at first too high expectations. This again was made clear by the many discussants of their paper indicating for example that formal tests for the null hypothesis of random missingness should be approached with caution, even though they are technically possible by contrasting a non-random model with its random submodel (e.g., by using the likelihood ratio test). On the other hand this model has received considerable attention and since it is available within the SPlus suite of functions termed Oswald (Smith, Robertson, and Diggle 1996), we have chosen to use this model to introduce a local influence approach, suggested by Cook (1986) to investigate the effect of extending the MAR model for dropout in the direction of non-random

dropout. A general theoretical treatment in full detail is discussed in Chapter 4 where we also compare these methodology with a more global approach within the selection modeling framework.

Precisely because of the issue of identifiability (Glynn, Laird and Rubin 1986), pattern-mixture models have gained renewed interest in recent years (Little 1993, 1994, Hogan and Laird 1997). Several authors have contrasted selection models and pattern-mixture models. This is done to either (1) answer the same scientific question, such as marginal treatment effect or time evolution, based on these two rather different modeling strategies, or (2) to gain additional insight by supplementing the selection model results with those from a pattern-mixture approach. Examples can be found in Verbeke, Lesaffre, and Spiessens (2001), Curran, Pignatti, and Molenberghs (2002), and Michiels *et al* (1999) for continuous outcomes. The categorical outcome case has been treated in Molenberghs, Michiels, and Lipsitz (1999), and Michiels, Molenberghs, and Lipsitz (1999). Further references include Ekholm and Skinner (1998), Molenberghs, Michiels, and Kenward (1998), Little and Wang (1996), Hedeker and Gibbons (1997), Cohen and Cohen (1983), Muthén, Kaplan, and Hollis (1987), Allison (1987), and McArdle and Hamagani (1992). An important issue is that pattern-mixture models are by construction under-identified. 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. Identifying restrictions are not the only way to overcome under-identification and we will discuss alternative approaches as well. All in all, while some authors perceive this under-identification as a drawback, we believe it is an asset since it forces one to reflect on the assumptions made. In Chapter 5 we will introduce several strategies to deal with pattern-mixture models and we will indicate how this under-identification can serve as a starting point for sensitivity analysis. Even considering the terminology concerning random versus non-random missingness it will be shown that this is still valid for pattern-mixture models.

An important issue now arising is whether a model for a given dropout pattern ought to be extended and, if the answer is affirmative, how this should be approached. In the framework of pattern-mixture models this is done rather explicitly by using ordinary extrapolation, sufficiently simplified models or one can consider identifying restrictions. But also using selection models one is implicitly drawing conclusions beyond the actual time of dropout. Unfortunately, the implications for the nature of the dropout mechanism are not always understood nor studied. Here fore we consider

it to be of interest to further characterize which mechanisms are more general than MAR or ACMV but still prevent missingness to depend on future, possibly unobserved outcomes. Therefore in Section 5.5 we provide an extra set of restrictions as a general characterization, term them non-future dependent (NFD) and illustrate, in a realistic setting, how these restrictions might be exploited in practice.

Since different missing data mechanisms are established in a selection model and a pattern-mixture model context one can choose one of the models based on the statistical and scientific merits of proposed missing value models on their own terms and one must consider the precise aim of sensitivity analysis to be reached. Furthermore, we advocate the use of pattern-mixture models as a tool to assess sensitivity of a selection model to the modeling assumptions, or vice versa (Molenberghs, Michiels and Lipsitz 1999, Michiels, Molenberghs and Lipsitz 1998). Explicitly, extra confidence in the conclusions can be gained if two analyzes, one within each framework, coincide in key aspects, such as covariate dependencies, strength of association between outcomes, etc. In Chapter 5, more precisely Section 5.4.1 we will compare Selection models and Pattern-mixture models applied to the Vorozole study introduced in the next chapter while in Chapter 6, a selection model together with methods of local and global influence will be applied to the milk protein trial, also introduced in the next chapter, in order to compare both analyzes.

A final topic of ongoing discussion is the fact that selection models as well as pattern-mixture models beit implicitly or explicitely respectively, draw conclusions beyond the actual time of dropout when dropout possibly is caused by death of the subject. We already touched this issue before related to the non-future dependent (NFD) restrictions but in Chapter 7 we will propose several methods how to deal with this problem: Time reversal, Accelerated Failure Time models and Random-effects models (Wu and Carroll 1988 and Wu and Bailey 1988, 1989). The Latter type of models are also referred to as shared-parameter models and Little (1995) combines them with the concepts of PMM. As an illustration we will again apply these new methods to a set of longitudinal data.

To conclude we will formulate in Chapter 8 some general findings and we will also indicate possibilities for further developments concerning the new methodology introduced throughout this thesis. We hope this will convince more statisticians of the problems related with missing data and the need for a more formal approach of **sensitivity analysis**.

Chapter 2

Key Examples

In this chapter, we introduce several typical longitudinal data sets which will be used throughout the text as key studies.

2.1 Mastitis in Dairy Cattle

A first example is a study concerning the occurrence of the infectious disease mastitis in dairy cows. This dataset was introduced in Diggle and Kenward (1994) and re-analyzed in Kenward (1998). Data were available of the milk yields in thousands of liters of 107 dairy cows from a single herd in two consecutive years: Y_{ij} $(i = 1, \ldots, 107; j = 1, 2)$. In the first year, all animals were supposedly free of mastitis, in the second year 27 became infected. Mastitis typically reduces milk yield, and the question of scientific interest is whether the probability of occurrence of mastitis is related to the yield that would have been observed had mastitis not occurred. A graphical representation of the complete data is given in Figure 2.1.

2.2 Rats Data

Secondly we consider data from a randomized experiment, designed to study the effect of the inhibition of the testosterone production in rats. The experiment was

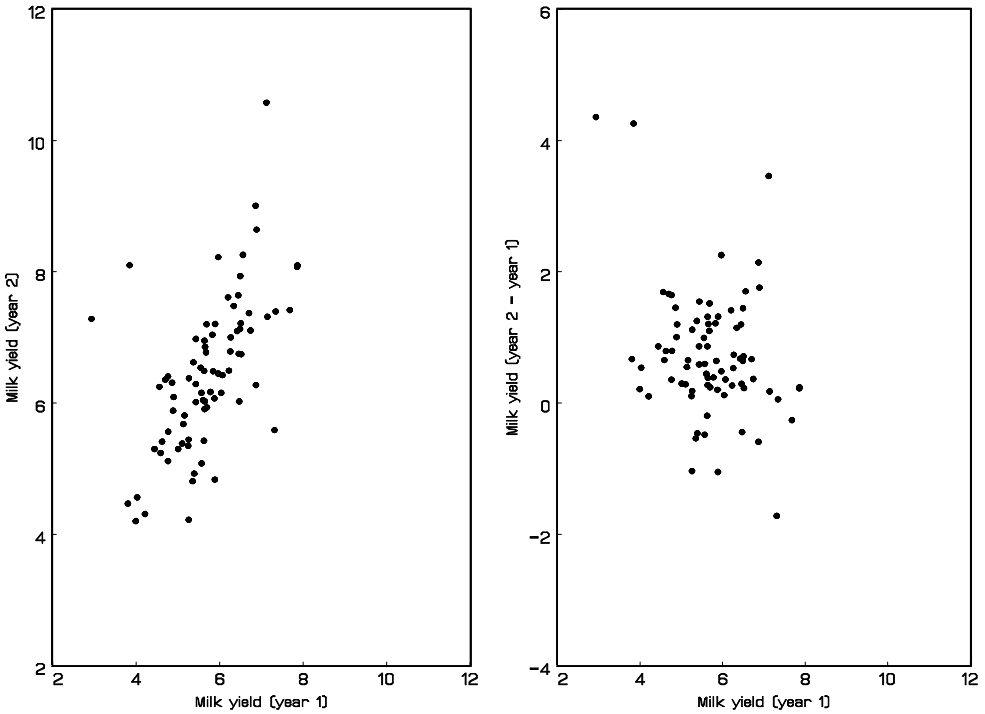

Figure 2.1: *Graphical representation of the Mastitis data. The first panel shows a scatter plot of the second measurement versus the first measurement. The second panel shows a scatter plot of the change versus the baseline measurement.*

conducted at the Department of Orthodontics of the Catholic University of Leuven (K.U.L.) in Belgium (1997) where a total of 50 male Wistar rats have been randomized to either a control group or one of the two treatment groups where treatment consisted of a low or high dose of the drug Decapeptyl which is an inhibitor for the testosterone production in the rats. The treatment started at the age of 45 days, and measurements were taken every 10 days, with the first observation taken at the age of 50 days. The responses of interest are distances (in pixels) between well-defined points on x-ray pictures of the skull of each rat, taken after the rat has been anesthetized. Unfortunately, many rats don't survive anesthesia implying that for only 22 (44%) rats all 7 designed measurements could have been taken. It is therefore very important to study how dropout affects the estimation of the treatment effects. Full details can be found in Verdonck et al. and the individual profiles of each rat are shown in Figure 2.2.

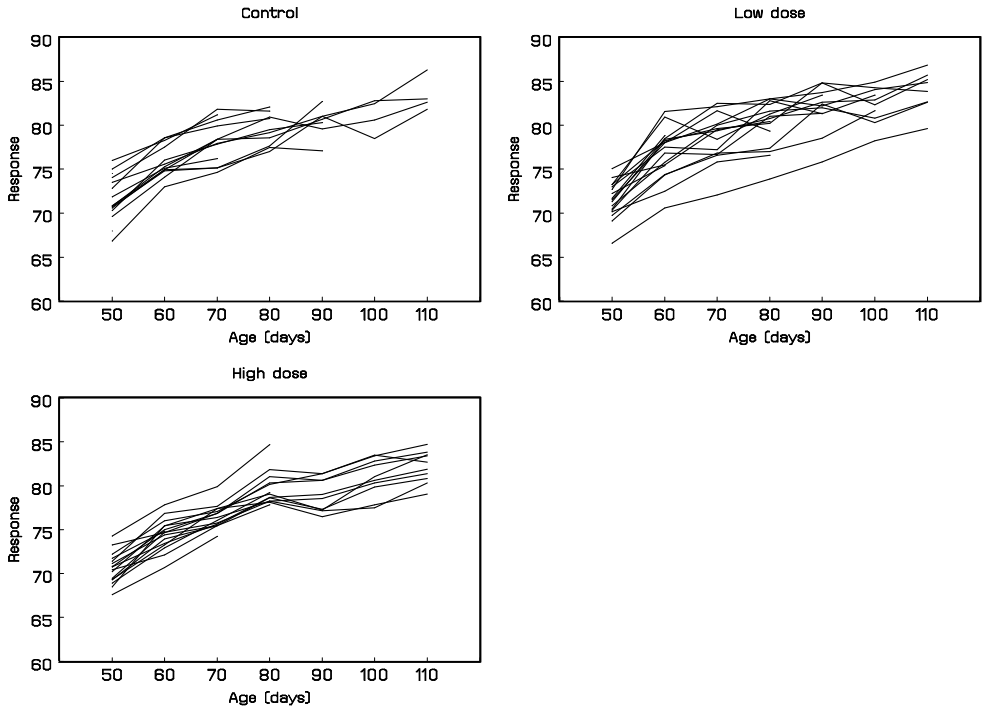

Figure 2.2: *Graphical representation of the Rats data. Individual growth curves for the three treatment groups separately.*

2.3 Milk Protein Trial

Another example is the milk protein data introduced by Verbyla & Cullis (1990) and re-analyzed by Diggle (1990) and Diggle and Kenward (1994). In this experiment 79 cows were randomized, after calving, to either of three diets: barley, lupins, or a mixture of both. The sampling plan envisaged to follow all 79 cows for 19 weeks and to determine protein content from a milk sample once in each study week. The time profiles for all 79 cows are plotted in Figure 2.3. All cows remained on study during the first fourteen weeks, whereafter the sample reduced to 59, 50, 46, 46 and 41 respectively, due to dropout.

The interest in these data arises from the fact that several analyzes have been performed before. Diggle (1990) for example, assumed random dropout whereas Diggle and Kenward (1994) concluded that dropout was non-random. However, several authors have remarked that the model of Diggle and Kenward (1994) should not be used to conclusively determine whether or not a dropout process is non-random.

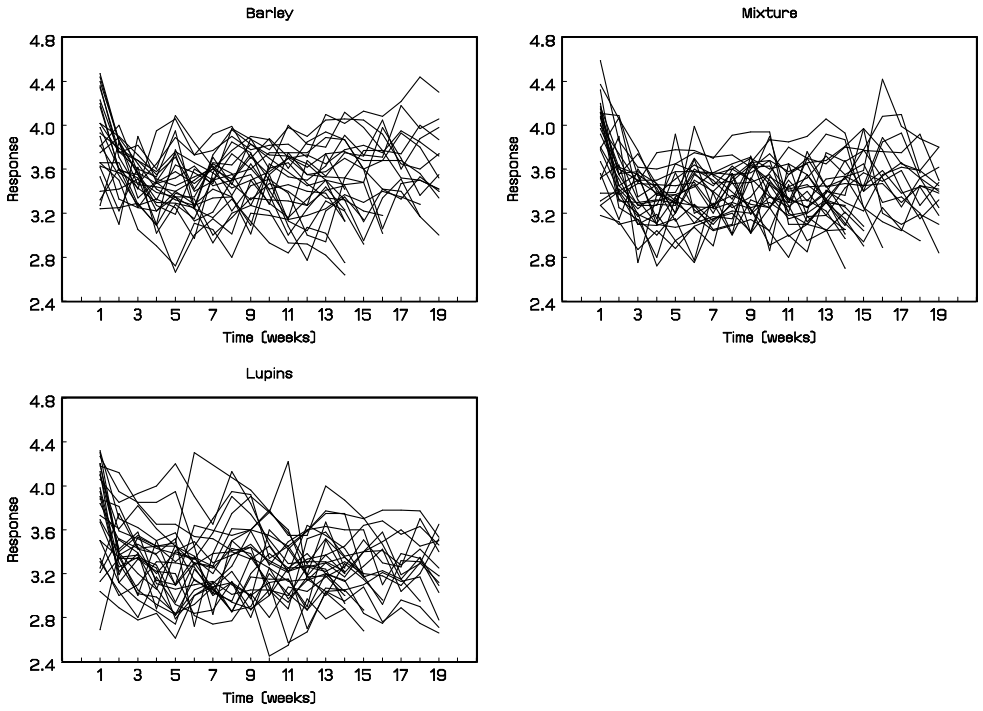

Figure 2.3: *Graphical representation of the Milk data. Individual profiles for the three diet groups separately.*

Indeed, Little (1995) says that "estimates rely heavily on normal assumptions and the correct specification of the dropout model, about which little is often known". Laird (1994) warns that "estimating the 'inestimable' can be accomplished only by making modeling *assumptions*, ... The consequences of model misspecification will probably be far more severe in the non-ignorable case". Rubin (1994) indicates that "even inferences for the data parameters generally depend on the posited missingness mechanism, a fact that typically implies greatly increased *sensitivity* of inference to reasonable model specifications". Molenberghs, Kenward and Lesaffre (1997) claim that "conclusions are conditional on the appropriateness of the assumed model, which in a fundamental sense is *not testable*".

In addition, serious doubts have been raised about even the appropriateness of the "dropout" concept in this study. Cullis (1994) warned that the conclusions inferred from the statistical model are very unlikely since usually there is no relation between dropout and a relatively low level of milk protein content. The real reason for dropout is human intervention. Cows entered the trial as they calved and the experiment was

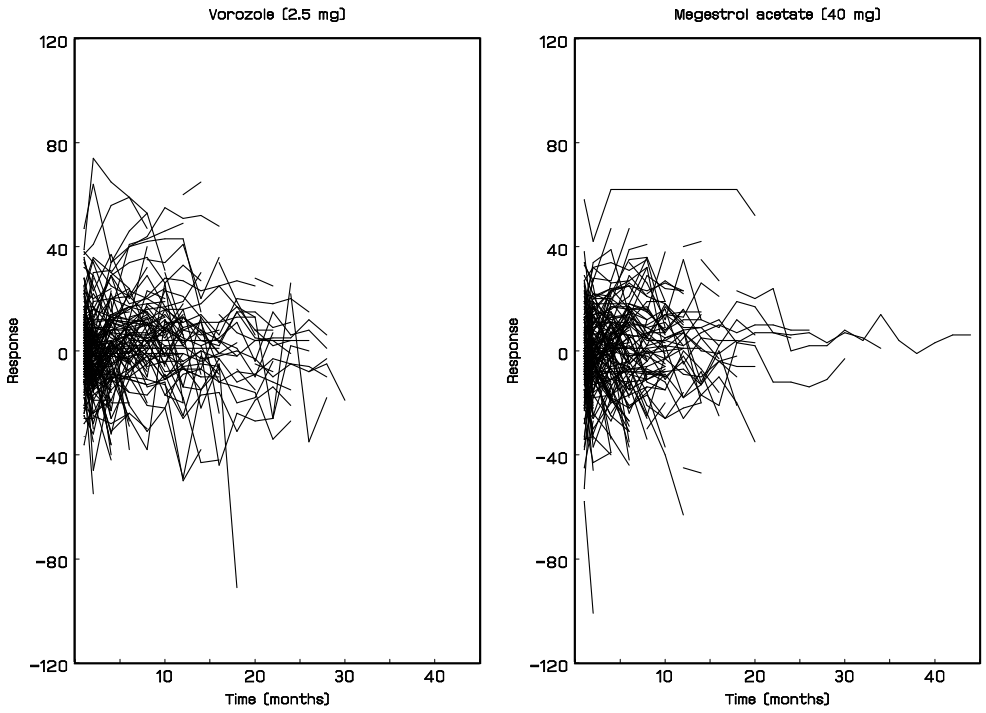

Figure 2.4: *Graphical representation of the Vorozole data. Individual profiles for the two treatment groups separately.*

terminated when feed availability declined in the paddock in which the animals were grazing. Thus, there are actually no dropouts but rather five cohorts representing the different starting times. Together with Cullis (1994), we conclude that especially with incomplete data a statistical analysis should not proceed without a thorough discussion with the experimenters.

2.4 Vorozole Study

The next set of data come from a randomized phase III trial comparing the new potent and selective third generation aromatase inhibitor *vorozole* (VOR) with *megestrol acetate* (MEG) in postmenopausal advanced breast cancer patients. This study was an open-label, multicenter, parallel group design conducted at 67 North American centers. Patients were randomized to either vorozole (225 patients, 2.5 mg taken once daily) or megestrol acetate (227 patients, 40 mg four times daily). The pa-

Table 2.1: *Vorozole study, means (standard deviations) per time (up to two years) and treatment arm for change in FLIC score versus baseline.*

	Vorozole			Megestrol Acetate		
Month	*N*	Mean	St. Dev.	*N*	Mean	St. Dev.
1	198	0.485	14.162	196	-1.622	15.706
2	176	-1.324	16.343	168	-1.268	16.988
4	130	1.031	17.808	136	0.971	16.825
6	94	4.883	17.425	104	1.808	19.038
8	77	7.519	18.506	76	2.737	19.315
10	68	6.309	16.312	60	2.733	16.808
12	58	4.207	21.079	39	2.821	21.738
14	42	3.857	19.806	32	2.219	20.789
16	37	0.189	18.590	22	1.409	18.940
18	26	1.423	25.942	15	2.533	23.086
20	24	0.750	14.405	11	5.909	21.422
22	20	-1.500	15.426	6	4.500	13.248
24	15	1.733	15.068	5	1.400	8.050

tient population consisted of postmenopausal patients with histologically confirmed estrogen-receptor positive metastatic breast carcinoma. All 452 randomized patients were followed until disease progression or death. The main objective was to compare the treatment group with respect to response rate while secondary objectives included a comparison relative to duration of response, time to progression, survival, safety, pain relief, performance status and quality of life. We will focus on overall quality of life, measured by the total Functional Living Index: Cancer (FLIC) (Schipper, Clinch and McMurray 1984). Precisely, a higher FLIC score is the more desirable outcome.

Patients underwent screening and for those deemed eligible a detailed examination at baseline (occasion 0) took place. Further measurement occasions were month 1, then from month 2 at bi-monthly intervals until month 44. The median age was 66 years for VOR, and 67 for MEG, and the means were respectively 65.1 (s.e. 9.8) and 65.6 (s.e. 10.0) years. The mean duration of breast cancer was 6.8 (s.e. 5.4) years for VOR, and 6.9 (s.e. 5.5) years for MEG. The average total FLIC score was 116.3 (s.e. 1.5) for VOR, and 117.1 (s.e. 1.3) for MEG. These total FLIC scores were calculated based on 199 resp. 213 patients. Goss *et al.* (1998) analyzed the data and found no significant differences: the response rate was 9.7% for VOR, versus 6.8% for MEG

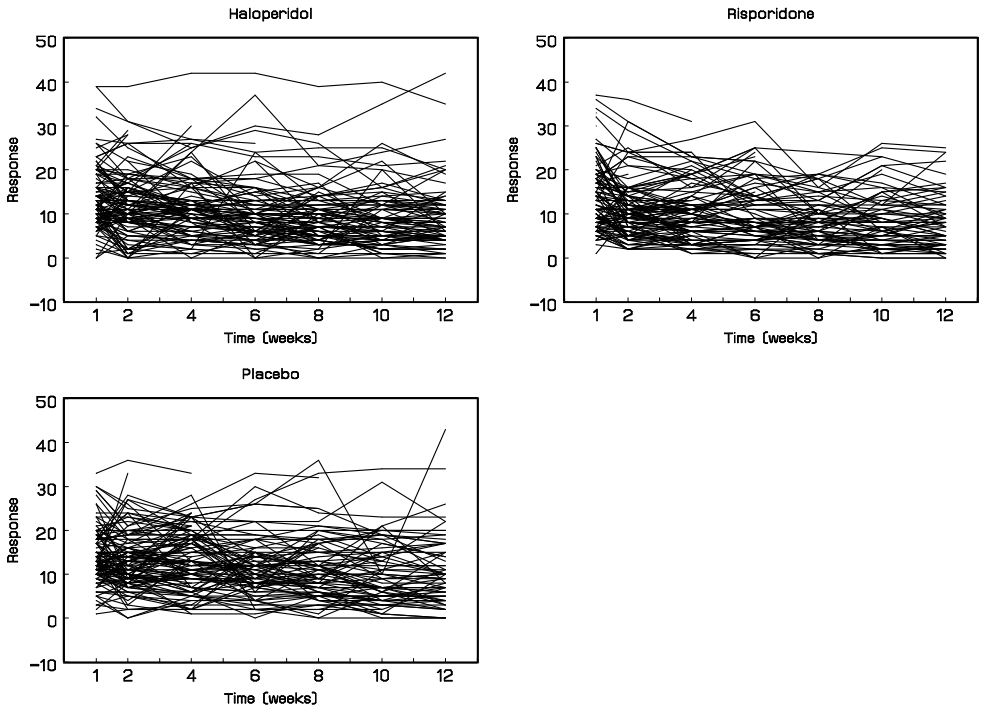

Figure 2.5: *Graphical representation of the Behave data. Individual profiles for the three treatment groups separately.*

(p=0.24); clinical benefit from treatment was demonstrated in 23.5% of VOR-treated patients versus 27.2% of MEG-treated patients (p=0.42). They also analyzed FLIC using a two-way ANOVA model with effects for treatment, disease status, as well as their interaction. Again, no significant difference was found. Figure 2.4 shows the individual profiles of all subjects separated over the two treatment groups and full details of this study are reported in Goss *et al.* (1998).

2.5 Behave Data

The last example concerns data from a three-armed clinical trial in patients with Alzheimer's disease (Reisberg et al. 1987). The outcome is a dementia score, ranging from 0 to 43. Treatment arm 1 is placebo (114 patients), while arms 2 (115 patients) and 3 (115 patients) involve active compounds. Of the patient population, 56.4% are female. There are 341 Caucasians, 2 orientals, and 1 black subject. Age ranges from 56 to 97 years with a median of 81 years. Measurements are taken at baseline, at

Table 2.2: *Behave data, sample size per treatment arm and dropout pattern.*

Pattern	1	2	3	4	5	6	7
Treatment 1	4	5	16	3	9	6	71
Treatment 2	4	9	7	6	3	5	81
Treatment 3	12	4	15	9	5	3	67

weeks 1, 2, and then every two weeks until week 12. Individual profiles are plotted in Figure 2.5. The study was conducted in 8 countries, and 50 investigators took part. The sample size per dropout pattern and per treatment arm is displayed in Table 2.2. In each of the arms, about 40% drop out before the end of the study.

Chapter 3

Modeling Incomplete Data

As is clear from the introduction, the problem of missing data is present in almost all longitudinal experiments. Some of these studies are designed in a way that the number of measurements planned to be taken per subject is variable and also the time points at which the measurements are taken can be random as well. These studies are called *unbalanced* and it is usually rather hard to define missingness within these settings. In *balanced* settings the number of measurements to be taken is fixed and the measurements are mostly taken even at a fixed set of time points approximately the same for all subjects. Because in the latter case missingness can be identified without ambiguity we will consider only this type of experiments and more precisely we will only take into account the case of *dropout* which means that a subject is completely observed until a certain point in time, where after no more measurements are taken. Moreover since dropout in balanced and unbalanced settings can be treated similar, restriction to balanced studies is no loss of generality.

In the following Section 3.1 we will introduce some general notation and terminology used to treat missing data while in Section 3.2 the difference between the selection models and the pattern-mixture frameworks based on a different factorization will become clear, Shared parameter models will be introduced and further details about the taxonomy introduced by Rubin (1976) and Little and Rubin (1987) will be discussed. Finally, Section 3.3 will outline two existing an rather standard approaches to treat missing data.

3.1 General Notation and Terminology

Let us introduce some necessary notation. We 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 $j = 1, \ldots, n_i$. The outcomes are grouped into a vector $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ and since we are dealing with balanced studies these vectors of measurements are all of the same length. 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 course, 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 a 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 $D_i = 1 + \sum_{j=1}^{n_i} R_{ij}$, 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 time of dropout are missing.

It is often necessary to split the vector $\boldsymbol{Y}_i$ into two subvectors $\boldsymbol{Y}_i^o$ and $\boldsymbol{Y}_i^m$ where $\boldsymbol{Y}_i^o$ contains those Y_{ij} for which $R_{ij} = 1$ and $\boldsymbol{Y}_i^m$ contains the remaining components. 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).

3.2 Missingness Process

In Modeling missing data it is thought to be necessary to consider a joint model for the measurement process together with the dropout process. In other words, interest is put into the joint density function

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

where $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$, and $\boldsymbol{W}_i$ are covariate matrices to be introduced later. We will use the parameter vector $\boldsymbol{\theta}$ to describe the measurement process while the vector $\boldsymbol{\psi}$ describes the missingness processes respectively. Later on $\boldsymbol{\theta}$ will be split into $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ representing the fixed effects and the variance parameters respectively. The above expression represents the joint distribution of the measurements Y_i and the dropout indicators D_i as defined before.

3.2.1 Different Factorizations

Starting from expression (3.1) we can factorize this density function in two ways leading each to a different framework as briefly discussed in the introduction. A first framework is based on following factorization:

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

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 this can be explained intuitively by considering the second factor to correspond to the (self-)selection of individuals into 'observed' and 'missing' groups. As an alternative one can consider so-called *pattern-mixture models* using the reversed factorization.

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

The factorized density (3.3) can be seen as a mixture of different populations, characterized by the observed pattern of missingness. After initial mention of these models

(Little and Rubin 1987, Glynn, Laird and Rubin 1986), they are receiving more attention lately (Little 1993, 1994a, 1995, Hogan and Laird 1997, Ekholm and Skinner 1998).

Although selection models and pattern-mixture models are interchangeable from a probabilistic point of view, in the sense that they represent different factorizations of the *same* joint distribution, in practice they encourage different kinds of simplifying assumptions. It is clear that the parameters in the selection models and these in the pattern-mixture models have a different meaning and also converting a model into one of the other framework is in general not straightforward, even not for normal measurement models. One attraction of selection models is that they 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)

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

and can be described as follows

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

  $$f(\boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{\alpha})$$

  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

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

  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). An informative process is allowed to depend on $\boldsymbol{Y}_s^o$.

Precisely because of this reason the selection Modeling framework is widely used and discussed. So it is argued by several authors that using a selection model, untestable assumptions have to be made about the mechanism of dropout (discussion of Diggle and Kenward 1994, Molenberghs, Kenward and Lesaffre 1997). Only recently, Little (1993) has suggested pattern-mixture models as a valuable alternative

to selection models. An early reference is Glynn, Laird, and Rubin (1986). The same reasoning concerning untestable assumptions is also true for pattern-mixture models but here every observed value of $\boldsymbol{R}_i$ leads to a different measurement process and more precisely each dropout pattern yields a different set of parameters. Because of this reason it is clear in the pattern-mixture settings which parameters cannot be identified. As a solution Little (1993) proposes the use of identifying restrictions in order to identify the non-identifiable to the identifiable parameters. It is still a problem to find evidence for the restrictions used through the data but an advantage is that separation between verifiable and unverifiable assumptions is clearly indicated.

While pattern-mixture models appear not to fit naturally into Little and Rubin's taxonomy, it is shown by Molenberghs, Michiels, Kenward and Diggle (1998) that pattern-mixture models can be classified similarly, and further that the intermediate category of "missing at random" is connected to particular kind of restrictions on the parameters of a pattern-mixture model in the case of monotone missingness. This suggests to us that a purely philosophical debate about the relative merits of the selection and pattern-mixture paradigms is unhelpful. Instead, the focus of debate should shift to a consideration of the statistical and scientific merits of proposed missing value models on their own terms. For example, if the question of scientific interest regards the treatment effect, averaged over all dropout patterns, then choosing an selection model seems to be obvious. On the other hand, if one is interested in the treatment effect, for various dropout patterns separately, then a pattern-mixture model is a natural choice.

3.2.2 Shared Parameter Models

Another alternative to jointly model the dropout and the measurement process is found by using a shared parameter model as they are discussed in Little (1995), Wu and Bailey (1989) and Mori, Woolson and Woodsworth (1992, 1994). Here one assumes that there exists a single random effect or a shared parameter, which is able to describe the dropout process as well as the measurement process in the sense that this latent variable splits the population in subgroups and conditional on this latent variable the dropout process and the measurement process are independent. This model can be represented as in Figure 3.1a. On the other, hand Figure 3.1b shows a possible extension by considering a categorical variable $q_i = (q_{i1}, ..., q_{ig})$ such that

$$Y_i \,|q_{ij} = 1, b_i \sim N\left(X_i\beta + Z_i b_i, \Sigma_i^{(j)}\right)$$

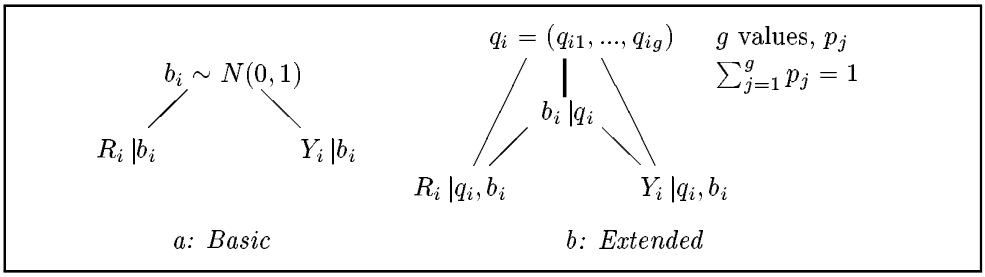

Figure 3.1: *Shared parameter model*

with

$$b_i\,|q_{ij} = 1 \sim N(\mu_j, D_j)$$

and

$$h\left[P\left(R_i = r\,|R_i > r, q_{ij} = 1, b_i\right)\right] = \frac{e^{W_i\gamma_j + \lambda b_i}}{1 + e^{W_i\gamma_j + \lambda b_i}}$$

Although we notice that these models can be very useful in describing missing data we will not take them into further account throughout our research. In Chapter 7 we will however, briefly suggest this type of models to deal with some unsolved issues within sensitivity analysis.

3.2.3 Ignorability

When a likelihood-based approach is used, one has to calculate the likelihood contribution of each subject i as follows

$$L^*(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i, \boldsymbol{y}_i, \boldsymbol{r}_i) \propto f(\boldsymbol{y}_i, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}). \tag{3.5}$$

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:

$$L(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i, \boldsymbol{y}_i, \boldsymbol{r}_i) \propto f(\boldsymbol{y}_i^o, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

with

$$f(\boldsymbol{y}_i^o, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = \int f(\boldsymbol{y}_i, \boldsymbol{r}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{W}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) d\boldsymbol{y}_i^m \tag{3.6}$$

$$= \int f(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m|\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{W}_i, \boldsymbol{\psi}) d\boldsymbol{y}_i^m \tag{3.7}$$

Under an MAR process, we obtain

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

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 then inference can be based on the marginal observed data density only. This technical requirement is referred to as the separability condition.

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 'informativeness' 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. Of course, ignorability is unhelpful when at least part of the scientific interest is directed towards the missingness process. Note that while the terminology introduced in the previous section is independent of the statistical framework chosen to analyze the data this is no longer the case with the terms *ignorable* and *non-ignorable* missingness. The latter terms depend crucially on the inferential framework (Rubin 1976).

Classical examples of the more stringent condition with frequentist methods are ordinary least squares and the generalized estimating equations approach of Liang and Zeger (1986). These GEE define an asymptotically unbiased estimator only under MCAR. Robins, Rotnizky and Zhao (1995) have established that some progress can be made under MAR and even under informative processes. Their method is based on including weights that depend on the missingness probability, proving the point that at least some information on the missingness mechanism should be included and thus that ignorability does not hold.

3.3 Standard Methodology for Incomplete Data

Missing data nearly always entail problems for the practicing statistician. First, inference will often be invalidated when the observed measurements do not constitute a simple random subset of the complete set of measurements. Secondly, even when correct inference would follow, it is not always an easy task to trick standard software

into operation on a ragged data structure.

Even in the simple case of a one-way ANOVA design and under an MCAR mechanism operating, problems occur since missingness destroys the balance between the sizes of the subsamples. This implies that a slightly more complicated least squares analysis has to be invoked. Of course, a regression module for the latter analysis is included in most statistical software packages. The trouble is that the researcher has to be aware which tool to choose for particular classes of incomplete data.

Little and Rubin (1987) give an extensive treatment of methods to analyze incomplete data, many of which are intended for continuous, normally distributed data. Some of these methods were proposed more than fifty years ago. Examples are Yates' (1933) iterated ANOVA and Bartlett's (1937) ANCOVA procedures to analyze incomplete ANOVA designs. The former method is an early example of the Expectation-Maximization (EM) algorithm (Dempster, Laird and Rubin 1977). This EM-algorithm is discussed in Section 3.3.1.

The computationally simplest technique is a complete case analysis, in which the analysis is restricted to the subjects for whom all intended measurements have been observed. A complete case analysis is popular because it maps a ragged data matrix into a rectangular one, by deleting incomplete cases. An alternative approach, with a similar effect on the applicability of complete data software, is based on imputing missing values. One distinguishes between single imputation and multiple imputation (Rubin 1987). 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. Also in the multiple imputation technique, 'holes' in the data set are filled, but to account for the uncertainty in filling in missing values, the imputation is done multiple times, and each time the complete data are analyzed. The theory of multiple imputation is explained in Section 3.3.2.

Another family is based on the principle of analyzing the incomplete data as such. A simple representative is the so-called available case analysis. A popular and very general technique to optimize incomplete data likelihoods under MAR is the EM algorithm (Dempster, Laird and Rubin 1977). Little and Rubin (1987) used the EM algorithm to analyze their incomplete version of the growth data. The principle ideas behind this method will be given in the next section.

3.3.1 EM-algorithm

The EM algorithm consists of two components, the *Expectation* and *Maximization* steps. Each step is completed once within each algorithm cycle. Cycles are repeated until a suitable convergence criterion is satisfied. In the expectation step the unobserved (or missing) data are estimated by their expectations given the observed data and current parameter values. In the maximization step the parameters are estimated using maximum likelihood applied to the observed data augmented by the estimates of the unobserved data. Effectively this maximizes, in each cycle, the expectation of the complete data log likelihood $\mathrm{E}[\log L(\boldsymbol{\theta})]$ where the expectation is taken with respect to the observed data and the current fitted values of $\boldsymbol{\theta}$. Dempster, Laird and Rubin (1977) show that on convergence the fitted parameters are equal to a local maximum of the likelihood function, which is the maximum likelihood estimate in the case of a unique maximum.

Two of the main drawbacks of the EM algorithm are its typically very slow rate of convergence and its lack of direct provision of a measure of precision for the maximum likelihood estimates. Both problems are in fact related and several proposals have been made to overcome them. We mention the technique suggested by Louis (1982), the EM-aided differentiation by Meilijson (1989), the "rate matrix" method of Meng and Rubin (1991), and the linear transformation method of Baker (1992). Standard errors and Wald statistics are computed directly from the observed information and score tests are also relatively simple to compute.

We will use Meilijson's (1989) proposal, which is based on the property that the derivative of the complete data score vector coincides with the observed information matrix. It leads to an easy numerical algorithm, using the classical finite differences of the score vector to approximate the derivative. Let the constant that defines the differences be ϵ. To compute the jth column of the information matrix, one changes $\boldsymbol{\theta}$ to $\boldsymbol{\theta}_j$, where all components remain the same, except for the jth one which is changed to $\theta_j + \epsilon$. Then one E step is carried out, yielding $\boldsymbol{Y}(\boldsymbol{\theta}_j)$. Next, the score vector $\boldsymbol{S}_j$ is computed. An approximation for the jth column is given by $(\boldsymbol{S}_j \boldsymbol{S})/\epsilon$, where $\boldsymbol{S}$ is the score vector at maximum. Replacing all quantities by their estimated values yields a convenient algorithm.

3.3.2 Multiple Imputation

The theory of multiple imputation is presented in Rubin (1987). Several other sources, such as Rubin and Schenker (1986), Little and Rubin (1987), Rubin (1996) and Schafer (1997), give an excellent account of the technique. As discussed by Rubin and Schenker (1986), the theoretical justification for multiple imputation is most easily understood using Bayesian methodology.

Suppose interest lies in estimating the vector $\boldsymbol{\beta}$, containing the parameters of interest. Rubin (1987) proposed using multiple imputation to "fill-in" the unobserved components of the outcome vectors using the observed data and then use the filled-in data to estimate $\boldsymbol{\beta}$. His method also yields a variance estimator. In order to be able to fill in values, we need the distribution of the missing data, given the observed data and a parameter vector $\boldsymbol{\gamma}$. Multiple imputation is most useful when $\boldsymbol{\gamma}$ is an easily estimated set of parameters, while $\boldsymbol{\beta}$ is complicated to estimate in the presence of missing data.

Recall that the observed data are $\boldsymbol{Y}^o$ and the complete data are $\boldsymbol{Y}$. Multiple imputation uses $\boldsymbol{Y}^o$ to fill in $\boldsymbol{Y}^m$, leading to the complete data $\boldsymbol{Y} = (\boldsymbol{Y}^o, \boldsymbol{Y}^m)$. If we knew the distribution of $\boldsymbol{Y}^m$, with parameter vector $\boldsymbol{\gamma}$, then we could impute $\boldsymbol{Y}^m$ by drawing from the conditional distribution $f(\boldsymbol{Y}^m|\boldsymbol{Y}^o, \boldsymbol{\gamma})$. Since $\boldsymbol{\gamma}$ is unknown, we estimate it from the data, yielding $\widehat{\boldsymbol{\gamma}}$, and use the distribution $f(\boldsymbol{Y}^m|\boldsymbol{Y}^o, \widehat{\boldsymbol{\gamma}})$. Because $\widehat{\boldsymbol{\gamma}}$ is a random variable, we must also take its variability into account in drawing imputations. In Bayesian terms, $\boldsymbol{\gamma}$ is a random variable of which the distribution depends on the data. So we first obtain the posterior distribution of $\boldsymbol{\gamma}$ from the data, a distribution which is a function of $\widehat{\boldsymbol{\gamma}}$.

After formulating the posterior distribution of $\boldsymbol{\gamma}$, we use the following imputation algorithm.

1. Draw $\boldsymbol{\gamma}^*$ from the posterior distribution of $\boldsymbol{\gamma}$, $f(\boldsymbol{\gamma}|\boldsymbol{X}, \boldsymbol{Y}^o)$. We approximate this posterior distribution by a normal.
2. Draw $\boldsymbol{Y}^m$ from $f(\boldsymbol{Y}^m|\boldsymbol{X}, \boldsymbol{Y}^o, \boldsymbol{\gamma}^*)$.
3. Use the completed data $\boldsymbol{Y}$ and the model to estimate the parameter of interest $\boldsymbol{\beta}^*$ and its variance $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*)$, called the within-imputation variance.

These three steps are repeated independently M times, resulting in $\boldsymbol{\beta}^*_m$, $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*_m)$,

$m = 1, \ldots, M$.

In case the data to be filled in are categorical, we use a uniform random number generator in step 2 (see Rubin 1987, pp. 169-170). Suppose the count Z is to be distributed over the cells Z_k^c , $k = 1, \ldots, c$. Then, the cumulative probabilities

$$\begin{aligned} \lambda_0 &= 0, \\ \lambda_k &= \frac{\sum_{k'=1}^{k} \nu_{k'}^c}{\nu}, \qquad k = 1, \ldots, c \end{aligned}$$

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

Finally, we combine the estimates obtained after M imputations. The overall estimated parameter vector is the mean of all individual estimates:

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

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

$$\boldsymbol{\Sigma}^* = \boldsymbol{W} + \frac{M+1}{M} \boldsymbol{B}$$

where

$$\boldsymbol{W} = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\Sigma}(\boldsymbol{\beta}_m^*),$$

the mean of the within-imputation variances, and

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

the between-imputations variance (Rubin 1987). Based on these variances, one can calculate approximate 95% confidence intervals. Finding an appropriate reference distribution is not an easy matter. Rubin (1987) proposes a multivariate T distribution. Shafer (1997, p. 113) suggests that the approximations by Li, Raghunathan and Rubin (1991) work well in practice. Since in our case the number of imputations will be large, we can certainly rely on the corresponding normal approximation.

Chapter 4

Selection Models

In this Chapter we will discuss the most popular selection model framework in more detail. Specifically we will consider a selection model similar to the one introduced by Diggle and Kenward (1994). While these authors consider a multivariate model for the measurements combined with a logistic regression model for the dropout process we will use the generalized mixed model as described in Laird and Ware (1982) and Verbeke and Molenberghs (1997, 2000) to model the measurement process keeping the dropout model to be based on a logistic regression. Section 4.1 treats this model in full detail and starting from this model we will define two formal tools to perform sensitivity analysis within the selection modeling framework. In section 4.2 Local and Global influence methodology will be introduced and compared with respect to similarity in results, advantages en disadvantages of both techniques. Finally section 4.3 list the results of an application of the local influence methodology to the rats dataset and the mastitis dataset introduced in Chapter 2. In Chapter 6 we will apply both influence tools to the milk protein trial and furthermore in section 5.4.1 we will contrast these methods with some techniques from the pattern-mixture framework introduced in Chapter 5. The methodology described in this chapter is also represented in full detail in some of our publications: Verbeke *et al* (1998), Thijs *et al* (2000) and Molenberghs *et al* (2001). For the categorical counterpart not discussed here we can refer to Van Steen *et al* (2002).

4.1 A Selection Model

While Diggle and Kenward (1994) combine a multivariate normal model for the measurement process with a logistic regression model for the dropout process we will generalize the measurement model slightly to be the linear mixed model (Laird and Ware 1982). In section 4.1.1 the linear mixed model will be introduced in full detail with specific emphasis on components of random variability that are typically encountered in longitudinal data while section 4.1.2 handles the logistic dropout model.

4.1.1 Measurement Model: The Linear Mixed Model

Linear mixed-effects models have been proposed by Laird and Ware (1982) and can be written as follows:

$$\boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{4.1}$$

where $\boldsymbol{Y_i}$ is the n_i dimensional response vector for subject i, $1 \leq i \leq N$, N is the number of subjects, X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ known design matrices, $\boldsymbol{\beta}$ is the p dimensional vector containing the fixed effects, $\boldsymbol{b}_i \sim N(\mathbf{0}, D)$ is the q dimensional vector containing the random effects , $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma_i)$ is a n_i dimensional vector of residual components, and $b_1, \ldots, b_N, \varepsilon_1, \ldots, \varepsilon_N$ are assumed to be independent. Finally, D is a general $(q \times q)$ covariance matrix with (i, j) element $d_{ij} = d_{ji}$ and Σ_i is a $(n_i \times n_i)$ covariance matrix.

- **Random Effects:** In model (4.1) the random effects stem from heterogeneity between individuals. This means that various aspects of their behavior may exhibit inter-individual random variation. For example, some subjects will be intrinsically high responders. The residual variability $\boldsymbol{\varepsilon}_i$ in (4.1) may be further refined and decomposed into the following qualitatively distinct components (Diggle, Liang and Zeger, 1994):

- **Serial correlation:** This component arises due to the fact that measurements closer in time often show a stronger similarity than measurements further apart.

- **Measurement error:** Measurement errors occur when the measurement process itself introduces an element of random variability. For instance, there might be substantial variation in results from bioassays of blood samples, even when two

measurements are taken at the same time from the same subject, or when a sample is split into two subsamples which are then analyzed separately.

This distinction leads to the decomposition $\boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}_i^{(1)} + \boldsymbol{\varepsilon}_i^{(2)}$ (Verbeke and Molenberghs, 2000), where

$$\begin{cases} \boldsymbol{\varepsilon}_i^{(1)} \sim N(\mathbf{0}, \tau^2 H_i), \\ \\ \boldsymbol{\varepsilon}_i^{(2)} \sim N(\mathbf{0}, \sigma^2 I_{n_i}). \end{cases} \tag{4.2}$$

In this formula, $\boldsymbol{\varepsilon}_i^{(1)}$ captures serial correlation. The serial covariance matrix H_i only depends on i through the number n_i of observations and through the time points t_{ij} at which measurements are taken.

The structure of the matrix H_i is determined through the autocorrelation function $\rho(t_{ij} - t_{ik})$. A first simplifying assumption is that it depends only on the time interval between two measurements Y_{ij} and Y_{ik}, i.e., $\rho(t_{ij} - t_{ik}) = \rho(\mid t_{ij} - t_{ik} \mid)$, where $u = \mid t_{ij} - t_{ik} \mid$ denotes time lag. This function decreases such that $\rho(0) = 1$ and $\rho(+\infty) = 0$. Two popular choices of this function ρ to capture serial correlation is by means of exponential or Gaussian decay. An exponential process is based on writing the correlation between two residuals at times t_{ij} and t_{ik} as

$$\text{Corr}(t_{ij}, t_{ik}) = \exp\left(-\frac{|t_{ij} - t_{ik}|}{\phi}\right) = \rho^{|t_{ij} - t_{ik}|}, \tag{4.3}$$

where $\rho = \exp(-1/\phi)$. The Gaussian counterpart is

$$\text{Corr}(t_{ij}, t_{ik}) = \exp\left(-\frac{(t_{ij} - t_{ik})^2}{\phi^2}\right) = \rho^{(t_{ij} - t_{ik})^2}, \tag{4.4}$$

where $\rho = \exp(-1/\phi^2)$.

It follows from (4.1) that, conditional on the random effect $\boldsymbol{b}_i$, $\boldsymbol{Y_i}$ is normally distributed with mean vector $X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i$ and with covariance matrix Σ_i. Therefore Inference can be based on the marginal distribution of the response $\boldsymbol{Y_i}$ which, after integrating over random effects, can be expressed as

$$\boldsymbol{Y_i} \sim N(X_i\boldsymbol{\beta}, Z_i D Z_i' + \Sigma_i). \tag{4.5}$$

Let $\boldsymbol{\alpha}$ denote the vector of all variance and covariance parameters (usually called variance components), i.e. $\boldsymbol{\alpha}$ consists of the $q(q+1)/2$ different elements in D and of

all parameters in Σ_i. Finally, let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ be the vector of all parameters in the marginal model for $\boldsymbol{Y_i}$.

When model (4.1) contains solely a random intercept between subjects, a serially correlated component and a measurement error, a useful aid to the formulation of an appropriate model for the covariance structure, especially the autocorrelation function, is the variogram (Diggle, 1990). For a stochastic process $Y(t)$, the variogram is defined as $V(u) = \frac{1}{2}E[Y(t) - Y(t-u)]^2$. Under the specified model, this reduces to $V(u) = \sigma^2 + \tau^2[1 - \rho(u)]$ (Diggle, Liang and Zeger, 1994).

Finally, decomposition (4.2) assumes that the variance of residual components $\boldsymbol{\varepsilon}_i$ is constant over time. This is not always the case and one way to accommodate variance heterogeneity is through a log-linear variance model producing exponential local effects, also called dispersion effects (Littell, Milliken, Stroup, and Wolfinger 1996; SAS Institute Inc. 1997). In this model, measurement errors take the form $\sigma^2 \text{diag}[\exp(U\boldsymbol{\delta})]$, where U is a design matrix and $\boldsymbol{\delta}$ a vector of dispersion parameters. This affords a way of modeling the variability in terms of effects to be specified, such as time.

4.1.2 Dropout Model

As stated before we consider only incompleteness due to dropout and furthermore we assume that the first measurement Y_{i1} is obtained for all subjects in the study. The model for the dropout process is based on a logistic regression for the probability of dropout at occasion j, given the subject is still in the study. We denote this probability by $g(\boldsymbol{h}_{ij}, y_{ij})$ in which $\boldsymbol{h}_{ij}$ is a vector possibly containing all responses observed up to but not including occasion j, as well as relevant covariates. We now can assume that $g(\boldsymbol{h}_{ij}, y_{ij})$ satisfies

$$\text{logit}[g(\boldsymbol{h}_{ij}, y_{ij})] = \text{logit}\left[\text{pr}(D_i = j | \boldsymbol{h}_{ij}, y_{ij})\right] = \boldsymbol{\Psi}\left(\boldsymbol{h}_{ij}, y_{ij}\right). \tag{4.6}$$

For simplicity we further assume the vector $\boldsymbol{h}_{ij}$ to contain only an intercept and the previous measurement y_{ij-1}. Finally we can rewrite expression (4.6) as follows:

$$\text{logit}[g(y_{ij-1}, y_{ij})] = \text{logit}\left[\text{pr}(D_i = j | y_{ij-1}, y_{ij})\right] = \psi_0 + \psi_1 y_{ij-1} + \psi_2 y_{ij}. \tag{4.7}$$

Based on expression (4.7) it is clear to understand the terminology introduced by Little and Rubin (1987) and described in Chapter 3. When both the parameters

ψ_1 and ψ_2 equal zero the dropout mechanism is called to be Missing Completely At Random (MCAR). When $\psi_1 \neq 0$ dropout is termed Missing At Random (MAR) and when $\psi_2 \neq 0$ we call the dropout mechanism to be Missing Not At Random (MNAR).

Expression (4.7) can now in a similar way be used to construct the dropout process:

$$f(d_i|\boldsymbol{y_i}, \boldsymbol{\psi}) = \begin{cases} \prod_{j=2}^{n_i}[1 - g(\boldsymbol{h}_{ij}, y_{ij})] & \text{completer } (d_i = n_i + 1), \\ \prod_{j=2}^{d-1}[1 - g(\boldsymbol{h}_{ij}, y_{ij})]g(\boldsymbol{h}_{id}, y_{id}) & \text{dropout } (d_i = d \leq n_i). \end{cases} \tag{4.8}$$

Since we now have the expressions for the measurement model as well as for the dropout model we are able to combine both mechanisms in one selection model and in doing so we can fit these models in order to analyze the data. In line with our discussion in the introduction, Rubin (1994) points out that such analyzes heavily depend on the assumed dropout process while it is impossible to find evidence for or against the model, unless supplemental information on the dropouts is available. Further, note that in practice, subjects may drop out for a variety of reasons, with several competing dropout processes operating simultaneously. This might lead some subjects to drop out at random while others drop out non-randomly. This possibility is not taken into account in the above model, which at best will capture a dominant trend, should it operate in a way that can be captured approximately by the proposed model form. While a general awareness of the need for sensitivity analysis has grown, only few actual proposals have been made. Moreover, many of these are to be considered as useful but ad hoc approaches. In our view, a more formal approach to sensitivity analyzes should be fruitful as well.

4.2 Sensitivity Analysis

George Box has a famous quote saying that all statistical models are wrong, but some are useful. Cook (1986) uses this idea to motivate his assessment of local influence. He suggests that more confidence can be put in a model which is relatively stable under small modifications. The best known perturbation schemes are based on case-deletion (Cook and Weisberg 1982) in which the effect is studied of completely removing cases from the analysis. This reasoning will form the basis for our global influence

methodology introduced in section 4.2.1 and in doing so it will be possible to determine which subjects might be influential for the analysis. On the other hand using case deletion all information from a single subject is deleted at once and therefore it is hard to tell whether that subject has some influence on a specific aspect of the model. A solution for the latter problem can be found in a quite different paradigm being a local influence approach where one again investigates how the results of an analysis are changed under small perturbations of the model but where these perturbations can have specific interpretations. In the framework of the linear mixed model Beckman, Nachtsheim and Cook (1987) used local influence to assess the effect of perturbing the error variances, the random-effects variances and the response vector. In the same context, Lesaffre and Verbeke (1998) have shown that the local influence approach is also useful for the detection of influential subjects in a longitudinal data analysis. In section 4.2.2 we will develop a similar methodology to detect influential subjects with respect to the dropout mechanism. Moreover, since the resulting influence diagnostics can be decomposed in interpretable components, these methods are particularly useful for gaining insight in the reasons why some subjects are more influential than others.

It can be argued that subjects with a large impact on the (dropout) model parameters are likely to be responsible for, e.g., false conclusions about the nature of the dropout mechanism. Indeed, due to the large sensitivity of conclusions to model assumptions, one or a few influential observations can drive the conclusions in selection models for incomplete longitudinal data. While such a statement would be broadly true in almost any regression setting, it is even more the case in this context. Kenward (1998) showed that two outlying subjects changed the dropout mechanism from random into non-random when analyzing the mastitis dataset, previously analyzed in Diggle and Kenward (1994). These results will be confirmed, using influence diagnostics. In addition, he showed that changing an appropriate conditional distribution from a normal to a t distribution with a low number of degrees of freedom, also changed the conclusions. These considerations, motivate the use of influence diagnostics to detect subjects that may distort conclusions based on selection models for incomplete longitudinal data. We can already conclude that the local and global approaches are complimentary, rather than competitors and can both take part in a complete sensitivity analysis.

4.2.1 Global Influence

A first tool to perform sensitivity analysis as stated before is by means of global influence starting from case-deletion. This methodology is based on the difference in log-likelihood between the model fitted to the dataset as a whole on the one hand and the dataset minus one subject on the other hand. Denote the likelihood function, corresponding to measurement model (4.5) and dropout model (4.8) as follows

$$\ell(\boldsymbol{\gamma}) \;=\; \sum_{i=1}^{N} \ell_i(\boldsymbol{\gamma}), \tag{4.9}$$

in which $\ell_i(\boldsymbol{\gamma})$ is the contribution of the i^{th} individual to the log-likelihood, and where $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\omega})$ is the s-dimensional vector, grouping the parameters of the measurement model and the dropout model. Further, we denote by

$$\ell_{(-i)}(\boldsymbol{\gamma}), \tag{4.10}$$

the log-likelihood function, where the contribution of the i^{th} subject has been removed. Cook's distances are based on measuring the discrepancy between either the maximized likelihoods (4.9) and (4.10) or (subsets of) the estimated parameter vectors $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\gamma}}_{(-i)}$, with obvious notation. Precisely, we will consider both

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

as well as

$$CD_{2i}(\boldsymbol{\gamma}) \;=\; 2\;(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(-i)})' \ddot{L}^{-1}\;(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(-i)}). \tag{4.11}$$

Formulation (4.11) easily allows to consider the global influence in a subvector of $\boldsymbol{\gamma}$, such as the dropout parameters $\boldsymbol{\psi}$, or the non-random parameter ω. This will be indicated using notation of the form $CD_{2i}(\boldsymbol{\psi})$, $CD_{2i}(\omega)$, etc.

Alternative global influence measures are possible. One could think of the behavior of a test statistic, such as a Wald test for treatment or time effect, under a case deletion scheme. A more formal study of such a quantity is topic of ongoing research.

In linear regression, global influence is conceptually simple, straightforward in computation and well studied. The latter two of these features do not carry over to more general settings. Indeed, the calculation of the Cook's distances requires N model fits (even though one-step approximations can reduce the burden somewhat). There is a more fundamental problem, however. It is hard to assess the influence that

can be ascribed to a specific cause, since by deleting a subject all types of influence stemming from it are lumped together and for this reason we will now develop the so called local influence tool where it is easier to interpret the perturbation.

4.2.2 Local Influence

As a second tool for sensitivity analysis the local influence method will now be described for the linear mixed-effects model with respect to the different variance components as introduced in previous section. More precisely, we are interested in the influence the of the assumption of non-random dropout on the parameters of interest, which will most often be the fixed-effect parameters, possibly supplemented with the variance components. This can be done in a meaningful way by extending (4.7) as follows:

$$\text{logit}[g(y_{ij-1}, y_{ij})] = \text{logit}\left[\text{pr}(D_i = j | y_{ij-1}, y_{ij})\right] \quad (4.12)$$

$$\text{logit}[g(y_{ij-1}, y_{ij})] = \psi_0 + \psi_1 y_{ij-1} + \omega_i y_{ij}. \quad (4.13)$$

in which different subjects give different *weights* to the response at time d to predict dropout at time d. If all ω_i equal zero, then the model reduces to a MAR model which cannot influence the measurement model parameters. Therefore (4.13) can be seen as an extension of the MAR model, which allows some individuals to drop out in a "less random" way ($|\omega_i|$ large) than others ($|\omega_i|$ small). It has to be noted that, even when ω_i is large, we still cannot conclude that the dropout model for these subjects is non-random. Rather, it is a way of pointing to subjects which, due to their strong influence, are able to distort the model parameters such that they can produce, for example, a dropout mechanism which is *seemingly* non-random. In reality, many different characteristics of such an individual's profile might be responsible for this effect. As mentioned earlier, such sensitivity has been alluded to by many authors, such as Laird (1994), Little (1994), and Rubin (1994).

Studying the effect of extending an MAR model to the non-random case on the parameters of interest (such as treatment effect, time evolution, variance components, dropout parameters, ...) can now be achieved by investigating the effect of perturbing the ω_i's around zero. This will be done using the local influence approach of Cook (1986). Of course, not all possible forms of impact, resulting from sensitivity to dropout model assumptions, will be found in this way. Therefore, the method proposed here should be viewed as one component of a sensitivity analysis, but ought

ideally to be supplemented with other methods. Furthermore, it is clear that the global influence method introduced earlier differs substantially from the local influence approach. Indeed, a global counterpart of the local influence method would allow one single subject at a time to drop out in a MNAR rather than a MAR way. Technically, such a model would yield infinite parameter estimates and therefore the local influence framework is the natural setting. The global influence method aimed at the effect of deleting one subject while in de local influence settings we are interested in changing the generating mechanism for one subject. Let us now study in full detail the local influence method.

We denote the log-likelihood function corresponding to model (4.13) by

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

in which $\ell_i(\boldsymbol{\gamma}|\boldsymbol{\omega})$ is the contribution of the i^{th} individual to the log-likelihood, and where $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\psi})$ is the s-dimensional vector, grouping the parameters of the measurement model and the dropout model, not including the $N \times 1$ vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)'$ of weights defining the perturbation of the MAR model. This expression arises from taking the logarithm of (3.5), the model components of which are described in Section 4.1. It is assumed that $\boldsymbol{\omega}$ belongs to an open subset Ω of $\mathbb{R}^N$. For $\boldsymbol{\omega}$ equal to $\boldsymbol{\omega_0} = (0, 0, \ldots, 0)'$, $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega_0})$ is the log-likelihood function which corresponds to a MAR dropout model.

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

$$LD(\boldsymbol{\omega}) \;=\; 2\left(\ell(\widehat{\boldsymbol{\gamma}}|\boldsymbol{\omega_0}) - \ell(\widehat{\boldsymbol{\gamma}}_\omega|\boldsymbol{\omega_0})\right).$$

This takes into account the variability of $\widehat{\boldsymbol{\gamma}}$. Indeed, $LD(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\gamma})$ is strongly curved at $\widehat{\boldsymbol{\gamma}}$ (which means that $\boldsymbol{\gamma}$ is estimated with high precision) and small otherwise. Therefore, a graph of $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ contains essential information on

the influence of perturbations. It is useful to view this graph as the geometric surface formed by the values of the $N + 1$ dimensional vector

$$\boldsymbol{x_i}(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix},$$

as $\boldsymbol{\omega}$ varies throughout Ω.

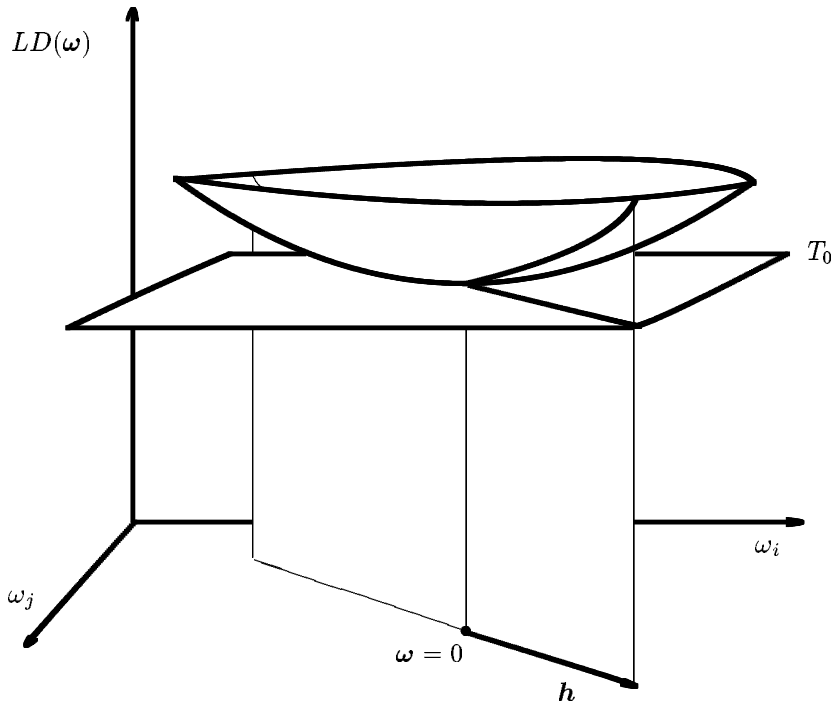

Figure 4.1: *Graphical representation of the likelihood surface in case of two subjects illustrating the principle of local influence.*

Since this so-called *influence graph* as shown in figure 4.1 can only be depicted when $N = 2$, Cook (1986) proposed to look at *local influence*, i.e., at the normal curvatures $C_{\boldsymbol{h}}$ of $\boldsymbol{x_i}(\boldsymbol{\omega})$ in $\boldsymbol{\omega_0}$, in the direction of some N dimensional vector $\boldsymbol{h}$ of unit length. Let $\boldsymbol{\Delta}_i$ be the s dimensional vector defined by

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

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

$$C_{\boldsymbol{h}} \;=\; 2 \;\Big|\; \boldsymbol{h}'\, \boldsymbol{\Delta}'\, \ddot{L}^{-1}\, \boldsymbol{\Delta}\, \boldsymbol{h} \;\Big|\,. \tag{4.14}$$

There are several ways in which (4.14) can be used to study $\boldsymbol{x_i}(\boldsymbol{\omega})$, each corresponding to a specific choice of the unit vector $\boldsymbol{h}$. One evident choice is the vector $\boldsymbol{h_i}$ containing 1 in the i^{th} position and 0 elsewhere, corresponding to the perturbation of the i^{th} weight only. This reflects the influence of allowing the i^{th} subject to drop out informatively, while the others can only drop out at random. It immediately follows from (4.14) that the corresponding local influence measure C_i is given by

$$C_i \;=\; 2 \;\Big|\; \boldsymbol{\Delta}_i'\, \ddot{L}^{-1}\, \boldsymbol{\Delta}_i \;\Big|\,. \tag{4.15}$$

Another important direction is the direction $\boldsymbol{h}_{\max}$ of maximal normal curvature $C_{\max}$. It shows how to perturb the MAR model to obtain the largest local changes in the likelihood displacement. It is readily seen that $C_{\max}$ is the largest eigenvalue of $-2\, \boldsymbol{\Delta}'\, \ddot{L}^{-1}\, \boldsymbol{\Delta}$, and that $\boldsymbol{h}_{\max}$ is the corresponding eigenvector.

When a subset $\boldsymbol{\gamma}_1$ of $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')'$ is of special interest, a similar approach can be used, replacing the log-likelihood by the profile log-likelihood for $\boldsymbol{\gamma}_1$, and the methods discussed above for the full parameter vector directly carry over. Let $\ddot{L}$ be partitioned as

$$\ddot{L} = \begin{pmatrix} \ddot{L}_{11} & \ddot{L}_{12} \\ \ddot{L}_{21} & \ddot{L}_{22} \end{pmatrix},$$

according to the dimensions of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$. Following Cook (1986), the local influence $C_{\boldsymbol{h}}(\boldsymbol{\gamma}_1)$ for $\boldsymbol{\gamma}_1$ in the direction of the unit vector $\boldsymbol{h}$ is given by

$$C_{\boldsymbol{h}}(\boldsymbol{\gamma}_1) \;=\; 2 \left| \boldsymbol{h}'\, \boldsymbol{\Delta}' \left[\ddot{L}^{-1} \;-\; \begin{pmatrix} 0 & 0 \\ 0 & \ddot{L}_{22}^{-1} \end{pmatrix} \right] \boldsymbol{\Delta} \boldsymbol{h} \right|. \tag{4.16}$$

It can be easily seen that $C_{\boldsymbol{h}}(\boldsymbol{\gamma}_1) \leq C_{\boldsymbol{h}}$ meaning that the normal curvature for $\boldsymbol{\gamma}_1$, in any direction $\boldsymbol{h}$, can never be larger than the normal curvature for the total vector $\boldsymbol{\gamma}$ in the same direction. Further, it immediately follows from (4.16) that, for $\ddot{L}_{12} = 0$,

$$C_{\boldsymbol{h}} \;=\; C_{\boldsymbol{h}}(\boldsymbol{\gamma}_1) \;+\; C_{\boldsymbol{h}}(\boldsymbol{\gamma}_2),$$

showing that, if $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are asymptotically independent, the normal curvature for γ is the sum of the normal curvatures for each of its components, which is in general not the case otherwise.

Finally, as for $C_{\boldsymbol{h}}(\gamma)$, there are many possible choices for the vector $\boldsymbol{h}$ in $C_{\boldsymbol{h}}(\gamma_1)$. For example, $C_i(\gamma_1)$, corresponding to $\boldsymbol{h} = \boldsymbol{h}_i$, defined above, expresses the local influence of allowing the i^{th} subject to drop out informatively, on the estimation of γ_1.

We will now in full detail derive the expressions for the local influence measurements. Here fore we consider complete and incomplete sequences in turn. The log-likelihood contribution for a complete sequence is

$$\ell_{i\omega} = \ln f(\boldsymbol{y_i}) + \sum_{j=2}^{n_i} \ln[1 - g(\boldsymbol{h}_{ij}, y_{ij})],$$

where the parameter dependencies are suppressed for notational ease. The density $f(\boldsymbol{y_i})$ is multivariate normal, described by (4.5) and the distribution assumptions of the random terms involved.

The mixed derivatives are particularly easy to calculate:

$$\begin{aligned} \frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i} &= 0, \\ \frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i} &= -\sum_{j=2}^{n_i} \boldsymbol{h}_{ij} y_{ij} g(\boldsymbol{h}_{ij}, y_{ij})[1 - g(\boldsymbol{h}_{ij}, y_{ij})]. \end{aligned}$$

Evaluating those under $\omega_i = 0$ merely results in replacing $g(\boldsymbol{h}_{ij}, y_{ij})$ by $g(\boldsymbol{h}_{ij}) = g(\boldsymbol{h}_{ij}, y_{ij})|_{\omega_i=0}$, which is the MAR version of the dropout model.

The contribution from incomplete sequences is more complicated. Its log-likelihood contribution is

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

of which the first component depends on $\boldsymbol{\theta}$ only, the second one on $\boldsymbol{\psi}$ only, while the third one contains both.

The mixed derivatives of the log-likelihood w.r.t. ω_i can be written as

$$\begin{aligned}\frac{\partial^2 \ell_{i\omega}}{\partial\boldsymbol{\theta}\partial\omega_i} &= \frac{\int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id},y_{id})dy_{id}\int \frac{\partial f(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\theta}}\frac{\partial g(\boldsymbol{h}_{id},y_{id})}{\partial\omega_i}dy_{id}}{\left[\int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id},y_{id})dy_{id}\right]^2} \\ &\quad - \frac{\int f(y_{id}|\boldsymbol{h}_{id})\frac{\partial g(\boldsymbol{h}_{id},y_{id})}{\partial\omega_i}dy_{id}\int \frac{\partial f(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\theta}}g(\boldsymbol{h}_{id},y_{id})dy_{id}}{\left[\int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id},y_{id})dy_{id}\right]^2}, \quad (4.17)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell_{i\omega}}{\partial\boldsymbol{\psi}\partial\omega_i} &= -\sum_{j=2}^{d-1}\boldsymbol{h}_{ij}y_{ij}g(\boldsymbol{h}_{ij},y_{ij})[1-g(\boldsymbol{h}_{ij},y_{ij})] \\ &\quad + \frac{\int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id},y_{id})dy_{id}\int f(y_{id}|\boldsymbol{h}_{id})\frac{\partial^2 g(\boldsymbol{h}_{id},y_{id})}{\partial\boldsymbol{\psi}\partial\omega_i}dy_{id}}{\left[\int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id},y_{id})dy_{id}\right]^2} \\ &\quad - \frac{\int f(y_{id}|\boldsymbol{h}_{id})\frac{\partial g(\boldsymbol{h}_{id},y_{id})}{\partial\omega_i}dy_{id}\int f(y_{id}|\boldsymbol{h}_{id})\frac{\partial g(\boldsymbol{h}_{id},y_{id})}{\partial\boldsymbol{\psi}}dy_{id}}{\left[\int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id},y_{id})dy_{id}\right]^2}. \quad (4.18)\end{aligned}$$

In order to evaluate these expressions under $\omega_i = 0$, we set $\omega_i = 0$ in the integrands and calculate the resulting simplified integrals.

$$\begin{aligned}\int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id},y_{id})dy_{id}\Big|_{\omega_i=0} &= \int f(y_{id}|\boldsymbol{h}_{id})g(\boldsymbol{h}_{id})dy_{id} \\ &= g(\boldsymbol{h}_{id}), \\ \int f(y_{id}|\boldsymbol{h}_{id})\frac{\partial g(\boldsymbol{h}_{id},y_{id})}{\partial\omega_i}dy_{id}\Big|_{\omega_i=0} &= g(\boldsymbol{h}_{id})[1-g(\boldsymbol{h}_{id})]\int y_{id}f(y_{id}|\boldsymbol{h}_{id})dy_{id} \\ &= g(\boldsymbol{h}_{id})[1-g(\boldsymbol{h}_{id})]\lambda(y_{id}|\boldsymbol{h}_{id}), \\ \int \frac{\partial f(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\theta}}g(\boldsymbol{h}_{id},y_{id})dy_{id}\Big|_{\omega_i=0} &= g(\boldsymbol{h}_{id})\int \frac{\partial f(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\theta}}dy_{id} \\ &= 0, \end{aligned} \quad (4.19)$$

$$\begin{aligned}
\int \frac{\partial f(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\theta}} \frac{\partial g(\boldsymbol{h}_{id}, y_{id})}{\partial \omega_i} dy_{id}\Big|_{\omega_i=0} &= g(\boldsymbol{h}_{id})[1-g(\boldsymbol{h}_{id})] \int y_{id} \frac{\partial f(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\theta}} dy_{id} \\
&= g(\boldsymbol{h}_{id})[1-g(\boldsymbol{h}_{id})] \frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\theta}}, \\
\int f(y_{id}|\boldsymbol{h}_{id}) \frac{\partial g(\boldsymbol{h}_{id}, y_{id})}{\partial \boldsymbol{\psi}} dy_{id}\Big|_{\omega_i=0} &= g(\boldsymbol{h}_{id})[1-g(\boldsymbol{h}_{id})]\boldsymbol{h}_{id}, \\
\int f(y_{id}|\boldsymbol{h}_{id}) \frac{\partial^2 g(\boldsymbol{h}_{id}, y_{id})}{\partial \boldsymbol{\psi} \partial \omega_i} dy_{id}\Big|_{\omega_i=0} &= g(\boldsymbol{h}_{id})[1-g(\boldsymbol{h}_{id})][1-2g(\boldsymbol{h}_{id})]\boldsymbol{h}_{id}\lambda(y_{id}|\boldsymbol{h}_{id}).
\end{aligned} \tag{4.20}$$

Combining (4.20) with (4.17)–(4.18) yields

$$\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i}\Big|_{\omega_i=0} = [1-g(\boldsymbol{h}_{id})] \frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\theta}}, \tag{4.21}$$

$$\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i}\Big|_{\omega_i=0} = -\sum_{j=2}^{d-1} \boldsymbol{h}_{ij} y_{ij} g(\boldsymbol{h}_{ij})[1-g(\boldsymbol{h}_{ij})] - \boldsymbol{h}_{id}\lambda(y_{id}|\boldsymbol{h}_{id}) g(\boldsymbol{h}_{id})[1-g(\boldsymbol{h}_{id})] \tag{4.22}$$

Let $V_{i,11}$ be the predicted covariance matrix for the observed vector $(y_{i1}, \ldots, y_{i,d-1})'$, $V_{i,22}$ is the predicted variance for the missing observation y_{id}, and $V_{i,12}$ is the vector of predicted covariances between the elements of the observed vector and the missing observation. It then follows from the linear mixed model (4.5) that the conditional expectation for the observation at dropout, given the history, equals

$$\lambda(y_{id}|\boldsymbol{h}_{id}) = \lambda(y_{id}) + V_{i,21} V_{i,11}^{-1} [\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]. \tag{4.23}$$

The derivatives of (4.23) w.r.t. the measurement model parameters are

$$\begin{aligned}
\frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\beta}} &= \boldsymbol{x}_{id} - V_{i,21} V_{i,11}^{-1} X_{i,(d-1)}, \\
\frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \alpha} &= \left[\frac{\partial V_{i,21}}{\partial \alpha} - V_{i,21} V_{i,11}^{-1} \frac{\partial V_{i,11}}{\partial \alpha} \right] V_{i,11}^{-1} [\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]
\end{aligned} \tag{4.24}$$

where $X_{i,(d-1)}$ indicates the first $(d-1)$ rows of the fixed-effect design for subject i, and $\boldsymbol{\alpha}$ indicates the subvector of covariance parameters within the vector $\boldsymbol{\theta}$.

The influence on the measurement model parameters only arises from those measurement occasions at which dropout occurs. This implies first that complete sequences cannot be influential and secondly that incomplete sequences only contribute at the actual dropout time. It is therefore interesting to compare two incomplete

sequences, with equal history, which drop out at the same time point. They then have the same contribution $1 - g(h_{id})$ to (4.21). Hence different influences on $\boldsymbol{\theta}$ can be ascribed to differences for the second factor of (4.21). For the fixed effects, we have that

$$\frac{\partial \lambda(y_{id}|h_{id})}{\partial \boldsymbol{\beta}} - \frac{\partial \lambda(y_{jd}|h_{jd})}{\partial \boldsymbol{\beta}} \quad = \quad x_{id} - x_{jd} - (V_{i,21} - V_{j,21}) V_{i,11}^{-1} X_{i,(d-1)}.$$

Hence, if the estimated covariance matrix for the complete data is the same for both sequences, the above expression reduces to $x_{id} - x_{jd}$ indicating that differences with respect to $C_i(\boldsymbol{\theta})$ can be entirely ascribed to differences in time-varying covariates for the mean structure. For the variance components, we get

$$\begin{aligned} \frac{\partial \lambda(y_{id}|h_{id})}{\partial \boldsymbol{\alpha}} - \frac{\partial \lambda(y_{jd}|h_{jd})}{\partial \boldsymbol{\alpha}} \quad &= \quad \left(\frac{\partial V_{i,21}}{\partial \boldsymbol{\alpha}} - \frac{\partial V_{j,21}}{\partial \boldsymbol{\alpha}} \right) V_{i,11}^{-1} (h_{id} - \lambda(h_{id})) \\ &\quad - (V_{i,21} - V_{j,21})\, V_{i,11}^{-1} \frac{\partial V_{i,11}}{\partial \boldsymbol{\alpha}} V_{i,11}^{-1} (h_{id} - \lambda(h_{id})), \end{aligned}$$

which equals zero when the estimated covariance matrix is the same for both complete sequences. Note that

$$V_{i,21} - V_{j,21} \quad = \quad (Z_{i2} - Z_{j2}) D Z_{i1}'$$

and that

$$\frac{\partial V_{i,21}}{\partial \boldsymbol{\alpha}} - \frac{\partial V_{j,21}}{\partial \boldsymbol{\alpha}} \quad = \quad (Z_{i2} - Z_{j2}) \frac{\partial D}{\partial \boldsymbol{\alpha}} Z_{i1}'$$

illustrating that

$$\frac{\partial \lambda(y_{id}|h_{id})}{\partial \boldsymbol{\alpha}} - \frac{\partial \lambda(y_{jd}|h_{jd})}{\partial \boldsymbol{\alpha}},$$

represents the difference in random-effects covariates.

4.2.3 Compound Symmetry

Upto this point the derivations are generally applicable and have lead to formal expressions. It is clear that depending on the chosen model structure these expressions can be simplified. Throughout the examples later in this Chapter different variance covariance structures are considered each using their own formulations. In this section we will only derive the full expressions for one special but important enlightening case being the random-intercept or compound symmetry model. In case other structures are requested the derivations are straightforward.

For the special case of compound symmetry, it arises from assuming that the only random coefficient in model (4.1) is a random intercept, i.e., $Z_i = \mathbf{1}_{n_i}$, a vector of ones, and $\boldsymbol{b_i}$ is scalar. Hence D reduces to τ^2. Assuming further that $\Sigma_i = \sigma^2 I_{n_i}$ the covariance matrix becomes $V_i = \sigma^2 I_{n_i} + \tau^2 J_{n_i}$, where J_{n_i} is an $(n_i \times n_i)$ matrix of ones.

It follows immediately that (4.23) and (4.24) reduce to

$$\begin{aligned}
\lambda(y_{id}|\boldsymbol{h}_{id}) &= \lambda(y_{id}) + \frac{\tau^2}{\sigma^2 + (d-1)\tau^2}\mathbf{1}_{d-1}[\boldsymbol{h}_{id}\lambda(\boldsymbol{h}_{id})], \\
\frac{\partial\lambda(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\beta}} &= \boldsymbol{x}_{id} - \frac{\tau^2}{\sigma^2 + (d-1)\tau^2}\mathbf{1}_{d-1}\boldsymbol{X}_{i,(d-1)}, \\
\frac{\partial\lambda(y_{id}|\boldsymbol{h}_{id})}{\partial\sigma^2} &= -\frac{\tau^2}{\sigma^2 + (d-1)\tau^2}\frac{\sigma^2}{\sigma^2 + (d-1)\tau^2}\mathbf{1}_{d-1}[\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})], \\
\frac{\partial\lambda(y_{id}|\boldsymbol{h}_{id})}{\partial\tau^2} &= \frac{1}{\sigma^2 + (d-1)\tau^2}\frac{\sigma^2}{\sigma^2 + (d-1)\tau^2}\mathbf{1}_{d-1}[\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})].
\end{aligned}$$

Instructive special cases arise by setting either $\sigma^2 = 0$ (no measurement error) or $\tau^2 = 0$ (no within-individual correlation). In the first case, we obtain

$$\begin{aligned}
\lambda(y_{id}|\boldsymbol{h}_{id}) &= \lambda(y_{id}) + \frac{1}{d-1}\mathbf{1}_{d-1}[\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})], \\
\frac{\partial\lambda(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\beta}} &= \boldsymbol{x}_{id} - \frac{1}{d-1}\mathbf{1}_{d-1}\boldsymbol{X}_{i,(d-1)},
\end{aligned}$$

while the other derivatives are equal to zero. In the uncorrelated case the second term drops in each of both expressions, while in addition the derivative w.r.t. τ^2 is non-zero. Exploring the derivative w.r.t. $\boldsymbol{\beta}$, it is seen that the case without measurement error produces the *within-series residual covariate* at time d, while the uncorrelated case produces the uncorrected covariate at time d.

Let us now derive an expression for C_i as given by (4.15). Using (4.21), we obtain:

$$C_i = 2[1 - g(\boldsymbol{h}_{id})]^2 \frac{\partial\lambda(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\theta}}' \ddot{L}^{-1} \frac{\partial\lambda(y_{id}|\boldsymbol{h}_{id})}{\partial\boldsymbol{\theta}}. \tag{4.25}$$

The first factor is large for a small dropout probability *at the time of dropout*, in other words for an unlikely event. This is intuitively appealing, since g_{id} then has potential of being improved by including dependence on y_{id}. For such a subject, "informativeness" would help.

The second factor of (4.25) involves $\ddot{L}^{-1}$ and is therefore harder to study. This is obvious, since then all measurements y_{ij} on a subject are equal, whence including y_{id} into the dropout model only adds a redundant covariate.

In the general compound symmetry setting, we can still make progress if we are prepared to make some approximations. The off-diagonal block of observed information matrix $\ddot{L}$ pertaining to the mixed derivatives w.r.t. $\boldsymbol{\beta}$ and α is not equal to zero and while the corresponding block of the expected information matrix is for a complete data problem, it is not so for an incomplete data set, unless the missing data are MCAR (Kenward and Molenberghs 1996). However, these authors also argue that in many practical settings the difference might be small. Therefore, we will assume that $\ddot{L}$ is block-diagonal, and assume that $C_i(\boldsymbol{\theta}) \simeq C_i^{\text{ap}}(\boldsymbol{\beta}) + C_i^{\text{ap}}(\sigma^2, \tau^2)$.

Let us consider $C_i(\boldsymbol{\beta})$ first. With some algebra we arrive at

$$\frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \boldsymbol{\beta}} = \xi_{id}\boldsymbol{x}_{id} + (1-\xi_{id})\boldsymbol{\rho}_{id},$$

with

$$\begin{aligned} \xi_{id} &= \frac{\sigma^2}{\sigma^2 + (d-1)\tau^2}, \\ \boldsymbol{\rho}_{id} &= \boldsymbol{x}_{id} - \frac{1}{d-1}\mathbf{1}X_{i(d-1)}. \end{aligned}$$

The matrix of second derivatives $\ddot{L}^{-1}(\boldsymbol{\beta})$ can generally be expressed as

$$\ddot{L}^{-1}(\boldsymbol{\beta}) = \sum_{i=1}^{N} X_{i(d-1)}^{'} V_{i,11}^{-1} X_{i(d-1)},$$

and since for compound symmetry,

$$V_{i,11}^{-1} = I_{d-1} + \frac{\tau^2}{\sigma^2 + (d-1)\tau^2} J_{d-1},$$

some straightforward algebra results in the following approximation to $C_i(\boldsymbol{\beta})$:

$$\begin{aligned} C_i^{\text{ap}}(\boldsymbol{\beta}) &= 2[1-g(\boldsymbol{h}_{id})]^2(\xi_{id}\boldsymbol{x}_{id} + (1-\xi_{id})\boldsymbol{\rho}_{id})' \\ &\quad \times\sigma^2\left[\sum_{i=1}^{N}\left(\xi_{id}X_{i(d-1)}^{'}X_{i(d-1)} + (1-\xi_{id})R_{i(d-1)}^{'}R_{i(d-1)}\right)\right]^{-1} \\ &\quad \times(\xi_{id}\boldsymbol{x}_{id} + (1-\xi_{id})\boldsymbol{\rho}_{id}), \end{aligned} \tag{4.26}$$

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

Expression (4.26) is the product of the factor which purely depends on the dropout probability, and a factor which has the structure of a leverage. When $\xi_{id} = 1$ for all

individuals, we have a classical leverage where each *measurement* is an independent contribution. When $\xi_{id} = 0$, each subject presents a single independent contribution. The general case is a weighted combination of the between- and within-individual contribution. These arguments motivate to call the second factor of $C_i^{\rm ap}(\boldsymbol{\beta})$ a generalized leverage, not only for compound symmetry, but also for general covariance structures.

Let us consider a similar approximation for the variance components (σ^2, τ^2). First note that

$$\begin{aligned}\frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \sigma^2} &= -\xi_{id}(1-\xi_{id})\overline{[\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]},\\ \frac{\partial \lambda(y_{id}|\boldsymbol{h}_{id})}{\partial \sigma^2} &= \frac{1}{\tau^2}\xi_{id}(1-\xi_{id})\overline{[\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]}.\end{aligned}$$

The (2×2) matrix of second derivatives $\ddot{L}^{-1}(\sigma^2, \tau^2)$ can be derived, using standard expressions for the inverse and the determinant of $\sigma^2 I_{d-1} + \tau^2 J_{d-1}$. This yields

$$\begin{aligned}C_i^{\rm ap}(\sigma^2, \tau^2) &= 2[1 - g(\boldsymbol{h}_{id})]^2 \xi_{id}^2 (1-\xi_{id})^2 \overline{[\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]}^2\\ &\quad \times \left(-1, \frac{1}{\tau^2}\right) \ddot{L}^{-1}(\sigma^2, \tau^2) \begin{pmatrix} -1 \\ \frac{1}{\tau^2} \end{pmatrix}, \qquad (4.27)\end{aligned}$$

where

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

It is important to note that, even though $\ddot{L}^{-1}(\sigma^2, \tau^2)$ has a somewhat complicated form, it occurs in (4.27) only through a scalar. Thus, $C_i^{\rm ap}(\sigma^2, \tau^2)$ can in practice be decomposed into three interpretable components:

- The first factor is shared with $C_i^{\rm ap}(\boldsymbol{\beta})$ and has the same interpretation.
- The second factor disappears when either the measurement error variance or the variance of the random intercept is reduced to zero. It is maximal when there is "balance" between both components of variability ($\xi_{id} = 0.5$).
- The third factor is large when the squared average residual of the history at the time of dropout is large.

For the dropout model parameters, there are no approximations involved. We derive in a similar fashion:

$$C_i(\boldsymbol{\psi}) = 2\left(\sum_{j=2}^{d} \boldsymbol{h}_{ij} y_{ij} V_{ij}\right)' \left(\sum_{i=1}^{N}\sum_{j=2}^{d} V_{ij}\boldsymbol{h}'_{ij}\boldsymbol{h}_{ij}\right)^{-1} \left(\sum_{j=2}^{d} \boldsymbol{h}_{ij} y_{ij} V_{ij}\right). \tag{4.28}$$

It is understood that $d = n_i$ for a complete case and that, for dropout, y_{id} has to be replaced with

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

Expression (4.28) bears some resemblance with the hat-matrix diagonal, used for diagnostic purposes in logistic regression. One of the differences is that the contributions from a single individual are summed in the first and third factor of (4.28), even though they contribute independent pieces of information to the logistic regression. This is because each individual is given a single weight ω_i. An even greater resemblance would be obtained by using an alternative weighting scheme which places a different weight on each measurement: ω_{ij}. This scheme does not imply any differences in the influence contributions for the measurement model However, for the dropout parameters, we obtain:

$$C_{ij}(\boldsymbol{\psi}) = 2(V_{ij}y_{ij}^2)\left\{V_{ij}\boldsymbol{h}'_{ij}\left(\sum_{i=1}^{N}\sum_{j=2}^{d} V_{ij}\boldsymbol{h}'_{ij}\boldsymbol{h}_{ij}\right)^{-1}\boldsymbol{h}_{ij}\right\}, \tag{4.29}$$

where the factor in curly braces equals the hat-matrix diagonal. In the case of dropout, the same replacement for y_{id} has to be made. When the length of a measurement sequence is restricted to 2, then (4.29) and (4.28) coincide. Later in this Chapter it will be shown that this type of expressions are easily interpretable and can be helpful in understanding the problems with the dropout mechanism.

4.2.4 Alternative Perturbation Schemes

As mentioned before, the perturbation scheme used has several elegant properties. The perturbation is around the MAR mechanism, which is often deemed a sensible starting point. Extra calculations are limited and free of numerical integration. Influence decomposes in a measurement and dropout part, the first of which is zero in

the case of a complete observation. Finally, if the special case of compound symmetry is assumed, the measurement part can approximately be written in interpretable components for the fixed effect and variance component parts.

However, other schemes are worthwhile considering as well. Most of the developments presented here can be adapted to such alternatives, although not all schemes will preserve the remarkable computational convenience. Also, interpretation of the influence expressions in an alternative scheme will require additional work.

Apart from MAR, often MCAR also is considered a useful model. It is then natural to consider departures from the MCAR model, rather than from the MAR model. This would change (4.7) to

$$\begin{aligned} \text{logit}(g(\boldsymbol{h}_{ij}, y_{ij})) &= \text{logit}\left[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)\right] \\ &= \boldsymbol{h}_{ij}\boldsymbol{\psi} + \omega_{i1} y_{i,j-1} + \omega_{i2} y_{ij}, \end{aligned} \tag{4.30}$$

with obvious change in the definition of $\boldsymbol{h}_{ij}$. This way, the perturbation parameter becomes a two-component vector $\boldsymbol{\omega}_i = (\omega_{i1}, \omega_{i2})$. As a result, the ith subject produces a pair (h_{i1}, h_{i2}), which is a normalized vector and hence main interest lies in its direction. Also, $C_{\boldsymbol{h}} = C_i$ is the local influence on $\widehat{\gamma}$ of allowing the ith subject to drop out randomly or non randomly. Figure 4.2 shows the result of this procedure, applied to the mastitis data. Pairs (h_{i1}, h_{i2}) are plotted. The main diagonal corresponds to the size direction, whereas the diagonal represents the purely incremental direction. The circles are used to indicate the minimal and maximal distances to the origin. Finally, squares rather than bullets are used for cows #4, #5, and #66.

Most cows lie in the *size direction*, but it is noticeable that #4, #5, and #66 tend toward the nonrandom direction. Further, no extremely large C_i are seen in this case.

Another extension would result from the observation that the choice of the incremental analysis may, although motivated by substantive insight, seem rather arbitrary. Hence, it would be desirable to have a more automatic, data-driven selection of a direction. One way of doing this is by considering

$$\begin{aligned} \text{logit}(g(\boldsymbol{h}_{ij}, y_{ij})) &= \text{logit}\left[\text{pr}(D_i = j | D_i \geq j, \boldsymbol{y}_i)\right] \\ &= \boldsymbol{h}_{ij}\boldsymbol{\psi} + \omega_i(\sin\theta y_{i,j-1} + \cos\theta y_{ij}). \end{aligned} \tag{4.31}$$

Now, it is possible to apply (4.31) for a selected number of angles θ, to range through a fine grid covering the entire circle, or to consider θ as another influence parameter. In the latter case, θ becomes subject-specific and the pair (ω_i, θ_i) is essentially a reparameterization of the pair $\boldsymbol{\omega}_i = (\omega_{i1}, \omega_{i2})$ in (4.31).

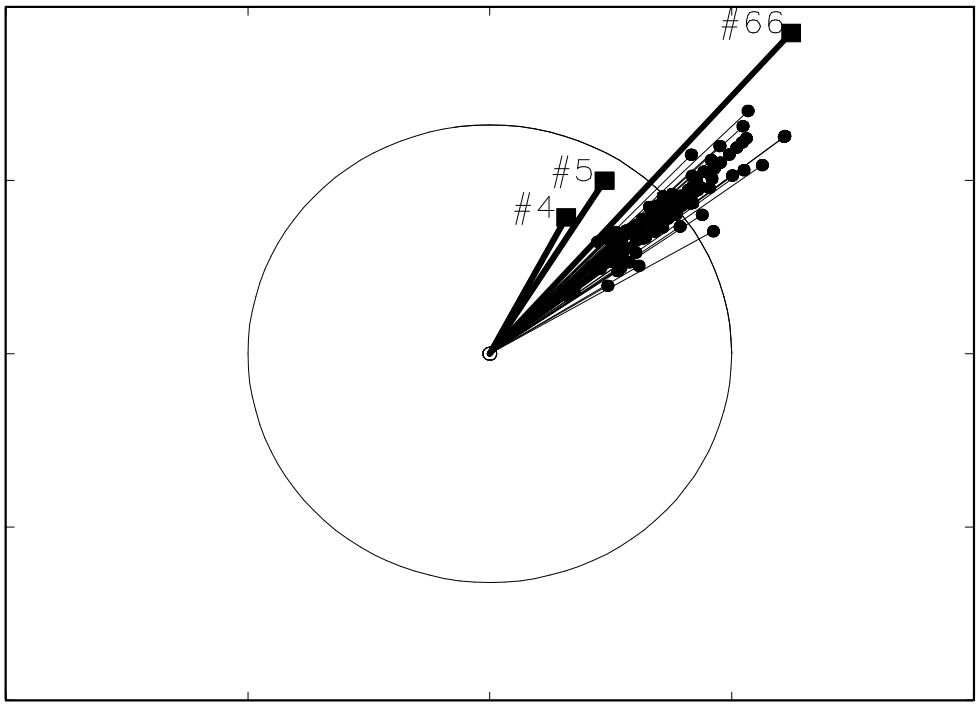

Figure 4.2: *Mastitis dataset, plot of the linear combination approach with influential subjects indicated.*

A completely different local influence approach would modify the general form (3.7) as follows:

$$\begin{aligned} & f(\boldsymbol{y}_i^o, \boldsymbol{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}, \omega_i) \\ &= \int f(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m | X_i, Z_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i | \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, X_i, \boldsymbol{\psi})^{\omega_i} d\boldsymbol{y}_i^m. \end{aligned} \tag{4.32}$$

Now, if $\omega_i = 0$, then the missing data process is considered ignorable and only the measurement process is considered. If $\omega_i = 1$, the posited, potentially nonrandom, model is considered. Other values of ω_i correspond to partial case weighting.

4.3 Application of Influence Tools

In this section we will use the mastitis dataset en de rats dataset as introduced in Chapter 2 to illustrate the methodology of local influence. For an application of the global influence methodology we refer to Chapter 6 where we will discuss the milk

protein trial with respect to a full sensitivity analysis also including a pattern-mixture approach which has to be introduced in the next Chapter.

4.3.1 Mastitis in Dairy Cattle

Informal approach

Diggle and Kenwad (1994) and Kenward (1998) performed several analyzes of these data and we will discuss these results first before considering the results of our sensitivity analysis. In Diggle and Kenward (1994), a separate mean for each group defined by the year of first lactation and a common time effect was considered, together with an unstructured 2×2 covariance matrix. The dropout model included both Y_{i1} and Y_{i2} and was reparameterized in terms of the size variable $(Y_{i1} + Y_{i2})/2$ and the increment $Y_{i2} - Y_{i1}$. It turned out that the increment was important, in contrast with a relatively small contribution of the size. If this model were assumed plausible, MAR would be rejected on the basis of a likelihood ratio test statistic of $G^2 = 5.11$ on 1 degree of freedom.

Kenward (1998) carried out what we could term a data driven sensitivity analysis. Let us describe his results in some detail in this section. He started from the original model in Diggle and Kenward (1994), albeit with a common intercept, since there was no evidence for a dependence on first lactation year. The fits are represented in Table 4.1. Using the likelihoods to compare the fit of the two models, we get a difference $G^2 = 5.11$ ($p = 0.02$). Kenward (1998) found that the corresponding Wald test yields $p = 0.002$ and concluded that this discrepancy might suggest that the asymptotic approximations on which these are based are not very accurate. Nevertheless there is a suggestion from the change in likelihood that ψ_2 is making a real contribution to the fit of the model. The dropout model parameters, estimated from the MNAR setting can be found in Table 4.1. Some insight into this fitted model can be obtained by re-writing it in terms of the milk yield totals $(Y_1 + Y_2)$ and increments $(Y_2 - Y_1)$:

$$\text{logit}[\text{P}(\text{mastitis})] = 0.37 - 0.145(y_1 + y_2) - 2.395(y_2 - y_1). \qquad (4.33)$$

The probability of mastitis (i.e., dropout) increases with larger negative increments, that is, those animals who showed (or would have shown) a greater decrease in yield over the two years have a higher probability of getting mastitis. The other differences in parameter estimates between the two models are consistent with this:

Table 4.1: *Mastitis dataset, maximum likelihood estimates (standard errors) of random and non-random dropout models fitted under several deletion schemes.*

			RANDOM DROPOUT			
Effect	Parameter	all	(53,54,66,69)	(4,5)	(66)	(4,5,66)
Measurement model:						
Intercept	β_0	5.77(0.09)	5.69(0.09)	5.81(0.08)	5.75(0.09)	5.80(0.09)
Time effect	β_d	0.72(0.11)	0.70(0.11)	0.64(0.09)	0.68(0.10)	0.60(0.08)
First variance	σ_1^2	0.87(0.12)	0.76(0.11)	0.77(0.11)	0.86(0.12)	0.76(0.11)
Second variance	σ_2^2	1.30(0.20)	1.08(0.17)	1.30(0.20)	1.10(0.17)	1.09(0.17)
Correlation	ρ	0.58(0.07)	0.45(0.08)	0.72(0.05)	0.57(0.07)	0.73(0.05)
Dropout model:						
Intercept	ψ_0	-2.65(1.45)	-3.69(1.63)	-2.34(1.51)	-2.77(1.47)	-2.48(1.54)
First measurement	ψ_1	0.27(0.25)	0.46(0.28)	0.22(0.25)	0.29(0.24)	0.24(0.26)
Second measurement	$\omega = \psi_2$	0	0	0	0	0
-2 loglikelihood		280.02	246.64	237.94	264.73	220.23
			NON-RANDOM DROPOUT			
Effect	Parameter	all	(53,54,66,69)	(4,5)	(66)	(4,5,66)
Measurement model:						
Intercept	β_0	5.77(0.09)	5.69(0.09)	5.81(0.08)	5.75(0.09)	5.80(0.09)
Time effect	β_d	0.33(0.14)	0.35(0.14)	0.40(0.18)	0.34(0.14)	0.63(0.29)
First variance	σ_1^2	0.87(0.12)	0.76(0.11)	0.77(0.11)	0.86(0.12)	0.76(0.11)
Second variance	σ_2^2	1.61(0.29)	1.29(0.25)	1.39(0.25)	1.34(0.25)	1.10(0.20)
Correlation	ρ	0.48(0.09)	0.42(0.10)	0.67(0.06)	0.48(0.09)	0.73(0.05)
Dropout model:						
Intercept	ψ_0	0.37(2.33)	-0.37(2.65)	-0.77(2.04)	0.45(2.35)	-2.77(3.52)
First measurement	ψ_1	2.25(0.77)	2.11(0.76)	1.61(1.13)	2.06(0.76)	0.07(1.82)
Second measurement	$\omega = \psi_2$	-2.54(0.83)	-2.22(0.86)	-1.66(1.29)	-2.33(0.86)	0.20(2.09)
-2loglikelihood		274.91	243.21	237.86	261.15	220.23
G^2 for NRD		5.11	3.43	0.08	3.57	0.005

the MNAR dropout model predicts a smaller average increment in yield β_d with larger second year variance and smaller correlation caused by greater negative imputed differences between yields.

To gain some additional insight into these two fitted models we now take a closer look at the raw data and the predictive behavior of the Gaussian MNAR model.

Under an MNAR model the predicted, or imputed, value of a missing observation is given by the ratio of expectations:

$$\hat{\boldsymbol{y}}_m = \frac{E_{\boldsymbol{Y}_m|\boldsymbol{Y}_o}\left[\boldsymbol{y}_{\text{mis}} P(\boldsymbol{r} \mid \boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}})\right]}{E_{\boldsymbol{Y}_m|\boldsymbol{Y}_o}\left[P(\boldsymbol{r} \mid \boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}})\right]}. \tag{4.34}$$

Recall that the fitted dropout model (4.33) implies that the probability of mastitis increases with decreasing values of the increment $Y_2 - Y_1$. We therefore plot the 27 imputed values of this quantity together with the 80 observed increments against the first year yield Y_1. This is presented in Figure 4.3, in which the imputed values are indicated with triangles and the observed values with crosses. Note how the imputed values are almost linear in Y_1: this is a well-known property of the ratio (4.34) within this range of observations. The imputed values are all negative, in contrast to the observed increments which are nearly all positive. With animals of this age, one would normally expect an increase in yield between the two years. The dropout model is imposing very atypical behavior on these animals and this corresponds to the statistical significance of the MNAR component of the model (ψ_2) but of course necessitates further scrutiny.

Another feature of this plot is the pair of outlying observed points circled in the top left hand corner. These two animals have the lowest and third lowest yields in the first year, but moderately large yields in the second, leading to the largest positive increments. It is likely that there is some anomaly, possibly illness, leading to their relatively low yields in the first year. One can conjecture that these two animals are the cause of the structure identified by the Gaussian MNAR model. Under the joint Gaussian assumption the MNAR model essentially "fills in" the missing data to produce a complete Gaussian distribution. To counterbalance the effect of these two extreme positive increments, the dropout model predicts negative increments for the mastitic cows, leading to the results observed. As a check on this conjecture these two animals were omitted from the data set and the MAR and MNAR Gaussian models were refitted. The resulting estimates are presented in the $(4, 5)$ column of Table 4.1. This procedure is similar to a global influence analysis by means of deleting two observations. It is clear that the influence on the measurement model parameters is small in the random dropout case, although the gap on the time effect β_d between the random and non-random dropout models is reduced when #4 and #5 are removed. The deviance is minimal and the MNAR model now shows no improvement in fit over MAR. The estimates of the dropout parameters, while still moderately large in an absolute sense, are of the same size as their standard errors. In the absence of the two anomalous animals the structure identified earlier in terms of the MNAR

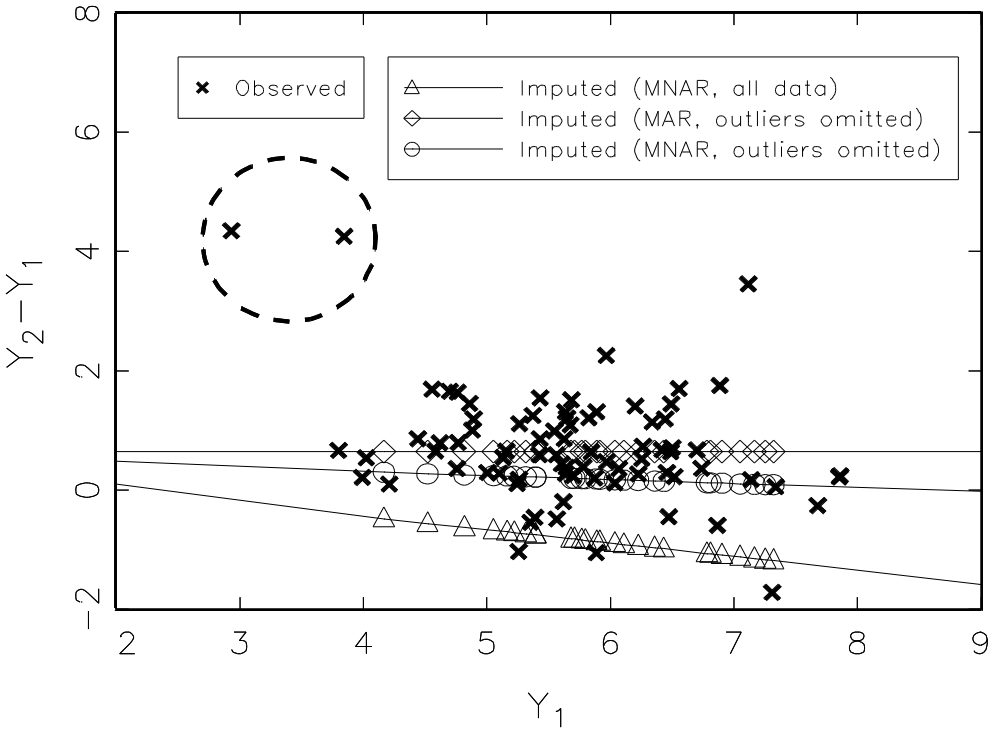

Figure 4.3: *Mastitis dataset, plot of observed and imputed year 2 - year 1 yield differences against year 1 yield. Two outlying points are circled.*

dropout model no longer exists. The increments imputed by the fitted model are also plotted in Figure 4.3, indicated by circles. While still lying among the lower region of the observed increments, these are now all positive and lie close to the increments imputed by the MAR model (diamonds). Thus, we have a plausible representation of the data in terms of joint Gaussian milk yields, two pairs of outlying yields and no requirement for an MNAR dropout process.

The two key assumptions underlying the MNAR model are, first, the form chosen for the relationship between dropout probability and response and, second, the distribution of the response or, more precisely, the conditional distribution of the possibly unobserved response given the observed response. In the current setting for the first assumption, if there is dependence of mastitis occurrence on yield, experience with logistic regression tells us that the exact form of the link function in this relationship is unlikely to be critical. In terms of sensitivity Kenward (1998) therefore considered the second assumption, the distribution of the response.

All the data from the first year are available, and a normal probability plot of these showed no great departures from the Gaussian assumption. Leaving this distribution unchanged, Kenward (1998) examine the effect of changing the conditional distribution of Y_2 given Y_1. One simple and obvious choice is to consider a heavy tailed t_m distribution. For the degrees of freedom the following choices were made: $m = 2, 10, 25$. Not surprisingly, his finding was that the heavier the tails of the t distribution, the better the outliers were accommodated. As a result, the difference between the MAR and MNAR models vanished (e.g., $G^2 = 1.08$ for a t_2 distribution). In addition, Kenward (1998) found that the estimated yearly increment in milk yield β_d from the MNAR model, increases to the value estimated under the MAR model.

The results observed here are consistent with those from the deletion analysis. The two outlying pairs of measurements identified earlier are not inconsistent with the heavy tailed t distribution so would require no "filling in" and hence no evidence for non-randomness in the dropout process under the second model.

It should be clear that interpreting a single non-random model, fitted to an incomplete set of data, should be avoided. Rather, a careful sensitivity assessment should supplement fitting of such models.

Local Influence Approach

While in the previous section we reported on the work done by Diggle and Kenward (1994) and Kenward (1998) we will now come to the applications of our methodology to the mastitis data. We performed a local influence analysis of these data.

Using the local influence methodology we are able to calculate influence measures C_i which are represented in Figure 4.4 in several ways and spit per subvector of the parameter vector γ. Considering this figure one notices that there are four influential subjects: #53, #54, #66 and #69, while #4 and #5 are *not* recovered. It is interesting to consider an analysis with these four cows removed. Unlike removing #4 and #5, the influence on the likelihood ratio test is rather small: $G^2 = 3.43$ instead of the original 5.11. The influence on the measurement model parameters under both random and non-random dropout is small.

It is very important to realize that one should not expect agreement between deletion and our local influence analysis. The latter focuses on the sensitivity of the results with respect to the assumed dropout model, more specifically how the results

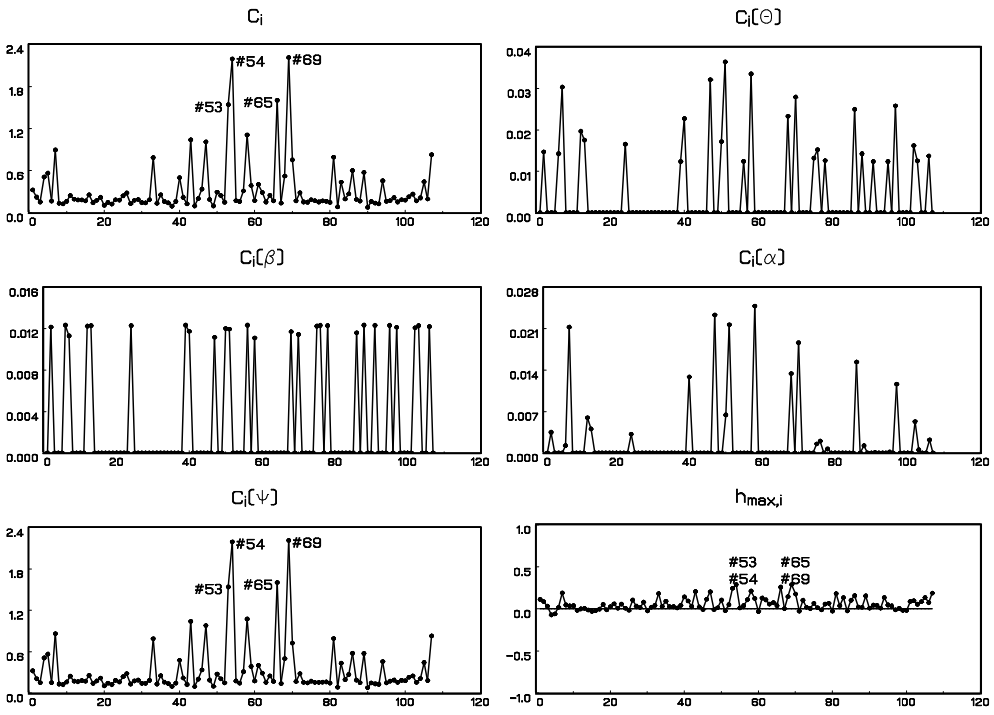

Figure 4.4: *Mastitis dataset, index plots of C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\psi})$ and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature, when the dropout model is parameterized in function of Y_{i1} and Y_{i2}.*

change when the MAR model is extended into the direction of non-random dropout. In particular, all subjects singled out so far are complete and hence $C_i(\boldsymbol{\theta}) \equiv 0$, placing all influence on $C_i(\boldsymbol{\psi})$ and $\boldsymbol{h}_{\max,i}$. This is confirmed from the $C_i(\boldsymbol{\beta})$ and $C_i(\boldsymbol{\alpha})$ panels where, certainly when the scale is compared to the one of $C_i(\boldsymbol{\psi})$, little or no influence is detected. Of course, a legitimate concern is precisely where one should place a cut-off between subjects that are influential and those that are not. Clearly, additional work studying the stochastic behavior of the influence measures would be helpful. In addition, informal guidelines can be used, such as studying 5% of the relatively most influential subjects.

More insight can also be obtained from studying (4.29). The contribution for subject i is made up of three factors. The first factor, V_i, is small for extreme dropout probabilities. The subjects having a very high probability to either remain in the study or disappear will be less influential. Cows #4 and #5 have dropout probabilities equal to 0.13 and 0.17 respectively. The 107 cows in the study span the dropout

probability interval $[0.13, 0.37]$. Thus, this component rather deflates the influence of subjects #4 and #5. Secondly, (4.29) contains a leverage factor in curly braces. Thirdly, a subject is relatively more influential when both milk yields are high. We now need to question whether this is plausible or relevant. Since both measurements are positively correlated, measurements with both milk yields high or low will not be unusual. Kenward (1998) observed that cows #4 and #5 are unusual on the basis of their *increment*. This is in line with several other applications of similar dropout models (Diggle and Kenward 1994, Molenberghs, Kenward and Lesaffre 1997) where it was found that a strong incremental component apparently yields a non-random dropout model. From our analysis, it is clear that such a conclusion may indeed only be apparent, since removing #4 and #5 leads to disappearance of the non-random component. In contrast, the size variable can often be replaced by just the history, and hence the corresponding model is very close to random dropout.

Even though a dropout model in the outcomes themselves, termed direct variables model, is equivalent to a model in the first variable Y_{i1} and the increment $Y_{i2} - Y_{i1}$, termed incremental variable representation, we will show that they lead to different perturbation schemes of the form (4.13). At first, this feature can be seen as both an advantage and a disadvantage. The fact that reparameterizations of the linear predictor of the dropout model leads to different perturbation schemes requires careful reflection based on substantive knowledge in order to guide the analysis, such as the considerations on the incremental variable made earlier.

We will present the results of the incremental analysis and then offer further comments on the rationale behind this particular transformation. From the diagnostic plots in Figure 4.5 it is obvious that we recover three influential subjects: #4, #5, and #66. While Kenward (1998) did not consider #66 to be influential, it appears to be somewhat distant from the bulk of the data. The main difference between both types is that the first two were likely sick during year 1, while this is not necessarily so for #66. An additional feature is that in all cases both $C_i(\boldsymbol{\psi})$ as well as $\boldsymbol{h}_{\max}$ show the same influential animals. In addition, $\boldsymbol{h}_{\max}$ suggests that the influence for #66 is different than for the others. It could be conjectured that the latter one pulls the coefficient ω in a different direction than the other two. The other values are all relatively small. This could indicate that for the remaining 104 subjects, MAR is plausible, while a deviation in the direction of the incremental variable, *with differing signs*, appears to be necessary for the other three subjects. At this point, a comparison between $\boldsymbol{h}_{\max}$ for the direct variable and incremental analyzes is useful. Since the

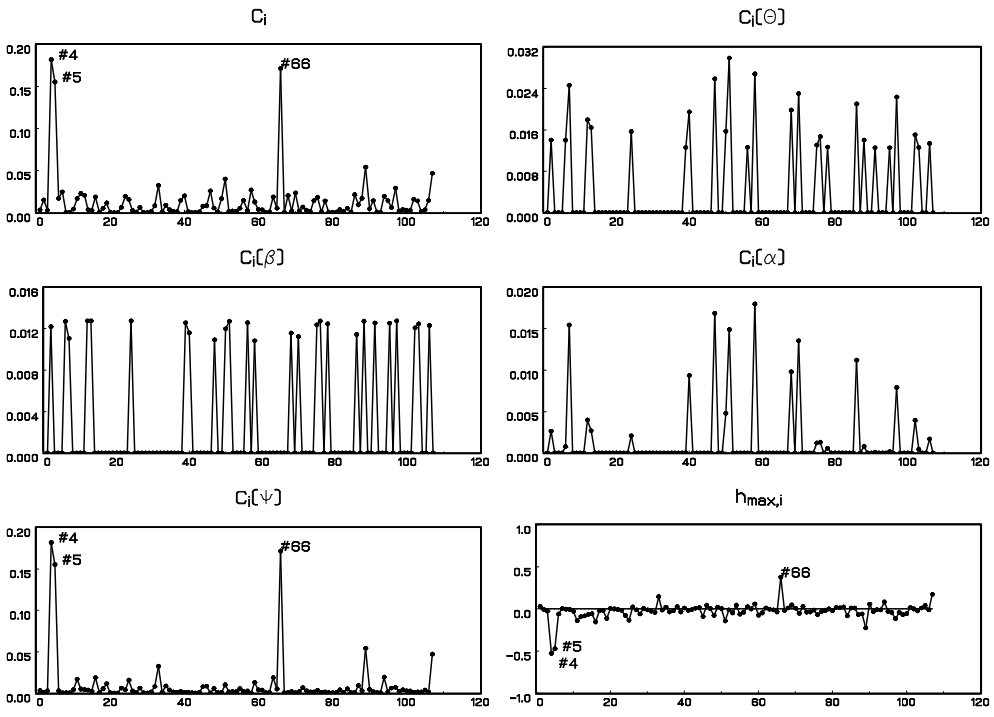

Figure 4.5: *Mastitis dataset, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\psi})$ *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, when the dropout model is parameterized in function of* Y_{i1} *and* $Y_{i2} - Y_{i1}$.

contributions h_i sum to one, these two plots are directly comparable. There is no pronounced influence indication in the direct variables case and perhaps only random noise is seen. A more formal way to distinguish between signal and noise in such plots is the subject of ongoing research.

In Figure 4.6, we have decomposed (4.29) in its three components: the variance of the dropout probability V_i, the incremental variable $Y_{i2} - Y_{i1}$, which is replaced by its predicted value for a dropout, and the hat-matrix diagonal. In agreement with the preceding discussion, the influence clearly stems from an unusually large increment, which survives the fact that V_i actually downplays the influence because Y_{41} and Y_{51} are comparatively small and dropout increases with the milk yield in the first year. Further, the sign difference of $h_{\max,4}$ and $h_{\max,5}$ versus $h_{\max,66}$ can be interpreted better.

We noted already that cows #4 and #5 have relatively small dropout probabili-

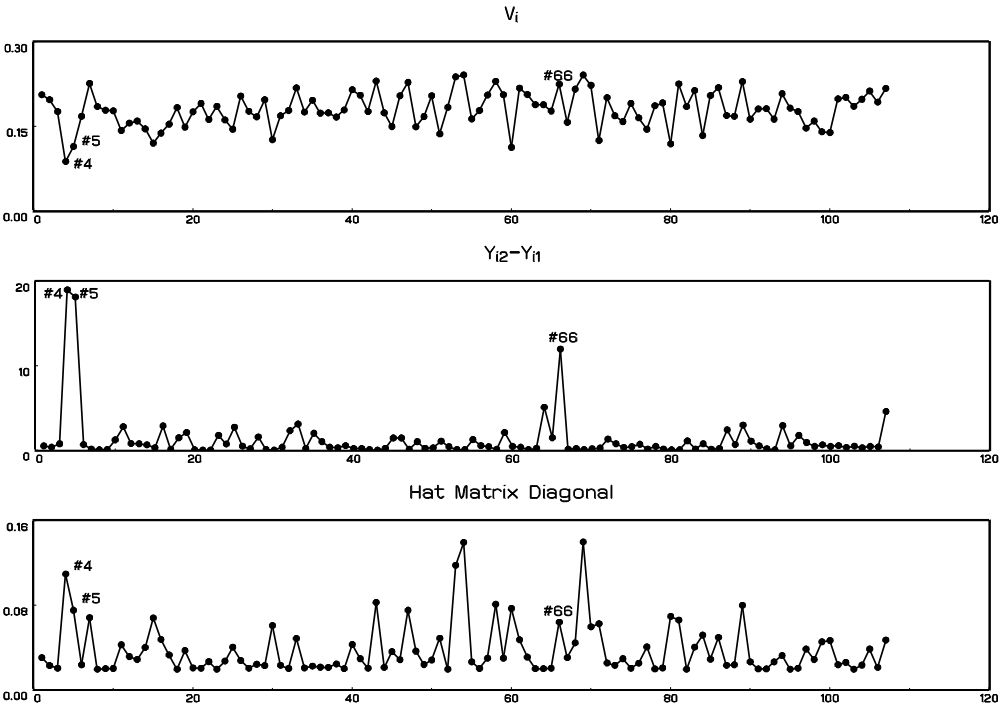

Figure 4.6: *Mastitis dataset, index plots of the three components of $C_i(\psi)$ when the dropout model is parameterized in function of Y_{i1} and $Y_{i2} - Y_{i1}$.*

ties. In contrast, the dropout probability of #66 is large within the observed range $[0.13; 0.37]$. Since for those subjects the increment is large, changing its "parameter" ω_i can have a large impact on the other dropout parameters ψ_0 and ψ_1. In order to avoid that the effects of the change for #4 and #5 will cancel with the effect for #66, the corresponding signs need to be opposite. Such a change implies either that all three dropout probabilities move towards the center of the range or are pulled away from it. (Note that $-\boldsymbol{h}_{\max}$ is another normalized eigenvector corresponding to the largest eigenvalue.)

Kenward (1998) considers extra analyzes with #4 and #5 removed. The resulting likelihood ratio statistic reduces to $G^2 = 0.08$. When only #66 is removed, the likelihood ratio for non-random dropout is $G^2 = 3.57$, very similar to the one when #53, #54, #66 and #69 were removed. Removing all three (#4, #5 and #66) results in $G^2 = 0.005$, i.e., complete disappearance of all evidence for non-random dropout. Details are given in Table 1.

We now provide insight into why the transformation of direct outcomes to incre-

ments is useful. We noted already that the associated perturbation schemes (4.13) are different. An important device in this respect is the equality

$$\psi_0 + \psi_1 Y_{i1} + \psi_2 Y_{i2} = \psi_0 + (\psi_1 + \psi_2) Y_{i1} + \psi_2 (Y_{i2} - Y_{i1}). \tag{4.35}$$

showing that the direct variables model checks the influence on the random dropout parameter ψ_1, whereas the random dropout parameter in the incremental model is $\psi_1 + \psi_2$. Not only is this a different parameter, it is estimated with more precision. One often observes that $\hat{\psi}_1$ and $\hat{\psi}_2$ exhibit a similar variance and negative correlation, in which case the linear combination with smallest variance is approximately in the direction of the sum $\psi_1 + \psi_2$. When the correlation is negative the difference direction $\psi_1 - \psi_2$ is obtained instead. Let us assess this in case all 107 observations are included. The estimated covariance matrix is

$$\begin{pmatrix} 0.59 & -0.54 \\ & 0.70 \end{pmatrix},$$

with correlation -0.84. The variance of $\hat{\psi}_1 + \hat{\psi}_2$ on the other hand is estimated to be 0.21. In this case, the direction of minimal variance is along $(0.74; 0.67)$ which is indeed close to the sum direction. When all three influential subjects are removed, the estimated covariance matrix becomes

$$\begin{pmatrix} 3.31 & -3.77 \\ & 4.37 \end{pmatrix},$$

with correlation -0.9897. Removing only #4 and #5 yields an intermediate situation of which the results are not shown. The variance of the sum is 0.15 which is a further reduction and still close to the direction of minimal variance. These considerations reinforce the claim that an incremental analysis is highly recommended. It might therefore be interesting to routinely construct a plot such as in Figure 1, even with longer measurement sequences. On the other hand, transforming the dropout model to a size variable $(Y_{i1} + Y_{i2})/2$ will worsen the problem since an insensitive parameter for Y_{i1} will result.

Finally, observe that a transformation of the dropout model to a size and incremental variable at the same time for the model with all three influential subjects removed gives a variance of the size and increment variables of 0.15 and 15.22 respectively. In other words, there is no evidence for an incremental effect, confirming that random dropout is plausible.

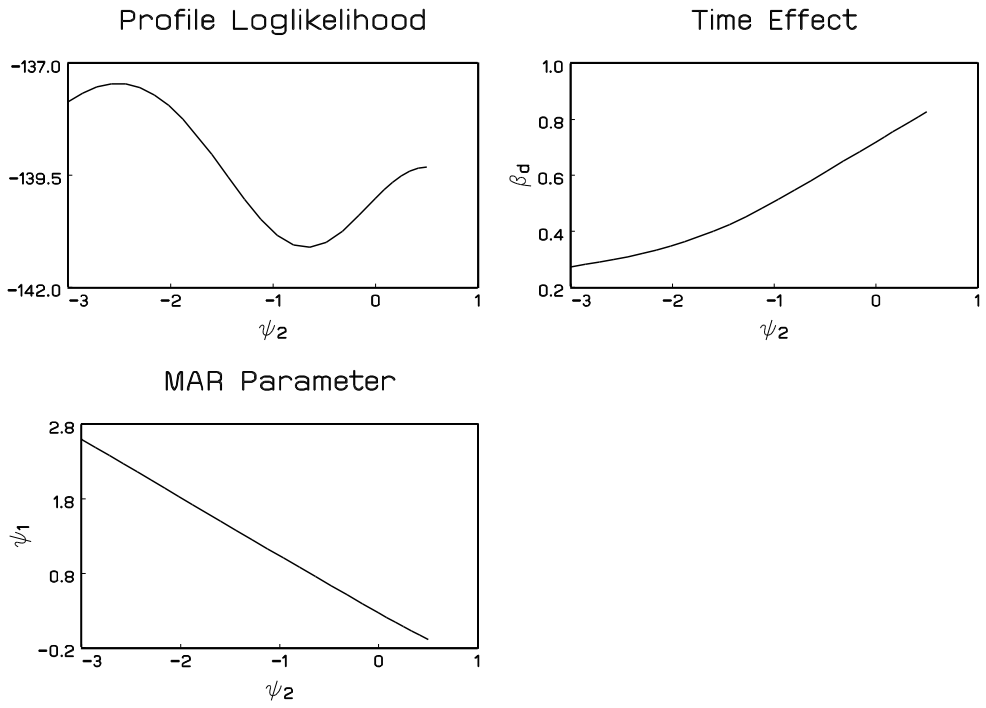

Figure 4.7: *Mastitis dataset, representation of the profile likelihood function in ψ_2 over a range which includes the MAR and MNAR models. In addition, the evolution of β_d and ψ_1 is given.*

Further insight into why the incremental analysis is useful can be found from a representation of the profile likelihood function in ψ_2 (Figure 4.7). The (non-convex) profile likelihood function is supplemented with a representation of the time effect β_d and ψ_1, as functions of ψ_2. In agreement with the considerations above, we observe once again that ψ_1 is almost exactly linear in ψ_2. More precisely, the size direction is constant, implying that its magnitude is nearly invariant to the missing data assumptions.

Although local and global influence are strictly speaking not equivalent, it is insightful to see how the global influence on $\boldsymbol{\theta}$ can be linked to the behavior of $C_i(\boldsymbol{\psi})$. We observed earlier that all locally influential subjects are completers and hence $C_i(\boldsymbol{\theta}) \equiv 0$. Yet, removing #4, #5 and #66 shows some effect on the discrepancy between the random dropout and non-random dropout estimates of the time effect β_d. In particular, random and non-random estimates with all three subjects removed are virtually identical (0.60 and 0.63). Since these subjects are influential in $C_i(\boldsymbol{\psi})$, the model could be improved by including incremental terms for these three subjects.

Such a model would still imply random dropout. In contrast, allowing a dependence on the increment in *all* subjects will influence $E(Y_{i2}|y_{i1}, \text{dropout})$ for all incomplete observations an hence the measurement model parameters under the informative assumption will change. In conclusion, this provides a way to assess the *indirect* influence of the dropout mechanism on the measurement model parameters through local influence methods. In the milk data set, this influence is likely due to the fact that an exceptional increment which is caused by a different mechanism, perhaps a diseased animal during the first year, is nevertheless treated on equal footing with the other observations within the dropout model. Such an analysis not possible with the case-deletion method because it is not possible to disentangle the various sources of influence.

4.3.2 Rats Data

Verbeke and Lesaffre (1999) already investigated the effect of dropout on the efficiency of the study, and derived methods for designing efficient longitudinal experiments, when dropout is to be expected. In this Section however, we will apply the local influence method in order to investigate how sensitive our inferences are with respect to modeling assumptions for the dropout process. The response of interest in this study is one of the parameters which can be used to characterize the height of the skull and the profiles are already shown in Figure 2.2. Notice that these profiles can be linearized by using the logarithmic transformation $t = \ln(1 + (Age - 45)/10))$ for the time scale and this is also the scale we will use in all statistical analyzes from now on. Note that the transformation was chosen such that $t = 0$ corresponds to the start of the treatment. A simple statistical model which can be used to describe these data (Verbeke and Lesaffre 1999) then assumes that y_{ij} satisfies a model of the form (4.1) with common average intercept β_0 for all three groups, with average slopes β_1, β_2 and β_3 for the three treatment groups respectively, and assuming compound symmetry covariance structure, with variance $\sigma^2 + \tau^2$ and covariance τ^2. These models are estimated under MCAR, MAR, and MNAR processes and the estimates are displayed in Table 4.2.

Figure 4.8 displays overall C_i, as well as influences for each of the relevant parameter subvectors. In addition, the direction $\boldsymbol{h}_{\max}$ corresponding to maximal local influence is given. As is clear from, e.g., (4.28), the absolute magnitude of $C_i(.)$ depends upon the scale on which the measurements are expressed, and hence each

Table 4.2: *Rats dataset, maximum likelihood estimates (standard errors) of completely random, random and non-random dropout models, with and without modification.*

Original Data				
Effect	Parameter	MCAR	MAR	MNAR
Measurement model:				
Intercept	β_0	68.61	68.61	68.61
Slope control	β_1	7.51	7.51	7.50
Slope low dose	β_2	6.87	6.87	6.86
Slope high dose	β_3	7.31	7.31	7.30
Compound symmetry	τ^2	3.44	3.44	3.44
Compound symmetry	σ^2	1.43	1.43	1.43
Dropout model:				
Intercept	ψ_0	-1.98	-8.48	-8.05
Prev. measurement	ψ_1		0.084	0.096
Curr. measurement	$\omega = \psi_2$			-0.017
-2 loglikelihood		1777.3	1774.5	1774.5
Modified Data				
Effect	Parameter	MCAR	MAR	MNAR
Measurement model:				
Intercept	β_0	70.20	70.20	70.26
Slope control	β_1	7.52	7.52	7.39
Slope low dose	β_2	6.97	6.97	6.88
Slope high dose	β_3	7.21	7.21	6.98
Compound symmetry	τ^2	40.38	40.38	40.83
Compound symmetry	σ^2	1.42	1.42	1.46
Dropout model:				
Intercept	ψ_0	-2.20	-0.79	3.23
Prev. measurement	ψ_1		-0.015	0.32
Curr. measurement	$\omega = \psi_2$			-0.38
-2 loglikelihood		1906.6	1894.6	1890.2

influence graph should be interpreted in a relative fashion.

The largest C_i are observed for rats #10, #16, #35, and #41, and virtually the same picture holds for $C_i(\psi)$. They are highlighted in Figure 4.9. All four belong to the low dose group. Arguably, their relatively large influence is caused by an interplay

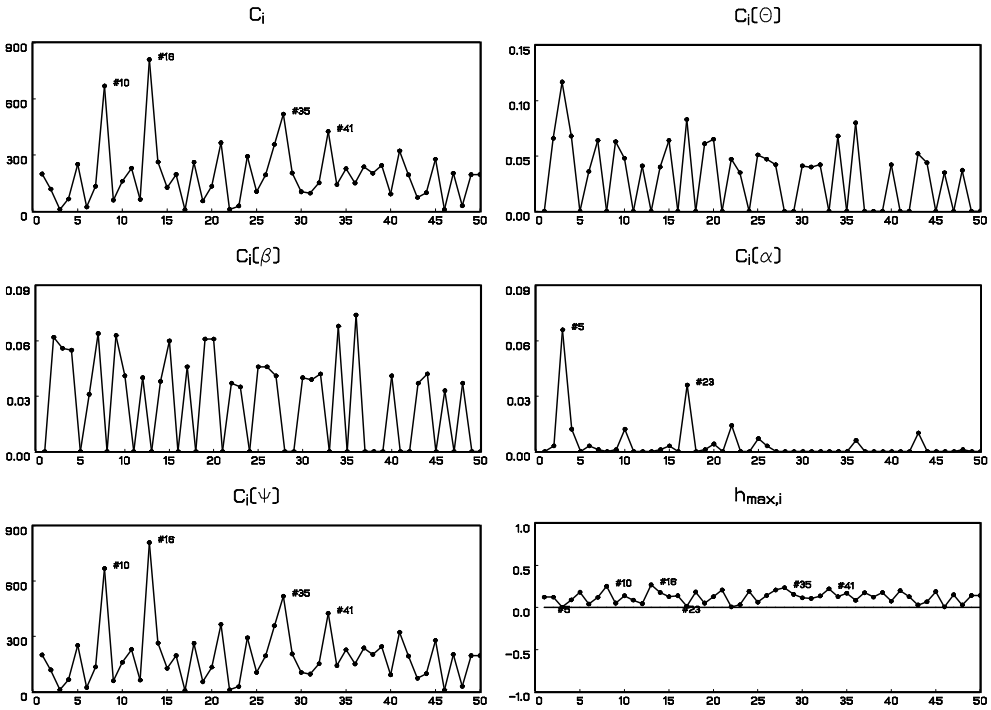

Figure 4.8: *Rats dataset, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, without modification.*

of three facts. First, the profiles are relatively high, and hence y_{ij} and h_{ij} in (4.28) are large. Secondly, since all four profiles are complete, the first factor in (4.28) contains a maximal number of large terms. Thirdly, the computed v_{ij} are relatively large, which is implied by the MAR dropout model parameter estimates in Table 4.2. Indeed, for these measurements the logit of the dropout probability is closest to 0 and hence v_{ij} is fairly close to its maximal value of 0.25.

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 4.9. To explain this, observe that the relative magnitude of $C_i(\boldsymbol{\alpha})$, approximately given by (4.27), 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 4.2). 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

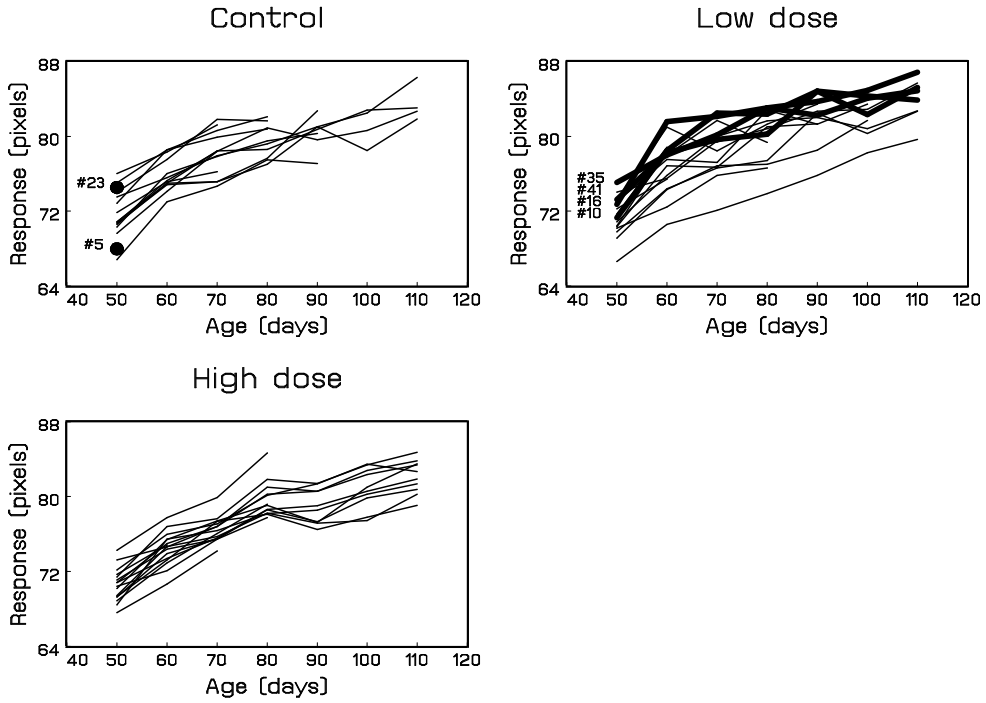

Figure 4.9: *Rats dataset, individual profiles with influential subjects highlighted, without modification.*

group by time mean.

All deviations discussed are fairly moderate. This conclusion is supported by the observation that the components of the normalized vector $\boldsymbol{h}_{\max}$ do not deviate much from $1/\sqrt{N}$, *and* it is consistent with the observation that likelihood ratio statistics for MAR versus MCAR, as well as for MNAR versus MAR, do not reject the null hypothesis.

To further explore the properties of the influence diagnostics, we consider a second analysis where all responses for rats #10, #16, #35, and #41 have been increased with 20 units. A graphical display is given in Figure 4.10. Table 4.2 contains the parameter estimates for all three models. The peaks observed earlier have become much clearer, in line with the observation that the test statistics for MAR versus MCAR, and for MNAR versus MAR, have become significant.

Graphical representations such as Figure 4.10 are sometimes judged misleading since the apparent magnitude of a subject is influenced by its neighbors. On the other hand, it preserves the order across all 6 index plots. One way to overcome

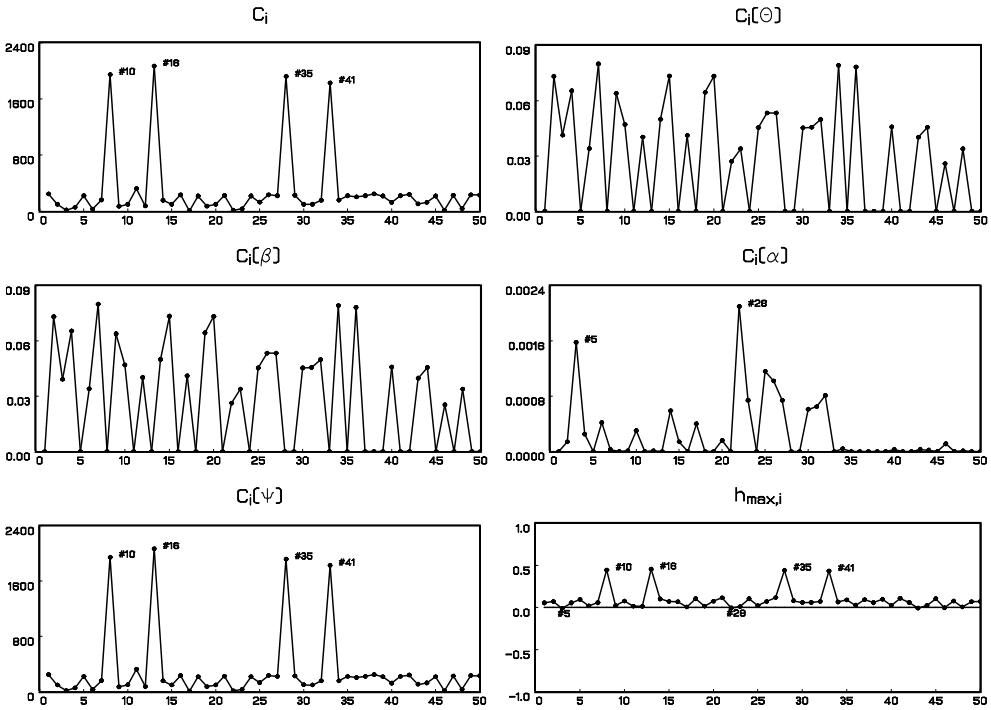

Figure 4.10: *Rats dataset, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, when 4 profiles have been shifted upward.*

this problem is by ordering one plot (e.g., according to C_i), and keeping this order across all six panels. This is done in Figure 4.12. Alternatively, scatter plots of (1) the measurement versus dropout components and (2) fixed-effects versus variance component elements can be used. An example of the latter is presented in Figure 4.13.

4.4 Concluding Remarks

We have applied local influence tools (Cook 1986, Lesaffre and Verbeke 1998) to the selection model for continuous data subject to non-random dropout, as presented in Diggle and Kenward (1994). In particular, it is shown how the impact of small perturbations around the null model of missing at random will affect the measurement model and dropout model parameters. In order to calculate the influence diagnostics it is not necessary to fit a non-random dropout model. All calculations have

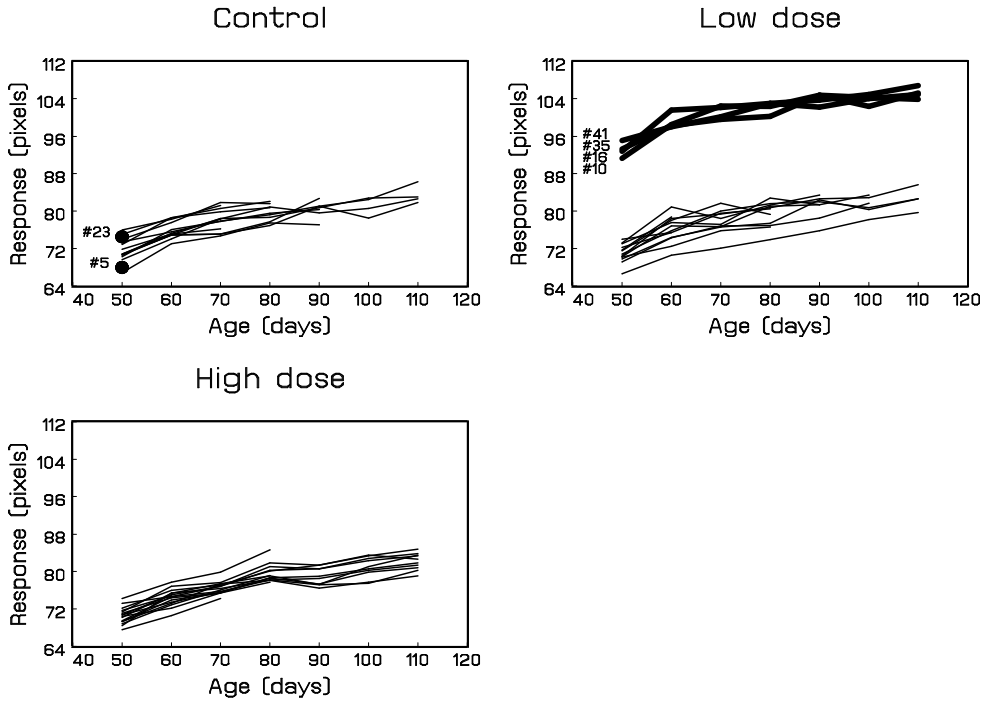

Figure 4.11: *Rats dataset, individual profiles with influential subjects highlighted, when 4 profiles have been shifted upward.*

been carried out in GAUSS and the programs are available but not yet very user friendly. In the special case of a compound-symmetry model, the influence measures are approximately decomposable into interpretable components.

First the method was applied to the mastitis data set, studied in Diggle and Kenward (1994) and Kenward (1998). We have illustrated that an informal sensitivity analysis based on substantive considerations, and a formal approach such as a local influence analysis, can usefully supplement each other. Kenward (1998) found that both removing subjects #4 and #5, as well as replacing the conditional distribution of the second milk yield given the first one, indicated a strong sensitivity of the conclusions about the incremental effect from year 1 to year 2 and the nature of the dropout process. For example, both these actions removed the evidence for MNAR. For the local influence analysis, it was deduced that an incremental variable representation of the dropout mechanism is beneficial over a direct variable representation. Contrasting our local influence approach with a case-deletion scheme as applied in Kenward (1998), we find the same two subjects, with in addition cow #66 being influential. One advantage of local influence is its ability to focus on direct and indirect influ-

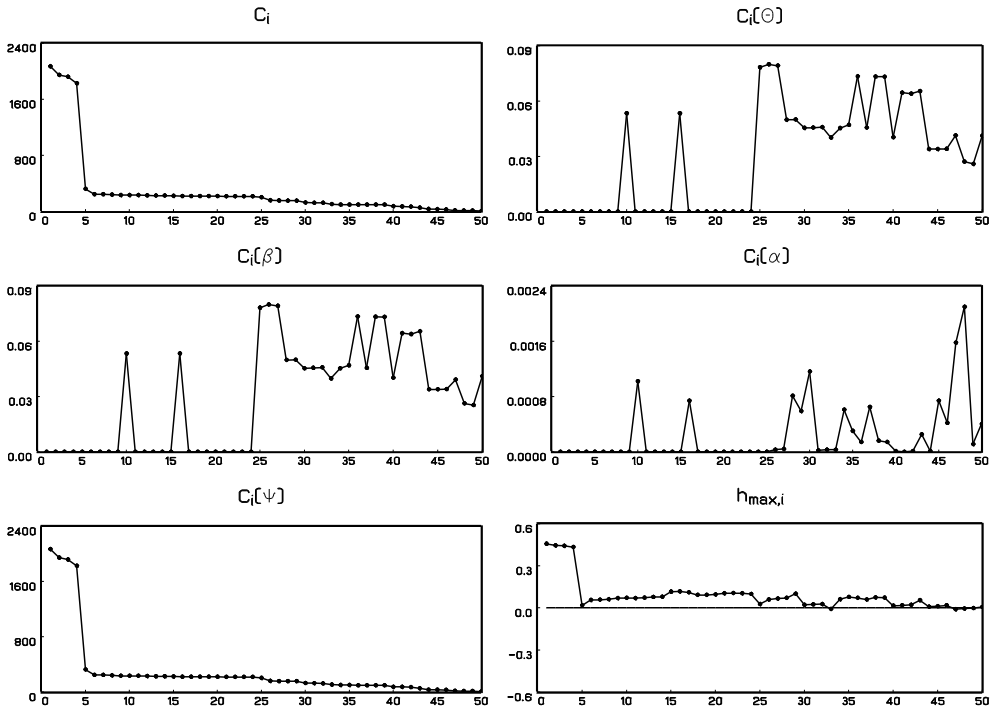

Figure 4.12: *Rats dataset, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, when 4 profiles have been shifted upward and the components have been ordered in decreasing order of* C_i.

ence on the dropout and measurement model parameters, stemming from perturbing the random dropout model in the direction of non-random dropout. In contrast, a case-deletion scheme combines all sources of influence, whether stemming from the dropout mechanism or not.

Second, the analysis of the rats data set (Verdonck et al. 1997) supports the claim that the influence measures are easy to interpret. In addition, studying the conditions under which the diagnostics are large can aid in judging when a model is appropriate. For example, reparameterizing the dropout model in terms of residuals rather than raw measurements will change the conditions under which such terms as g_{ij}, v_{ij}, or $C_i(\boldsymbol{\psi})$ are large. Thus, study of these conditions can help judging the appropriateness of the selection model chosen.

While all of these parameterizations lead to perturbation schemes that are members of the family (4.13), it is clear that other schemes are worthwhile considering as well. However, not all schemes will lead to expressions that are both fairly easy

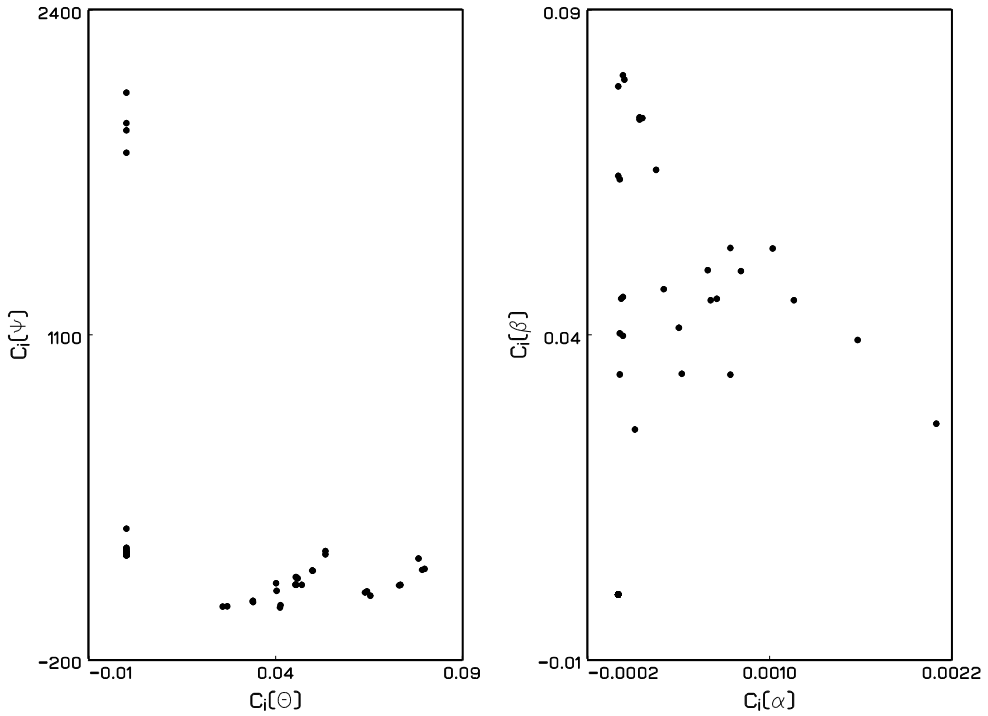

Figure 4.13: *Rats dataset, scatter plots of (1) $C_i(\boldsymbol{\theta})$ versus $C_i(\boldsymbol{\psi})$ and (2) $C_i(\boldsymbol{\beta})$ versus $C_i(\boldsymbol{\alpha})$, when 4 profiles have been shifted upward.*

to interpret and to calculate, at least in special cases such as compound symmetry. Further research in this area is needed. Finally, the ideas outlined in this Chapter are not confined to the selection model of Diggle and Kenward (1994). Currently, work is done to explore this route for categorical responses, and for missingness models of the pattern-mixture type (Little 1993, 1994).

We have studied one single but interesting case, with its specific aspects but also with its limitations. The perturbation scheme chosen here has several elegant properties. The perturbation is around the often considered MAR mechanism. Extra calculations are limited and free of numerical integration. Influence decomposes in a measurement and dropout part, the first of which is zero in the case of a complete observation. Finally, if the special case of compound symmetry is assumed, the measurement part can approximately be written in interpretable components for the fixed effect and variance component parts. More general we can state that in different situations, local influence may be able to reveal different aspects of influence. For example, rather than detecting one or a few cases, a cluster of influential subjects

could be revealed. This is likely to happen in situations with more structure, such as hierarchical studies (multi-center trials, stratified samples, etc.). Limited simulations set up to explore such capabilities have been done and are reported in Verbeke *et al* (2001). The results are encouraging in the sense that, when sufficiently different in key aspects, an entire group of subjects can be revealed with the proposed method.

Of course, the local influence parameterizations presented here are not the only ones possible or even desirable. It is clear that other schemes are worthwhile considering as well. Should other perturbation schemes be deemed more interesting for a particular application, then the methods outlined in this Chapter adapt in a straightforward fashion. A possible extension in that sense might be the inclusion of covariates into the dropout mechanism. This has not been possible here given the relatively simple data structure. However, GAUSS code has been developed which allows for this possibility. Further more we have been considering a perturbation scheme using two different weights for dependence on the previous and the current measurement in the dropout probability. This idea has only been studied briefly and the results are not discussed here but may be topic of further research in this area. Another idea which needs further investigation is the need for more formal rules to decide whether a subject is clearly influential, clearly not influential, or borderline, additional research is required. Presently, such rules of thumb as exploring the, say, 5% most influential subjects in more detail can be used.

Finally, we have tried to convey our conviction that sensitivity assessment for MNAR models is imperative. Both the formal and informal sensitivity analyzes have illustrated that mechanical discrimination between MAR and MNAR, based on hypothesis testing, is a dead end street, since MNAR models are very sensitive to such aspects as modeling assumptions, outlying and otherwise influential observations, etc. Insight into whether MAR or rather MNAR is to be preferred should therefore be the subject of an integrated sensitivity analysis.

Chapter 5

Pattern Mixture Models

The high sensitivity of selection modeling results to the correct specification of the measurement model as well as the dropout model, about which little is often known, has been extensively documented as already indicated in Chapter 4. This has lead to growing interest in pattern-mixture modeling, based on the factorization (3.3) (Little 1993, Glynn, Laird and Rubin 1986, Hogan and Laird 1997). After initial mention of pattern-mixture models (Glynn, Laird, and Rubin 1986, Little and Rubin 1987), they are receiving more attention lately (Little 1993, 1994a, 1995, Hogan and Laird 1997, Ekholm and Skinner 1998, Molenberghs, Michiels, Kenward, and Diggle 1998, Molenberghs, Michiels, and Kenward 1998). Concerning the results introduced in this chapter we refer also to Michiels *et al* (2001), Thijs *et al* (2002) and Kenward, Molenberghs and Thijs (2002).

We will first illustrate the idea of pattern-mixture modeling using a simple setting. Let us adopt pattern-mixture decomposition (3.3) and suppress dependence on covariates:

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

with notation as laid out in Chapter 3. Restricting attention to dropout (Section 3.2.1), we obtain,

$$f(\boldsymbol{y}_i, d_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|d_i, \boldsymbol{\theta}) f(d_i|\boldsymbol{\psi}). \tag{5.1}$$

Consider a continuous response at three times of measurement which will be mo-

deled using a trivariate Gaussian distribution. Assume that there may be dropout at time 2 or 3, and let the dropout indicator T_i take the values 1 and 2 to indicate that the last observation occurred at these times and 3 to indicate no dropout. Then, in the first instance, the model implies a different distribution for each time of dropout. We can write

$$\boldsymbol{y}_i \mid t_i \quad \sim \quad N(\boldsymbol{\mu}(t_i), \Sigma(t_i)), \tag{5.2}$$

where

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

for $t = 1, 2, 3$. Recall that t indicates length of sequences, rather than time points of measurements actually taken. 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

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

Its variance can be derived by application of the delta method (see Section 5.4.2).

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 comprise all the parameters from the completers plus those from the following two submodels. For $t = 2$

$$N\left(\begin{pmatrix} \mu_1(2) \\ \mu_2(2) \end{pmatrix}; \begin{pmatrix} \sigma_{11}(2) & \sigma_{21}(2) \\ \sigma_{21}(2) & \sigma_{22}(2) \end{pmatrix}\right),$$

and for $t = 1$

$$N\left(\mu_1(1); \sigma_{11}(1)\right).$$

This is a *saturated* pattern-mixture model and the representation makes it very clear what information each dropout group provides and, consequently, the assumptions that need to be made if we are to predict the behavior of the unobserved responses, and so obtain marginal models for the response. If the three sets of parameters $\boldsymbol{\mu}(t)$ are simply equated, with the same holding for the corresponding variance components, then this implies that dropout is completely random. Progress can be made

with less stringent restrictions however. Little (1993) introduces so-called *complete case missing value* (CCMV) restrictions. These can be defined in terms of conditional distributions. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$. Then the CCMV restrictions imply that for any $T = t < j$

$$f(y_j \mid y_1, \ldots, y_{j-1}, T = t) \quad = \quad f(y_j \mid y_1, \ldots, y_{j-1}, T = n).$$

Little (1993) shows how these constraints can be used to identify all the parameters in the model and so obtain estimates for these and the marginal probabilities. The CCMV restrictions essentially equate conditional distributions beyond time t (i.e., those unidentifiable from this dropout group), with the same conditional distributions from the completers. Another restriction is to identify the former conditional distributions and all conditional distributions from those who drop out after t. This has been called the *available case missing value* (ACMV) restrictions and it has been shown (Molenberghs, Michiels, Kenward, and Diggle 1998) that for dropout, these conditions are equivalent to MAR in the selection model framework. Again, such constraints can be used to develop methods of estimation or to set up schemes for sensitivity analysis. A detailed account is given further in this chapter.

In practice, choice of restrictions will need to be guided by the context. In addition, the form of the data will typically be more complex, requiring, for example, a more structured model for the response with the incorporation of covariates. Hence, models for $f(t_i|\boldsymbol{\psi})$ can be constructed in many ways. Most authors assume the dropout process is fully observed and that T_i satisfies a parametric model (Wu and Bailey 1988, 1989, Little 1993, DeGruttola and Tu 1994). Hogan and Laird (1997) extend this to cases where the dropout time is allowed to be right censored and no parametric restrictions are put on the dropout times. Their conditional model for $\boldsymbol{y}_{\text{obs}}$ given T_i is a linear mixed model with dropout time as one of the covariates in the mean structure. Due to the right censoring, the estimation method must handle incomplete covariates. Hogan and Laird (1997) use the EM algorithm (Dempster, Laird, and Rubin 1977) for ML estimation.

At this point, a distinction between so-called *outcome-based* and *random-coefficient-based* models is useful. In the context of the former, Little (1995) and Little and Wang (1996) consider the restrictions implied by a selection dropout model in the pattern-mixture framework. For example, with two time points and a Gaussian response, Little proposes a general form of dropout model:

$$P(\text{dropout} \mid \boldsymbol{y}) \quad = \quad g(y_1 + \lambda y_2), \tag{5.3}$$

with the function $g(\cdot)$ left unspecified. In a selection modeling context, (5.3) is often assumed to have a logistic form as in (4.6). This relationship implies that the conditional distribution of Y_1 given $Y_1 + \lambda Y_2$ is the same for those who drop out and those who do not. With this restriction and given λ, the parameters of the full distribution of the dropouts is identified. The "weight" λ can then be used as a sensitivity parameter, its size determining dependence of dropout on the past and present, as in the selection models. Such a procedure can be extended to more general problems (Little 1995, Little and Wang 1996). It is instructive in this very simple setting to compare the sources of identifiability in the pattern-mixture and selection models. In the former, the information comes from the assumption that the dropout probability is some function of a linear combination of the two observations with known coefficients. In the latter, it comes from the shape of the assumed conditional distribution of Y_2 given Y_1 (typically Gaussian), together with the functional form of the dropout probability. The difference is highlighted if we consider a sensitivity analysis for the selection model that varies λ in the same way as with the pattern-mixture model. Such sensitivity analysis is much less convincing because the data can, through the likelihood, distinguish between the fit associated with different values of λ.

Therefore, identifiability problems in the selection context tend to be masked. Indeed, there are always unidentified parameters, although a related "problem" seems absent in the selection model. This apparent paradox has been observed by Glynn, Laird, and Rubin (1986). Let us discuss this paradox in some detail.

Assume we have two measurements where Y_1 is always observed and Y_2 is either observed ($t = 2$) or missing ($t = 1$). Let us further simplify the notation by suppressing dependence on parameters and additionally adopting the following definitions:

$$\begin{aligned} g(t|y_1, y_2) &:= f(t|y_1, y_2), \\ p(t) &:= f(t), \\ f_t(y_1, y_2) &:= f(y_1, y_2|t). \end{aligned}$$

Equating the selection model and pattern-mixture model factorizations yields

$$\begin{aligned} f(y_1, y_2)g(d = 2|y_1, y_2) &= f_2(y_1, y_2)p(t = 2), \\ f(y_1, y_2)g(d = 1|y_1, y_2) &= f_1(y_1, y_2)p(t = 1). \end{aligned}$$

Since we have only two patterns, this obviously simplifies further to

$$\begin{aligned} f(y_1, y_2)g(y_1, y_2) &= f_2(y_1, y_2)p, \\ f(y_1, y_2)[1 - g(y_1, y_2)] &= f_1(y_1, y_2)[1 - p], \end{aligned}$$

of which the ratio yields

$$f_1(y_1, y_2) = \frac{1 - g(y_1, y_2)}{g(y_1, y_2)} \frac{p}{1-p} f_2(y_1, y_2).$$

All selection model factors are identified, as are the pattern-mixture quantities on the right-hand side. However, the left-hand side is not entirely identifiable. We can further separate the identifiable from the nonidentifiable quantities:

$$f_1(y_2|y_1) = f_2(y_2|y_1) \frac{1 - g(y_1, y_2)}{g(y_1, y_2)} \frac{p}{1-p} \frac{f_2(y_1)}{f_1(y_1)}. \tag{5.4}$$

In other words, the conditional distribution of the second measurement given the first one, *in the incomplete first pattern*, about which there is no information in the data, is identified by equating it to its counterpart from the complete pattern, modulated via the ratio of the "prior" and "posterior" odds for dropout $[p/(1-p)$ and $g(y_1, y_2)/(1 - g(y_1, y_2))$, respectively] and via the ratio of the densities for the first measurement.

Thus, although an identified selection model is seemingly less arbitrary than a pattern-mixture model, it incorporates *implicit* restrictions. Indeed, precisely these are used in (5.4) to identify the component for which there is no information and again this clearly illustrates the need for sensitivity analysis.

In Section 5.1, we will describe a general strategy for fitting pattern-mixture models. The remainder of this chapter is devoted to a formal juxtaposition of several strategies for pattern-mixture modeling.

5.1 Pattern-Mixture Models

As indicated before, this family is based on factorization (3.3). 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. For example, dropout can vary with treatment arm and age of the respondent, whereas the measurement model can depend on treatment arm, sex, and measurement time.

The measurement model has to reflect dependence on dropout. Thus, the parameters in (4.5) become pattern-dependent: $\boldsymbol{\beta}(d_i)$ and $\Sigma(d_i)$. The dependence of

parameters on dropout can be done in several ways, as will be outlined in Section 5.2. In its most general form, this implies that (4.1) is replaced by

$$\begin{cases} \boldsymbol{Y_i} = \boldsymbol{X_i}\boldsymbol{\beta}(\boldsymbol{d_i}) + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} \\ b_i \sim N(0, D(d_i)), \\ \varepsilon_i \sim N(0, \Sigma_i(d_i)), \end{cases} \tag{5.5}$$

Thus, the fixed effects as well as the covariance parameters are allowed to change with dropout pattern and a priori no restrictions are placed on the structure of this change.

The dropout process simplifies to $f(d_i|W_i, \boldsymbol{\psi})$ which is a, possibly covariate-corrected, model for the probability to belong to a particular pattern. Its components, $g(h_{ij})$, containing only covariates now, describe the dropout rate at each occasion. Thus, the fixed effects as well as the covariance parameters are allowed to change with dropout pattern and a priori no restrictions are placed on the structure of this change.

It immediately follows from (3.3) that the likelihood contribution of the ith subject, based on the observed data $(\boldsymbol{y}_{\text{obs}}, t_i)$, is proportional to

$$f(\boldsymbol{y}_{\text{obs}}, t_i) \quad = \quad f(t_i) f(\boldsymbol{y}_{\text{obs}}|t_i),$$

which only requires specifying a marginal model for the dropout process and a conditional model for the observed outcomes, given the dropout pattern as in (5.5). Further, as for ignorable selection models, both models can be fitted separately, provided separability of their parameters.

Model family (5.5) contains underidentified members since it describes the full set of measurements in pattern t_i, even though there are no measurements after occasion t_i, as was pointed out using the paradox for the simple case of two measurements. 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 routes can be taken to solve this problem. They are described in detail in Section 5.2 but let us briefly sketch them. Focusing on fitting Pattern-mixture models we will describe several strategies. A first strategy is based on Little (1993, 1994a), who advocated the use of identifying restrictions which works well in relatively simple settings. Molenberghs, Michiels, Kenward, and Diggle (1998) proposed a particular set of restrictions for the monotone case which correspond to MAR and in Thijs, Molenberghs, Verbeke, Michiels, and Curran (2001) we introduce a formal way how to deal with these kind of restrictions. Alternatively, as a second strategy, 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 are more obvious. For example, it is not possible to assume an unstructured time trend in incomplete patterns, except if one restricts attention to the time range from onset until dropout. In contrast, assuming a linear time trend allows estimation in all patterns containing at least two measurements. In general, we distinguish between two types of simplification to identify pattern-mixture models. First, functional model forms can be restricted to those which are supported by the information available within a pattern. For example, a linear time trend with a fixed treatment effect, together with a compound symmetry covariance structure, is identifiable as soon as there are two time points. Second, one can let the parameters vary across patterns in a parametric way. Thus, rather than estimating a separate time trend in each pattern, one could assume that the time evolution is unstructured within a pattern, but parallel across patterns. The available data can be used to assess whether such simplifications are supported *within the range of the observed data*. Using the so-obtained profiles past the time of dropout still requires extrapolation or, in other words, a leap of faith. Both strategies are discussed in detail in Section 5.2 and will be applied to the vorozole dataset in Section 5.4.1. On the other hand we will indicate a first tool to study sensitivity of the model assumptions by comparing selection models and pattern-mixture models. In the literature one can consult Michiels, Molenberghs, Bijnens, Vangeneugden and Thijs (2001) for an overview on the results and also in Section 5.4.1 we will discuss

the results related to the Vorozole data.

While in a missing data context, the choice of modeling framework needs careful consideration, the simplicity of the classical MCAR, MAR, and MNAR taxonomy is no longer a feature particular to the selection modeling approach, since, in the case of monotone missing data, the same taxonomy can be developed for pattern-mixture models. For the latter, the interpretation is equally instructive as MAR. The intermediate case corresponds to an explicit and reasonably natural set of restrictions on the unidentifiable components of the full data distribution. Since we are able to fully identify the missing information we now can further distinguish between dropout mechanisms that are depending on future, possibly unobserved measurements. Again this can be done in the selection and the pattern-mixture framework were we define MNF (Missing Non-Future dependent) and NFMV (Non-Future dependent Missing Values) restrictions respectively. Section 5.5 describes both approaches in full detail.

5.2 Fitting Pattern-mixture models and Sensitivity Analysis

Sensitivity analysis for pattern-mixture models can be conceived in many different ways. Crucial aspects are whether pattern-mixture and selection modeling are to be contrasted with one another or rather the pattern-mixture modeling is the central focus of interest.

In the latter case, it is natural to conduct sensitivity analysis *within* the pattern-mixture family. The key area where sensitivity analysis should be focused is on the unidentified components of the model and the way(s) in which this is handled. We will explicitly consider three strategies to deal with under-identification.

- **Strategy 1.** Little (1993, 1994) advocated the use of identifying restrictions and presented a number of examples. We will outline a general framework for identifying restrictions in Section 5.3, with CCMV (introduced by Little 1993), ACMV, and neighboring case missing value restrictions (NCMV) as important special cases. Recall that ACMV is the natural counterpart of MAR in the Pattern-mixture modeling framework. This provides a way to compare ignorable selection models with their counterpart in the pattern-mixture setting. Molenberghs, Michiels, and Lipsitz (1999) and Michiels, Molenberghs, Lipsitz

(1999) took up this idea in the context of binary outcomes, with a marginal global odds ratio model to describe the measurement process (Molenberghs and Lesaffre 1994).

- **Strategy 2.** As opposed to identifying restrictions, model simplification can be done in order to identify the parameters. The advantage is that the number of parameters decreases, which is desirable since the length of the parameter vector is a general issue with pattern-mixture models. Indeed, 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 occurs sufficiently often. Broadly, we distinguish between two types of simplifications.
  - **Strategy 2a.** Trends can be restricted to functional forms supported by the information available within a pattern. For example, a linear or quadratic time trend is easily extrapolated beyond the last obtained measurement. One only needs to provide an ad hoc solution for the first or the first few patterns. In order to fit such models, one simply has to carry out a model building exercise within each of the patterns separately.
  - **Strategy 2b.** Next, one can let the parameters vary across patterns in a controlled parametric way. Thus, rather than estimating a separate time trend within each pattern, one could for example assume that the time evolution within a pattern is unstructured, but parallel across patterns. This is effectuated by treating pattern as a covariate. The available data can be used to assess whether such simplifications are supported within the time ranges for which there is information.

While the second strategy is computationally simple, it is important to note that there is a price to pay. Indeed, simplified models, qualified as "assumption rich" by Sheiner, Beal and Dunne (1997), are also making untestable assumptions, just as in the selection model case. Indeed, using the fitted profiles to predict the evolution, within a pattern, past the time of dropout is based on extrapolation. Still, the need of assumptions and their implications are more obvious. It is, for example, not possible to assume an unstructured time trend in incomplete patterns, except if one restricts attention to the time range from onset until dropout. In contrast, assuming a linear time trend allows estimation in all patterns containing at least two measurements. However, it is less obvious what the precise nature of the dropout mechanism is. An obvious modeling approach, in particular for normally distributed outcomes, is

to specify the dropout mechanism as a polytomous regression. In the identifying restrictions setting on the other hand, the assumptions are clear from the start.

A final observation, applying to both strategies, is that pattern-mixture models do not always automatically provide estimates and standard errors of marginal quantities of interest, such as overall treatment effect or overall time trend. Hogan and Laird (1997) provided a way to derive selection model quantities from the pattern-mixture model. Several authors have followed this idea to formally compare the conclusions from a selection model with the selection model parameters in a pattern-mixture model (Verbeke, Lesaffre, and Spiessens 1998, Curran, Pignatti, and Molenberghs 1998, Michiels *et al* 1999).

5.3 Identifying Restriction Strategies

In line with the results obtained by Molenberghs *et al* (1998), we restrict attention to monotone patterns. In general, let us assume we have $t = 1, \ldots, T$ dropout patterns where the dropout indicator is $d = t + 1$. For pattern t, the complete data density is given by

$$f_t(y_1, \ldots, y_T) = f_t(y_1, \ldots, y_t) f_t(y_{t+1}, \ldots, y_T | y_1, \ldots, y_t). \tag{5.6}$$

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, 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 say, is identified. A general expression for this is

$$f_t(y_s | y_1, \ldots y_{s-1}) = \sum_{j=s}^{T} \omega_{sj} f_j(y_s | y_1, \ldots y_{s-1}), \quad s = t+1, \ldots, T. \tag{5.7}$$

We will use $\boldsymbol{\omega}_s$ as shorthand for the set of ω_{sj}'s used. Every $\boldsymbol{\omega}_s$ which sums to one

provides a valid identification scheme. Let us incorporate (5.7) into (5.6):

$$f_t(y_1,\ldots,y_T) = f_t(y_1,\ldots,y_t)\prod_{s=0}^{T-t-1}\left[\sum_{j=T-s}^{T}\omega_{T-s,j}f_j(y_{T-s}|y_1,...,y_{T-s-1})\right] \quad (5.8)$$

Expression (5.8) 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) proposes CCMV which uses the following identification:

$$f_t(y_s|y_1,\ldots y_{s-1}) = f_T(y_s|y_1,\ldots y_{s-1}), \quad s = t+1,\ldots,T.$$

In other words, information which is unavailable is always borrowed 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:

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

We will refer to these restrictions as *neighboring case missing values* or NCMV.

The third special case of (5.7) will be ACMV. Thus, ACMV is reserved for the counterpart of MAR in the Pattern-mixture context. Let us derive the corresponding $\boldsymbol{\omega}_s$ vectors. Expression (5.7) can be restated as

$$f_t(y_s|y_1,\ldots,y_{s-1}) = f_{(\geq s)}(y_s|y_1,\ldots,y_{s-1}), \quad (5.9)$$

for $s = t+1,\ldots,T$. 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 (5.9) as follows:

$$\begin{aligned} f_t(y_s|y_1,\ldots,y_{s-1}) &= f_{(\geq s)}(y_s|y_1,\ldots,y_{s-1}) \\ &= \frac{\sum_{j=s}^{T}\alpha_j f_j(y_1,\ldots,y_s)}{\sum_{j=s}^{T}\alpha_j f_j(y_1,\ldots,y_{s-1})} \\ &= \sum_{j=s}^{T}\frac{\alpha_j f_j(y_1,\ldots,y_{s-1})}{\sum_{j=s}^{T}\alpha_j f_j(y_1,\ldots,y_{s-1})} f_j(y_s|y_1,\ldots,y_{s-1}). \quad (5.10) \end{aligned}$$

Next, comparing (5.10) to (5.7) yields:

$$\omega_{sj} = \frac{\alpha_j f_j(y_1,...,y_{s-1})}{\sum_{\ell=s}^{T}\alpha_\ell f_\ell(y_1,...,y_{s-1})}. \quad (5.11)$$

We have now derived two equivalent explicit expressions of the MAR case. Expression (5.10) is the conditional density of a mixture, whereas (5.7) with (5.11) is a mixture of conditional densities. Clearly, ω defined by (5.11) consists of components which are nonnegative and sum to one. In other words, a valid density function is defined.

Finally we can incorporate the restrictions (5.7), with the CCMV, NCMV, and ACMV forms as special cases in a comprehensive strategy to fit Pattern-mixture models.

5.3.1 Strategy Outline

We will briefly sketch the strategy. Several points which require further specification will be discussed in subsequent sections using a simple example with only three measurements.

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

2. Select an identification method of choice.

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

$$f_t(y_{t+1}, \ldots, y_T | y_1, \ldots, y_t). \tag{5.12}$$

4. Using standard multiple imputation methodology (Rubin 1987, Schafer 1997, Verbeke and Molenberghs 2000), draw multiple imputations for the unobserved components, given the observed outcomes and the correct pattern-specific density (5.12).

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).

5.3.2 Special Case: 3 Measurements

In this case, there are only three patterns and identification (5.8) takes the following form:

$$
\begin{aligned}
f_3(y_1, y_2, y_3) &= f_3(y_1, y_2, y_3), && (5.13)\\
f_2(y_1, y_2, y_3) &= f_2(y_1, y_2) f_3(y_3|y_1, y_2), && (5.14)\\
f_1(y_1, y_2, y_3) &= f_1(y_1) \left[\omega f_2(y_2|y_1) + (1-\omega) f_3(y_2|y_1)\right] \\
&\quad \times f_3(y_3|y_1, y_2). && (5.15)
\end{aligned}
$$

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 (5.7), the only place where a choice has to be made in pattern 1. Setting $\omega = 1$ corresponds to NCMV, while $\omega = 0$ implies CCMV. Using (5.11) in this particular case, ACMV corresponds to

$$\omega = \frac{\alpha_2 f_2(y_1)}{\alpha_2 f_2(y_1) + \alpha_3 f_3(y_1)}. \tag{5.16}$$

The conditional density $f_1(y_2|y_1)$ in (5.14) can be rewritten as

$$f_1(y_2|y_1) = \frac{\alpha_2 f_2(y_1, y_2) + \alpha_3 f_3(y_1, y_2)}{\alpha_2 f_2(y_1) + \alpha_3 f_3(y_1)}.$$

5.3.3 Drawing from the Conditional Densities

In the previous section, we have seen how general identifying restrictions (5.7), with CCMV, NCMV, and ACMV as special cases, lead to the conditional densities for the unobserved components, given the observed ones. This came down to deriving expressions for $\boldsymbol{\omega}$, such as in (5.11) for ACMV. This endeavor corresponds to items 2 and 3 of the strategy outline (5.3.1). In order to carry out item 4, we need to draw imputations from these conditional densities.

Let us proceed by studying the special case of three measurements first. To this end, we consider identification scheme (5.13)–(5.15) and we start off by avoiding the specification of a parametric form for these densities. The following steps are required:

1. Estimate the parameters of the identifiable densities: $f_3(y_1, y_2, y_3)$, $f_2(y_1, y_2)$, and $f_1(y_1)$. Then, for each of the m imputations, we have to execute the following steps.

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 it will be assumed that in all densities from which we draw, this drawn 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) The proportions ω need to be chosen or determined. Every ω in the unit interval is valid. Specific cases are:

   - For NCMV, $\omega = 1$.
   - For CCMV, $\omega = 0$.
   - For ACMV, ω is calculated from (5.16). Note that, given y_1, this is a constant, depending on α_2 and α_3.

   In order 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. Then continue with (b) and (c).

   (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, to obtain the same number of imputed datasets. Inference then proceeds as outlined Rubin (1987), Schafer (1997) and Verbeke and Molenberghs (2000).

Let us expand on steps 1 and 2 and assume that the observed densities are estimated using linear mixed models. Then, $f_3(y_1, y_2, y_3)$, $f_2(y_1, y_2)$, and $f_1(y_1)$ produce fixed-effect and variance parameters. Let us group all of them in γ and assume a draw is made from their distribution, γ^* say. To this end, their precision estimates need to be computed. These are easily obtained in most standard software packages, such as SAS, rendering this step a very straightforward one.

Let us illustrate this procedure for (5.14). Let us assume that the ith subject has only two measurements, and hence belongs to the second pattern. Let its design

matrices be X_i and Z_i for the fixed effects and random effects, respectively. Its mean and variance for the *third* pattern are:

$$\boldsymbol{\mu}_i(3) = X_i\boldsymbol{\beta}^*(3), \tag{5.17}$$

$$V_i(3) = Z_iD^*(3)Z_i' + \Sigma_i(3), \tag{5.18}$$

where (3) indicates that the parameters are specific to the third pattern.

Now based on (5.17)–(5.18), and the observed values $y_i = (y_{i1}, y_{i2})'$, the parameters for the conditional density follow immediately:

$$\begin{aligned}\boldsymbol{\mu}_{i,2|1}(3) &= \boldsymbol{\mu}_{i,2}(3) + V_{i,21}(3)[V_{i,11}(3)]^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{i,2}(3)),\\ V_{i,2|1}(3) &= V_{i,22}(3) - V_{i,21}(3)[V_{i,11}(3)]^{-1}V_{i,12}(3),\end{aligned}$$

where a subscript 1 indicates the first two components and a subscript 2 refers to the third component. Draws from every other conditional density is entirely similar.

In several cases, the conditional density is a mixture of normal densities. Then, drawing from (5.7) consists of two steps:

- Draw a random uniform variate U to determine which of the $n-s+1$ components one is going to draw from. Specifically, the kth component is chosen if

$$\sum_{j=s}^{k-1}\omega_{sj} \leq U < \sum_{j=s}^{k}\omega_{sj},$$

  where $k = s, \ldots, n$. Note that, if $k = 1$, the left hand sum is set equal to zero.

- Draw from the kth component.

A few comments are in place. Except for in cases with only a few time points, the number of ω parameters proliferates quite rapidly. There are several ways to deal with it. First, special but important restrictions such as NCMV, CCMV, and ACMV do not suffer from this problem since each of the ω's involved is then determined by the choice of restriction. Second, one might envisage partial but important sensitivity analysis by letting all ω's be equal to a fixed quantity, which is chosen as, for example, a member of a grid filling the unit interval. Third, one could put prior distributions on the ω's, perhaps governed by simple hyperpriors. The first solution is followed in this Chapter. The other ones require further exploration.

In addition, determining the conditional distribution of the unobserved outcomes, given the observed ones, is easy in the Gaussian case. For categorical outcomes this is easy as well since it comes down to determining conditional multinomial probabilities which are again multinomial. However, for other distributional forms, this can be quite burdensome. In that case, the conditional distributions will have to be replaced by the corresponding ratio of marginal distributions. While this will change the algebra a bit, the methodology will not undergo fundamental changes.

5.4 The Vorozole Study

We now have arrived at the point where we can apply our own methods to a real life example being the vorozole study. In this section we will discuss our findings with respect to this dataset and the results can mainly be split into two parts. In a first part the main emphasis is put on the comparison of selection models with pattern-mixture models in order to study the sensitivity of the results while a second part focuses on pattern mixture models and more precisely the different strategies to deal with the pattern-mixture framework.

5.4.1 Selection Models versus Pattern-mixture Models

We advocate the use of pattern-mixture models as a tool to assess sensitivity of a selection model to the modeling assumptions, or vice versa. Explicitly, it will be argued that extra confidence in the conclusion can be gained if two analyzes, one within each framework, coincide in key aspects, such as covariate dependencies, strength of association between outcomes, etc. We will outline ways to fit both selection and pattern-mixture models, based on linear mixed models for the measurement process. Virtually all models will be fitted using standard statistical software.

Exploratory Analysis

Most books on longitudinal data discuss exploratory analysis. See, for example, Diggle, Liang, and Zeger (1994). However, most effort is spent to model building and formal aspects of inference. In this section, we present a selected set of plots to underpin the model building. We distinguish between two modes of display: (1) averaged

over (sub)populations and (2) individual profiles. Both ways are used to present three fundamental aspects of the longitudinal structure: (1) the average evolution; (2) the variance function, (3) the correlation structure. Each of those will be discussed in turn. In addition, the variogram will be discussed.

The Average Evolution

The average evolution describes how the profile for a number of relevant subpopulations (or the population as a whole), evolves over time. The results of this exploration will be useful in order to choose a fixed-effects structure for the linear mixed model.

The individual profiles are displayed in Figure 5.1, while the mean profiles per treatment arm, as well as their 95% confidence intervals, are plotted in Figure 5.2. The average profiles indicate an increase over time which is slightly stronger for the vorozole group until month 14, and afterwards, the megestrol acetate group shows a slightly higher FLIC score. As can be seen from the confidence intervals, these differences are clearly not significant.

The individual profiles augment the averaged plot with a suggestion of the variability seen within the data. The thinning of the data towards the later study times suggests that trends at later times should be treated with caution. Therefore, we decided to restrict attention to the first 2 years only. This leads to a maximum of 13 observations per subject (month 1, 2, 4, 6, ..., 24). While these plots also give us some indications about the variability at given times and even about the correlation between measurements of the same individual, it is easier to base such considerations on residual profiles and standardized residual profiles.

The Variance Structure

In addition to the average evolution, the evolution of the variance is important to build an appropriate longitudinal model. Clearly, one has to correct the measurements for the fixed-effects structure and hence detrended values have to be used. These detrended values are merely the outcome values (change in FLIC-score), subtracted by the mean change, calculated at each time point separately. Again, two plots are of interest. The first one pictures the average evolution of the variance as function of time, the second one merely produces the individual residual plots. The detrended profiles

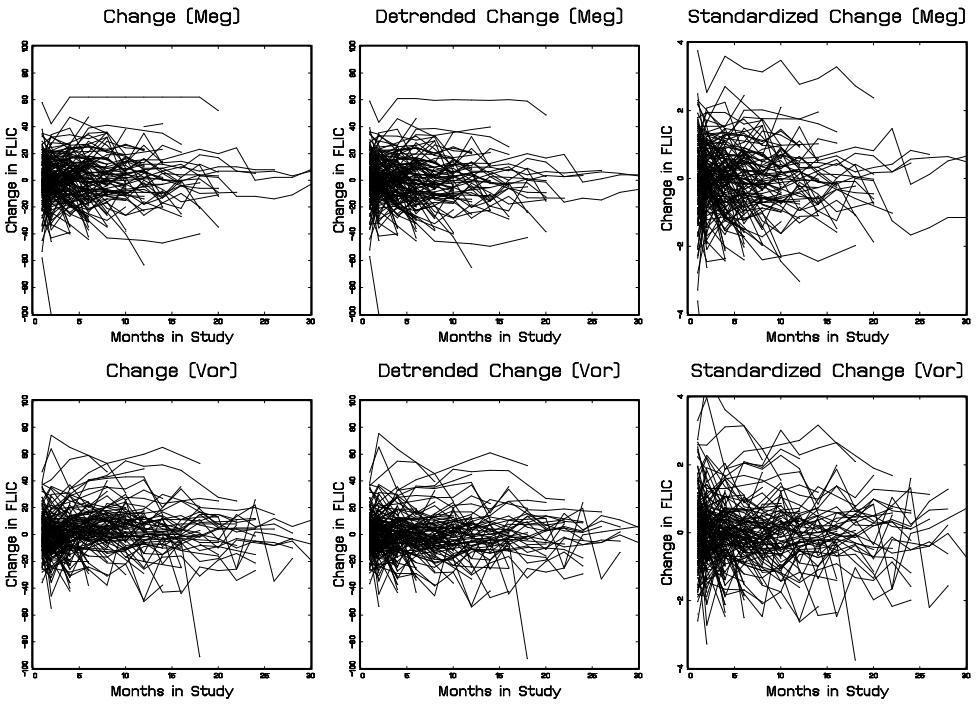

Figure 5.1: *Vorozole study, individual profiles for change (raw, detrended, and standardized)*

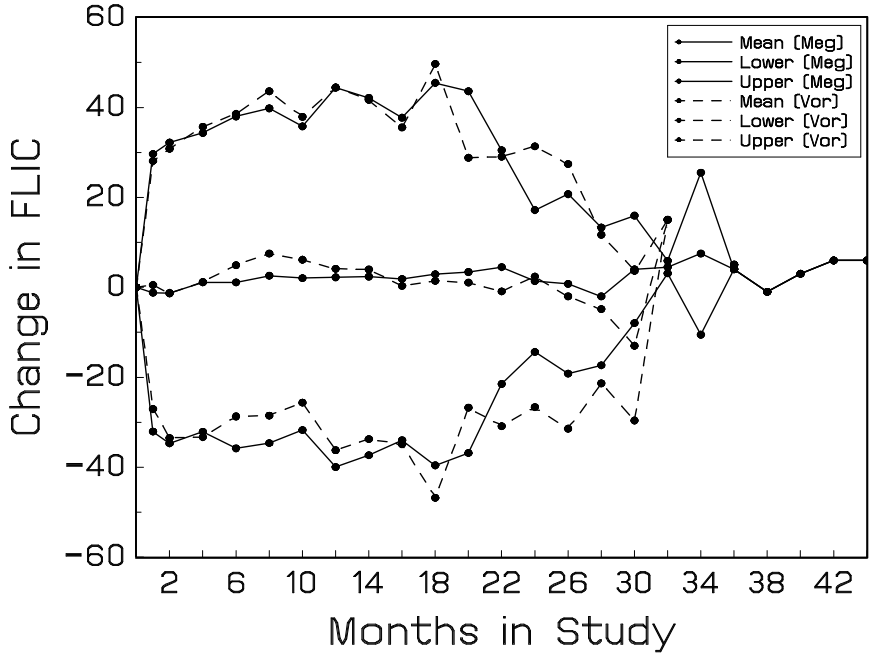

Figure 5.2: *Vorozole study, mean profiles and 95% confidence intervals*

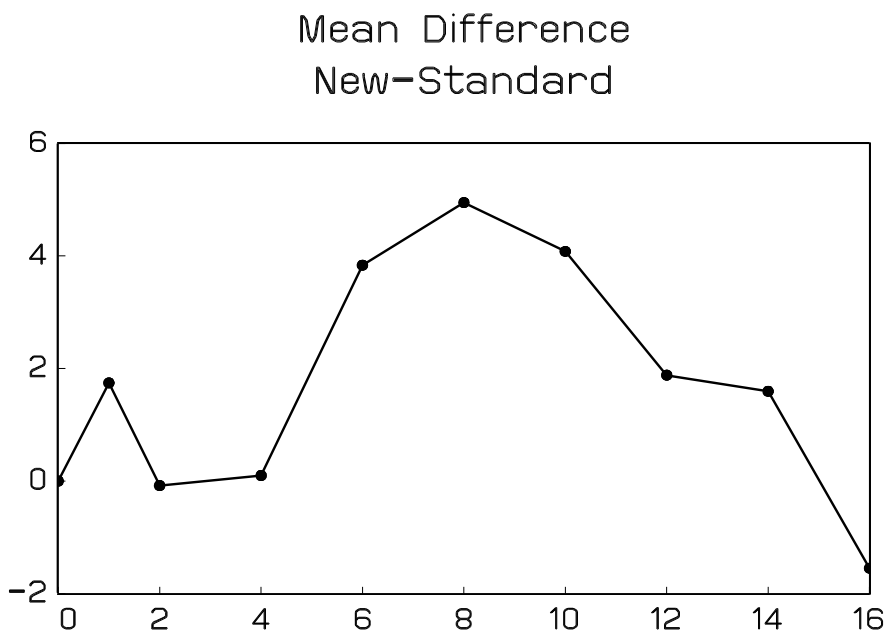

Figure 5.3: *Vorozole study, treatment difference*

are displayed in Figure 5.1, while the corresponding variance function is plotted in Figure 5.4.

The variance function seems to be relatively stable, except for a sharp decline near the end (at which point there are large dropout rates), and hence a constant variance model is a plausible starting point. The individual detrended profiles show subjects' tendency, most clearly in the vorozole group, to decrease immediately before leaving the study. In addition, the detrended profiles also suggest that the variance decreases over time.

The Correlation Structure

The correlation structure describes how measurements within a subject correlate. The correlation function depends on a pair of times and only under the assumption of stationarity does this pair of times simplify to the time lag only. This is important since many exploratory and modeling tools are based on this assumption. A plot of standardized residuals is useful in this respect (Figure 5.1). The picture is not radically different from the previous individual plots, which can be explained by the relative flatness of both mean profile and variance functions. If one or both struc-

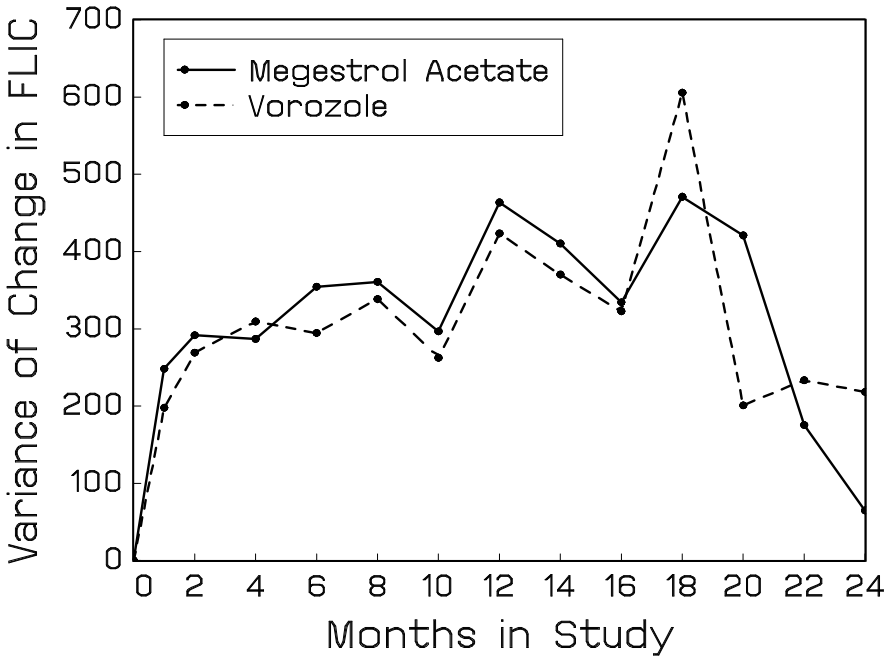

Figure 5.4: *Vorozole study, variance function*

tures is varying with time, the standardized residuals will contribute useful additional information.

A different way of displaying the variance structure is using a scatterplot matrix, such as in Figure 5.5. The off-diagonal elements picture scatterplots of standardized residuals obtained from pairs of measurement occasions. The decay of correlation with time is studied by considering the evolution of the scatters with increasing distance to the main diagonal. Stationarity on the other hand implies that the scatterplots remain similar within diagonal bands *if measurement occasions are approximately equally spaced.* In addition to the scatterplots, we place histograms on the diagonal, capturing the variance structure including such features as skewness. If the axes are given the same scales, it is very easy to capture the attrition rate as well.

The Variogram

Model (4.1) distinguishes between three components of variability. The first one groups traditional random effects (as in a random-effects ANOVA model) and random coefficients (Longford 1993). It stems from inter-individual variability, i.e., he-

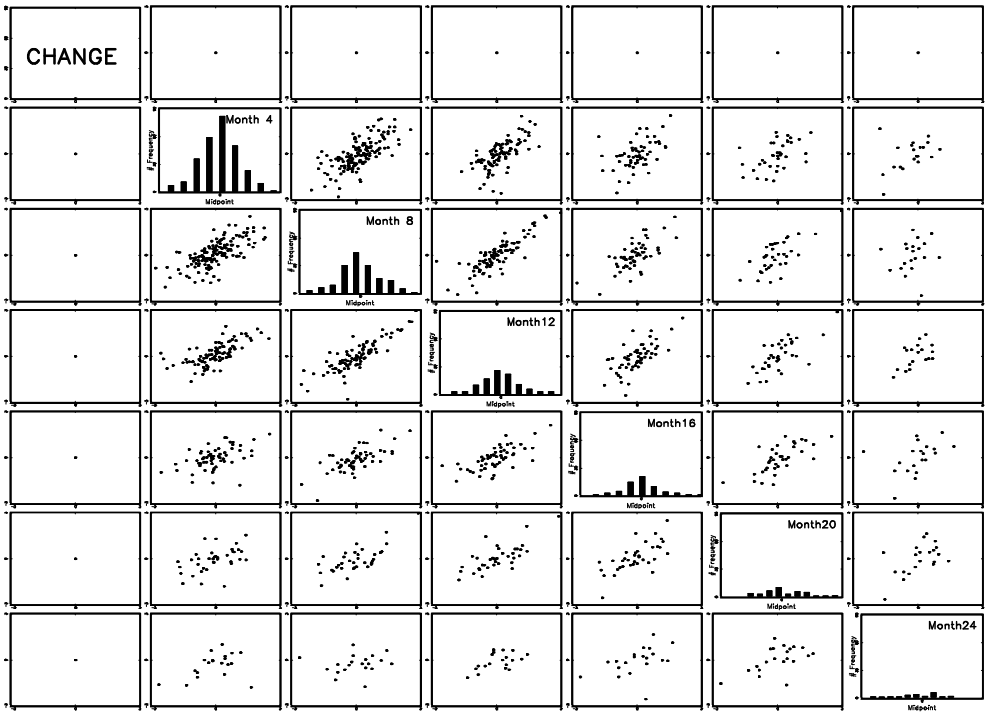

Figure 5.5: *Vorozole study, scatterplot matrix*

terogeneity between individual profiles. The second component, serial association, is present when residuals close to each other in time are more similar than residuals further apart. This notion is well-known from the time-series literature (Ripley 1981, Diggle 1983, Cressie 1991). Finally, on top of the other two components, there is potentially also measurement error. This results from the fact that for delicate measurements (e.g., laboratory essays), even immediate replication will not be able to avoid considerable variation. In longitudinal data, these three components of variability can be distinguished by virtue of both *replication* as well as a clear *distance* concept (time).

Diggle (1991) and Diggle, Liang and Zeger (1994) promote the so-called semivariogram to picture the variance components. It is easily estimated even with irregular observation times (but might require some amount of smoothing). Given a stationary mean-zero stochastic process $Y(t)$ with constant variance, the variogram is defined as

$$V(u) = \frac{1}{2}E\left\{[Y(t) - Y(t-u)]^2\right\}.$$

Specializing (4.1) to random intercept only, D simplifies to a scalar, δ^2 say, and it is shown by Diggle (1990) that the variogram equals

$$V(u) = \sigma^2 + \tau^2(1 - \rho(u)),$$

where $u = t_{ij} - t_{ik}$ is the time lag between both measurements and $\rho(u)$ is the serial correlation between two measurements with the specified lag, calculated for example from (4.3) or (4.4). Note that $V(0) = \sigma^2$ and $V(\infty) = \sigma^2 + \tau^2$. Plotting the process variance,

$$\text{Var}(Y_{ij}) = \delta^2 + \sigma^2 + \tau^2,$$

as a horizontal line and the variogram as a curve, the three components of variability are easy to retrieve. The measurement error is $V(0)$, the random intercept variance is the difference between the process variance and $V(\infty)$, and the variance of the serial process is seen as the band, occupied by the variogram, which increases from $V(0)$ to $V(\infty)$. With irregularly spaced data, it is usually necessary to smooth the variogram. The shape of the variogram conveys information about the structure of the serial correlation function.

The variogram for this study is given in Figure 5.6. The dots correspond to the observed variogram. The fitted variogram, where $V(u)/\text{Var}(Y_{ij})$ is plotted with respect to u, will be explored further later on.

Selection Models for the Vorozole Study

For the measurement model, we start by ignoring the dropout mechanism. This choice will turn out to be justified at the end of this section. Since we are modeling change versus baseline, all models are forced to pass through the origin. This is done by allowing the main covariate effects, but only through their interactions with time.

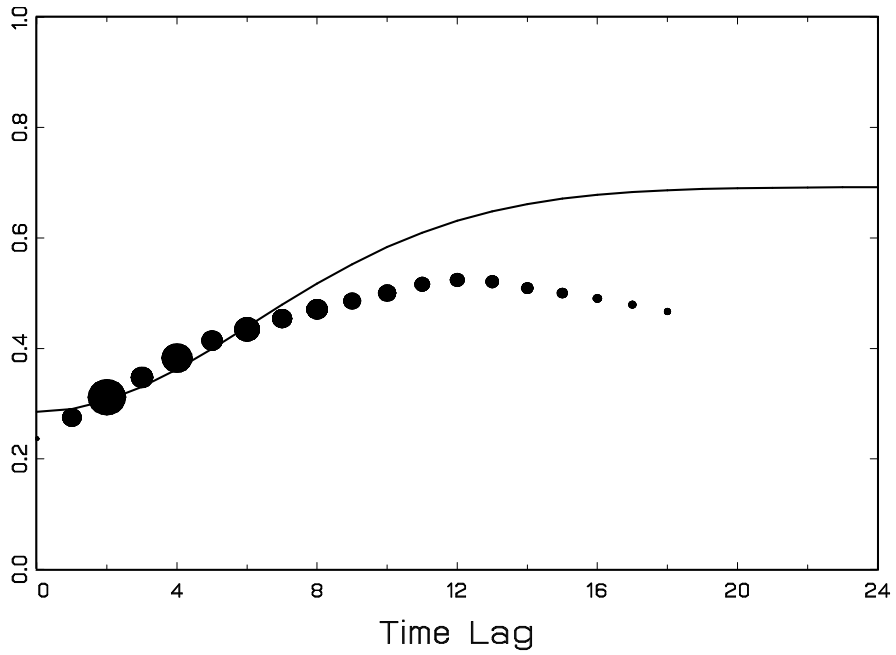

Figure 5.6: *Vorozole study, observed and fitted Variogram*

The following covariates were considered for the measurement model: baseline value, treatment, dominant site, and time in months (up to a cubic time trend). Second order interactions were considered as well. Then, a backwards selection procedure was performed. For design reasons, treatment was kept in the model in spite of its non-significance. An F test for treatment effect produces a p value of 0.5822. Apart from baseline, no other time-stationary covariates were kept. A quadratic time effect provided an adequate description of the time trend. Based on the variogram, we confined the random-effects structure to random intercepts, and supplemented this with a spatial Gaussian process and measurement error. The final model is presented in Table 5.1. The fitted variance structure is represented by means of the fitted variogram, which is given in Figure 5.6. The total correlation between two measurements, one month apart, equals 0.696. The residual correlation, which remains after accounting for the random effects, is still equal to 0.491. The serial correlation, obtained by further ignoring the measurement error, equals $\rho = \exp(-1/7.22^2) = 0.981$.

Fitted profiles are displayed in Figure 5.7 and Figure 5.8. In Figure 5.8, empirical Bayes estimates of the random effects are included whereas in Figure 5.7 the purely marginal mean is used. For each treatment group, we obtain three sets of profiles. The fitted complete profile is the average curve that would be obtained, had all individuals

Table 5.1: *Vorozole study, estimates of the selection model*

Effect	Estimate (s.e.)
Fixed-Effect Parameters:	
time	7.78 (1.05)
time∗baseline	-0.065 (0.009)
time∗treatment	0.086 (0.157)
time^2	-0.30 (0.06)
time^2∗baseline	0.0024 (0.0005)
Variance Parameters:	
random intercept (δ^2)	105.42 (21.304)
serial variance (τ^2)	77.96 (18.537)
serial association (ϕ)	7.22 (1.319)
measurement error (σ^2)	77.83 (4.067)

been completely observed. If we use only those predicted values that correspond to occasions at which an observation was made, then the fitted incomplete profiles are obtained. The latter are somewhat above the former when the random effects are included, and somewhat below when they are not, suggesting that individuals with lower measurements are more likely to disappear from the study. In addition, while the fitted complete curves are very close (the treatment effect was not significant), the fitted incomplete curves are not, suggesting that there is more dropout in the standard arm than in the treatment arm. This is in agreement with the dropout rate, displayed in Figure 5.10, and should not be seen as evidence of a bad fit. Finally, the observed curves, based on the measurements available at each time point, are displayed. These are higher than the fitted ones, but this should be viewed with the standard errors of the observed means in mind (see Figure 5.2).

Next, we will study factors which influence dropout. A logistic regression model, described by (4.7) and (4.8) is used. To start, we restrict attention to MAR processes, whence $\psi_d = 0$. The first model includes treatment, dominant site, baseline value, and the previous measurement but only the last two are significant, producing

$$\begin{aligned} \text{logit}[g(\boldsymbol{h}_{ij})] &= 0.080(0.341) - 0.014(0.003)\text{base}_i \\ &\quad -0.033(0.004)y_{i,j-1}. \end{aligned} \tag{5.19}$$

Diggle and Kenward (1994) and Molenberghs, Kenward, and Lesaffre (1997) con-

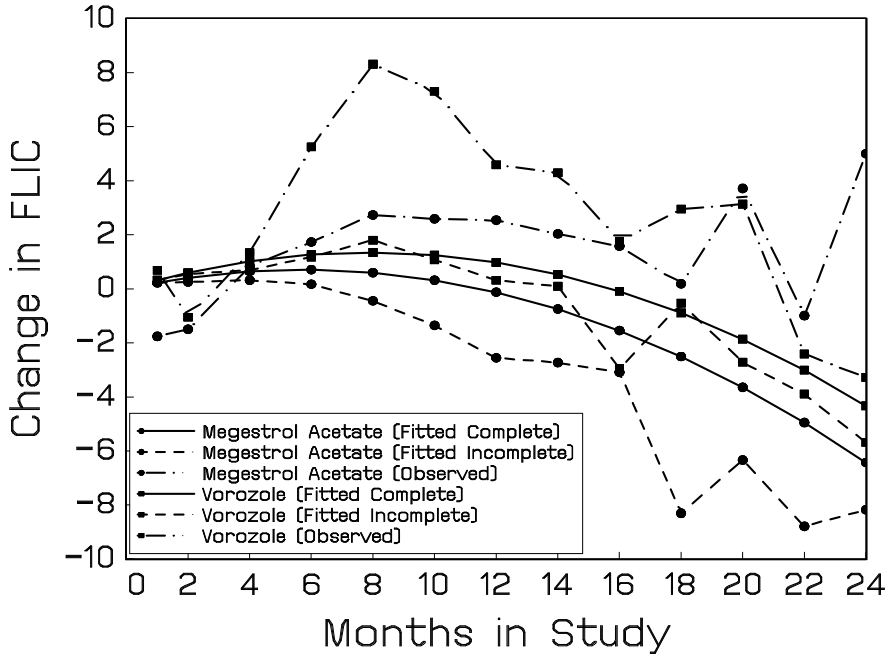

Figure 5.7: *Vorozole study, fitted profiles, averaging the predicted means for the incomplete and complete measurement sequences, without the random effects*

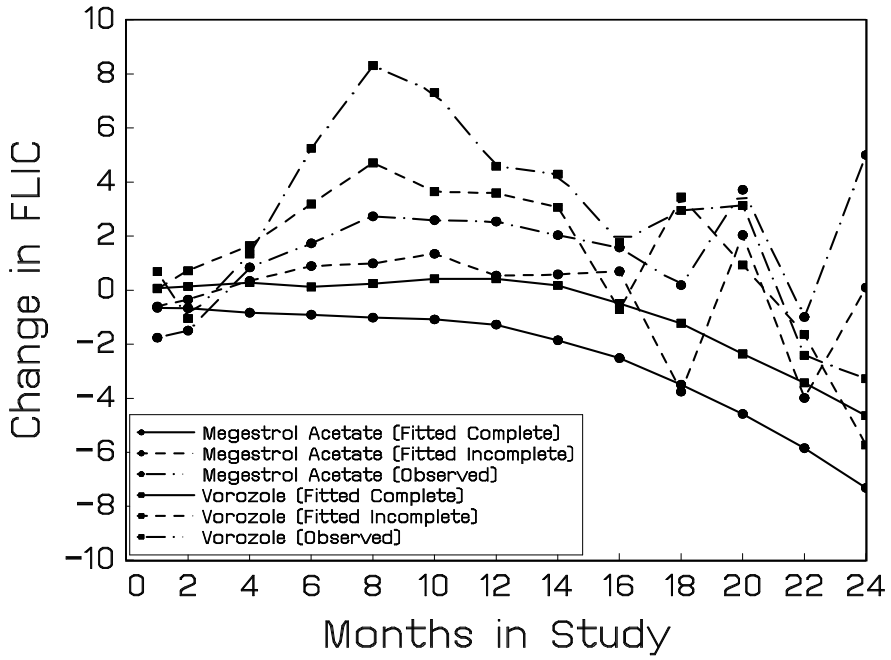

Figure 5.8: *Vorozole study, fitted profiles, averaging the predicted means for the incomplete and complete measurement sequences, including the random effects*

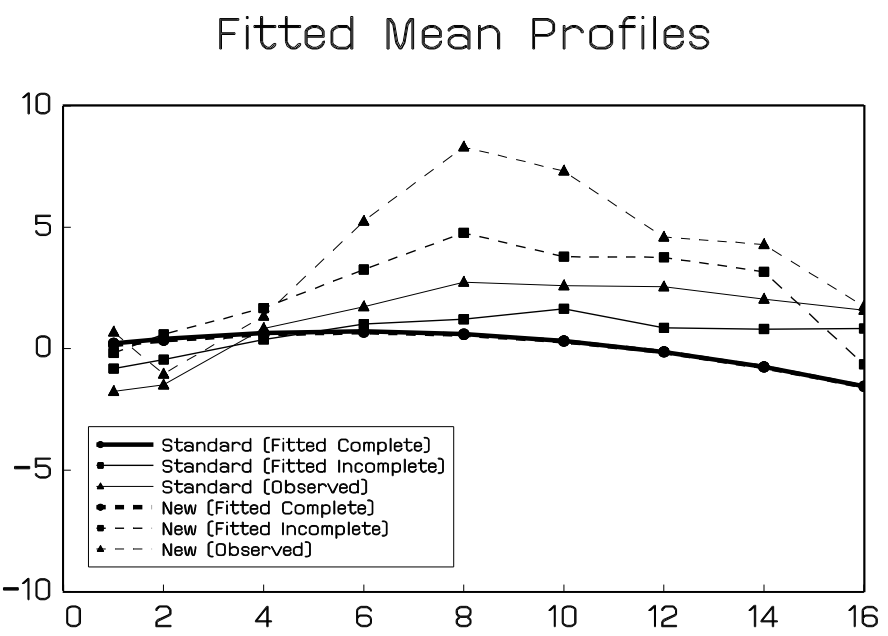

Figure 5.9: *Vorozole study, fitted profiles, averaging the predicted means for the incomplete and complete measurement sequences*

sidered non-random versions of this model by including the current, possible unobserved measurement, such as in (4.7). This requires more elaborate fitting algorithms, since the missing data process is then non-ignorable. Diggle and Kenward (1994) used the simplex algorithm (Nelder and Mead 1965), while Molenberghs, Kenward, and Lesaffre (1997) fitted their models with the EM algorithm (Dempster, Laird and Rubin 1977). The algorithm of Diggle and Kenward is implemented in Oswald (Smith, Robertson and Diggle 1996). With larger datasets such as this one, convergence can be painstakingly difficult and one has to worry about apparent convergence. Therefore, we first proceed in an alternative way. Both Diggle and Kenward (1994) and Molenberghs, Kenward, and Lesaffre (1997) observed that in informative models, dropout tends to depend on the increment, i.e., the difference between the current and previous measurements $y_{ij} - y_{i,j-1}$. Clearly, a very similar quantity is obtained as $y_{i,j-1} - y_{i,j-2}$, but a major advantage of such a model is that it fits within the

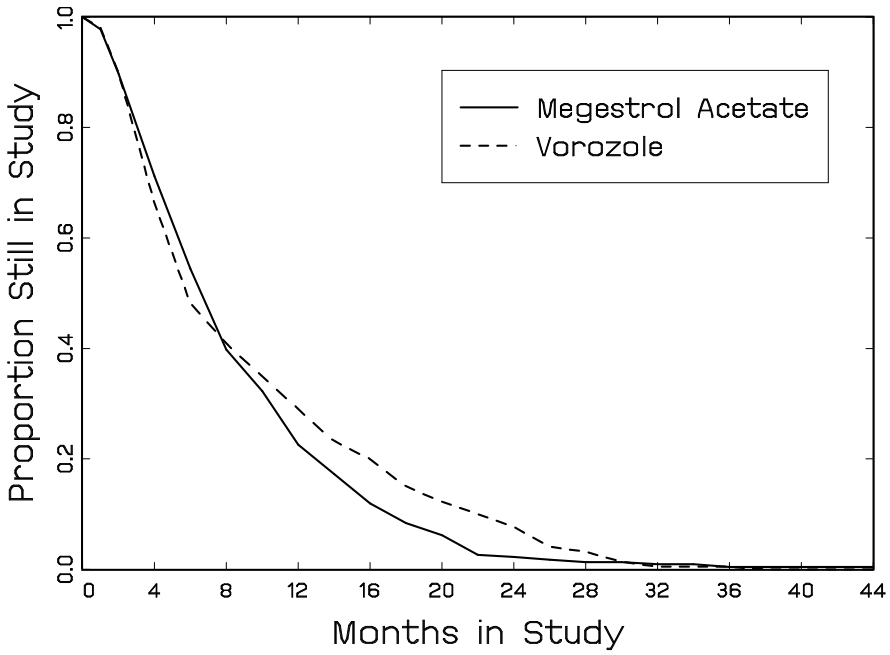

Figure 5.10: *Vorozole study, observed dropout per treatment arm*

MAR framework. In our case, we obtain

$$\begin{aligned}
\text{logit}[g(\boldsymbol{h}_{ij})] &= 0.033(0.401) - 0.013(0.003)\text{base}_i \\
&\quad +0.012(0.006)y_{i,j-2} - 0.035(0.005)y_{i,j-1} \\
&= 0.033(0.401) - 0.013(0.003)\text{base}_i \\
&\quad -0.023(0.005)\frac{y_{i,j-2}+y_{i,j-1}}{2} \\
&\quad -0.047(0.010)\frac{y_{i,j-1}-y_{i,j-2}}{2}. \qquad (5.20)
\end{aligned}$$

indicating that both size and increment are significant predictors for dropout. We conclude that dropout increases with a decrease in baseline, in overall level of the outcome variable, as well as with a decreasing evolution in the outcome.

Using Oswald, both dropout models (5.19) and (5.20) can be compared with their non-random counterparts, where y_{ij} is added to the linear predictor. The first one becomes

$$\text{logit}[g(\boldsymbol{h}_{ij}, y_{ij})] = 0.53 - 0.015\text{base}_i - 0.076y_{i,j-1} + 0.057y_{ij} \qquad (5.21)$$

while the second one becomes

$$\begin{aligned} \text{logit}[g(\boldsymbol{h}_{ij}, y_{ij})] &= 1.38 - 0.021\text{base}_i \\ &\quad -0.0027 y_{i,j-2} - 0.064 y_{i,j-1} + 0.035 y_{ij}. \end{aligned} \tag{5.22}$$

Formal testing of dropout models (5.21) versus (5.19) and for (5.22) versus (5.20) are possible in principle, but will not be carried out for two reasons. First, the likelihood function tends to be very flat for non-random dropout models and therefore the determination of the likelihood ratio is often computationally non-trivial. More fundamentally, Rubin (1994), Little (1994), Laird (1994), and Molenberghs, Kenward and Lesaffre (1997) point out that formal testing for non-random dropout faces philosophical objections. Indeed, non-random dropout models are identified only due to strong but unverifiable assumptions. Hogan and Laird (1997) suggest pattern-mixture models as a viable alternative.

Pattern-mixture models for the Vorozole Study

In analogy with the exploration in the selection model context, it is natural to explore the data from a pattern-mixture point of view. To this end, plots per dropout pattern can be constructed. Figures 5.11 and 5.12 display the individual and averaged profiles per pattern.

Figure 5.12 clearly shows that pattern-specific profiles are of a quadratic nature with in most cases a sharp decline prior to dropout. Note that this is in line with the fitted dropout mechanism (5.20). Therefore, this feature needs to be reflected in the pattern-mixture model. In analogy with our selection model, the profiles are forced to pass through the origin. This is done by allowing only time main effect and interactions of other covariates with time in the model.

The most complex pattern-mixture model we consider includes a different parameter vector for each of the observed patterns. This is done by including the interaction of all effects in the model with *pattern*, a factor variable calculated as 2+ the number of observations after baseline. We then proceed by backward selection in order to simplify the model. First, we found that the covariance structure is common to all patterns, encompassing random intercept, a serial exponential process, and measurement error.

For the fixed effects we proceeded as follows. A backward selection procedure,

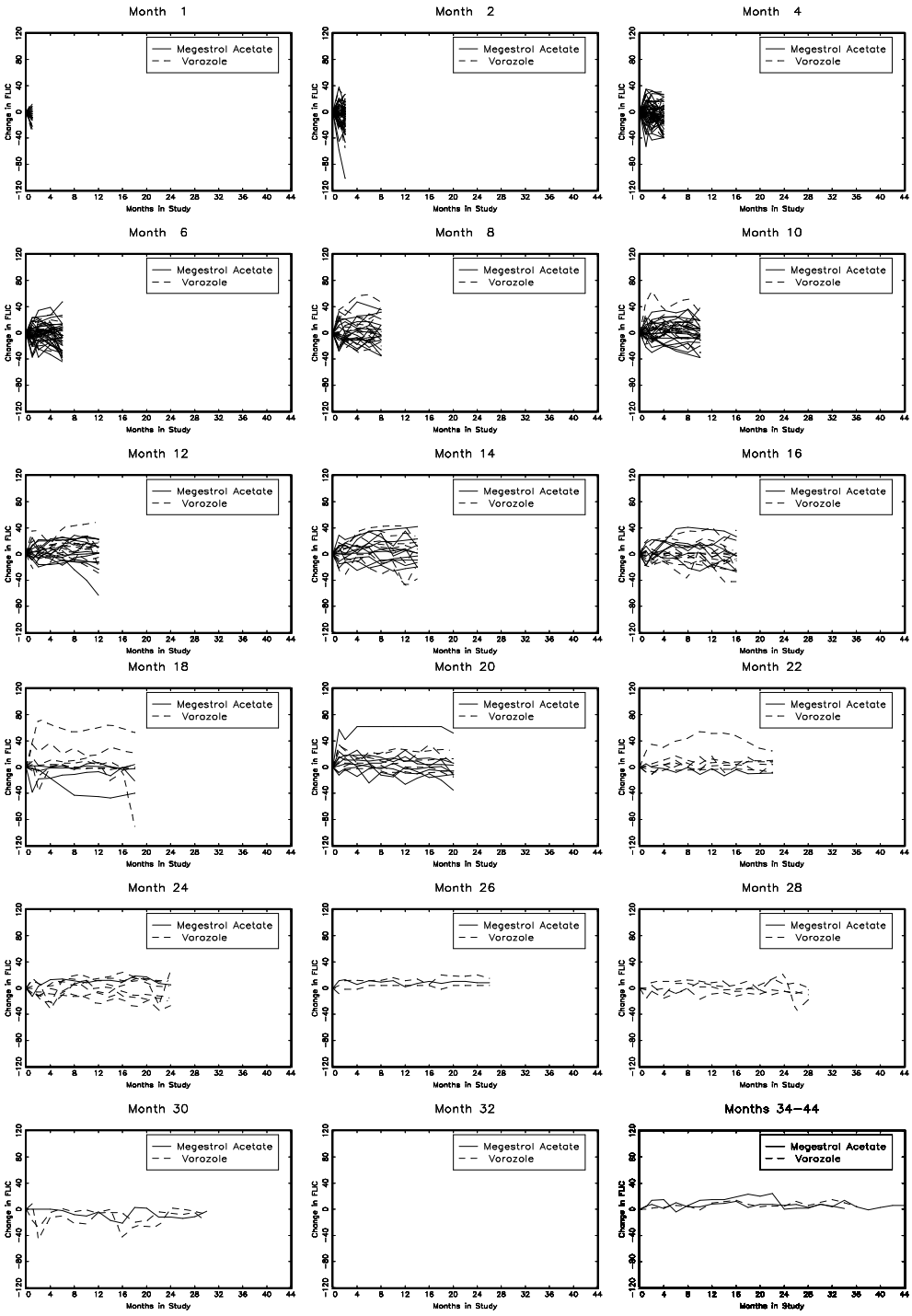

Figure 5.11: *Vorozole study, individual profiles, per dropout pattern*

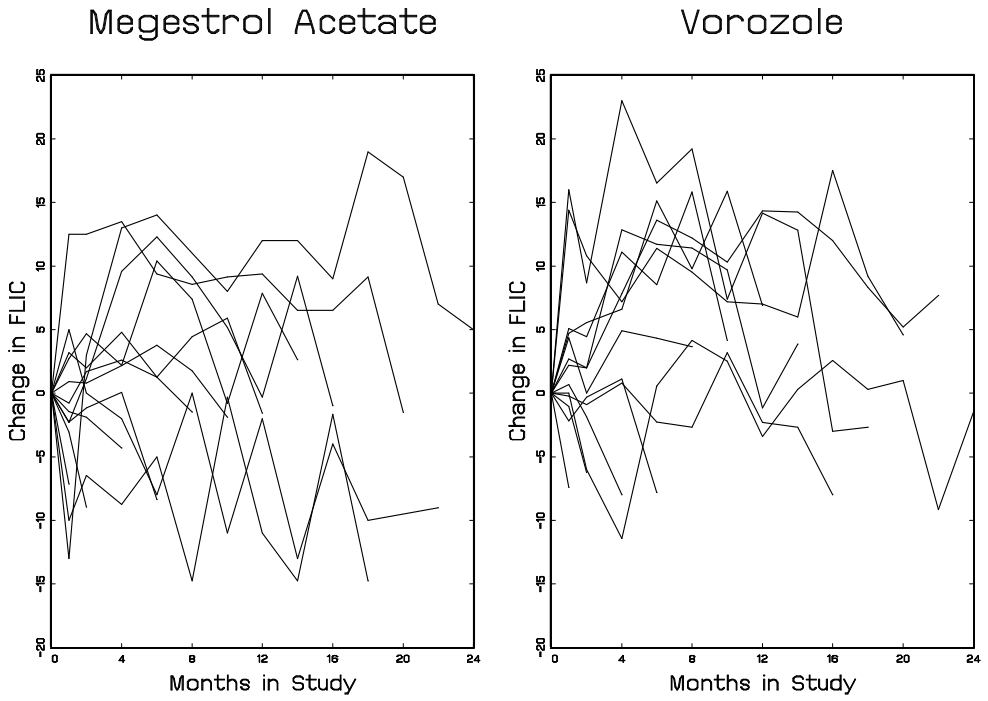

Figure 5.12: *Vorozole study, mean profiles, per dropout pattern*

starting from a model that includes a main effect of time and time^2, as well as interactions of time with baseline value, treatment effect, dominant site and pattern, and the interaction of pattern with time^2. This procedure revealed main effects of time and time^2, as well as interactions of time with baseline value, treatment effect, and pattern, and the interaction of pattern with time^2. This reduced model can be found in Table 5.2. As was the case with the selection model in Table 5.1, treatment effect is non-significant. Indeed, a single degree of freedom F test yields a p value of 0.6868. Note that such a test is possible since treatment effect does not interact with pattern, in contrast to the model which we will describe later. The fitted profiles are displayed in Figure 5.13. We observe that the profiles for both arms are very similar. This is due to the fact that treatment effect is not significant but perhaps also because we did not allow a more complex treatment effect. For example, we might consider an interaction of treatment with the square of time and, more importantly, an treatment effect which is pattern-specific. Some evidence for such an interaction is seen in Figure 5.12.

Our second, expanded model, allowed for up to cubic time effects, the interaction of time with dropout pattern, dominant site, baseline value and treatment, as well

Table 5.2: *Vorozole study, estimates of the first pattern-mixture model*

Fixed-Effect Parameters (Estimate (s.e.)):				
Pattern	Time	Time*Baseline	$Time^2$	Time*Group (0)
main	4.671 (0.844)	-0.031 (0.004)	-0.034 (0.029)	-0.067 (0.166)
3	-8.856 (2.739)			
4	-0.796 (2.958)		-1.918 (1.269)	
5	-1.959 (1.794)		-0.145 (0.365)	
6	1.600 (1.441)		-0.541 (0.197)	
7	0.292 (1.295)		-0.107 (0.133)	
8	1.366 (1.035)		-0.181 (0.080)	
9	1.430 (1.045)		-0.132 (0.071)	
10	1.176 (1.025)		-0.118 (0.061)	
11	0.735 (0.934)		-0.083 (0.049)	
12	0.797 (1.078)		-0.078 (0.055)	
13	0.274 (0.989)		-0.023 (0.046)	
14	0.544 (1.087)		-0.026 (0.049)	
15				
Variance Parameters:				
Random intercept (δ^2)		78.45		
Serial variance (τ^2)		95.38		
Serial association (ϕ)		8.85		
Measurement error (σ^2)		73.77		

Table 5.3: *Vorozole study, estimates of the second pattern-mixture model*

Fixed-Effect Parameters (Estimate (s.e.)):				
Pattern	Time	Time∗Baseline	$Time^2$	$Time^2$∗Baseline
main	5.468 (5.089)	-0.034 (0.040)	-0.271 (0.206)	0.002 (0.002)
3	7.616 (21.908)	-0.119 (0.175)		
4	44.097 (17.489)	-0.440 (0.148)	-18.632 (7.491)	0.1458 (0.0644)
5	22.471 (10.907)	-0.218 (0.089)	-5.871 (2.143)	0.0484 (0.0178)
6	10.578 (9.833)	-0.055 (0.079)	-1.429 (1.276)	0.0080 (0.0107)
7	14.691 (8.424)	-0.123 (0.069)	-1.571 (0.814)	0.0127 (0.0069)
8	7.527 (6.401)	-0.061 (0.052)	-0.827 (0.431)	0.0058 (0.0036)
9	-12.631 (7.367)	0.086 (0.058)	0.653 (0.454)	-0.0065 (0.0038)
10	14.827 (6.467)	-0.126 (0.053)	-0.697 (0.343)	0.0052 (0.0029)
11	5.667 (6.050)	-0.049 (0.049)	-0.315 (0.288)	0.0021 (0.0023)
12	12.418 (6.473)	-0.093 (0.051)	-0.273 (0.296)	0.0016 (0.0024)
13	1.934 (6.551)	-0.022 (0.053)	-0.049 (0.289)	0.0003 (0.0024)
14	6.303 (6.426)	-0.052 (0.050)	-0.182 (0.259)	0.0015 (0.0021)
15				
Pattern	Time∗Group (0)	Time∗Domsite (1)	Time∗Domsite (2)	Time∗Domsite (3)
main		-0.873 (1.073)	0.941 (0.845)	0.023 (0.576)
3	0.445 (5.095)	-5.822 (17.401)	-9.320 (9.429)	1.431 (9.878)
4	0.867 (1.552)	2.024 (3.847)	4.393 (2.690)	5.681 (2.642)
5	-1.312 (0.808)	2.937 (2.596)	0.940 (1.697)	1.414 (1.633)
6	-0.249 (0.686)	-1.378 (2.699)	-4.366 (2.367)	-3.237 (2.289)
7	-0.184 (0.678)	-0.547 (1.917)	-1.099 (1.456)	-1.015 (1.344)
8	0.527 (0.448)	1.302 (1.130)	-0.914 (0.811)	
9	0.782 (0.502)	3.881 (1.485)	1.733 (1.226)	4.548 (1.218)
10	-0.809 (0.464)	2.359 (1.241)	-0.436 (0.843)	
11	-0.080 (0.443)	1.138 (1.128)	-0.326 (0.753)	
12	0.331 (0.579)		-3.595 (0.996)	
13	-0.679 (0.492)	0.317 (1.152)	0.182 (0.825)	
14	0.433 (0.688)		-1.694 (0.972)	
15	-1.323 (0.706)			
Variance Parameters:				
Random intercept (δ^2)		98.93		
Serial variance (τ^2)		38.86		
Serial association (ϕ)		6.10		
Measurement error (σ^2)		73.65		

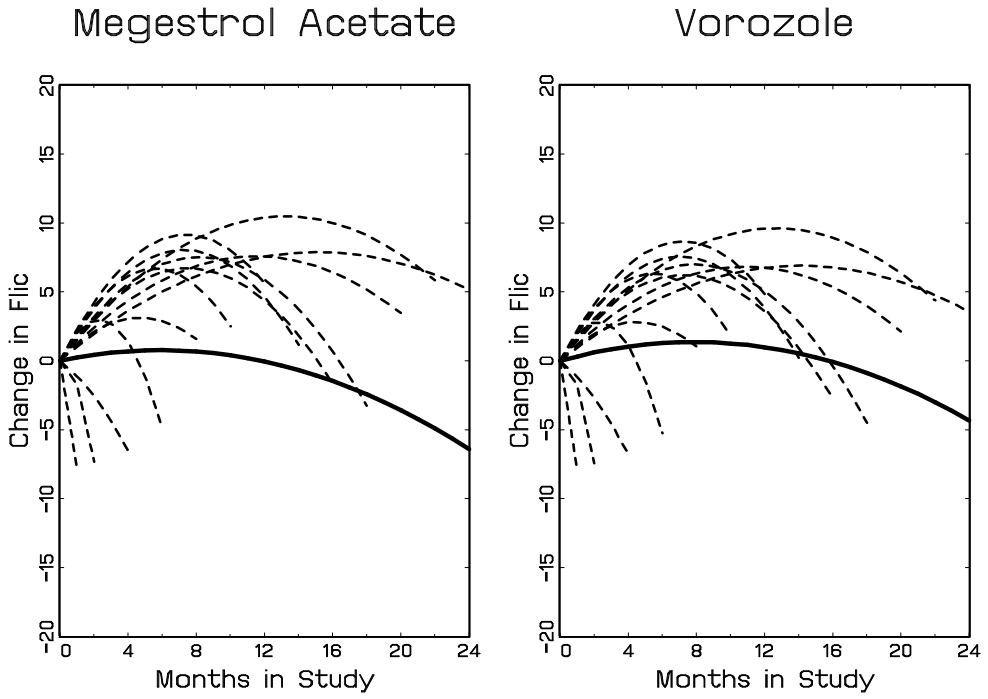

Figure 5.13: *Vorozole study, fitted selection model and first pattern-mixture model*

as their two- and three-way interactions. After a backward selection procedure, the effects included are time and time^2, the two-way interaction of time and dropout pattern, as well as three factor interactions of time and dropout pattern with (1) baseline, (2) group, and (3) dominant site. Finally, time^2 interacts with dropout pattern and with the interaction of baseline and dropout pattern. No cubic time effects were necessary, which is in agreement with the observed profiles in Figure 5.12. The parameter estimates of this model are displayed in Table 5.3. The model is graphically represented in Figure 5.14.

Because a pattern-specific parameter has been included, we have several options for the assessment of treatment. Since there are 13 patterns (remember we cut off the patterns at 2 years), one can test the global hypothesis, based on 13 degrees of freedom, of no treatment effect. We obtain $F = 1.25$, producing $p = 0.2403$, indicating that there is no overall treatment effect. Each of the treatment effects separately is at a non-significant level. Alternatively, the *marginal* effect of treatment can be calculated, which is the weighted average of the pattern-specific treatment effects, with weights given by the probability of occurrence of the various patterns. Its standard error is calculated using a straightforward application of the delta method

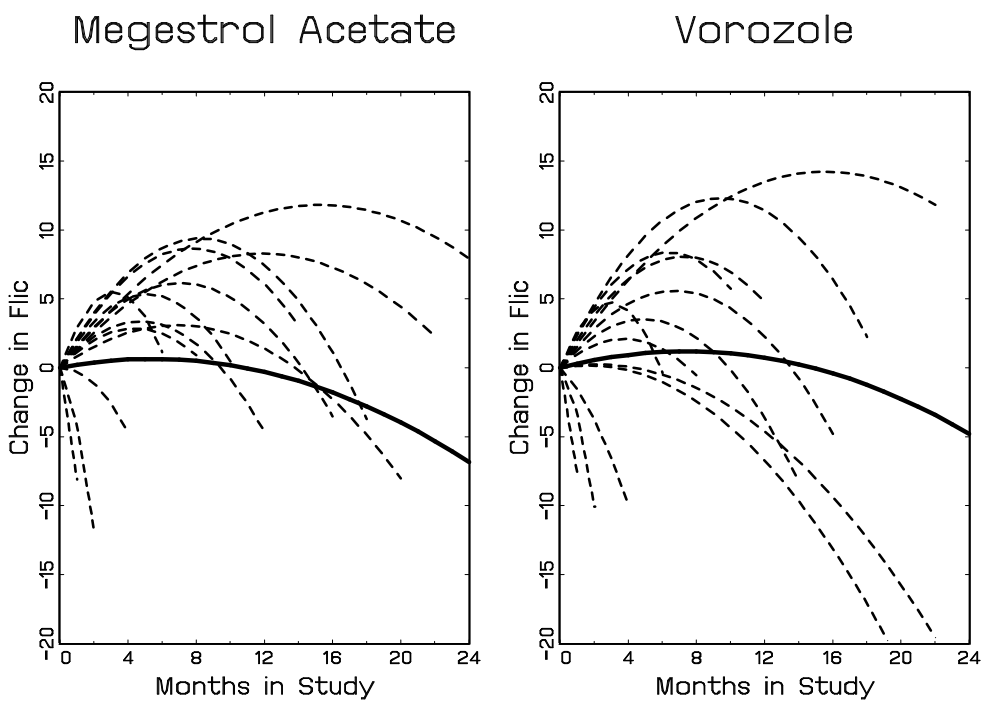

Figure 5.14: *Vorozole study, fitted selection model and second pattern-mixture model*

(see Section 5.4.2). This effect is equal to $-0.286(0.288)$ producing a p value of 0.3206, which is still non-significant.

In summary, we obtain a non-significant treatment effect from all our different models, which gives more weight to this conclusion.

Concluding Remarks

In this paper we have concentrated on total FLIC (i.e., change of the score versus baseline), a quality of life score measured in a multi-centric two arm study in postmenopausal women suffering from metastatic breast cancer. Since virtually all patients were followed up until disease progression or death, the amount of dropout is large. A very large group of patients drops out after just a couple of months.

While classically only selection models are fitted, pattern-mixture models can be seen as a viable alternative. We analyzed the data using both, leading to a sensitivity analysis. More confidence in the results can be gained if both models lead to similar conclusions.

The average profile in the selection model depends on the baseline value, as well as on time. The latter effect is mildly quadratic. There is no evidence for a treatment difference. However, it should be noted that the average profile found is the one that *would* have been observed, had no subjects dropped out, and under the additional assumption that the MAR assumption is correct. Fitting non-random dropout models, in the sense of Diggle and Kenward (1994) is possible, but computationally difficult for a fairly large trial like this one. A separate study of the dropout mechanism revealed that dropout increases with three elements: (1) an unfavorable baseline score, (2) an unfavorable value at the previous month, as well as (3) an unfavorable change in value from the penultimate to the last obtained value.

A pattern-mixture model is fitted by allowing at first a completely separate parameter vector for each observed dropout pattern, which is then simplified by using standard model selection procedures, by considering whether effects are common to all patterns. A first pattern-mixture model features a common treatment effect, of which the assessment is then straightforward. A second model includes a separate treatment effect for each dropout pattern. This leads to two distinct test. The first one tests for equality of the whole treatment vector to be zero. The second one first calculates the marginal treatment effect from the vector of effects, by composing a weighted sum, where the weights are the multinomially estimated probabilities of the various patterns. In all cases, there is no treatment effect. However, a graphical display of the fitted profiles per pattern is enlightening, since it clearly confirms the trend detected in the selection models, that patients tend to drop out when their quality of life score is declining. Since this feature is usually coupled to an imminent progression or death, it should not come as a surprise. An important advantage of pattern-mixture models is that fitting them is more straightforward than non-random selection models. The additional calculations needed for the marginal treatment effect and its associated precision can be done straightforwardly using the delta method.

5.4.2 Pattern-Mixture Models using New Methodology

While previous discussion considered a comparison between selection models and pattern-mixture models we will now apply the methodology introduced earlier in this chapter to the vorozole data. In order to study the impact of the modeling choices, we will first focus on an analysis, restricted to those subjects with 1, 2, and 3 follow up measurements, respectively. Thereafter, we will conduct a sensitivity analysis on

the entire set of patients and follow up times. For the first analysis, 190 subjects are included, with subsample sizes 35, 86, and 69, respectively. The pattern probabilities are

$$\widehat{\boldsymbol{\pi}} = (0.184, 0.453, 0.363)', \tag{5.23}$$

and asymptotic covariance matrixes

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}) = \begin{pmatrix} 0.000791 & -0.000439 & -0.000352 \\ -0.000439 & 0.001304 & -0.000865 \\ -0.000352 & -0.000865 & 0.001217 \end{pmatrix}. \tag{5.24}$$

These figures, apart from giving a feel for the relative importance of the various patterns, will be needed to calculate the marginal treatment effect and to test for its importance, which was the primary goal of the analysis.

It is of interest to study the treatment arm specific pattern probabilities as well. For the vorozole arm, the subsample sizes are 18, 48, and 36, producing probabilities $\widehat{\boldsymbol{\pi}}_v = (0.177, 0.471, 0.354)'$ with asymptotic covariance matrix

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}_v) = \begin{pmatrix} 0.001425 & -0.000814 & -0.000611 \\ -0.000814 & 0.002442 & -0.001628 \\ -0.000611 & -0.001628 & 0.002239 \end{pmatrix}.$$

For the megestrol acetate arm, the subsample sizes are 17, 38, and 33, giving probabilities $\widehat{\boldsymbol{\pi}}_m = (0.193, 0.432, 0.375)'$ and asymptotic covariance matrix

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\pi}}_m) = \begin{pmatrix} 0.001771 & -0.000948 & -0.000823 \\ -0.000948 & 0.002788 & -0.001840 \\ -0.000823 & -0.001840 & 0.002663 \end{pmatrix}.$$

The treatment arm specific probabilities are not significantly different from each other. A classical χ^2 test produces $p = 0.864$. Hence, we will work with expressions (5.23) and (5.24).

We will apply each of the three strategies, show how a model can be fitted and we will indicate how the appropriate hypotheses can be tested.

Table 5.4: *Vorozole study, multiple imputation estimates and standard errors for CCMV, NCMV, and ACMV restrictions (strategy 1), pattern 1.*

Effect	initial	CCMV	NCMV	ACMV
Time	3.40(13.94)	13.21(15.91)	7.56(16.45)	4.43(18.78)
Time*base	-0.11(0.13)	-0.16(0.16)	-0.14(0.16)	-0.11(0.17)
Time*treat	0.33(3.91)	-2.09(2.19)	-1.20(1.93)	-0.41(2.52)
$Time^2$		-0.84(4.21)	-2.12(4.24)	-0.70(4.22)
$Time^2$ *base		0.01(0.04)	0.03(0.04)	0.02(0.04)
σ_{11}	131.09(31.34)	151.91(42.34)	134.54(32.85)	137.33(34.18)
σ_{12}		59.84(40.46)	119.76(40.38)	97.86(38.65)
σ_{22}		201.54(65.38)	257.07(86.05)	201.87(80.02)
σ_{13}		55.12(58.03)	49.88(44.16)	61.87(43.22)
σ_{23}		84.99(48.54)	99.97(57.47)	110.42(87.95)
σ_{33}		245.06(75.56)	241.99(79.79)	286.16(117.90)

Table 5.5: *Vorozole study, multiple imputation estimates and standard errors for CCMV, NCMV, and ACMV restrictions (strategy 1), pattern 2.*

Effect	initial	CCMV	NCMV	ACMV
Time	53.85(14.12)	29.78(10.43)	33.74(11.11)	28.69(11.37)
Time*base	-0.46(0.12)	-0.29(0.09)	-0.33(0.10)	-0.29(0.10)
Time*treat	-0.95(1.86)	-1.68(1.21)	-1.56(2.47)	-2.12(1.36)
$Time^2$	-18.91(6.36)	-4.45(2.87)	-7.00(3.80)	-4.22(4.20)
$Time^2$ *base	0.15(0.05)	0.04(0.02)	0.07(0.03)	0.05(0.04)
σ_{11}	170.77(26.14)	175.59(27.53)	176.49(27.65)	177.86(28.19)
σ_{12}	151.84(29.19)	147.14(29.39)	149.05(29.77)	146.98(29.63)
σ_{22}	292.32(44.61)	297.38(46.04)	299.40(47.22)	297.39(46.04)
σ_{13}		57.22(37.96)	89.10(34.07)	99.18(35.07)
σ_{23}		71.58(36.73)	107.62(47.59)	166.64(66.45)
σ_{33}		212.68(101.31)	264.57(76.73)	300.78(77.97)

Fitting a Model

The patients in this study drop out mainly because they relapse or die. This in itself poses specific challenges that can be addressed within the pattern-mixture framework much easier than in the selection model framework. Indeed, if one is prepared to make

Table 5.6: *Vorozole study, multiple imputation estimates and standard errors for CCMV, NCMV, and ACMV restrictions (strategy 1), pattern 3.*

Effect	initial	CCMV	NCMV	ACMV
Time	29.91(9.08)	29.91(9.08)	29.91(9.08)	29.91(9.08)
Time*base	-0.26(0.08)	-0.26(0.08)	-0.26(0.08)	-0.26(0.08)
Time*treat	0.82(0.95)	0.82(0.95)	0.82(0.95)	0.82(0.95)
Time^2	-6.42(2.23)	-6.42(2.23)	-6.42(2.23)	-6.42(2.23)
Time^2 *base	0.05(0.02)	0.05(0.02)	0.05(0.02)	0.05(0.02)
σ_{11}	206.73(35.86)	206.73(35.86)	206.73(35.86)	206.73(35.86)
σ_{12}	96.97(26.57)	96.97(26.57)	96.97(26.57)	96.97(26.57)
σ_{22}	174.12(31.10)	174.12(31.10)	174.12(31.10)	174.12(31.10)
σ_{13}	87.38(30.66)	87.38(30.66)	87.38(30.66)	87.38(30.66)
σ_{23}	91.66(28.86)	91.66(28.86)	91.66(28.86)	91.66(28.86)
σ_{33}	262.16(44.70)	262.16(44.70)	262.16(44.70)	262.16(44.70)

the assumption that a patient who dies is representative of a slice of the population with the same characteristics, and with a certain probability to die, then identifying restrictions (i.e., extrapolation beyond the time of death) is meaningful. In case one does not want to extrapolate beyond the moment of death, one can restrict modeling to the observed data only. The former viewpoint refers to Strategy 1, while the latter refers to Strategy 2. An intermediate approach would be to allow for extrapolation beyond relapse and not beyond death. (For the current dataset, the information needed in order to do so is unavailable.) Note that, while this may seem a disadvantage of pattern-mixture models, we believe it is an asset, because this framework not only forces one to think about such issues, it also provides a modeling solution, no matter which point of view is adopted. This contrasts with selection models where extrapolation is always done, be it explicitly by modeling the profile, averaged over all patterns. Precisely on this issue we will spend some extra attention in Section 5.5.

For Strategy 1, we start with fitting a model to the observed data, including time and time^2 effects, as well as their interactions with baseline value. Further, time by treatment interaction is included, for consistency with the original analysis plan. All effects interact with time, in order to force profiles to pass through the origin, since we are studying change versus baseline. An unstructured 3×3 covariance matrix is assumed for each pattern. Parameter estimates are presented in Tables 5.4–5.6, in

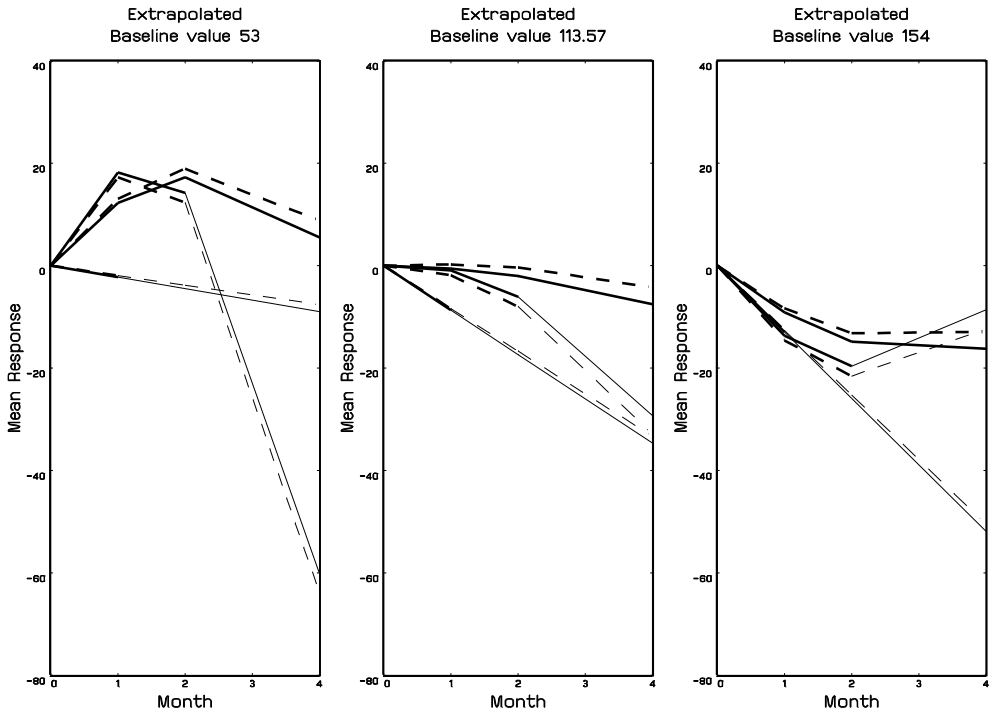

Figure 5.15: *Vorozole study, extrapolation based on model fitted to observed data (Strategy 2a), for three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented, the bold portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type, the dashed line refers to megestrol acetate, the solid line is the vorozole arm.*

the "initial" column. Obviously, not all effects are estimable in this initial model.

Let us present this model graphically. Since there is one binary (treatment arm) and one continuous covariate (baseline level of FLIC score), insight can be obtained by plotting the models for selected values of baseline. Precisely, we chose the minimum, average, and maximum values (Figure 5.15). Note that the extrapolation can have surprising effects, even with these relatively simple models. Thus, while this form of extrapolation is simple, its plausibility can be called into question.

This initial model provides a basis, and its graphical representation extra motivation, to consider identifying restriction models. The methodology of Section 5.3 is then applied, and results are presented in Tables 5.4–5.6, for CCMV, NCMV, ACMV respectively. In all Figures 5.16–5.18, the same mean response scale as in Figure 5.15

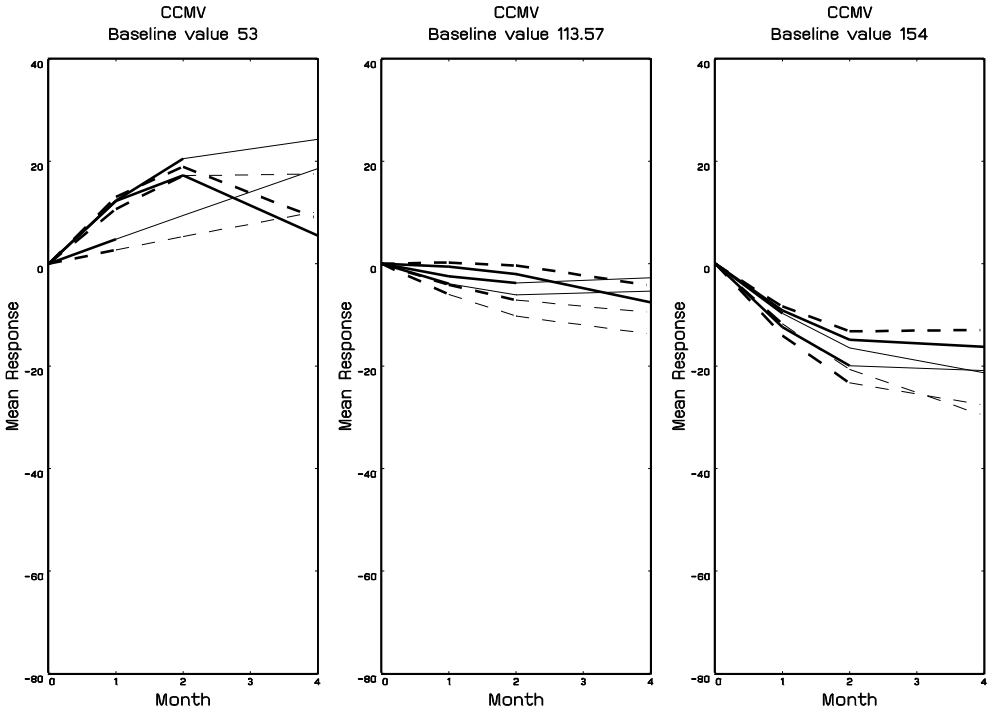

Figure 5.16: *Vorozole study, complete case missing value restrictions analysis, for three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented, the bold portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type, the dashed line refers to megestrol acetate, the solid line is the vorozole arm.*

was retained, illustrating that the identifying restriction strategies extrapolate much closer to the observed data mean responses. There are some differences among the identifying restriction methods. Roughly speaking, CCMV extrapolates rather towards a rise whereas NCMV seems to predict more of a decline, at least for baseline value 53. Further, ACMV indicates rather a steady state. For the other baseline levels, a status quo or a mild increase is predicted. This conclusion needs to be considered carefully. Since these patients drop out mainly because they relapse or die, it seems unlikely to expect a rise in quality of life. Hence, it is well possible that the dropout mechanism is not CCMV, since this strategy always refers to the "best" group, i.e., the one with the best prognosis. ACMV, which compromises between all strategies may be more realistic, but here NCMV is likely to be better since information is borrowed from the nearest pattern.

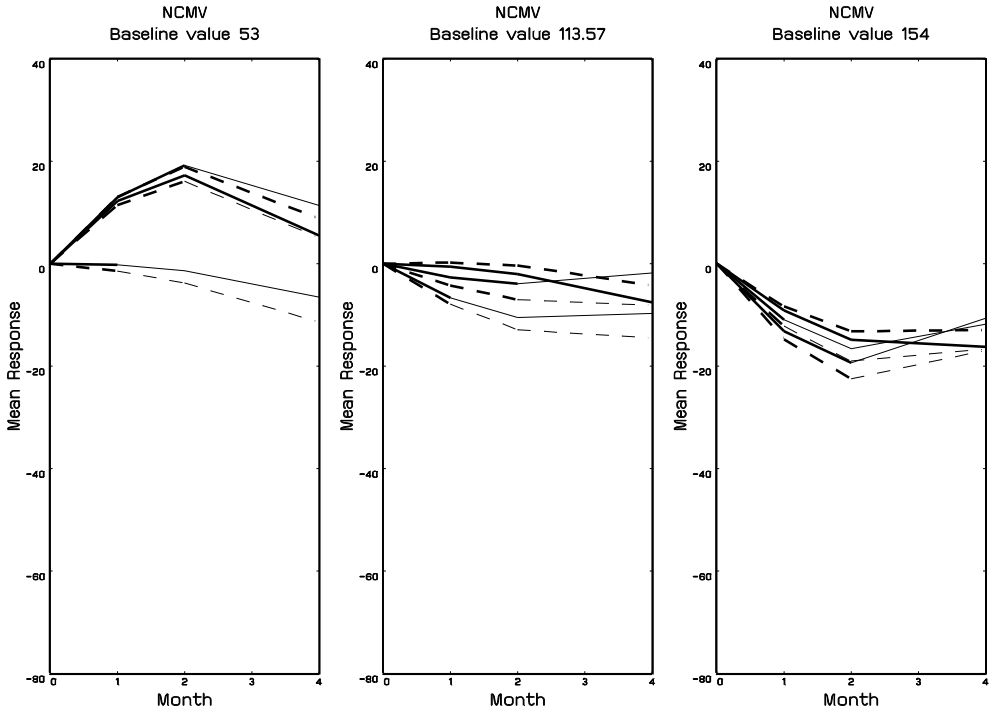

Figure 5.17: *Vorozole study, neighboring case missing value restrictions analysis, for three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented, the bold portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type, the dashed line refers to megestrol acetate, the solid line is the vorozole arm.*

Nevertheless, the NCMV prediction looks more plausible since the worst baseline value shows declining profiles, whereas the best one leaves room for improvement. Should one want to explore the effect of assumptions beyond the range of expression (5.7), one can allow ω_s to include components outside of the unit interval. In that situation, one has to ensure that the resulting density is still non-negative over its entire support.

A possible option concerning strategy 2a is presented in Table 5.7 where we specify solely a different coefficient for the quadratic time trend.

In Strategy 2b, *pattern* is included as a covariate. An initial model is considered with the following effects: time, the interaction between time and treatment, baseline value, pattern, treatment∗baseline, treatment∗pattern, and baseline∗pattern. Fur-

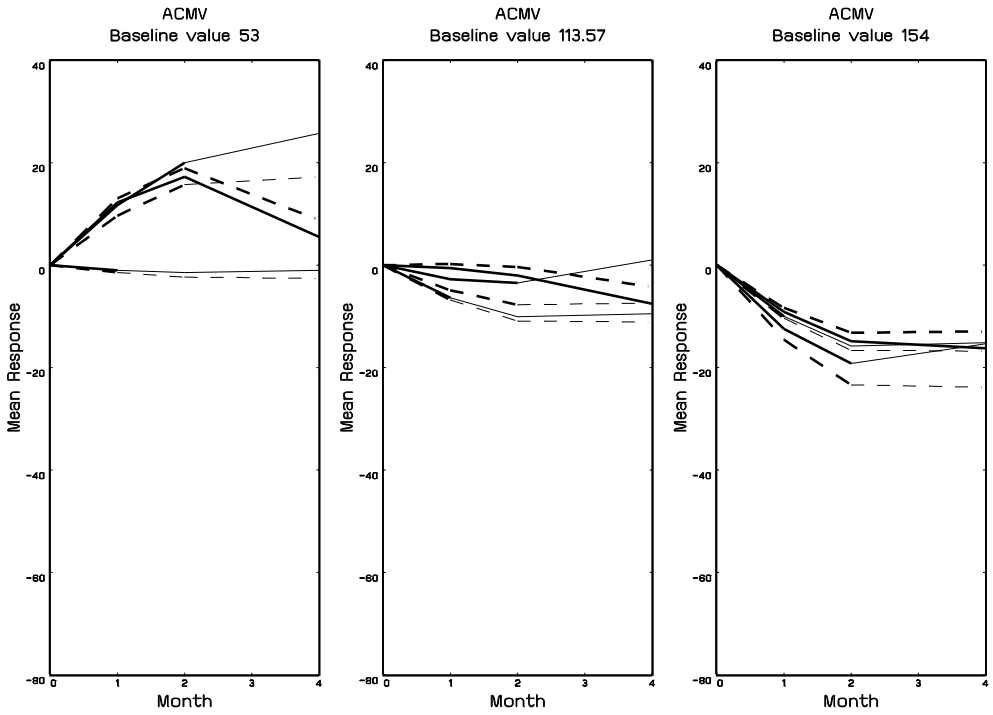

Figure 5.18: *Vorozole study, available case missing value restrictions analysis, for three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented, the bold portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type, the dashed line refers to megestrol acetate, the solid line is the vorozole arm.*

ther, time^2 is included, as well as its interaction with baseline, treatment, and pattern. No interactions beyond the third order are included, and unstructured covariance matrix is common to all three patterns. This implies that the current model is *not* equivalent to a Strategy 1 model, where all parameters are pattern-specific. The estimated model parameters are presented in Table vor strategy2b and the graphical representation is given in Figure 5.19. Early dropouts decline immediately, whereas those who stay longer in the study first show a rise and then decline thereafter. However, this is less pronounced for higher baseline values. On the other hand, the extrapolation based on the fitted model is very unrealistic, rendering this deceptively simple approach a bad choice.

These findings suggest, again, that a more careful reflection on the extrapolation method is required. This is very well possible in a pattern-mixture context, but then

Table 5.7: *Vorozole study, estimates (and standard errors) of a reduced simple model (strategy 2a).*

Effect	Pattern	Estimate (s.e.)
Time		33.06(6.67)
Time*treat		0.40(0.84)
Time*base		-0.29(0.06)
Time^2	1	-16.71(3.46)
Time^2	2	-8.56(1.90)
Time^2	3	-7.09(1.78)
Time^2 *base		0.06(0.01)
σ_{11}		178.02(18.46)
σ_{12}		121.75(18.30)
σ_{22}		238.31(26.98)
σ_{13}		88.75(24.94)
σ_{23}		121.10(34.70)
σ_{33}		274.58(48.32)

the first strategy, rather than the second strategy, has to be used.

Hypothesis Testing

Focusing on treatment effect, a by-product of the pattern-mixture strategy is that a separate treatment effect will likely be estimated for each pattern. This is the case for all five models in Tables 5.4–5.6 and Table 5.8. Let us note in passing that this does not need to be the case. Since the main scientific interest is placed on the estimation of the *marginal treatment effect*, we can combine the pattern-specific effects into a pattern-averaged effect, as follows.

Let $\beta_{\ell d}$ represent the treatment-effect parameter estimates $\ell = 1, \ldots, g$ (assuming there are g groups) in pattern $d = 1, \ldots, T$ and let π_d be the proportion of subjects in pattern d. Then, the estimates of the marginal treatment effects β_ℓ are:

$$\beta_\ell = \sum_{d=1}^{T} \beta_{\ell d} \pi_d, \qquad \ell = 1, \ldots, g. \tag{5.25}$$

The variance is obtained using the delta method. Precisely, it assumes the form

$$\text{Var}(\beta_1, \ldots, \beta_g) = AVA', \tag{5.26}$$

Table 5.8: *Vorozole study, estimates (and standard errors) of a model with pattern as a covariate (strategy 2b).*

Effect	Pattern	Estimate (s.e.)
Time	1	7.29(15.69)
Time	2	37.05(7.67)
Time	3	39.40(9.97)
Time*treat	1	5.25(6.41)
Time*treat	2	3.48(5.46)
Time*treat	3	3.44(6.04)
Time*base	1	-0.21(0.15)
Time*base	2	-0.34(0.06)
Time*base	3	-0.36(0.08)
Time*treat*base		-0.06(0.04)
Time^2	1	*-9.18(2.47)*
Time^2	2	-9.18(2.47)
Time^2	3	-7.70(2.29)
Time^2 *treat		1.10(0.74)
Time^2 *base		0.07(0.02)
σ_{11}		173.63(18.01)
σ_{12}		117.88(17.80)
σ_{22}		233.86(26.61)
σ_{13}		89.59(24.56)
σ_{23}		116.12(34.27)
σ_{33}		273.98(48.15)

where

$$V = \left(\begin{array}{c|c} \text{Var}(\beta_{\ell d}) & 0 \\ \hline 0 & \text{Var}(\pi_d) \end{array}\right)$$

$$A = \frac{\partial(\beta_1, \ldots, \beta_g)}{\partial(\beta_{11}, \ldots, \beta_{Tg}, \pi_1, \ldots, \pi_T)}.$$

The estimate of the variance-covariance matrix of the $\hat{\beta}_{\ell d}$ is obtained from statistical software. The multinomial quantities are easy to obtain from the pattern-specific sample sizes. In the case of the vorozole data, these quantities are presented in (5.23)

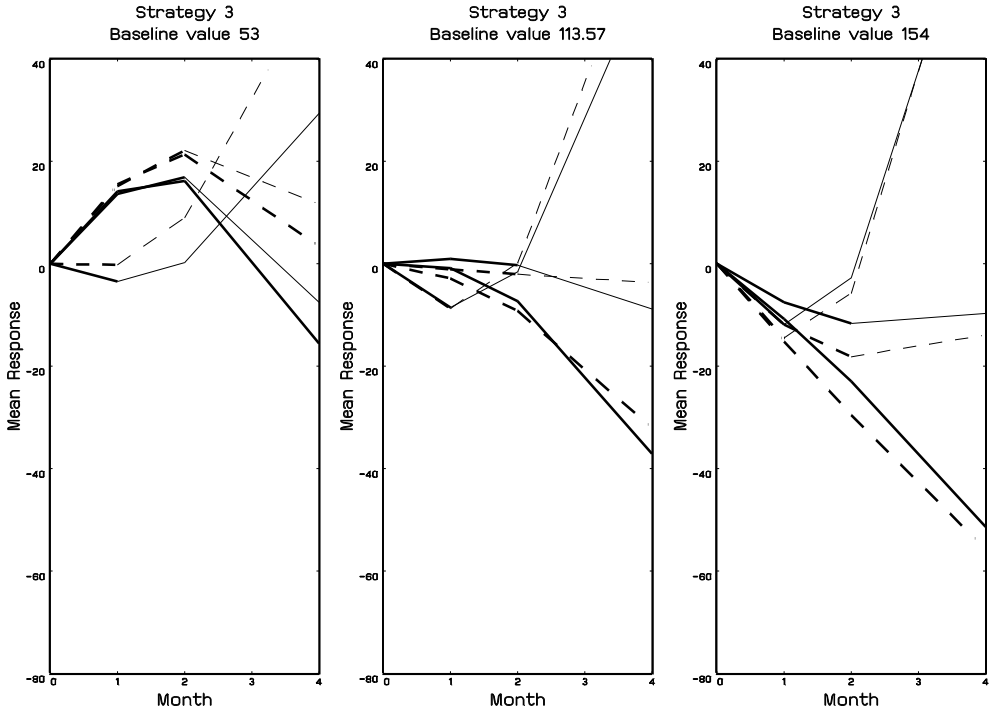

Figure 5.19: *Vorozole study, models with* pattern *used as a covariate (Strategy 2b), for three levels of baseline value (minimum, average, maximum), plots of mean profiles over time are presented, the bold portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in thin line type, the dashed line refers to megestrol acetate, the solid line is the vorozole arm.*

and (5.24). A Wald test statistic for the null hypothesis $H_0 : \beta_1 = \cdots = \beta_g = 0$ is then given by

$$\boldsymbol{\beta}_0' AVA' \boldsymbol{\beta}_0, \tag{5.27}$$

where

$$\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_g)'.$$

We now apply this approach to the models in Tables 5.4–5.6 and Table 5.8. All three pattern-mixture strategies will be considered. Since the identifying restriction strategies are slightly more complicated than the others, we will consider the other strategies first.

For Strategy 2a, recall that the parameters are presented in Tables 5.4–5.6 as the

initial model. The treatment effect vector is $\beta = (0.33, -0.95, 0.82)'$ with, since the patterns are analyzed separately, diagonal covariance matrix:

$$V = \begin{pmatrix} 15.28 & & \\ & 3.44 & \\ & & 0.90 \end{pmatrix}.$$

This leads to the test statistic $\beta' V^{-1} \beta = 1.02$ on 3 degrees of freedom, producing $p = 0.796$. In order to calculate the marginal treatment effect, we apply (5.26)–(5.27). The marginal effect is estimated as $\hat{\beta}_0 = -0.07$ (s.e. 1.16). The corresponding asymptotic p value is $p = 0.95$. Both approaches agree on the non-significance of the treatment effect.

For Strategy 2b, the parameters are presented in Table 5.8. The treatment effect vector is $\beta = (5.25, 348, 3.44)'$ with non-diagonal covariance matrix:

$$V = \begin{pmatrix} 41.12 & 23.59 & 25.48 \\ 23.59 & 29.49 & 30.17 \\ 25.48 & 30.17 & 36.43 \end{pmatrix}.$$

The correlation between them is quite substantial. The reason is that some parameters, in particular the other treatment effects (three-way interaction with baseline and time, interaction with time^2), are common to all three patterns, hence inducing dependence across patterns. This leads to the test statistic $\beta' V^{-1} \beta = 0.70$ on 3 degrees of freedom, producing $p = 0.874$.

Calculating the marginalized treatment effect, we obtain $\hat{\beta}_0 = 3.79$ (s.e. 5.44). The corresponding asymptotic p value is $p = 0.49$. The different numerical value of the treatment effects, as compared to those obtained with the other strategies, is entirely due to the presence of a quadratic treatment effect which, for ease of exposition, is left out of the picture in testing here. It is straightforward to add this parameter to the contrast(s) being considered, should one want to do so.

For Strategy 1, we will consider several approximate ways of inference. The CCMV case will be discussed in detail. The two other restriction types are entirely similar.

There are three treatment effects, one for each pattern. Hence, multiple imputation produces a vector of treatment effects and the within, between, and total covariance matrices:

$$\beta_{CC} = (-2.09, -1.68, 0.82)', \tag{5.28}$$

$$W_{CC} = \begin{pmatrix} 1.67 & 0.00 & 0.00 \\ 0.00 & 0.59 & 0.00 \\ 0.00 & 0.00 & 0.90 \end{pmatrix},$$

$$B_{CC} = \begin{pmatrix} 2.62 & 0.85 & 0.00 \\ 0.85 & 0.72 & 0.00 \\ 0.00 & 0.00 & 0.00 \end{pmatrix}, \tag{5.29}$$

and

$$T_{CC} = \begin{pmatrix} 4.80 & 1.02 & 0.00 \\ 1.02 & 1.46 & 0.00 \\ 0.00 & 0.00 & 0.90 \end{pmatrix}.$$

In a stratified analysis, we want to test the hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$. Using (5.28)–(5.29), we can apply multiple imputation methodology.

Note that, even though the analysis is done per pattern, the between and total matrices have non-zero off-diagonal elements. This is because the imputation is based on information from *other* patterns, hence introducing inter-pattern dependence. Results are presented in Table 5.9. All p values are highly non-significant, in line with earlier evidence from Strategies 2a and 2b, although a bit more extreme.

For the marginal parameter, the situation is more complicated here than with Strategies 2a and 2b. Indeed, classical theory often assumes inference is geared towards the original vector, or linear contrasts thereof. Formula (5.25) represents a non-linear transformation of the parameter vector and therefore needs further development. First, consider $\boldsymbol{\pi}$ to be part of the parameter vector. Since there is no missingness involved in this part, it contributes to the within matrix, but not to the between matrix. Then, using (5.26), the approximate within matrix for the marginal treatment effect is

$$W_0 = \boldsymbol{\pi}' W \boldsymbol{\pi} + \boldsymbol{\beta}' \mathrm{Var}(\boldsymbol{\pi}) \boldsymbol{\beta},$$

with, for the between matrix, simply

$$B_0 = \boldsymbol{\pi}' B \boldsymbol{\pi}.$$

The results are presented in the second panel of Table 5.9. All three p values are in between those obtained for Strategies 2a and 2b. Of course, all five agree on

Table 5.9: *Vorozole study, tests of treatment effect for CCMV, NCMV, and ACMV restrictions.*

Parameter	CCMV	NCMV	ACMV
Stratified analysis:			
k	3	3	3
τ	12	12	12
denominator d.f. w	28.41	17.28	28.06
r	1.12	2.89	1.14
F statistic	0.044	0.022	0.030
p value	0.988	0.995	0.993
Marginal Analysis:			
Marginal effect (s.e.)	-0.85(0.77)	-0.63(1.22)	-0.74(0.85)
k	1	1	1
τ	4	4	4
denominator d.f. w	4	4	4
r	1.49	4.57	1.53
F statistic	0.072	0.018	0.054
p value	0.801	0.900	0.828

the non-significance of the treatment effect. The reason for the differences is to be found in the way the treatment effect is extrapolated beyond the period of observation. Indeed, the highest p value is obtained for Strategy 2a and, from Figure 5.15, we learn that virtually no separation between both treatment arms is projected. On the other hand, wider separations are seen in Figure 5.19. Finally, we note that all conclusions are conditional upon the unverifiable assumption that the posited restrictions (and hence, dropout mechanisms) are correct. Therefore, they should preferably be used in conjunction, within a sensitivity analysis.

Analysis of All Patterns

From our initial analysis it is clear that the apparently simpler Strategy 2 should be avoided. In the current analysis, aimed to use all data, we therefore restrict ourselves to the more involved but also more satisfactory identifying-restrictions Strategy 1. The distribution of subjects over patterns is described in Table 5.10. Based on this information, we chose to remove the three subjects in the last patterns, due to

Table 5.10: *Vorozole study, distribution of the subjects over the patterns.*

Pattern	Number of measurements	Number of subjects	Proportion	Percentage
3	1	35	0.087281796	8.7281796
4	2	86	0.214463840	21.4463840
5	3	69	0.172069830	17.2069830
6	4	45	0.112219450	11.2219450
7	5	29	0.072319202	7.2319202
8	6	33	0.082294264	8.2294264
9	7	22	0.054862843	5.4862843
10	8	17	0.042394015	4.2394015
11	9	19	0.047381546	4.7381546
12	10	9	0.022443890	2.2443890
13	11	10	0.024937656	2.4937656
14	12	6	0.014962594	1.4962594
15	13	8	0.019950125	1.9950125
16	14	3	0.007481297	0.7481297
17	15	4	0.009975062	0.9975062
18	16	3	0.007481297	0.7481297
20	18	1	0.002493766	0.2493766
21	19	1	0.002493766	0.2493766
25	23	1	0.002493766	0.2493766

sparseness. Covariates kept in the models are time, time interactions with baseline and treatment group, time^2 and the interaction between time^2 and baseline. The three identifying restrictions are chosen, and the assessment of the marginal treatment effect is: $p = 0.9407$ for NCMV, $p = 0.7570$ for CCMV, and $p = 0.0487$ for ACMV. Clearly, the impact of the identifying restrictions on the main conclusion is dramatic. The MAR-based ACMV restrictions yield a significant effect, albeit borderline, whereas the others are far from significance. A graphical representation is given in Figure 5.20.

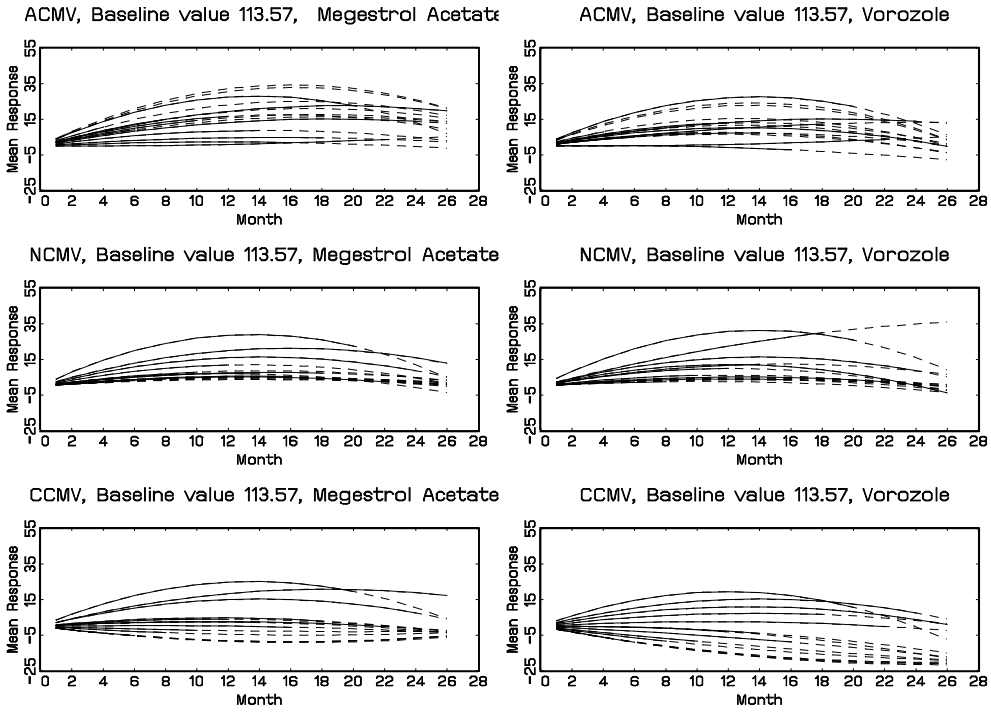

Figure 5.20: *Vorozole study, models using all patterns (strategy 1), for the average level of baseline value and for both treatment arms, plots of mean profiles over time are presented, the solid portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in dashed type.*

5.4.3 Concluding Remarks

We now have illustrated three distinct strategies to fit pattern-mixture models. In this way, we have brought together several existing practices. Little (1993, 1994) has proposed identifying restrictions, which we here formalized using the connection with MAR and multiple imputation. Strategy 2 refers to fitting a model per pattern and using pattern as a covariate.

By contrasting these strategies on a single set of data, one obtains a range of conclusions rather than a single one, which provides insight into the sensitivity to the assumptions made. Especially with the identifying restrictions, one has to be very explicit about the assumptions and moreover this approach offers the possibility to consider several forms of restrictions. Special attention should go to the ACMV restrictions, since they are the MAR counterpart within the pattern-mixture context.

In addition, a comparison between the selection and pattern-mixture modeling approaches is useful to obtain additional insight into the data and/or to assess sensitivity.

The identifying restrictions strategy provides further opportunity for sensitivity analysis. Indeed, since CCMV and NCMV are extremes for the $\boldsymbol{\omega}_s$ vector in (5.7), it is very natural to consider the idea of *ranges* in the allowable space of $\boldsymbol{\omega}_s$. Clearly, any $\boldsymbol{\omega}_s$ which consists of non-negative elements that sum to one is allowable, but also the idea of extrapolation could be useful, where negative components are allowed, given they provide valid conditional densities.

As is clear from previous Chapters, the modeling of incomplete data from longitudinal studies has been approached by many authors through the pattern-mixture modeling framework. An important issue is whether a model for a given dropout pattern ought to be extended and, if the answer is affirmative, how this should be approached. A possible solution is ordinary extrapolation, based on sufficiently simplified models while secondly also identifying restrictions can be used. Unfortunately, the implications for the nature of the dropout mechanism are not always understood nor studied. In the subsequent sections we will describe necessary and sufficient conditions ensuring that dropout does not depend on future, unobserved observations.

We already indicated that extrapolation beyond the time of dropout can be very easy by means of simple models considering simple polynomial time trends within patterns (Hogan and Laird 1997) or when *pattern* is used as a covariate. In the latter case, rather than estimating a separate time trend within each pattern, one could assume that the time evolution within a pattern is unstructured, but parallel across patterns. While these strategies are computationally simple, there is a non-trivial price to pay: simplified models also make untestable assumptions, just as in the selection model case. To overcome this problem we deemed the identification-restrictions strategy to be the more promising one and therefore we provided a general framework with complete case missing value restrictions (CCMV, Little 1993), where information is borrowed from the completers, available case missing values (ACMV), equivalent to MAR (Molenberghs et al. 1998), where information is borrowed from all patterns for which the unobserved measurement in the pattern under study is observed, and neighboring case missing value restrictions (NCMV), where information is borrowed from the closest available pattern, as important special cases. ACMV provides a way to compare MAR under both the selection model and the PMM frameworks. Even though the use of such restrictions leads the modeler to careful

reflections on the choices made, not all implications of such a choice are obvious. In the light of this, it is useful to know MAR and ACMV are equivalent, implying that, under ACMV, missingness does not depend on unobserved measurements.

We consider it to be of interest to further characterize which mechanisms are more general than ACMV (MAR) but still prevent missingness to depend on future, possibly unobserved outcomes. Therefore in the next section we will provide a general characterization of such restrictions, term them "non-future dependent" (NFD) and illustrate, in a realistic setting, how these restrictions might be exploited in practice. The formulation of this characterization does not imply prejudice against those PMM strategies that do allow for dependence on future outcomes. Rather, it should be viewed as a step in refining the missingness mechanism taxonomy. Indeed, the classification based on MCAR, MAR (ACMV), and MNAR is expanded by further splitting MNAR in non-future dependent and future-dependent mechanisms. Important examples exist of mechanisms that do not satisfy NFD, such as the models proposed by Wu and Bailey (1989). Also, Little (1995) allows for PMM where the missingness mechanism depends on unknown random coefficients, implying a dependence on future, possibly unobserved outcomes. Thus, our characterization enables one to select an appropriate missingness mechanism. If a dependence on an underlying latent variable is deemed plausible, a shared-parameter model can be chosen, whereas if missingness is believed to depend on current and past measurements, the proposed family is appropriate.

5.5 Future Dependence

Within the selection model context as an analogue to the terminology before we can define the NFD condition as "Missing Non Future dependent" (MNF) and the formulation is as follows.

$$f(r = t|y_1, \cdots, y_T) = f(r = t|y_1, \cdots, y_{t+1}),$$

or, alternatively, as

$$f(D = d|y_1, \cdots, y_T) = f(D = d|y_1, \cdots, y_d).$$

Taking into account the standard terminology as introduced by Little and Rubin (1987) it can be helpful to note that MAR is a special case of MNF and that MNF is a subclass of MNAR.

In the PMM framework, we can now define the NFD condition as a new type of identification restrictions termed "Non Future dependent Missing Value restrictions" (NFMV) and here the formulation is as follows:

$$\forall t \geq 2, \forall j < t-1 : f(y_t|y_1, \cdots, y_{t-1}, r = j) = f(y_t|y_1, \cdots, y_{t-1}, r \geq t-1). \tag{5.30}$$

NFMV is not a single set of restrictions, but rather leaves one conditional distribution per incomplete pattern unidentified:

$$f(y_t|y_1, \cdots, y_{t-1}, r = t-1). \tag{5.31}$$

In other words, the distribution (5.31) of the current outcome, given the previous ones is unconstrained. A possible choice for this distribution is the one in agreement with (5.30), in which case ACMV is obtained as a special case. On the other hand, if (5.31) is chosen in line with either NCMV or CCMV, then the combination of this choice with (5.30) does not yield NCMV or CCMV overall, showing that these restrictions are not consistent with MNF, as is clear from the following theorem.

Theorem 1 *For longitudinal data with dropouts, MNF* $\Longleftrightarrow$ *NFMV.*

The proof of this theorem is similar to the proof of Theorem 1 in Molenberghs, Michiels, Kenward and Diggle (1998): The MNF assumption states

$$f(r = t|y_1, \cdots, y_T) = f(r = t|y_1, \cdots, y_{t+1})$$

and the NFMV assumption reads

$$\forall t \geq 2, \forall j < t-1 : f(y_t|y_1, \cdots, y_{t-1}, r = j) = f(y_t|y_1, \cdots, y_{t-1}, r \geq t-1).$$

First, a lemma will be established.

Lemma 1 *In a longitudinal setting with dropout, NFMV* $\Longleftrightarrow \forall t \geq 2, \forall j < t-1 :$ $f(y_t|y_1, \cdots, y_{t-1}, r = j) = f(y_t|y_1, \cdots, y_{t-1})$.

Note that this lemma assumes strictly one less identification than the corresponding lemma in Molenberghs et al. (1998). Let us prove the lemma first.

Take $t \geq 2, j < t-1$, then NFMV leads to:

$$\begin{aligned}
& f(y_t|y_1,\cdots,y_{t-1}) \\
&= \sum_{i=1}^{t-2} f(y_t|y_1,\cdots,y_{t-1},r=i)f(r=i) + f(y_t|y_1,\cdots,y_{t-1},r\geq t-1)f(r\geq t-1) \\
&= \sum_{i=1}^{t-2} f(y_t|y_1,\cdots,y_{t-1},r=j)f(r=i) + f(y_t|y_1,\cdots,y_{t-1},r=j)f(r\geq t-1) \\
&= f(y_t|y_1,\cdots,y_{t-1},r=j)\left[\sum_{i=1}^{t-2} f(r=i) + f(r\geq t-1)\right] \\
&= f(y_t|y_1,\cdots,y_{t-1},r=j).
\end{aligned}$$

Conversely, again take $t \geq 2, j < t-1$:

$$\begin{aligned}
& f(y_t|y_1,\cdots,y_{t-1},r\geq t-1)f(r\geq t-1) \\
&= f(y_t|y_1,\cdots,y_{t-1}) - \sum_{i=1}^{t-2} f(y_t|y_1,\cdots,y_{t-1},r=i)f(r=i) \\
&= f(y_t|y_1,\cdots,y_{t-1}) - \sum_{i=1}^{t-2} f(y_t|y_1,\cdots,y_{t-1})f(r=i) \\
&= f(y_t|y_1,\cdots,y_{t-1})\left[1 - \sum_{i=1}^{t-2} f(r=i)\right] \\
&= f(y_t|y_1,\cdots,y_{t-1},r=j)\left[1 - \sum_{i=1}^{t-2} f(r=i)\right] \\
&= f(y_t|y_1,\cdots,y_{t-1},r=j)f(r\geq t-1).
\end{aligned}$$

This completes the proof of the lemma. We are now able to prove Theorem 1.

MNF $\Rightarrow$ NFMV

Consider the ratio Q of the complete data density to the density involving only the previous and current measurements. This gives, under the MNF assumption:

$$Q = \frac{f(y_1,\cdots,y_T)f(r=i|y_1,\cdots,y_{i+1})}{f(y_1,\cdots,y_{i+1})f(r=i|y_1,\cdots,y_{i+1})} = f(y_{i+2},\ldots,y_T|y_1,\ldots,y_{i+1}). \tag{5.32}$$

Further, one can always write:

$$\begin{aligned}
Q &= \frac{f(y_{i+2},\cdots,y_T|y_1,\cdots,y_i,r=i)f(y_1,\ldots,y_{i+1}|r=i)f(r=i)}{f(y_1,\ldots,y_{i+1}|r=i)f(r=i)} \\
&= f(y_{i+2},\ldots,y_T|y_1,\ldots,y_{i+1},r=i).
\end{aligned} \tag{5.33}$$

Equating (5.32) and (5.33):

$$f(y_{i+2}, \cdots, y_T | y_1, \cdots, y_{i+1}, r = i) = f(y_{i+2}, \cdots, y_T | y_1, \cdots, y_{i+1}). \tag{5.34}$$

Now, to show that (5.34) implies the NFMV conditions (5.30), we will proceed by induction on t. In case $t = 2$, NFMV imposes no restrictions and the result trivially holds.

Suppose by induction NFMV holds $\forall t \leq i$. We will now prove the hypothesis for $t = i + 1$. Choose $j \leq i$, arbitrary but fixed. Then from the induction hypothesis and Lemma 1, it follows that

$$\forall j < t - 1 \leq i - 1 : f(y_t | y_1, \cdots, y_{t-1}, r = j) \overset{\text{induction}}{=} f(y_t | y_1, \cdots, y_{t-1}, r \geq t - 1) \overset{\text{lemma 1}}{=} f(y_t | y_1, \cdots, y_{t-1}).$$

Taking the product over $t = j + 2, \cdots, i$ then gives

$$f(y_{j+2}, \cdots, y_i | y_1, \cdots, y_{j+1}, r = j) = f(y_{j+2}, \cdots, y_i | y_1, \cdots, y_{j+1}). \tag{5.35}$$

After integration over $y_{i+2}, \cdots, y_T$, equation (5.34) leads to

$$f(y_{j+2}, \cdots, y_{i+1} | y_1, \cdots, y_{j+1}, r = j) = f(y_{j+2}, \cdots, y_{i+1} | y_1, \cdots, y_{j+1}). \tag{5.36}$$

Dividing (5.36) by (5.35) and equating the left and right hand sides, we find that

$$f(y_{i+1} | y_1, \cdots, y_i, r = j) = f(y_{i+1} | y_1, \cdots, y_i).$$

This holds $\forall j \leq i - 1$, and Lemma 1 shows this is equivalent with NFMV.

NFMV $\Rightarrow$ MNF

Starting from the NFMV assumption and Lemma 1, we have

$$\forall t \geq 2, \forall j < t - 1 : f(y_t | y_1, \cdots, y_{t-1}, r = j) = f(y_t | y_1, \cdots, y_{t-1}). \tag{5.37}$$

We now factorize the full data density as

$$\begin{aligned} f(y_1, \cdots, y_T, r = i) &= f(y_1, \cdots, y_{i+1}, r = i) f(y_{i+2}, \cdots, y_T | y_1, \cdots, y_{i+1}, r = i) \\ &= f(y_1, \cdots, y_{i+1}, r = i) \prod_{t=i+2}^{T} f(y_t | y_1, \cdots, y_{t-1}, r = i). \end{aligned}$$

Using (5.37), it follows that

$$
\begin{aligned}
& f(y_1, \cdots, y_T, r = i) && (5.38)\\
&= f(y_1, \cdots, y_{i+1} | r = i) f(r = i) \prod_{t=i+2}^{T} f(y_t | y_1, \cdots, y_{t-1}) \\
&= f(y_1, \cdots, y_{i+1} | r = i) f(r = i) f(y_{i+2}, \cdots, y_T | y_1, \cdots, y_{i+1}) \\
&= \frac{f(y_1, \cdots, y_{i+1} | r = i) f(r = i)}{f(y_1, \ldots, y_{i+1})} f(y_1, \cdots, y_{i+1}) f(y_{i+2}, \cdots, y_T | y_1, \cdots, y_{i+1}) \\
&= \frac{f(y_1, \cdots, y_{i+1} | r = i) f(r = i)}{f(y_1, \ldots, y_{i+1})} f(y_1, \cdots, y_T) \\
&= f(r = i | y_1, \ldots, y_{i+1}) f(y_1, \ldots, y_T) && (5.39)
\end{aligned}
$$

An alternative factorization of $f(y, r)$ gives

$$f(y_1, \cdots, y_T, r = i) = f(r = i | y_1, \cdots, y_T) f(y_1, \cdots, y_T). \tag{5.40}$$

It follows from (5.38) and (5.40) that

$$f(r = i | y_1, \cdots, y_T) = f(r = i | y_1, \cdots, y_{i+1}),$$

completing the proof of Theorem 1.

5.6 NFMV Identifying Restrictions

For pattern t, the complete data density is given by

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

As before, 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, modeled using the observed data. Then, identifying restrictions are applied in order to identify the second component. Let us outline how NFMV restrictions can be implemented:

- From the data, identify $f_t(y_1, \cdots, y_t)$.
- The user has full freedom to choose $f_t(y_{t+1} | y_1, \cdots, y_t)$. Substantive considerations can be used to identify this density. Or a family of densities can be considered by way of sensitivity analysis.

- Using (5.30), the densities $f_t(y_j|y_1, \cdots, y_{j-1})$, $(j \geq t+2)$ are identified. This identification involves not only the patterns for which y_j is observed, but also the pattern for which y_j is the current, the first unobserved measurement.

From the first and the second item, it follows that $f_t(y_1, \cdots, y_{t+1})$ is identified from modeling and choice. Next, NFMV states that

$$f_t(y_s|y_1, \ldots, y_{s-1}) = f_{(\geq s-1)}(y_s|y_1, \ldots, y_{s-1}), \tag{5.41}$$

for $s = t+2, \ldots, T$. Now, we can transform (5.41) as follows:

$$\begin{aligned} & f_t(y_s|y_1, \ldots, y_{s-1}) \\ &= f_{(\geq s-1)}(y_s|y_1, \ldots, y_{s-1}) \\ &= \frac{\sum_{j=s-1}^{T} \alpha_j f_j(y_1, \ldots, y_s)}{\sum_{j=s-1}^{T} \alpha_j f_j(y_1, \ldots, y_{s-1})} \\ &= \sum_{j=s-1}^{T} \frac{\alpha_j f_j(y_1, \ldots, y_{s-1})}{\sum_{j=s-1}^{T} \alpha_j f_j(y_1, \ldots, y_{s-1})} f_j(y_s|y_1, \ldots, y_{s-1}). \end{aligned}$$

Thus, a general expression is:

$$f_t(y_s|y_1, \ldots y_{s-1}) = \sum_{j=s-1}^{T} \omega_{sj} f_j(y_s|y_1, \ldots y_{s-1}), \quad s = t+2, \ldots, T, \tag{5.42}$$

with

$$\omega_{sj} = \frac{\alpha_j f_j(y_1, ..., y_{s-1})}{\sum_{\ell=s-1}^{T} \alpha_\ell f_\ell(y_1, ..., y_{s-1})}.$$

Choosing ω_{sj} different from the ones specified above yields missing data mechanisms that do depend on future observations. In an integrated sensitivity analysis, it can be envisaged that the impact of such departures on substantive conclusions might be explored. NFMV is promising for sensitivity analysis. Indeed, in the general MNAR case, the conditional distribution of the unobserved measurements given the observed ones needs to be determined by means of assumptions. Under NFMV, only the conditional distribution of the *first* ("current") unobserved outcome given the observed ones needs to be identified by assumption. Thus, when MNF is deemed plausible, one combines the flexibility of a broad class of models with a sensitivity space that is reasonably easy to manage. In the special case of MAR, the conditional distributions of the unobserved outcomes are completely identified by means of ACMV and there is no further room for sensitivity analysis.

As an illustration, let us again consider the special case of three measurements and three patterns. As we have seen before general identifying restrictions take the following form:

$$\begin{aligned} f_3(y_1,y_2,y_3) &= f_3(y_1,y_2,y_3), && (5.43)\\ f_2(y_1,y_2,y_3) &= f_2(y_1,y_2)f_3(y_3|y_1,y_2), && (5.44)\\ f_1(y_1,y_2,y_3) &= f_1(y_1)\left[\omega f_2(y_2|y_1)+(1-\omega)f_3(y_2|y_1)\right] \\ &\quad \times f_3(y_3|y_1,y_2). && (5.45) \end{aligned}$$

and we can completely identify $f_3(y_1,y_2,y_3)$ from the data while $f_2(y_1,y_2,y_3)$ can be identified by one possible restriction only. Therefore, only the identification of $f_1(y_1,y_2,y_3)$ leaves room for choice. Setting $\omega = 1$ corresponds to NCMV, while $\omega = 0$ implies CCMV. ACMV corresponds to

$$\omega = \frac{\alpha_2 f_2(y_1)}{\alpha_2 f_2(y_1)+\alpha_3 f_3(y_1)}. \qquad (5.46)$$

For NFMV, we obtain:

$$\begin{aligned} f_3(y_1,y_2,y_3) &= f_3(y_1,y_2,y_3),\\ f_2(y_1,y_2,y_3) &= f_2(y_1,y_2)g_2(y_3|y_1,y_2),\\ f_1(y_1,y_2,y_3) &= f_1(y_1)g_1(y_2|y_1)\left[\omega g_2(y_3|y_1,y_2)+(1-\omega)f_3(y_3|y_1,y_2)\right], \quad (5.47) \end{aligned}$$

where the notation $g_2(y_3|y_1,y_2)$ and $g_1(y_2|y_1)$ is used to indicate a free choice by the modeler. In contrast to 5.44), the mixture, weighted by ω, is on the distribution of

$$f_1(y_3|y_1,y_2) = f_{(\geq 2)}(y_3|y_1,y_2).$$

Explicitly,

$$\omega = \frac{\alpha_2 f_2(y_1,y_2)}{\alpha_2 f_2(y_1,y_2)+\alpha_3 f_3(y_1,y_2)}$$

and hence, the full conditional can be written as:

$$f_{(\geq 2)}(y_3|y_1,y_2) = \frac{\alpha_2 f_2(y_1,y_2,y_3)+\alpha_3 f_3(y_1,y_2,y_3)}{\alpha_2 f_2(y_1,y_2)+\alpha_3 f_3(y_1,y_2)}.$$

In case we opt for ACMV for the two distributions that can be chosen freely, i.e., $g_2(y_3|y_1,y_2) = f_3(y_3|y_1,y_2)$ and $g_1(y_2|y_1)$ as in (5.44) with ω as given by (5.46), it is clear that (5.47)–(5.47) coincide with (5.43)–(5.45). Obviously, choosing the free distributions to be either NCMV or CCMV, does not imply that (5.47)–(5.47) as a whole follow NCMV or CCMV. In other words, this illustrates that in general NCMV and CCMV do not satisfy the MNF definition.

5.7 The Behave Dataset

A linear mixed model (Verbeke and Molenberghs 2000) is fitted to the outcomes. Effects considered in the model were, apart from treatment effect: age, time, investigator, and country, as well as 2- and 3-way interactions. From an initial model selection, only main effects of age, time, time^2 and treatment group were retained. The variance structure is modeled using a random subject effect, an exponential serial correlation process, and measurement error.

We first consider the selection model approach. As in Diggle and Kenward (1994), we combine the measurement model with a logistic regression for dropout with either only an intercept (MCAR), also an effect for previous outcome (MAR), or even an effect of the current possibly unobserved measurement (MNAR). The fitted average profiles are plotted in Figure 5.21. The treatment effect is not significant, with $p = 0.245$ under MCAR and MAR, and $p = 0.262$ under MNAR. Comparing the MAR and MNAR models, the likelihood ratio statistic is 5.4 on 2 degrees of freedom. While this might be taken as some evidence for non-random dropout, such a conclusion is strongly model dependent (Verbeke and Molenberghs 2000, Ch. 19). More reasonably it can be taken as some evidence against this particular MAR model.

Next, we turn attention to the pattern-mixture strategies. The SAS procedure MIXED was used for fitting these pattern-specific models.

Three families of identification strategies are considered:

- ACMV, which is equivalent to MAR and therefore also MNF.
- CCMV and NCMV are considered, as was done in Thijs et al. (2000) for a different set of data. These mechanisms are incompatible with MNF.
- Two MNF mechanisms are considered. The first one, FD1, considers CCMV for the conditional distribution of the current measurement, given the past. The second one, FD2, considers NCMV instead.

Focusing first on the overall, marginal treatment effect, we obtain the following set of p values:

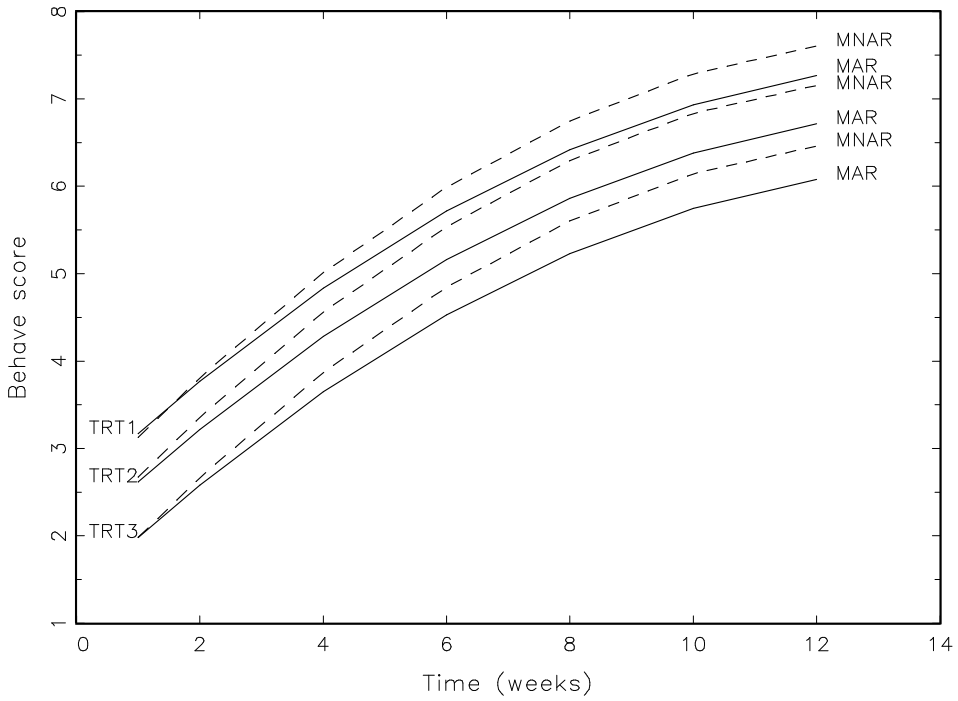

Figure 5.21: *Behave dataset, fitted average profiles using selection models.*

$$\begin{aligned} &\text{ACMV} : p = 0.0407 \\ &\text{CCMV} : p = 0.0002 \quad \text{MNF1} : p = 0.0413 \\ &\text{NCMV} : p = 0.0024 \quad \text{MNF2} : p = 0.0245. \end{aligned}$$

All p values are significant, in contrast to the selection model setting. Of course, here we are considering the treatment effect, corrected for dropout pattern. Further, even though all p values are significant, there is substantial difference between the strategies. Mechanisms such as CCMV and NCMV, may be overly optimistic. The three MNF mechanisms (ACMV, MNF1, MNF2) provide reasonably similar evidence.

Figures 5.22–5.24 graphically summarize the fit of these models. Clearly, the identifying restrictions chosen have a strong impact, especially for the patterns with earlier dropout. Of course, from Table 5.11 it is clear that the earlier patterns are rather sparsely filled. It is striking to see that the MNF patterns are not all grouped together.

Table 5.11: *Vorozole study, sample size per treatment arm and dropout pattern.*

Pattern	1	2	3	4	5	6	7
Treatment 1	4	5	16	3	9	6	71
Treatment 2	4	9	7	6	3	5	81
Treatment 3	12	4	15	9	5	3	67

5.8 Concluding Remarks

In conclusion we now have argued, in accordance with Thijs et al. (2000), that PMM, with in addition identifying restrictions to specify the conditional distribution of the unobserved measurements, given the observed ones in a given pattern, is a potentially useful way to model incomplete longitudinal data. In particular, such restrictions allow one to reflect carefully on the nature of the assumptions made. For example, a particular set of restrictions, termed ACMV, corresponds to MAR. Here, we established a family of MNAR models, termed MNF (non-future dependent), which avoid dependence of dropout on future, unobserved outcomes. Not only does this family embed, again, MAR, it provides a sensibly yet wide space within which sensitivity analysis can be conducted.

We believe that our approach can play a useful role, as a member of a collection of sensitivity tools. Of course, a sensitivity analysis can be conducted within different frameworks, and there are times where the setting will determine which framework is the more appropriate one (for example Bayesian or frequentist), in conjunction with technical and computational considerations. Draper (1995) has considered ways of dealing with uncertainty in the very natural Bayesian framework and developments in the missing value setting are ongoing. A thorough comparison between the various frameworks will be interesting and worth undertaking in the future.

The SAS and GAUSS macros which have been used to carry out the multiple imputation related tasks are available on the internet and therefore it has become relatively simple with respect to computations to apply a similar type of analyzes to any other dataset.

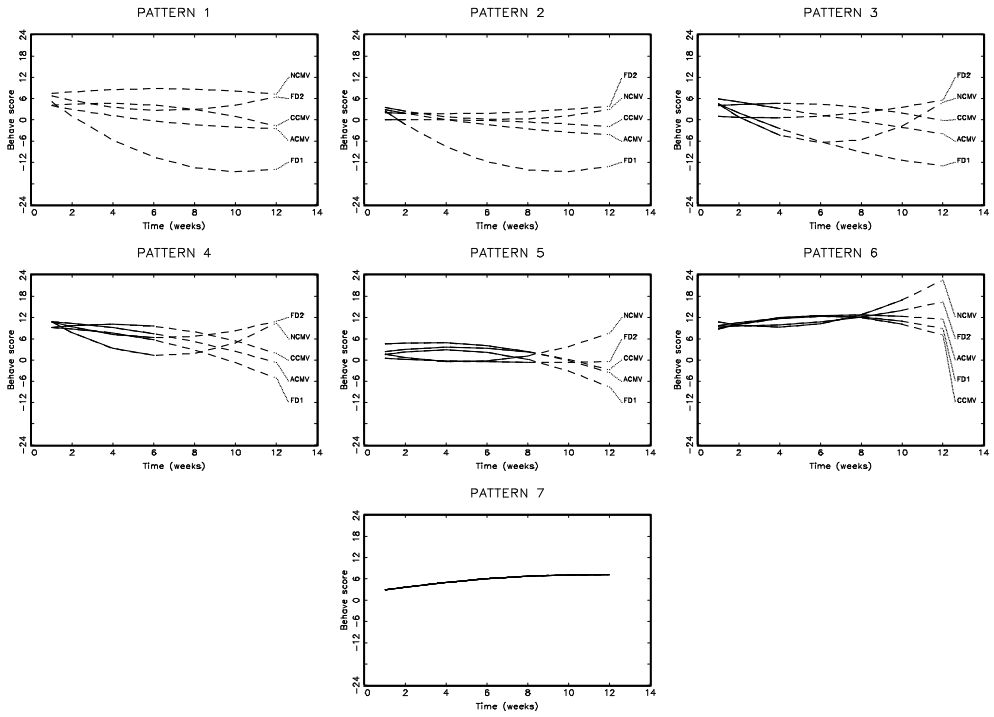

Figure 5.22: *Behave dataset, pattern-mixture models, fitted average profiles for each of the five identification strategies are presented, the solid portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in the dashed line, treatment arm 1.*

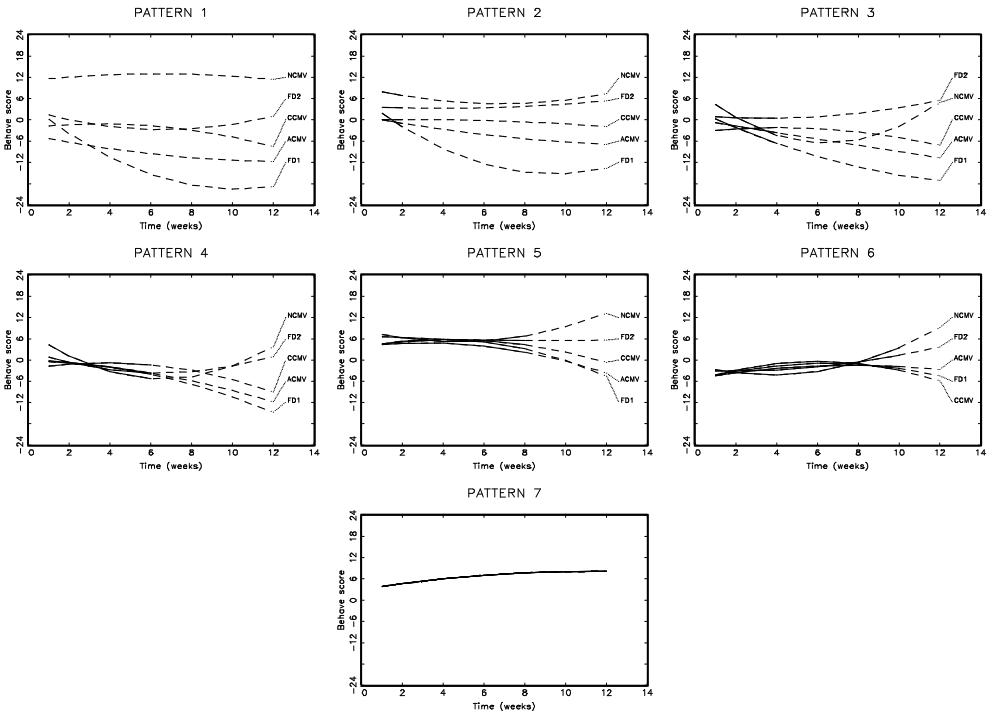

Figure 5.23: *Behave dataset, pattern-mixture models, fitted average profiles for each of the five identification strategies are presented, the solid portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in the dashed line, treatment arm 2.*

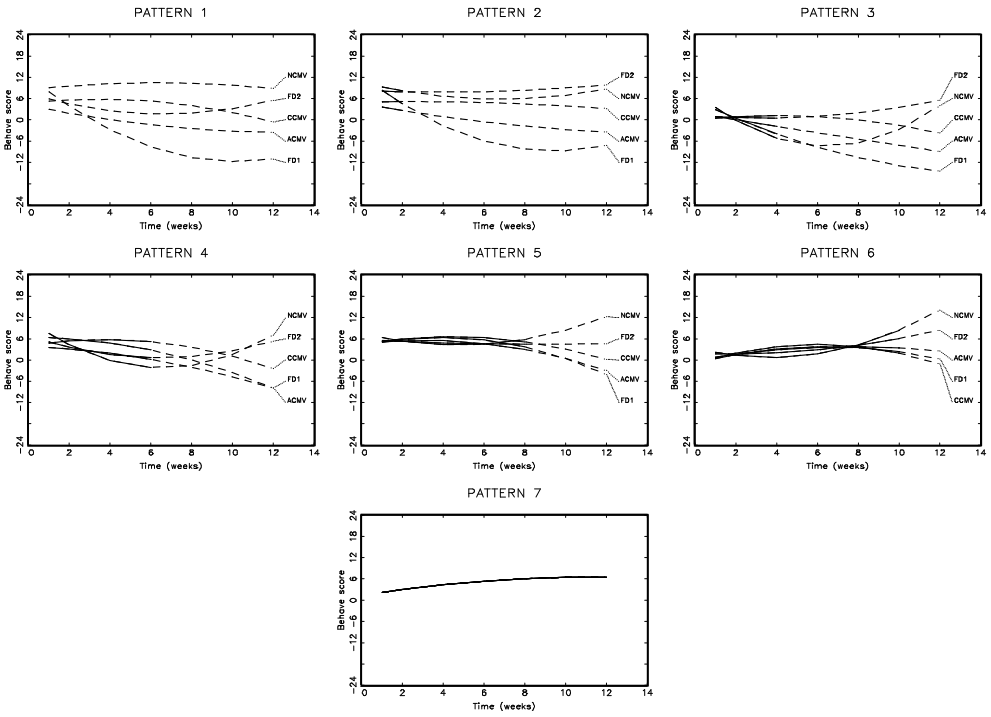

Figure 5.24: *Behave dataset, pattern-mixture models, fitted average profiles for each of the five identification strategies are presented, the solid portion of the curves runs from baseline until the last obtained measurement, while the extrapolated piece is shown in the dashed line, treatment arm 3.*

Chapter 6

Case Study

In this chapter we will discuss the Milk protein trial as introduced in Chapter 2 in full detail. We will start by briefly reconsider the analyses done by Diggle and Kenward (1994) and Kenward (1998) where a first attempt was made towards sensitivity analysis. Thereafter we will apply our own methodology of local and global influence and we will compare the results of both analyses in order to give an as complete as possible overview of the sensitivity of the results to the model assumptions made.

6.1 Previous Work

Diggle and Kenward (1994) considered a measurement model similar to model 4.1 including separate intercepts for the barley (μ_1), mixed (μ_2) and lupins (μ_3) groups, and a common time effect (β) which is linear during the first three weeks and constant thereafter. The covariance structure is described by a random intercept, an exponential serial process, and measurement error. For example, for the barley diet group:

$$Y_{ij} = \mu_1 + b_i + \beta t_{ij} I(t_{ij} \leq 3) + w_{ij} + \varepsilon_{ij},$$

where $b_i \sim N(0, d)$, the w_{ij} have variance τ^2 and serial correlation ρ, and $\varepsilon_{ij} \sim N(0, \sigma^2)$. The dropout model includes dependence on the previous and current, possibly unobserved, measurements. Since dropout only happens from week 15 onwards, Diggle and Kenward (1994) chose to set the dropout probability for earlier occasions

equal to zero. Thereafter, they allowed separate intercepts per time point, but common dependencies on previous and current measurements. Precisely, their model can be expressed as follows. Equation (4.6) specializes to:

$$\text{logit}[g(y_{i,j-1}, y_{ij})] = \psi_{0j} + \psi_1 y_{i,j-1} + \psi_2 y_{ij}, \quad (j = 15, 16, 17, 19).$$

To acknowledge the fact that dropout starts from week 15 onwards, the product in (4.8) is over $j = 15, \ldots, n_i$ instead of over $j = 2, \ldots, n_i$.

Diggle and Kenward (1994) found that the dropout model is non-random. In view of the comments by Cullis (1994) and the sensitivity of conclusions to model assumptions, great care is needed. Curran, Pignatti, and Molenberghs (1998) assessed sensitivity of this conclusion by means of two alternative modeling strategies. First, they acknowledged the possibility of ragged entry and a fixed termination data, rather than more conventional dropout. To this end, the individual profiles were reversed and right aligned. The conclusions thus obtained do not contradict those from Diggle and Kenward (1994). Second, they considered dropout occasion as an (imperfect) surrogate for paddock to which diet was assigned. This leads naturally to a stratified analysis based on dropout occasion (weeks 15, 16, 17, or 19). In other words, a *pattern-mixture* analysis (Little 1993, 1994) was conducted. One then obtains dropout pattern specific diet effects, which can be combined to yield the marginal diet effects. Again, these results were in good agreement with those from Diggle and Kenward (1994). However, it is still possible that a set of observations is responsible for, e.g., conclusions about the dropout mechanism, in all analyses performed thus far. This provides additional motivation for an influence analysis.

6.2 Sensitivity Analysis

We will now introduce two models which use the same measurement model as Diggle and Kenward (1994) but different dropout models. This will allow us to illustrate how the choice of dropout model can have an important impact on the substantial conclusion.

A first dropout model is closely related to the one of Diggle and Kenward (1994) who defined occasion-specific intercepts ψ_{0j} $(j = 15, 16, 17, 19)$, assumed common slopes and set the dropout probability equal to zero at other occasions. We also model dropout from week 15 onwards but we will keep the intercepts constant for

occasions 15 to 19. Precisely, our first model contains three parameters (intercept ψ_0, dependence on the previous measurement ψ_1, and dependence on the current measurement ψ_2), which produces the following version of (4.6):

$$\text{logit}[g(y_{i,j-1}, y_{ij})] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij}. \tag{6.1}$$

As in the Diggle and Kenward (1994) model, the product in (4.8) is over $j = 15, \ldots, n_i$ instead of over $j = 2, \ldots, n_i$.

Parameter estimates for this model under both MAR and MNAR, are listed in Tables 6.1 – 6.4. A number of additional analyses, also presented in these tables, will be discussed in section 6.2.4. The fitted model is qualitatively equivalent to the model used by Diggle and Kenward (1994), who concluded overwhelming evidence for non-random dropout (likelihood ratio statistic 13.9). In line with these results we also could decide in favor of a non-random process (likelihood ratio statistic 14.59).

In our second dropout model we allow dropout to start from the second week. More precisely, model (6.1) is retained, while the product in (4.8) is over $j = 2, \ldots, n_i$, in agreement with the original definition. Careful reflection on the status of this model is needed. While on the one hand it may seem a natural choice, given also the availability of this model in standard software packages such as Oswald (Smith, Robertson, and Diggle 1996), it may raise doubts since no dropout was observed during the first 14 weeks. Therefore, it is interesting to study this model and its impact on model parameters as well as on the conclusions from an influence analysis.

The fitted model is listed in Table 6.5. A striking difference with the previous analysis is that the MAR assumption is borderline not rejected (likelihood ratio statistic 3.63). Apparently, the onset of dropout is a major source of sensitivity, to be explored further. As results from theory, the measurement model parameters do not change under the MAR model, compared to those displayed in Tables 6.1 – 6.4. The measurement model obtained under MNAR has changed only slightly.

Which of the two analyses is to be preferred is debatable and depends on substantive considerations also, rather than on statistical ones only. Recall that the first analysis accounts for the *post hoc observation* that no dropout occurred prior to week 15. However, there is a, perhaps small, chance for the experiment to terminate in a field prior to week 15, and our second model acknowledges this possibility. Nevertheless, should dropout occur prior to week 15 (e.g., when the experiment is repeated), it is likely to occur at a lower rate than later in the sequence. The second model is not able to acknowledge this, since it assumes a constant dropout rate, and hence may

Table 6.1: *Milk protein trial, maximum likelihood estimates (standard errors) of the random dropout model, dropout starts from week 15 onwards, the entire set of data is contrasted with several deletion schemes, (1) removal of #51, #59, and #68; (2) removal of #1, #7, #38, #43, #51, #59, #65, #68, and #74.*

		RANDOM DROPOUT		
Effect	Par.	all	set 1	set 2
Measurement model:				
Barley	μ_1	4.147(0.053)	4.134(0.052)	4.132(0.053)
Mixed	μ_2	4.046(0.052)	4.020(0.052)	4.042(0.053)
Lupins	μ_3	3.935(0.052)	3.950(0.052)	3.957(0.053)
Time effect	β	-0.226(0.015)	-0.221(0.015)	-0.226(0.015)
Rand. int. var.	d	-0.001(0.010)	-0.007(0.011)	-0.005(0.010)
Meas. err. var.	σ^2	0.024(0.002)	0.024(0.002)	0.023(0.002)
Ser. var.	τ^2	0.073(0.012)	0.074(0.013)	0.069(0.012)
Ser. corr.	ρ	0.152(0.037)	0.145(0.037)	0.152(0.039)
Dropout model:				
Intercept	ψ_0	17.87(3.15)	28.69(4.97)	
Prev. meas.	ψ_1	-6.02(1.00)	-9.39(1.58)	
-2 loglikelihood		51.844	14.575	-43.894
Wald (diet) (2 d.f.)		17.27	14.42	12.25
p value		0.0002	0.0007	0.0022

fail to provide an adequate description. The first model on the other hand, makes the reasonable assumption that dropout is absent during the first period, and occurs at an approximately constant rate thereafter. Therefore, this model should deserve our preference. In any case, it is clear that there is an enormous sensitivity of the results due to this model choice and hence substantial reflection on the structure of the dropout process is necessary.

6.2.1 Global Influence

Global influence results are shown in Figures 6.1–6.4. They are based on fitting a MNAR model for each cow deleted. The Cook's distances for the first and the second model are shown in Figures 6.2 and 6.4 respectively. The individual curves with influential subjects highlighted are plotted in Figure 6.1 where subject #38 pertains

Table 6.2: *Milk protein trial, maximum likelihood estimates (standard errors) of the random dropout model, dropout starts from week 15 onwards, the entire set of data is contrasted with several deletion schemes, (b,m); barley and mixed diets only; (b,l): barley and lupins diets only; (m,l): mixed and lupins diets only.*

		RANDOM DROPOUT		
Effect	Par.	(b,m)	(b,l)	(m,l)
Measurement model:				
Barley	μ_1	4.163(0.059)	4.104(0.061)	–
Mixed	μ_2	4.062(0.058)	–	4.071(0.058)
Lupins	μ_3	–	3.893(0.060)	3.959(0.058)
Time effect	β	-0.232(0.018)	-0.210(0.018)	-0.235(0.018)
Rand. int. var.	d	0.001(0.011)	-0.003(0.018)	-0.005(0.011)
Meas. err. var.	σ^2	0.023(0.003)	0.023(0.003)	0.025(0.003)
Ser. var.	τ^2	0.067(0.013)	0.081(0.021)	0.073(0.014)
Ser. corr.	ρ	0.159(0.045)	0.128(0.046)	0.166(0.045)
Dropout model:				
Intercept	ψ_0	17.87(3.15)	28.69(4.97)	33.41(6.15)
Prev. meas.	ψ_1	-6.02(1.00)	-9.39(1.58)	-10.84(1.95)
-2 loglikelihood		-14.365	29.516	76.801
Wald (diet) (2 d.f.)		–	–	–
p value		–	–	–

to the first model only.

There is very little difference in some of the Cook's distance plots, when Figures 6.2 and 6.4 are compared. Precisely, CD_{1i}, $CD_{2i}(\gamma)$, $CD_{2i}(\theta)$ are virtually identical. The three others are similar in the sense that there is some overlap in the subjects indicated as peaks, but with varying magnitudes. Subject #38 is influential on the dropout measures $CD_{2,38}(\psi,\omega)$, $CD_{2,38}(\psi)$, and $CD_{2,38}(\omega)$. This is not surprising since #38 is rather low in the middle portion of the measurement sequence, while it is very high from week 15 onwards. Therefore, this sequence is picked up in the second analysis only. By studying plots with the evolution of the parameters separately during the deletion process (not shown here) we can conclude that subject #38 has some impact on the serial correlation parameter while #65 is rather influential for the measurement error. In view of the fairly smooth deviation from a straight line of the former and the abrupt peaks in the latter, this is not a surprise.

Table 6.3: *Milk protein trial, maximum likelihood estimates (standard errors) of the non-random dropout model, dropout starts from week 15 onwards, the entire set of data is contrasted with several deletion schemes, (1) removal of #51, #59, and #68; (2) removal of #1, #7, #38, #43, #51, #59, #65, #68, and #74.*

		NON-RANDOM DROPOUT		
Effect	Par.	all	set 1	set 2
Measurement model:				
Barley	μ_1	4.152(0.053)	4.138(0.052)	4.136(0.053)
Mixed	μ_2	4.050(0.052)	4.022(0.051)	4.046(0.053)
Lupins	μ_3	3.941(0.052)	3.954(0.052)	3.961(0.053)
Time effect	β	-0.224(0.015)	-0.219(0.015)	-0.225(0.015)
Rand. int. var.	d	0.002(0.009)	-0.004(0.010)	-0.002(0.010)
Meas. err. var.	σ^2	0.025(0.002)	0.025(0.002)	0.024(0.002)
Ser. var.	τ^2	0.067(0.011)	0.070(0.012)	0.064(0.011)
Ser. corr.	ρ	0.163(0.039)	0.151(0.039)	0.162(0.042)
Dropout model:				
Intercept	ψ_0	15.64(3.54)	25.30(5.06)	30.87(6.63)
Prev. meas.	ψ_1	-10.72(2.02)	-11.99(2.26)	-15.06(3.31)
Curr. meas.	$\omega \equiv \psi_2$	5.18(1.49)	3.56(1.60)	4.84(2.12)
-2 loglikelihood		37.257	9.620	-50.210
Wald (diet) (2 d.f.)		17.31	14.55	12.27
p value		0.0002	0.0007	0.0022
G^2 for MNAR (1 d.f.)		14.59	4.96	6.32
p value		0.0001	0.0260	0.0120

Figure 6.3 considers the CD_{2i} measures for the diet group contrasts. While Figure 6.2 revealed some influence on the measurement model parameters, it is clear this is not affecting the diet contrasts. Of course, there is virtually no influence coming from the cows in the diet group which does not contribute to the corresponding contrast.

Based on our second model all forms of $CD_{2i}(.)$, whether based on the entire parameter vector $\boldsymbol{\gamma}$, the dropout parameters (ψ_0, ψ_1, ω), or subsets of the latter, indicate that subjects #51, #59, and #68 are influential. In contrast, CD_{1i} which is based directly on the likelihood, does not reveal these subjects, but rather subject #65 jumps out. Thus, while the former three subjects have a substantial impact on the parameter estimates, they do not change the likelihood in a noticeable way. From

Table 6.4: *Milk protein trial, maximum likelihood estimates (standard errors) of the non-random dropout model, dropout starts from week 15 onwards, the entire set of data is contrasted with several deletion schemes, (b,m); barley and mixed diets only; (b,l): barley and lupins diets only; (m,l): mixed and lupins diets only.*

		NON-RANDOM DROPOUT		
Effect	Par.	(b,m)	(b,l)	(m,l)
Measurement model:				
Barley	μ_1	4.166(0.059)	4.112(0.061)	–
Mixed	μ_2	4.064(0.058)	–	4.075(0.058)
Lupins	μ_3	–	3.901(0.060)	3.965(0.058)
Time effect	β	-0.230(0.018)	-0.208(0.018)	-0.232(0.018)
Rand. int. var.	d	0.002(0.010)	0.004(0.016)	0.001(0.010)
Meas. err. var.	σ^2	0.024(0.003)	0.024(0.003)	0.026(0.003)
Ser. var.	τ^2	0.064(0.012)	0.072(0.017)	0.065(0.012)
Ser. corr.	ρ	0.163(0.046)	0.143(0.051)	0.183(0.049)
Dropout model:				
Intercept	ψ_0	23.11(5.45)	12.70(4.15)	15.25(4.45)
Prev. meas.	ψ_1	-10.74(2.34)	-11.02(2.65)	-11.20(2.64)
Curr. meas.	$\omega \equiv \psi_2$	3.10(1.75)	6.31(2.14)	5.68(1.84)
-2 loglikelihood		-17.459	17.196	64.805
Wald (diet) (2 d.f.)		–	–	–
p value		–	–	–
G^2 for MNAR (1 d.f.)		3.09	12.32	12.00
p value		0.0786	0.0004	0.0005

a plot of the dropout parameter estimates for each deleted case (not shown here) it is very clear that upward peaks in $\widehat{\psi}_{0(-i)}$ for subjects #51 and #59 are compensated with downward peaks in $\widehat{\omega}_{(-i)}$. An explanation for this phenomenon can be found in the variance-covariance matrix of the dropout parameters (correlations shown in the lower triangle):

$$\begin{pmatrix} 8.22 & 0.43 & -2.85 \\ (0.14) & 1.14 & -1.18 \\ (-0.71) & (-0.79) & 1.94 \end{pmatrix}.$$

From a principal components analysis it follows that more than 90% of the variation

Table 6.5: *Milk protein trial, maximum likelihood estimates (standard errors) of random and non-random dropout models.*

Effect	Parameter	MAR	MNAR
Measurement model:			
Barley	μ_1	4.147 (0.053)	4.152 (0.053)
Mixed	μ_2	4.046 (0.052)	4.050 (0.052)
Lupins	μ_3	3.935 (0.052)	3.941 (0.052)
Time effect	β	-0.226 (0.015)	-0.224 (0.015)
Random intercept variance	d	-0.001 (0.010)	0.002 (0.009)
Measurement error variance	σ^2	0.024 (0.002)	0.025 (0.002)
Serial variance	τ^2	0.073 (0.012)	0.067 (0.011)
Serial correlation	ρ	0.152 (0.037)	0.163 (0.040)
Dropout model:			
Intercept	ψ_0	10.483 (2.010)	6.477 (2.867)
Previous measurement	ψ_1	-4.326 (0.651)	-5.917 (1.069)
Current measurement	$\omega \equiv \psi_2$		2.732 (1.396)
-2 loglikelihood		194.316	190.691

is captured by the linear combination $0.93\psi_0 - 0.37\omega$. Hence, there is mass transfer between these two parameters, of course with sign reversal, with little impact on the likelihood value, and little effect on the MAR parameter ψ_1.

Let us now turn to the subjects which are globally influential. A first and common reason for those subjects to show up is the fact that they all have a rather strange profile. Remember the overall trend to be sloping downwards during the first three weeks and constant thereafter. Subject #65 appears with large $CD_{65,1}$ and large $CD_2(\theta)$. The reason for this can be found in the fact that its profile shows extremely low and high peaks. Subjects #51, #59 and #68 on the other hand only show large values for $CD_2(\psi,\omega), CD_2(\psi), CD_2(\omega)$. This means that these subjects are influential for the dropout parameters. For subject #51 this can be explained by the fact that it drops out in spite of the rather high profile. Subjects #59 and #68 on the contrary, stay in the experiment though they both have rather low profiles.

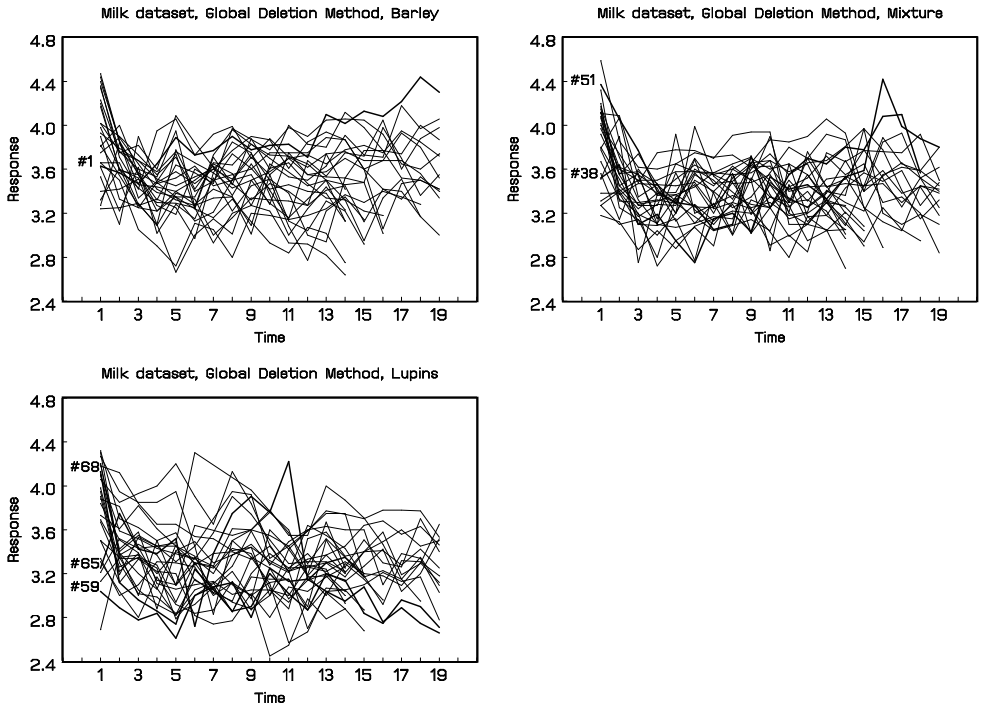

Figure 6.1: *Milk protein trial, individual profiles, with globally influential subjects highlighted, dropout modeled from week 15.*

6.2.2 Local Influence

Local influence plots and individual profiles, with the influential subjects highlighted, for the first model, respectively, are depicted in Figures 6.5–6.8. Corresponding graphs for our second model are shown in Figures 6.9–6.12. It is slightly easier to discuss results of the second model up front and then compare them to the first model. Two versions are considered, based on two equivalent forms of the dropout model linear predictor

$$\psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij} = \lambda_0 + \lambda_1 y_{i,j-1} + \lambda_2 (y_{ij} - y_{i,j-1}). \tag{6.2}$$

The standard analysis, corresponding to the left hand side, is termed *raw analysis*, while the right hand side refers to an *incremental* parameterization.

Observe that the plots for C_i and $C_i(\boldsymbol{\psi})$ are virtually identical in Figure 6.5. This is due to the relative magnitudes of the $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ components. Profiles #51, #59, and #66–#68 are highlighted in Figure 6.10. An explanation for the influence in $\boldsymbol{\psi}$ is found by studying (4.28) in more detail. Let us therefore simplify this expression

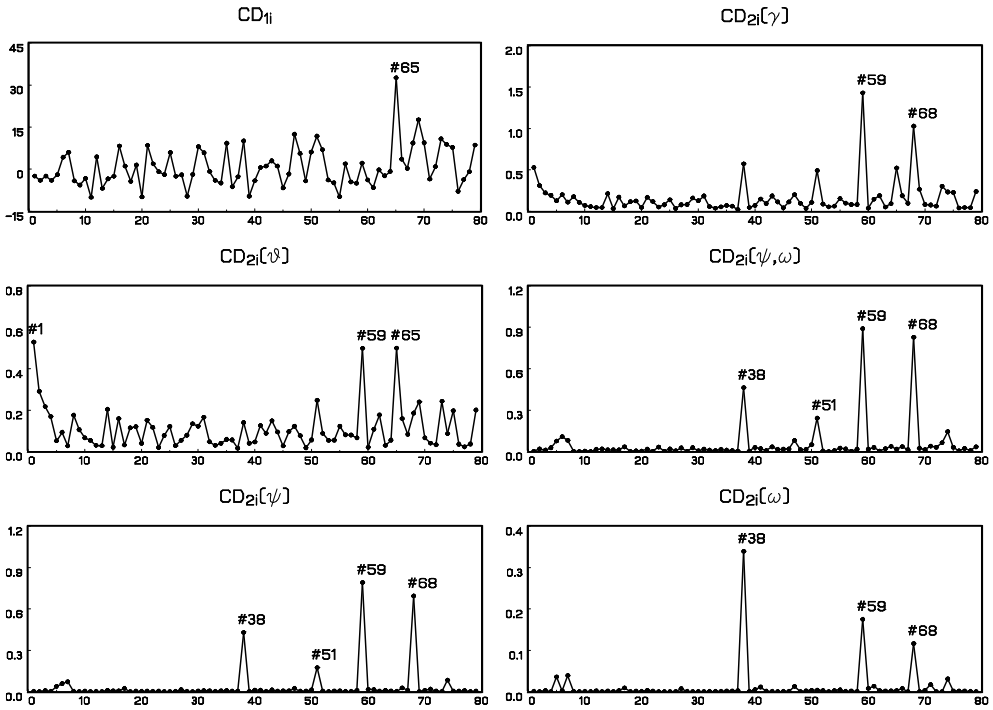

Figure 6.2: *Milk protein trial, index plots of* CD_{1i}, $CD_{2i}(\boldsymbol{\gamma})$, $CD_{2i}(\theta)$, $CD_{2i}(\boldsymbol{\psi}, \omega)$, $CD_{2i}(\boldsymbol{\psi})$, $CD_{2i}(\omega)$, *dropout modeled from week 15.*

to

$$F(y) = y^2 g(1-g), \tag{6.3}$$

which is based on the assumption that previous and current measurements are approximately equal. Given estimates for $\boldsymbol{\psi}$ it is easy to determine numerically when this function is maximal. Apparently, for $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ as in Table 6.5, the maximum is obtained for $y = 2.51$, exactly as seen in the influential profiles, which are all in the lupins group (Figure 6.10). Further note that there is some agreement between the locally and globally influential subjects though there is no compelling need for the two approaches to be identical (#51 appears in different influential components in the two approaches). Indeed, while global influence lumps together all sources of influence, our local influence approach is designed to detect subjects which, due to several causes, tend to have a strong impact on ω and therefore on the conclusion about the nature of the dropout mechanism.

Observe that one factor in (6.3) is the square of the response. This is a direct consequence of our raw parameterization of the dropout process, the logit of which is

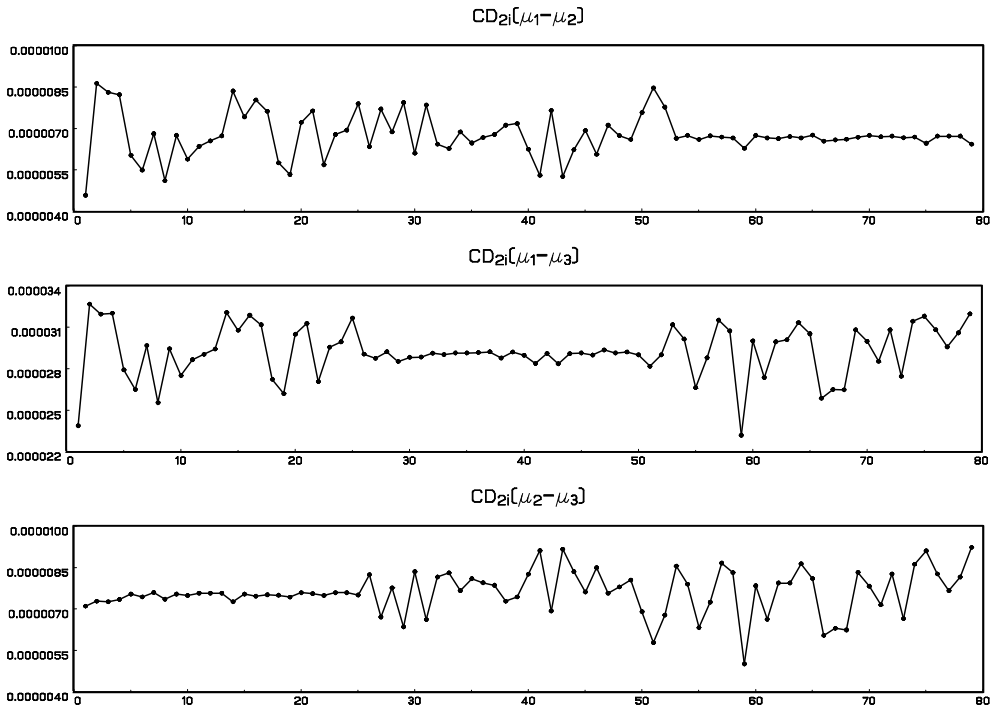

Figure 6.3: *Milk protein trial, index plots of* $CD_{2i}(\mu_1-\mu_2)$, $CD_{2i}(\mu_1-\mu_3)$, $CD_{2i}(\mu_2-\mu 3)$, *dropout modeled from week 15.*

in terms of the previous and current outcomes, to which no transformation is applied. Molenberghs *et al* (1999) argued that, since two subsequent measurements are usually positively correlated, it is not unusual for both of them to be high, and suggested to reparameterize the dropout model (6.1) in terms of the *increment*, i.e., y_{ij} is replaced by $y_{ij} - y_{i,j-1}$. This is related to the approach of Diggle and Kenward (1994) who reparameterized their dropout model in term of the increment just introduced and the size (the average of both measurements). Even though the raw and incremental parameterization in (6.2) are equivalent for model fitting purposes, Molenberghs *et al* (1999) showed that they lead to different perturbation schemes of the form (6.1). Thus, local influence is now focusing on a different set of parameters and one should not expect it to give the same answer. Therefore, it is crucial to guide the parameterization by careful substantive knowledge. In a sense, dependence on the increment is most dramatic since at the time of dropout there is no information about the increment, whereas size can be assessed reasonably well from $Y_{i,j-1}$, especially if the correlation is sufficiently high. The results of this analysis are presented in Figure 6.11 and the most influential profiles are highlighted in Figure 6.12. A slightly dif-

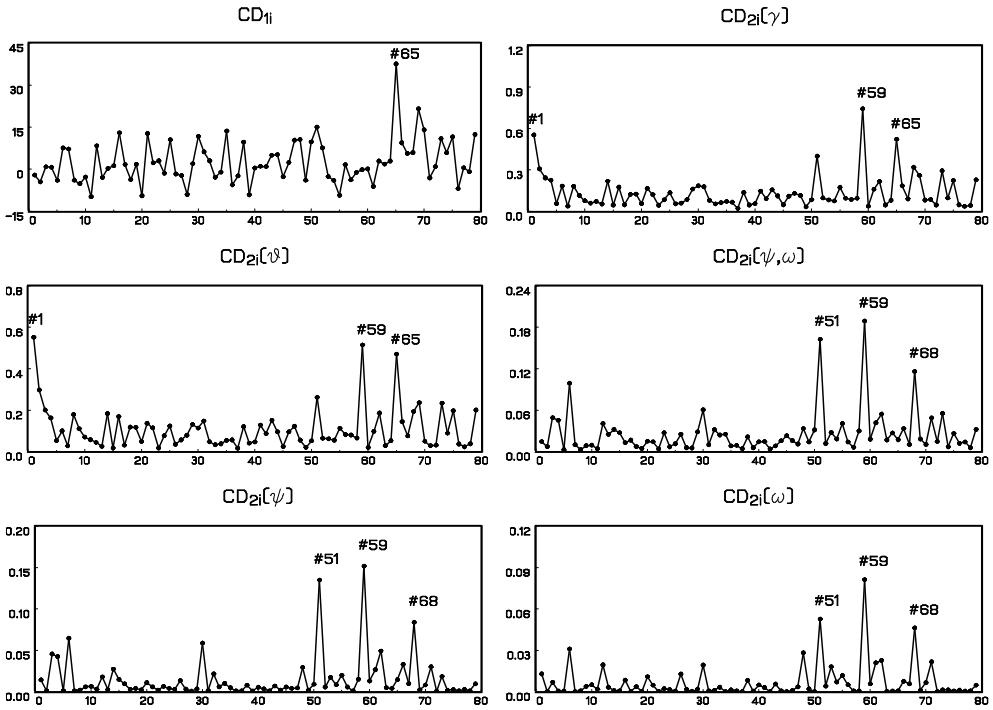

Figure 6.4: *Milk protein trial, index plots of* CD_{1i}, $CD_{2i}(\gamma)$, $CD_{2i}(\theta)$, $CD_{2i}(\psi,\omega)$, $CD_{2i}(\psi)$, $CD_{2i}(\omega)$, *dropout modeled from week 2.*

ferent but overlapping set of profiles is responsible for the influence now. The most important feature is that the influence is very minor. The components of the direction of maximal curvature $\boldsymbol{h}_{\max}$ shows virtually no peaks.

Finally, we will compare both models. The direct-variable results found in Figure 6.5 agree fairly well with those in Figure 6.9, the differences being the absence of #66 and #67 and the appearance of #43. The latter profile is extremely low at the end of the period, where dropout is modeled, and therefore yields a large value for (6.3). For #66 and #67, there is a logical explanation for their disappearance. Indeed, these profiles are very low during the first part of the experimental period, in spite of which they do not drop out. However, during the latter part, their profile is still low *and* they drop out, which is totally plausible behavior and hence their influence was marked in the second but not in this analysis.

For the incremental analysis, there is a larger discrepancy between both models as one can notice from comparing Figures 6.7 and 6.11. While the direction of maximal curvature still shows no unusual subjects, C_i shows somewhat different subjects to be

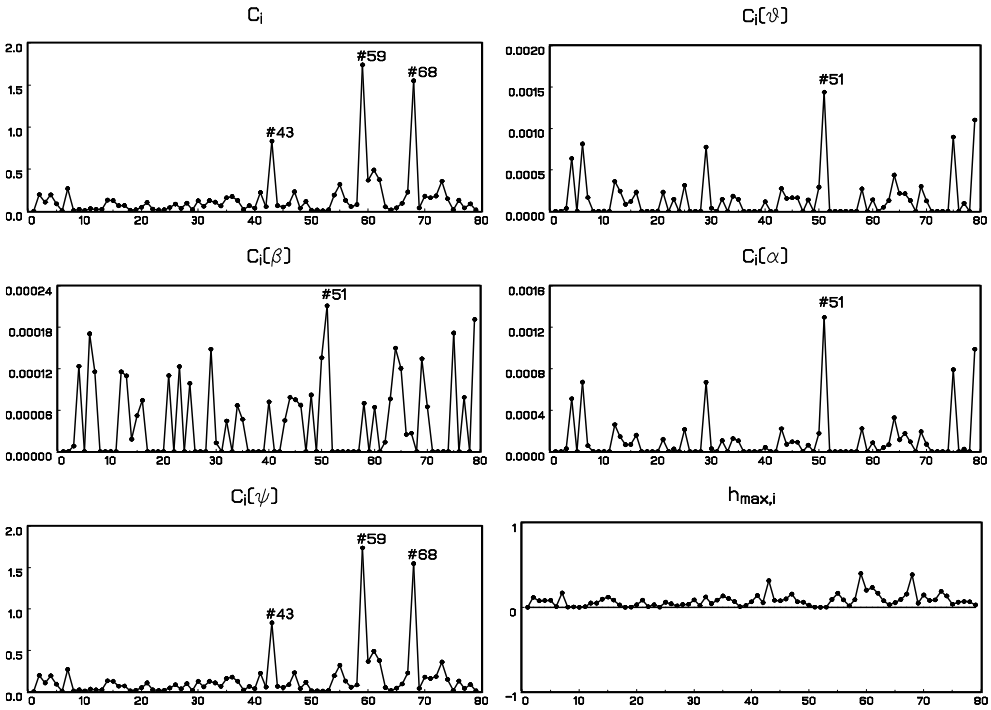

Figure 6.5: *Milk protein trial, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, dropout modeled from week 15.*

influential. Precisely, subjects #7, #51, and #74 are highly influential for the first model whereas subjects #51 (again), #66, #67 and #73 are the ones detected with the second model. It is noteworthy that #51 appears as the subject with largest C_i and $C_i(\theta)$ for the first model, indicating that the measurement model influence $C_i(\theta)$ is of the same order of magnitude as the dropout model influence $C_i(\psi)$, which is in contrast to the other analysis. Both #7 and #74 are *on average* not particularly low profiles, but they are among the lowest ones during the last month of the experiment and, while there are some others with the same feature, these two have a low overall level, but a *high* increment, which is very unusual.

6.2.3 Overview

Table 6.6 summarizes the subjects which are found to be influential in the analyses performed. While it can be argued that the various influence analyses serve different purposes, it is of some importance to distinguish between those subjects who are

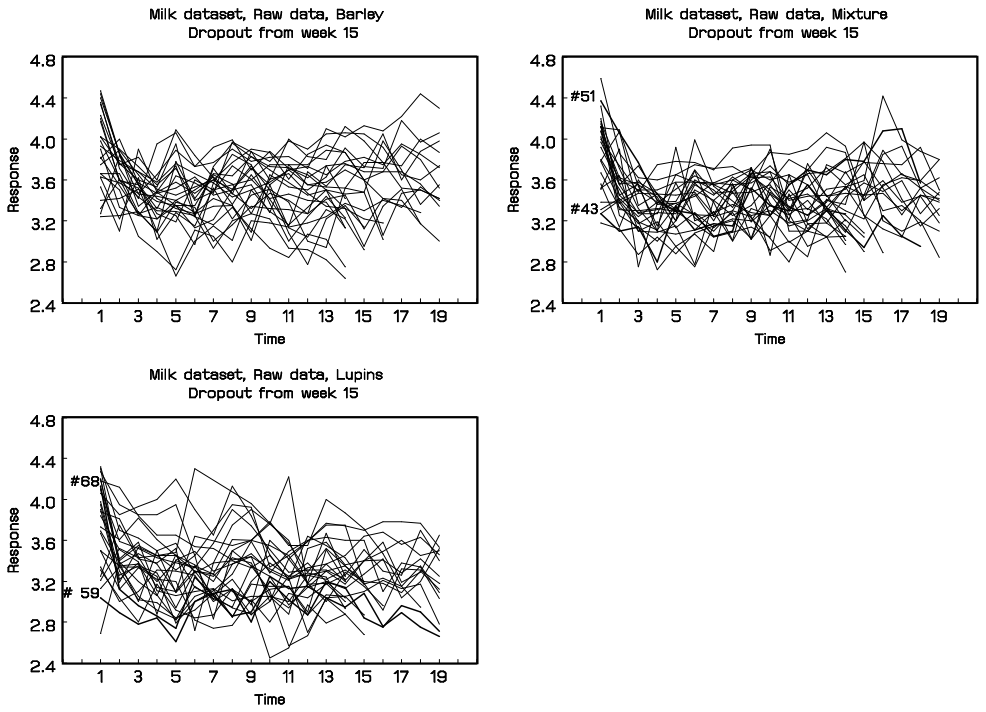

Figure 6.6: *Milk protein trial, individual profiles, with locally influential subjects highlighted, dropout modeled from week 15.*

influential overall and others which turn up in one or a few analyses. Cow #51 is highlighted in all six analyses and cows #59 and #68 show up 4 times, all others being seen three times or less. Clearly, #51 shows up unambiguously in the global influence plots and it yields the highest $C_i(\theta)$, $C_i(\beta)$ and $C_i(\alpha)$ values in the local influence analysis, even though one might argue that in some local influence plots it is closely followed by slightly lower peaks. Inspecting its profile more closely, we conclude that it deviates from the typical profile in a number of ways. First, it is among the highest profiles during the period of initial drop, whereafter it is fairly low during the first half of the period, followed by a period of almost linear increase until the end of the study. The other two, #59 and #68, are on average the lowest profiles, not only within their group, but overall.

Whereas global influence, as stated before, starts from deleting one subject completely, local influence only changes the dropout process for one subject from random dropout to non random dropout. Because of the completely different approach there is no need for both methods to yield similar results. Though by looking at the influential subjects for all cases studied above we notice some overlap.

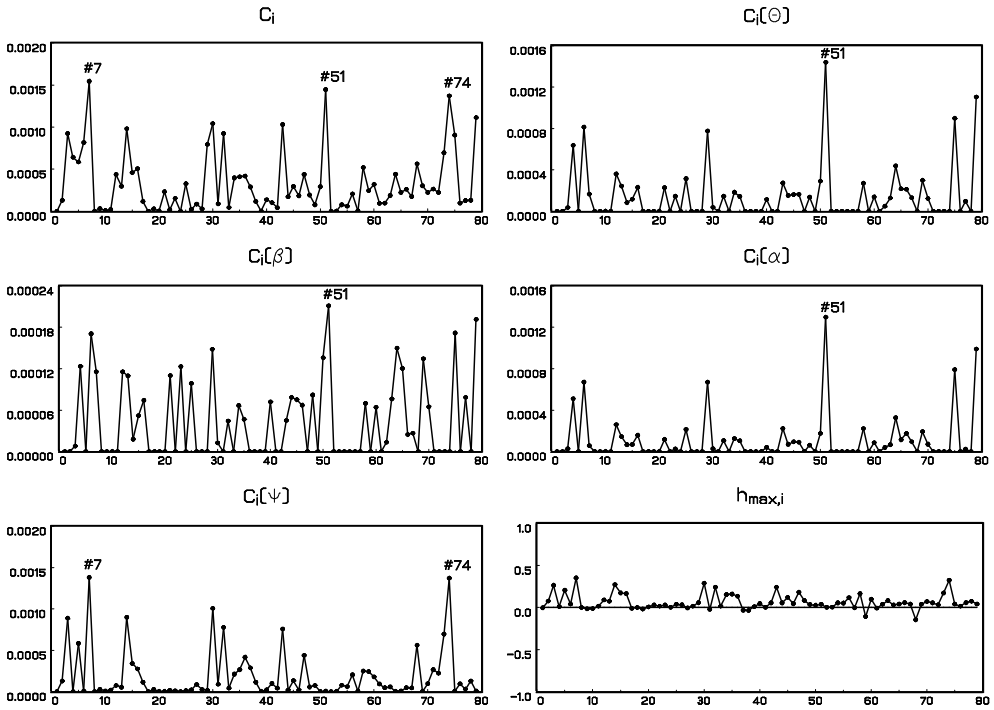

Figure 6.7: *Milk protein trial, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, dropout modeled from week 15, incremental analysis.*

Table 6.6: *Milk protein trial, summary of influential subjects.*

	Drop From Week 15			Drop From Week 2		
Subject	Global	Loc.(Raw)	Loc.(Inc)	Global	Loc.(Raw)	Loc.(Inc)
1	*			*		
7			*			
38	*					
43		*				
51	*	*	*	*	*	*
59	*	*		*	*	
65	*			*		
66					*	*
67					*	*
68	*	*		*	*	
73						*
74			*			

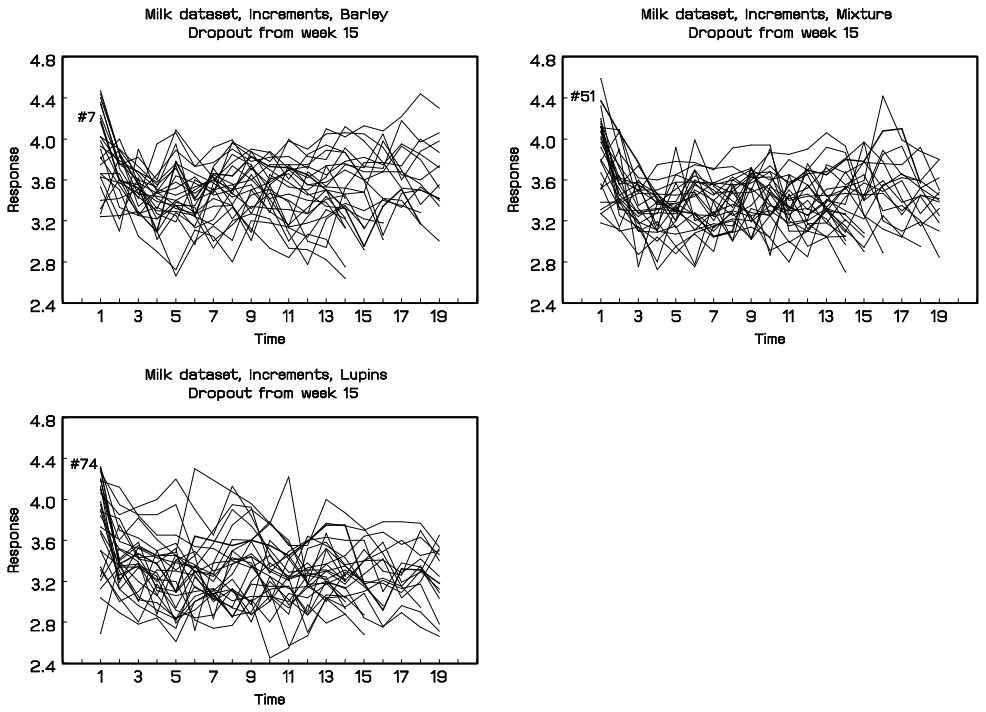

Figure 6.8: *Milk protein trial ,individual profiles, with locally influential subjects highlighted, dropout modeled from week 15, incremental analysis.*

6.2.4 Deleting Selected Subgroups

Focusing on the analyses where dropout is starting from week 15 onwards, we can explore the impact of a *group* of influential subjects further by removing such a group from the data. We define two sets to be removed. The first one consists of #51, #59, and #68, i.e., those subjects that are found to be influential at least twice (see Table 6.6). The second set consists of all subjects found to be influential: #1, #7, #38, #43, #51, #59, #65, #68, and #74. Results are given in Tables 6.1 – 6.4. Clearly, the impact on the parameter estimates and their standard errors is relatively small for the measurement model parameters, but is much larger for the dropout model parameters. This is reflected in the likelihood ratio test for MNAR. Indeed, while this test is significant in all cases, removing the three most influential subjects seriously reduces the evidence for non-random dropout. Thus, the strong evidence for MNAR, stemming from the original analyses, may well have been an overstatement.

Let us study the impact on the measurement model parameters further. An important research question is directed towards differences in diet. To this effect, a two

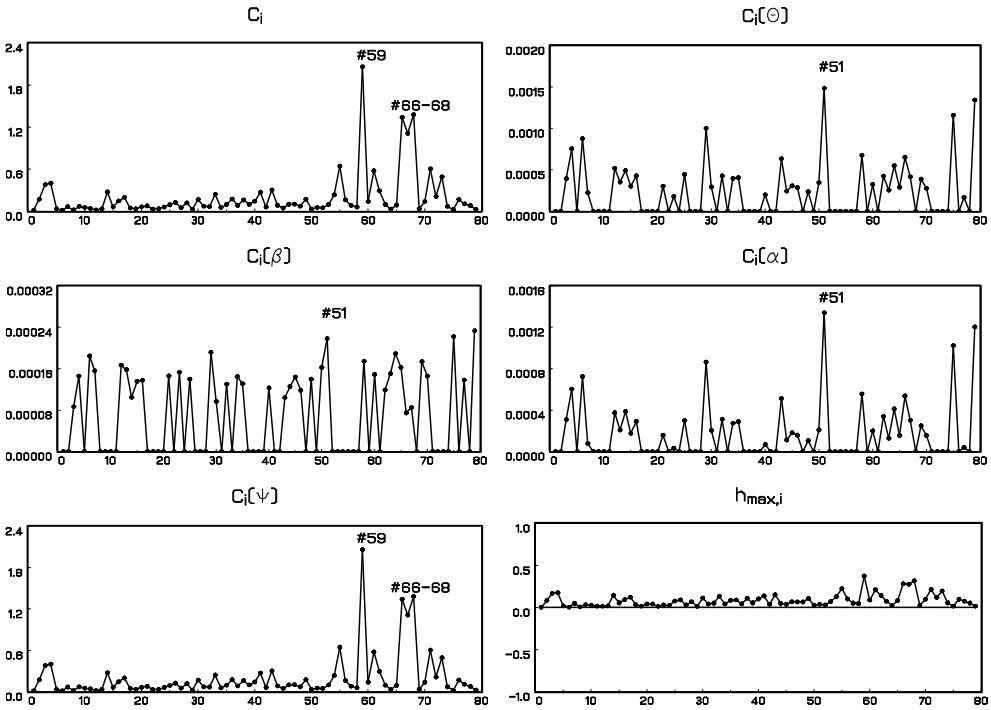

Figure 6.9: *Milk protein trial, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature, dropout modeled from week 2.*

degree of freedom Wald test is computed. We observe little or no difference between the two analyses (MAR and MNAR) for a given deletion scheme, although there is somewhat of a reduction of the evidence when removing sets of subjects. However, these differences do not change the magnitude of the evidence.

Further, we can study the impact of an entire diet group by confining the analyses to two out of three groups. This is done in Tables 6.2 and 6.4. The impact of the lupins group on the dropout mechanism is clear in this respect. Indeed, in the analysis where lupins is removed, the evidence for non-random dropout is non-significant. It is reassuring that this group contains two of the most influential subjects: #59 and #68 (#51 belongs to the mixed group).

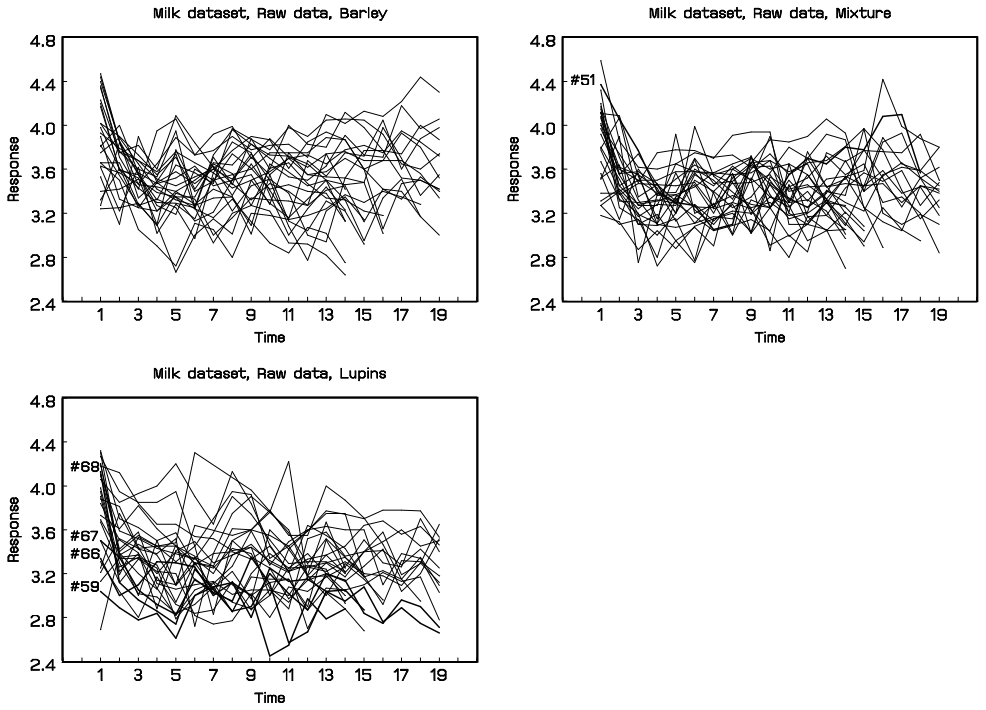

Figure 6.10: *Milk protein trial, individual profiles, with locally influential subjects highlighted, dropout modeled from week 2.*

6.3 Concluding Remarks

Since the model of Diggle and Kenward (1994) in general, and its application to the milk protein trial in particular, has received considerable criticism, we have argued it is useful to perform a sensitivity analysis. To this end, we used the complementary methods of local (Verbeke *et al* 1998) and global influence (Chatterjee and Hadi 1988). We argue that influential subjects can have a large impact on the substantive conclusions, especially in the context of selection models for incomplete data, due to the well known sensitivity to model assumptions, and therefore formal tools for their detection are to be welcomed.

We introduced two different models both based on the one used by Diggle and Kenward. The first one models dropout from week 15 onwards. A second one allows dropout starting from week 2. When dropout is based on the last 5 weeks, the model fitting results are, as expected, very close to those of Diggle and Kenward (1994), with a highly significant indication for non-random dropout. When the dropout model is

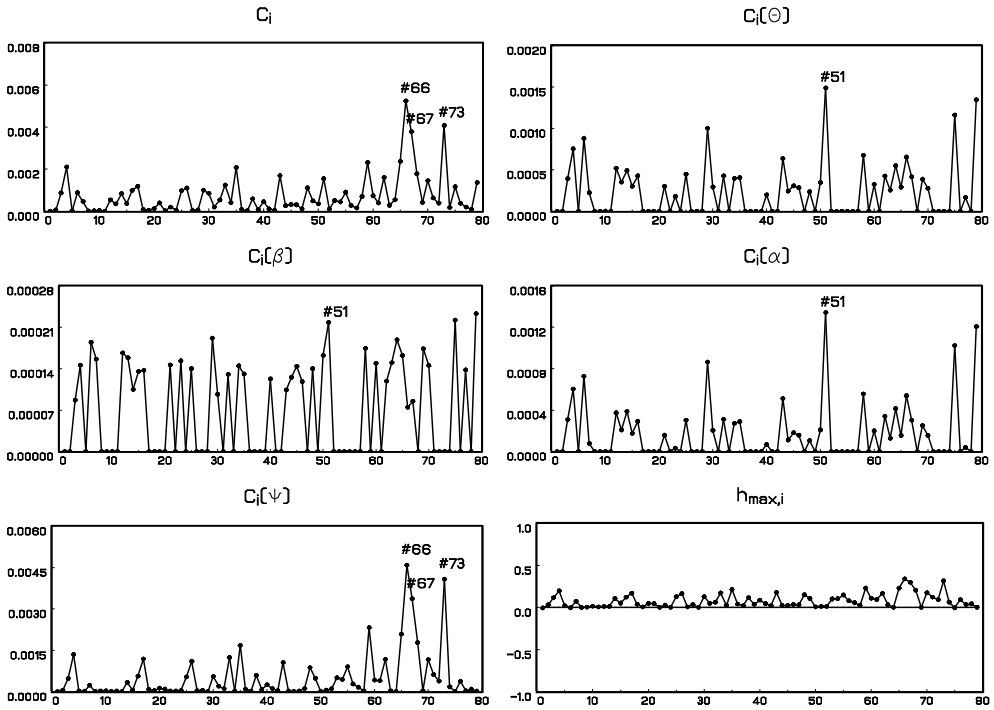

Figure 6.11: *Milk protein trial, index plots of* C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, *and of the components of the direction* $\boldsymbol{h}_{\max}$ *of maximal curvature dropout modeled from week 2, incremental analysis.*

based on the entire period, there is little evidence for non-random dropout. Moreover, influential subjects in the two approaches are entirely different. Both analyses concentrate on behavior in the period during which dropout is modeled. The latter indicates that the choice of period to which dropout applies is crucial.

Finally, we compared the direct variable analysis with an incremental one where dropout depends on the difference between the current and previous measurement. Again, each analysis leads to different subjects to be influential indicating that one should carefully discuss which analysis is preferable. While both model formulations in (6.2) are equivalent, they lead to a different influence analysis, simply because the parameters at which the influence is targeted are different. Which one is chosen may depend on substantive considerations as well as on the observation made by Molenberghs *et al* (1999) that the parameter of $Y_{i,j-1}$ is the most efficiently calculated in the incremental model, provided $\hat{\psi}_1$ and $\hat{\psi}_2$ are negatively correlated. The latter condition is satisfied in many longitudinal applications, as was already noted by Diggle and Kenward (1994).

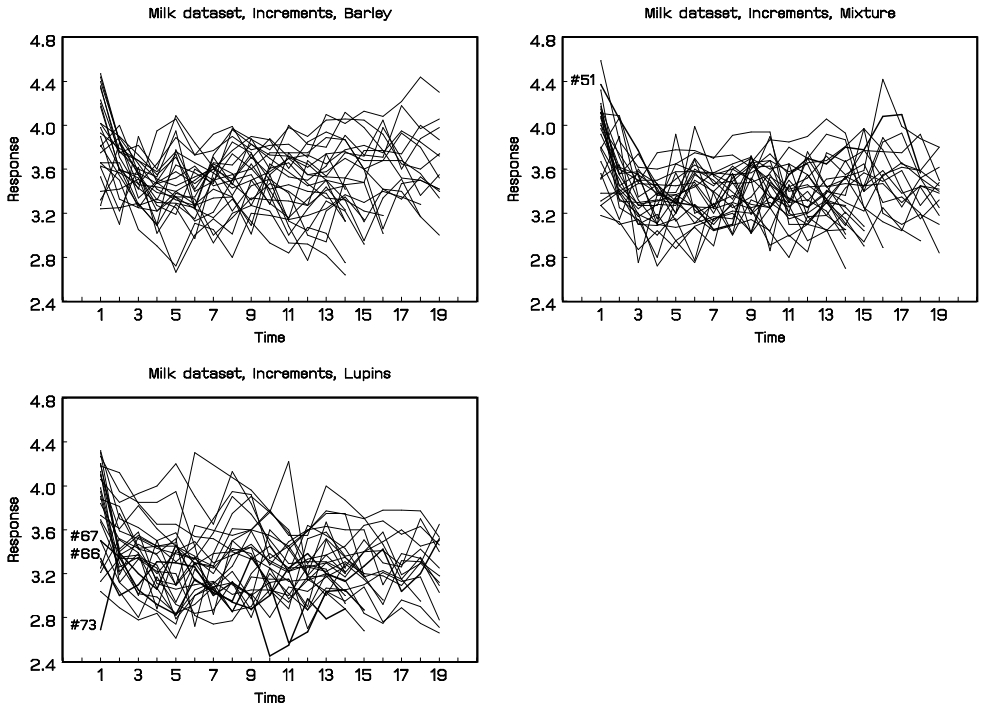

Figure 6.12: *Milk protein trial, individual profiles, with locally influential subjects highlighted, dropout modeled from week 2, incremental analysis.*

Clearly, the influence analyses performed here are not the only ones possible. For example, in local influence one may study different perturbation schemes. However, the ones considered here focus in a clear way on the impact of the informative dropout parameter and lead to computationally very tractible expressions hereby avoiding the need for cumbersome integration.

In order to have more formal rules to decide whether a subject is clearly influential, clearly not influential, or borderline, additional research is required. Presently, such rules of thumb as exploring the, say, 5% most influential subjects in more detail can be used.

Chapter 7

Drawing Conclusions Beyond Dropout

In the previous chapters we have discussed several ways to consider dropout in the modeling process as well as to the major concern about the fact that models are often based on strong assumptions concerning the exact process of dropout. Precisely because of this latter reason, there has been a growing awareness of the need for methods that investigate the sensitivity of the results with respect to the model assumptions (Little 1994, Rubin 1994, Laird 1994, Fitzmaurice, Molenberghs and Lipsitz 1995, and Molenberghs, Goetghebeur and Lipsitz 1998) and several proposals have been made to deal with missing data using selection models, Pattern-mixture models and shared parameter models. A general concensus has emerged that, where all three modeling strategies need very strong and untestable assumptions, in the pattern-mixture framework it is clearer what assumptions are to be made and therefore the pattern-mixture models are quoted as "*more obvious models*" (Glynn, Laird and Rubin 1986, Little 1993, 1994, Hogan and Laird 1997). Next to this discussion concerning sensitivity there is another potentially important issue in all modeling frameworks, one implicitly (selection modeling, shared parameter modeling) or explicitly (pattern-mixture modeling) imputes data and draws conclusions beyond the time of dropout. However, in dealing with quality of life data, it is thought to be useless to take into account time points beyond dropout time when dropout might be due to death. In this chapter we will tackle this problem and we propose some new approaches which will be applied

to the vorozole dataset.

7.1 Time Reversal

Since several studies are specifically interested in what happens just prior to dropout it might be a simple but effective solution to reverse the time period of each patient in a sense that time of dropout is considered as a fixed point in time and the investigator is able to have a general idea about findings prior to dropout.

To illustrate this idea we will consider the vorozole dataset. In the original time scale patients are measured 1 month after they entered the study and then starting from month 2 bi-monthly until month 44. Since all patients were followed until death only few patients are observed at the last time points and therefore we will consider only data prior to month 30. In reversing the time scale we must take into account the fact that the time interval between the first and the second measurement is only one month while elsewhere this is two months. Therefore the reversed timescale will have a time interval of one month between the last and the previous measurement prior to *dropout* where dropout here is the actual starting point of the patient.

We will now first analyze the vorozole data using a selection model. For the linear mixed model we use fixed effects *time*, $time^2$, *treatment group* and *baseline value* and an autoregressive correlation structure. The dropout model will be fitted assuming MAR and MNAR and both can be compared. The results are given in Table 7.1. Using the likelihood-ratio test one might conclude the dropout mechanism to be MNAR. Taking into account the fact that this is actually the starting point of the patient which should be random this conclusion is strange. On the other hand, ongoing simulations studies are showing that the real distribution of this likelihood ratio statistic based on unobserved measurements is not to be interpreted as a true likelihood ratio statistic and therefore need to be treated with care.

Secondly we fit a pattern mixture model using the three types of identification restrictions (CCMV, NCMV and ACMV) as described in Thijs *et al* (2001). The main idea is to combine information available to impute the missing observations. The method can be described as follows

(1) We fit a linear mixed model per pattern and use the same model as in de selection framework.

Table 7.1: *Maximum likelihood estimates (standard errors) of random and non-random dropout models, fitted to the Vorozole data.*

Effect	Par.	MAR	MNAR
Measurement model:			
Intercept	β_0	35.5077 (3.6769)	34.8704(4.0145)
Time	β_1	1.5586(0.2207)	2.8706(0.2344)
Treatment	β_2	-1.3747(1.1761)	-0.5853(2.6441)
baseline value	β_3	-0.3552(0.0301)	-0.3816(0.0308)
Time^2 effect	β_4	-0.0481(0.0091)	-0.0719(0.0090)
Meas. err.	σ^2	241.4754(11.2979)	269.6956(13.4729)
Autoregressive	$AR(1)$	0.6962(0.0161)	0.6656(0.0180)
Dropout model:			
Intercept	ψ_0	-1.2328(0.0572)	-2.4507(0.2117)
Prev. meas.	ψ_1	-0.0077(0.0035)	-0.1008(0.0103)
Curr. meas.	ψ_1		0.1351(0.0138)
-2 loglikelihood		8224.9339	8189.0928

(2) Based on these model-parameters and the choice of identification restrictions we create 5 new completed datasets.

(3) These new datasets are then analyzed again per pattern.

(4) The results per pattern and per imputed dataset can finally be combined using multiple imputation methodology from Schafer (1997) as introduced in Section 5.4.2.

Following this strategy we can draw conclusions concerning the overall treatment effect for all three types of identification restrictions. An overview is given in Table 7.2 and there we can see that in all three cases there is no significant treatment effect. The mean profiles per pattern and per restriction are shown in Figures 7.1–7.3 and also here one notices rather small difference between the three identification restrictions and between the two treatment groups.

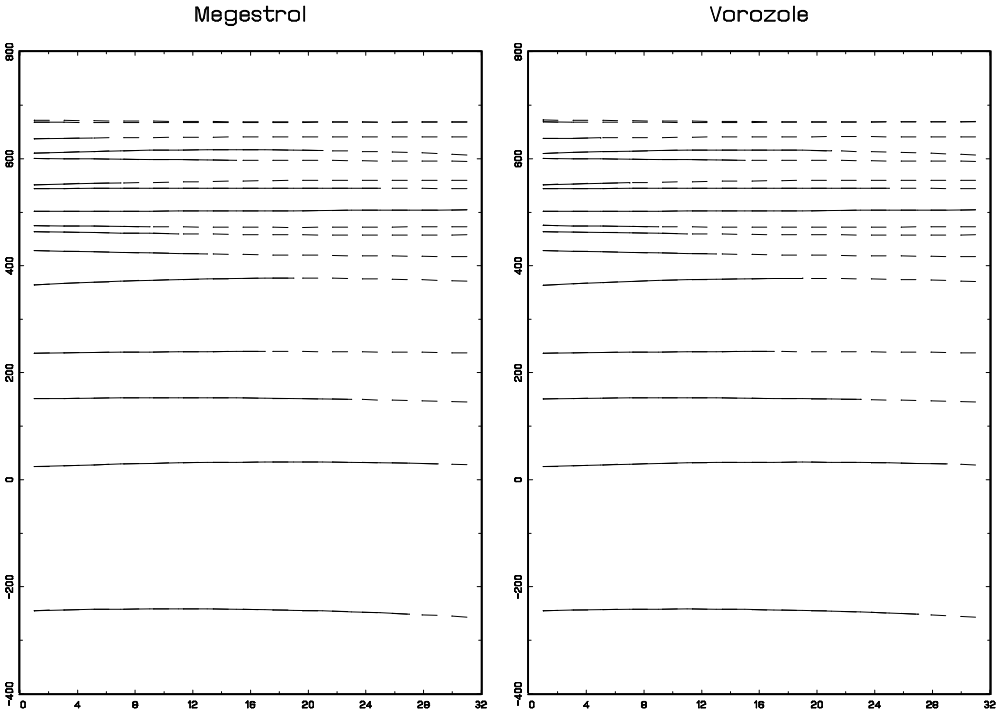

Figure 7.1: *Vorozole dataset, plots of mean profiles over reversed time are presented using the CCMV restrictions, the solid portion of the curves runs from the last obtained measurement until the baseline measurement, while the extrapolated piece is shown in dashed type.*

7.2 Accelerated Failure Time Models

A second approach is based on the idea that in medical research one could consider every patient to have some internal time axis. In order to stretch this *life-time* of every patient to a common time axis we can make use of accelerated failure time models and we can write such a model as follows.

$$log(T) = \alpha_0 + \alpha^t h(\mathbf{U}, \mathbf{V}) + \epsilon$$

with continuous and categorical covariate vectors $\mathbf{U}$ and $\mathbf{V}$ and error term $\epsilon \sim logF(2a, 2b)$ ($Log - F$ Accelerated Failure Time model). Using the right choice of degrees of freedom for the F-distribution also well known error distributions as Weibull, generalized gamma, log-normal, log-logistic,... are included. The combination of this model with a general location model to describe a survival process is described in Cho and Schenker (1999). Similar ideas to describe the time of dropout can be used in

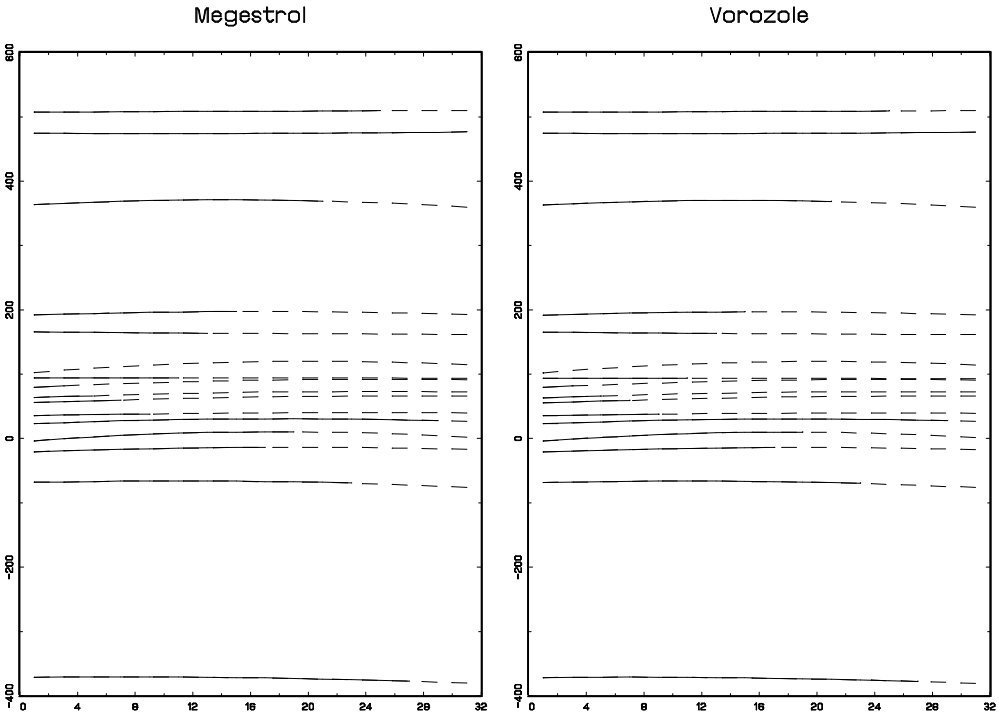

Figure 7.2: *Vorozole dataset, plots of mean profiles over reversed time are presented using the NCMV restrictions, the solid portion of the curves runs from the last obtained measurement until the baseline measurement, while the extrapolated piece is shown in dashed type.*

combination with a linear mixed model to describe the measurement process.

A possibility is to rescale this *life-time* to a $0-1$ interval using a acceleration factor. The model mentioned above can now be rewritten as follows.

$$\begin{aligned} \log(T_i) &= \alpha_0 + \alpha_1 X_i + \epsilon_T, \\ Y_{ij} &= \beta_0 + \beta_1 X_i + \beta_2 \frac{t_{ij}}{T_i} + \epsilon_Y. \end{aligned}$$

In doing so we still have to be careful because we make an underlying assumption that all patients drop out of the study due to death and this may not be the case. When other reasons are present or some profiles are censored we still miss some relevant information and possible solutions to deal with this problem are topic of further research. Finally we can think of another approach by combining both the principle of the accelerated failure time and time reversal. Again the latter idea is a topic of further research.

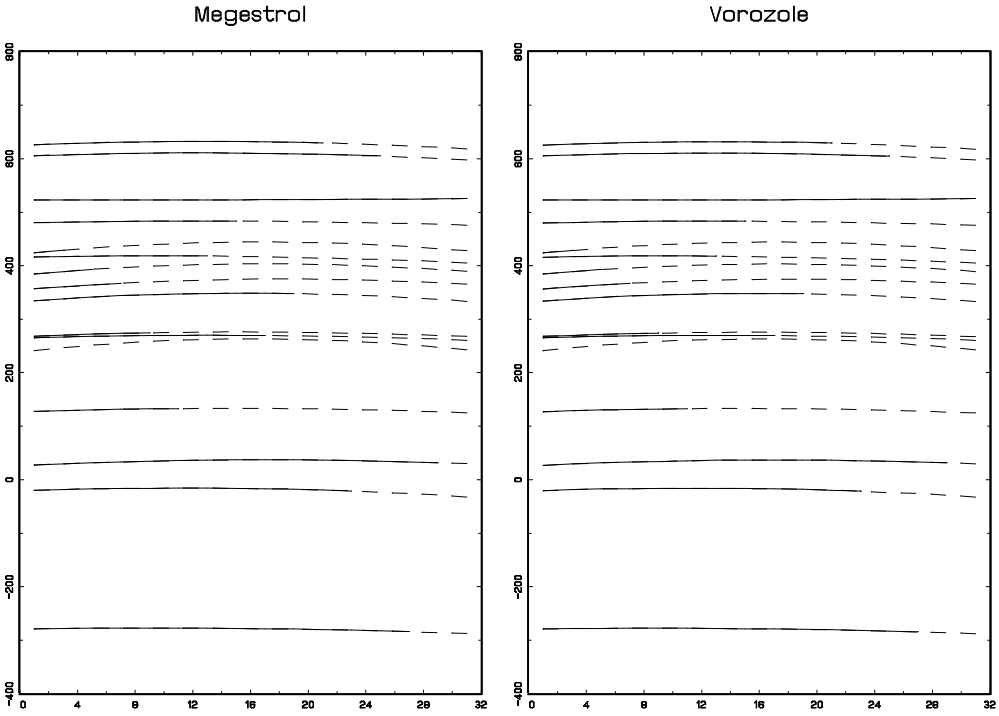

Figure 7.3: *Vorozole dataset, plots of mean profiles over reversed time are presented using the ACMV restrictions, the solid portion of the curves runs from the last obtained measurement until the baseline measurement, while the extrapolated piece is shown in dashed type.*

Finally we think that also shared parameter models as introduced in Section 3.2.2 can yield some extra insight in this problem and we would like to consider this approach in the future. In conclusion we believe that the methods described above can give new perspectives to handling missing data without having to impute data in a senseless way. The results listed here are already indicating that these methods can be performed successful and we hope that future research will be as promissing.

Table 7.2: *Vorozole dataset, overview of pattern mixture methodology.*

		Treatment effect per pattern		
		CCMV	NCMV	ACMV
Pattern =	1	13.500	1.360	4.386
	2	13.331	0.835	7.645
	3	12.725	0.455	6.922
	4	11.133	0.368	6.638
	5	9.166	-0.312	4.341
	6	8.973	0.827	1.969
	7	8.047	2.119	7.244
	8	11.166	2.630	8.323
	9	4.429	-1.155	4.780
	10	6.169	-1.379	5.110
	11	10.453	5.290	10.525
	12	2.315	-2.230	-1.120
	13	9.927	9.068	10.978
	14	-5.645	-8.268	-6.335
	15	0.000	0.000	0.000
	16	9.495	8.934	9.828
F-statistic for treatment		1.552	0.783	1.703
P-value for treatment		0.102	0.706	0.048

Chapter 8

General Conclusion

In this final chapter we will briefly reflect on the methodology introduced in the previous chapters and we will mainly discuss possible extensions as well as topics of ongoing and future research. It is clear that we believe our methods to be a first step in the direction of a more formal sensitivity analysis.

As already stated in Chapters 4 and 6, local and global influence tools are very useful and easy to apply but on the other hand they incorporate a wide range of possible extensions which we will briefly summarize here. From the results of the Milk Protein Trial we have seen that the overlap of influential subjects revealed by local and by both techniques although not expected, is rather large. For this reason we may believe that both techniques are valuable because they confirm each other therefore both local and global influence are considered as complementary tools within the scope of a full sensitivity analysis. Taking this into account it can be a broad topic of further research how we might combine general sensitivity tools within the selection modeling framework.

Within the influence tools and more specific local influence, we also indicated some possible extensions. While we have applied local influence tools (Cook 1986, Lesaffre and Verbeke 1997) to a rather specific selection model for continuous data as presented in Diggle and Kenward (1994) one could aim to develop similar tools for other model families. Currently also categorical data are being considered using different model families as the one introduced by Baker, Rosenberger and Dersimonian (1992), or the Dale model described in Molenberghs and Lesaffre (1994) and Molenberghs,

Kenward and Lesaffre (1997). In particular, we have shown how the impact of small *perturbations* around the null model of missing at random will affect the *model parameters* and how these specific influence measurements can be *interpreted*. Again we found some interesting routes for further exploration. First, we can discuss the perturbation scheme chosen here because it is not the only one possible. We already noticed that a direct variable or an incremental variable notation can yield important differences and although we did not report on this we also have looked at a possibility with a double omega notation leading to nice interpretable results. It therefore is clear that this field is not fully explored. Secondly once we have calculated these influence measurements we still did a rather ad hoc selection of the so called influential subjects. One might want to investigate the influence measurements closer in order to determine some distributional characteristics or specify a formal test to spot influential subjects. Finally we only concentrated on the influence on the model parameters but one might be interested in the influence on specific test statistics. Depending on which statistic is of interest this can be reached by reformulating the model (Wald Statistics) or it can be necessary to develop a new variation of influence tools.

All previous extensions still pop up within influence analysis for selection models but can be easily expanded to other frameworks as pattern-mixture models. Furthermore the comparison between selection models and pattern-mixture models can also be fruitful in performing a sensitivity analysis. A valuable tool in this comparison according to us might be the random effects model where specific choices of the *random effect* also incorporate selection models and pattern-mixture models in one single framework as extremes. This actually is just a wild thought and further research might explain whether this is a possible approach or not and if so a general influence methodology can be developed.

Finally concerning the pattern-mixture models we have considered several strategies and again by contrasting these strategies one obtains a range of conclusions rather than a single one, which provides insight into the sensitivity to the assumptions made. Especially with the identifying restrictions, one has to be very explicit about the assumptions and moreover this approach offers the possibility to consider several forms of restrictions. Special attention should go to the ACMV restrictions, since they are the MAR counterpart within the pattern-mixture context. The identifying restrictions strategy provides further opportunity for sensitivity analysis. Indeed, since CCMV and NCMV are extremes for the $\boldsymbol{\omega}$ vector, it is very natural to consider the idea of *ranges* in the allowable space of $\boldsymbol{\omega}$. Clearly, any $\boldsymbol{\omega}$ which consists of non-

negative elements that sum to one is allowable, but also the idea of extrapolation could be useful, where negative components are allowed, given they provide valid conditional densities. This idea of ranges can also be explored in future research. Furthermore identifying restrictions is a potentially useful way to model incomplete longitudinal data and such restrictions allow one to reflect carefully on the nature of the assumptions made. While a particular set of restrictions corresponds to MAR we now established a family of MNAR models which avoid dependence of dropout on future, unobserved outcomes. Not only does this family embed, again, MAR, it provides a sensibly yet wide space within which sensitivity analysis can be conducted. If a dependence on an underlying latent variable is deemed plausible, afore mentioned shared-parameter models can be chosen but then again further research is required.

The SAS and GAUSS macros which have been implemented to carry out the sensitivity analysis based on local and global influence as well as the multiple imputation related tasks are available from the internet but are written on a rather ad hoc basis. Here it might be an interesting option the contact computer scientists dealing with the development of similar macros within these software packages.

Acknowledgment

We gratefully acknowledge financial support from FWO-Vlaanderen Research Project G.0002.98 "Sensitivity Analysis for Incomplete and Coarse Data" and from Vlaams Instituut voor de Bevordering van het Weten schappelijk-Technologisch Onderzoek in de Industrie.

We also would like to thank the Janssen Research Foundation for the kind permission to use their data at first but also for the fruitful coöperation we received throughout this work.

Samenvatting

In een longitudinale (klinische of epidemiologische) studie of experiment wordt een individu herhaald gemeten in de tijd. Het is niet ongebruikelijk dat sommige meetreeksen vroegtijdig afgebroken worden om redenen buiten de controle van de onderzoeker (de patiënt verhuist, weigert verdere medewerking omdat hij/zij vindt dat het middel niet werkt of te veel nevenwerkingen heeft, enz.). In deze situatie spreekt men van dropout. In de praktijk wordt jammer genoeg nog steeds op een zeer ad hoc manier gepoogd het probleem van onvolledige gegevens op te lossen, zoals het zich beperken tot de volledige meetreeksen (*complete case analysis*) of, erger nog, het invullen van de ontbrekende gegevens (*imputation*). De gevaren hiervan worden onderstreept in Molenberghs, Bijnens en Shaw (1997).

Om hieraan tegemoet te komen is het noodzakelijk van dropout expliciet in het statistisch model op te nemen en er is dan ook recent veel werk verricht m.b.t. modelbouw voor onvolledige gegevens (Kenward, Lesaffre, en Molenberghs 1994; Michiels en Molenberghs 1995, 1997; Molenberghs, Michiels, Kenward en Diggle 1998; Goetghebeur en Molenberghs 1996a, 1996b; Molenberghs en Goetghebeur 1997). Menig auteur doet suggesties voor de analyse van non-random nonrespons (Nordheim 1984; Baker en Laird 1988; Baker 1995; Park en Brown 1994; Diggle en Kenward 1994; Molenberghs, Kenward en Lesaffre 1997). Men blijft geïnteresseerd aan de volledige gegevens, maar heeft een model met onbekende parameters nodig om geobserveerde en niet-geobserveerde gegevens "aan elkaar te lijmen". Hoewel deze complexere modellen ontbrekende gegevens selectief toelaten en dus meer kans maken om de primaire parameters onvertekend te schatten blijft er wel de noodzaak om impliciet of expliciet onderstellingen te maken betreffende het droupt mechanisme. Rubin (1976) en Little en Rubin (1987, Ch. 6) maken onderscheid tussen verscheidene nonrespons of dropout processen. Een dropout proces wordt *completely random* (volledig random, MCAR)

genoemd als het onafhankelijk is van geobserveerde en niet geobserveerde waarnemingen en *random* (MAR) als het, gegeven de geobserveerde data, onafhankelijk is van de niet geobserveerde data. In alle andere gevallen spreekt men van *informatieve* dropout. Voor een random dropout proces kan een geldige analyse verkregen worden via maximum likelihood, zonder het dropout proces te modelleren, indien de parameters die het meetproces beschrijven functioneel onafhankelijk zijn van de parameters die het dropout proces beschrijven. Men spreekt dan van *ignorable* dropout (Rubin 1976) en de analyse vereenvoudigt in belangrijke mate. Vaak zijn de redenen voor dropout echter veelvuldig en is de MAR veronderstelling moeilijk vol te houden. Het is dan mogelijk om, op basis van de geobserveerde gegevens, een model voor informatieve dropout te fitten. Het is echter moeilijk om de gekozen modelvorm helemaal te verantwoorden en de data bevatten soms weinig informatie voor de dropout model parameters. In zulke gevallen is sensitiviteitsanalyse aangewezen. Diggle en Kenward (1994) stelden zo'n model voor in de context van continue herhaalde metingen. Andere modellen werden voorgesteld door Schluchter (1988), Laird, Lange en Stram (1987), Wu en Bailey (1988, 1989), en Wu en Carroll (1988). Deze laatsten gebruikten random effecten om het nonrespons proces te beschrijven. Verder hebben wij ons voornamelijk toegelegd op continue longitudinale gegevens maar ook categorische gegevens kregen heel wat aandacht. Baker en Laird (1988) breiden het werk van Fay (1986) uit en geven een duidelijk overzicht van modellen voor onvolledige kruistabellen. Stasny (1986), Baker, Rosenberger, en Dersimonian (1992), Conaway (1992, 1993), Park en Brown (1994) en Molenberghs, Kenward en Lesaffre (1997) beschouwen informatieve modellen voor herhaalde categorische metingen. Met dit groeiend volume aan literatuur voor non-random nonrespons, is ook de bezorgdheid toegenomen over het feit dat modellen vaak steunen op sterke onderstellingen waar betrekkelijk weinig informatie voor bestaat in de data zelf. Dit werd reeds opgemerkt door Glynn, Laird en Rubin (1986) die aangaven dat dit probleem zeer typisch is voor zogenaamde *selectiemodellen*, waar de gemeenschappelijke verdeling van meetproces en dropout proces gefactoriseerd wordt als de marginale verdeling van het meetproces en de conditionele verdeling van het dropout proces, gegeven de metingen, terwijl dit minder sterk speelt bij *pattern-mixture modellen* waar de andere factorisatie wordt gebruikt (Little 1993, 1994). Molenberghs, Michiels, Kenward en Diggle (1998) bestuderen formele connecties tussen beide paradigma's.

Tenslotte tonen Molenberghs, Goetghebeur, Lipsitz en Kenward(1999) de nood voor sensitiviteitsanalyse aan door een overzicht te geven van de problemen die zich voordien bij modellen voor informatieve dropout:

1. Modellen met gelijke of vergelijkbare fit op het niveau van de geobserveerde onvolledige gegevens, kunnen sterk verschillen in termen van predictie en interpretatie van de (hypothetische) volledige gegevens. Afhankelijk van de gestelde onderzoeksvraag, kan dit tweede aspect van een model een belangrijke rol spelen bij de wetenschappelijke besluitvorming. Het is dan uiteraard van belang te weten in hoeverre de conclusies van het gekozen model afhangen.

2. Wanneer informatieve modellen gefit worden, is een oplossing in het inwendige van de parameterruimte niet gegarandeerd. Dit geldt zelfs voor op het eerste gezicht "reguliere" gegevens, bijv. kruistabellen waar alle aantallen strikt groter zijn dan nul. Men stelt vast dat sommige modellen aanleiding geven tot randoplossingen, bijv. onder de vorm van voorspelde aantallen gelijk aan nul, of parameters op de rand van de parameterruimte (wat in bepaalde gevallen $\pm\infty$ kan impliceren). In zulke gevallen is klassieke asymptotische theorie niet meer gegarandeerd. Een ernstig probleem is ook wanneer *negatieve* aantallen voorspeld worden. Het blijkt dat men dit te weinig controleert.

3. Sommige modellen zijn overgespecifieerd in de zin dat de bijhorende likelihood geen unieke oplossing heeft, doch tot een hele familie oplossingen leidt. Het detecteren van overspecificatie is niet steeds eenvoudig, en heeft recent wat aandacht gekregen (Catchpole en Morgan 1997). Ook hier stellen we vast dat elk lid van de familie dezelfde fit induceert op het niveau van de onvolledige gegevens, maar tot een andere fit kan leiden op het niveau van de (hypothetische) volledige gegevens.

Ondanks het feit dat een algemeen besef is ontstaan van de nood aan sensitiviteitsanalyse zijn slechts enkele voorstellen gedaan die telkens ad hoc zijn. Vandaar dat wij hier ook proberen wiskundige instrumenten uit te bouwen om Sensitiviteit op een meer formele basis te kunnen *meten*. Zoals reeds eerder vermeld zullen we ons beperken tot het geval waar we continue longitudinale gegevens verzamelen en Hoofdstuk 2 beschrijft vijf typische gegevensverzamelingen waarvoor de door ons ontwikkelde methodologie nuttig is en waarop ze ook zal worden toegepast. Een algemeen kader voor het modelleren van onvolledige gegevens, voornamelijk gebaseerd op Rubin (1976) en Little en Rubin (1987), wordt geschetst in Hoofdstuk 3. Dit raamwerk zal de discussie in de daaropvolgende hoofdstukken vergemakkelijken. In Hoofdstuk 4 beschrijven we een eerste familie van modellen die zullen gebruikt worden met name de selectiemodellen. Ook gaan we hier meer specifiek kijken naar het selectie model van Diggle en Kenward (1994) omdat dit model reeds uitgebreid besproken is geworden

door diverse auteurs. vertrekkende van dit model kunnen we dan ook een eerste tool realiseren om sensitiviteitsanalyses uit te voeren door de invloed van kleine perturbaties in het vooropgestelde model te onderzoeken. Hiertoe is *local influence* (Cook 1986) uitermate geschikt en deze methode zal dan ook uitgebreid bestudeerd worden. Aan de hand van diverse perturbatieschema's zijn we in staat de invloed van het nonrespons mechanisme op modelparameters te karakteriseren. Vergelijkbaar met deze methode beschrijven we ook globale invloedsmaten en beide methodes worden toegepast op typisch longitudinale gegevens zoals ingevoerd in Hoofdstuk 2. Een tweede familie zal worden besproken in Hoofdstuk 5 en binnen deze familie is het onze bedoeling om aan te geven welke de mogelijke strategieën zijn en hoe deze met elkaar kunnen vergeleken worden. Meer bepaald bespreken we hoe binnen deze nieuwe familie op een eenvoudige manier hetzelfde raamwerk betreffende *random* versus *non-random* dropout (Rubin 1976 en Little en Rubin 1987) kan worden toegepast en we beschrijven uitvoerig welke vorm van identificatierestricties (Little 1993, 1994, 1995) hiermee gepaard gaan om Pattern-mixture modellen te beschouwen. Als speciale vorm bekijken we *non future dependent missing value* restricties waarbij het mechanisme van dropout enkel kan afhangen van de huidige mogelijks niet geobserveerde meting maar niet meer van de in de toekomst geplande metingen. Ter illustratie beschouwen we in Hoofdstuk 6 de Milk protein trial en hierop passen we de methoden van lokale en globale invloed toe alsook de verschillende strategieën binnen het pattern-mixture framework. In dit hoofdstuk merken we op dat de verschillende methoden tot nogtoe ontwikkeld wat betreft de resultaten behoorlijk goed met elkaar vergelijkbaar zijn wat erop wijst dat het betrouwbare tools zijn. Hoofdstuk 7 beschouwt enkele problemen die kunnen optreden wanneer we zonder meer gebruik maken van selectiemodellen of pattern-mixture modellen en meer bepaald wanneer er conclusies worden getrokken na het tijdstip van dropout wanneer dropout mogelijk wordt veroorzaakt door sterfte en tenslotte zal in Hoofdstuk 8 een algemene conclusie worden geformuleerd en tevens zullen we daar aangeven welke problemen nog eventueel verder onderzoek vereisen.

Nieuwe statistische technieken, hoe bruikbaar ook, worden slechts écht gebruikt indien ze breed toegankelijk zijn voor de gemeenschap van potentiële gebruikers (zoals bijv. de officiële en farmaceutische statistici). Deze toegankelijkheid impliceert bruikbare software. Zo zijn het lineair gemengd model (voor continue herhaalde gegevens) en *generalized estimating equations* (Liang en Zeger 1986) (voor discrete gegevens) pas echt breed verspreid geraakt met de respectieve implementaties in de SAS procedures MIXED en GENMOD. Hieraan is een periode van semi-professionele software vooraf gegaan met een waaier van macro's.

Het kan uiteraard geen realistische ambitie zijn om de door ons voorgestelde technieken meteen in commerciële software om te zetten. We menen echter wel te mogen zeggen dat we een belangrijke bijdrage hebben geleverd door de voorgestelde methoden te implementeren in macro's in breed verspreide paketten. Voor de door ons beoogde doelgroep van biostatistici in industrie en overheid zouden verdere implementaties in SAS (als macro of in de interactieve matrix taal IML), SPlus en GAUSS dan ook expliciete voordelen kunnen opleveren.

References

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. (1990) *Categorical Data Analysis*, Nw York: Wiley.

Allison, P.D. (1987) Estimation of linear models with incomplete data. *Sociology Methodology*, 71–103.

Arnold, B.C., Castillo, E. and Sarabia, J.M. (1992) *Conditionally Specified Distributions, Lecture Notes in Statistics* **73**, New York: Springer Verlag.

Arnold, B.C. and Strauss, D. (1991) Pseudolikelihood estimation: some examples. *Sankhya: the Indian Journal of Statistics - Series B*, **53**, 233–243.

Baker, S.G. (1992) A simple method for computing the observed information matrix wxhen using the EM algorithm with categorical data. *Journal of Computational and Graphical Statistics*, **1**, 63–76.

Baker, S.G. and Laird, N.M. (1988) Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association*, **83**, 62–69.

Bartlett, M.S. (1937) Some examples of statistical methods of research in agriculture and applied botany. *Journal of the Royal Statistical Society - Series B*, **4**, 137–170.

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

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society - Series B*, **36**, 192–225.

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

Brown, C.H. (1990) Protecting against nonrandomly missing data in longitudinal studies. *Biometrics*, **46**, 143–155.

Chatterjee, S. and Hadi, A.S. (1988) *Sensitivity analysis in linear regression.* New York: John Wiley and Sons.

Cho, M. and Schenker, N. (1999) Fitting the log-F accelerated failure time model with incomplete covariate data. *Biometrics*, **55**, 826–833.

Clayton, D. and Hills, M. (1993) *Statistical Models in Epidemiology.* Oxford: Oxford University Press.

Cohen, J. and Cohen, P. (1983) *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conaway, M.R. (1992) The analysis of repeated categorical measurements subject to nonignorable nonresponse. *Journal of the American Statistical Association*, **87**, 817–824.

Conaway, M.R. (1993) Non-ignorable non-response models for time-ordered categorical variables. *Applied Statistics*, **42**, 105–115.

Conaway, M.R., Waternaux, C., Allred, E., Bellinger, D. and Leviton, A. (1992) Pre-natal blood lead levels and learning difficulties in children: an analysis of non-randomly missing categorical data. *Statistics in Medicine*, **11**, 799–811.

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.

Cox, D.R.(1972) The analysis of multivariate binary data. *Applied Statistics*, **21**, 113–120.

Cressie, N.A.C. (1991) *Statistics for Spatial Data.* New York: John Wiley and Sons.

Cullis, B.R. (1994): Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 79–80.

Curran, D., Pignatti, F. and Molenberghs, G. (2002) Milk protein trial: informative dropout versus random drop-in. *Submitted.*

Dale, J.R. (1986) Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.

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

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

Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns.* London: Mathematics in Biology, Academic Press.

Diggle, P.J. (1988) An approach to the analysis of repeated measures. *Biometrics*, **44**, 959–971.

Diggle, P.J. (1990) *Time Series: A Biostatistical Introduction.* Oxford: Oxford University Press.

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

Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994) *Analysis of Longitudinal Data.* Oxford: Oxford Science Publications, Clarendon Press.

Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B*, **57**, 45–97.

Ekholm, A. and Skinner, C. (1998) The muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics*, **47**, 251–263.

Fay, R.E. (1986) Causal models for patterns of nonresponse. *Journal of the American Statistical Association*, **81**, 354–365.

Fitzmaurice, G.M., Molenberghs, G. and Lipsitz, S.R. (1995) Regression models for longitudinal binary responses with informative dropouts. *Journal of the Royal Statistical Society, Series B*, **57**, 691–704.

Fowler, F. J. (1988) *Survey Research Methods*. Newbury Park, CA: Sage.

Friedman, L.M., Furberg, C.D. and DeMets, D.L. (1998) *Fundamentals of Clinical Trials*. New York: Springer-Verlag.

Fuchs, C. (1982) Maximum likelihood estimation and model selection in contingency tables with missing data. *Journal of the American Statistical Association*, **77**, 270–278.

Gelman, A. and Speed, T.P. (1993) Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society - Series B*, **55**, 185–188.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Geys, H., Molenberghs, G. and Ryan, L.M. (1997) Pseudo-likelihood Inference for Clustered Binary Data. *Communications in Statistics: Theory and Methods*, **26**, 2743–2767.

Geys, H., Molenberghs, G. and Ryan, L.M. (1999) Pseudo-likelihood Modelling of Multivariate Outcomes in Developmental Toxicology. *Journal of the American Statistical Association*, **94**, 734–745.

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

Glynn, R.J., Laird, N.M. and Rubin, D.B. (1993) Multiple Imputation in Mixture Models for Nonignorable Nonresponse With Follow-ups. *Journal of the American Statistical Association*, **88**, 984–993.

Goss, P.E., Winer, E.P., Tannock, I.F., Schwartz, L.H. and Kremer, A.B. (1999) A randomized phase III trial comparing the new potent and selective third-generation aromatase inhibitor vorozole with megestrol acetate in postmenopausal advanced breast cancer patients. *Journal of Clinical Oncology*, **17**, 52–63.

Green, S., Benedetti, J. and Crowley, J. (1997) *Clinical Trials in Oncology.* London: Chapman & Hall.

Greenlees, W.S., Reece, J.S. and Zieschang, K.D. (1982) Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, **77**, 251–261.

Hand, D.J. and Crowder, M.J. (1995) *Practical Longitudinal Data Analysis.* London: Chapman and Hall.

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

Heckman, J.J. and Robb, R. (1986) lternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples* (pp. 63–107). New York: Springer.

Hedeker, D. and Gibbons, R.D. (1997) Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, **2**, 64–78.

Heitjan, D.F. and Rubin, D.B. (1991) Ignorability and coarse data. *Annals of Statistics*, **19**, 2244–2253.

Heyting, A., Tolboom, J.T.B.M. and Essers, J.G.A. (1992) Statistical handling of dropouts in longitudinal clinical trials. *Statistics in Medicine*, **11**, 2043–2061.

Hogan, J.W. and Laird, N.M. (1997) Mixture Models for the Joint Distribution of Repeated Measures and Event Times. *Statistics in Medicine*, **16**, 239–258.

Holland, P.W. (1986) A Comment on Remarks by Rubin and Hartigan. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples* (pp. 149–151). New York: Springer.

Kahn, H. and Sempos, C.T. (1989) *Statistical Methods in Epidemiology.* New York: Oxford University Press.

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., Lesaffre, E. and Molenberghs, G. (1994) An Application of Maximum Likelihood and Generalized 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*, **13**, 236–247.

Kenward, M.G. and Molenberghs, G. (1999) Parametric models for incomplete continuous and categorical longitudinal studies data. *Statistical Methods in Medical Research*, **8**, 51–83.

Kenward, M.G., Molenberghs, G. and Thijs, H. (2002) Pattern-Mixture Models With Proper Time Dependence. *Submitted to Biometrika*

Koch, G., Singer, J., Stokes, M., Carr, G., Cohen, S. and Forthofer, R. (1991) Some aspects of weighted least-squares analysis for longitudinal categorical data. In: J.H. Dwyer, Feinleib, M., Lipper, P. and Hoffmeister, H. (Eds.), *Statistical Models for Longitudinal Studies of Health* (pp. 215–260). Oxford: Oxford University Press.

Laird, N.M. (1988) Missing data in longitudinal studies. *Statistics in Medicine*, **7**, 305–315.

Laird, N.M. (1994) Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 84.

Laird, N.M., Lange, N. and Stram, D. (1987) Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association*, **82**, 97–105.

Laird, N.M. and Ware, J. H. (1982) Random effects models for longitudinal data. *Biometrics*, **38**, 963–974.

Lang, J.B. and Agresti, A. (1994) Simultaneously modelling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625–632.

Lawrance, A.J. (1995) Deletion influence and masking in regression. *Journal of the Royal Statistical Society, Series B*, **57**, 181–190.

Le Cessie, S. and Van Houwelingen, J.C. (1994) Logistic regression for correlated binary data. *Applied Statistics*, **43**, 95-108.

Lehnen, R.G. and Koch, G.G. (1974) Analyzing panel data with uncontrolled attrition. *Public Opinion Quarterly*, **38**, 40–56.

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*, **53**, 570–582.

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

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

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

Lilienfeld, D.E. and Stolley, P.D. (1994) *Foundations of Epidemiology.* New York: Oxford University Press.

Lindsey, J.K. (1993) *Models for Repeated Measurements.* Oxford: Oxford University Press.

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

Littell, R.C., Milliken, G.A., Stroup, W.W. and Wolfinger, R.D. (1996) *SAS System for Mixed Models.* SAS Institute Inc, Cary, NC, USA.

Little, R.J.A. (1976) Inference about means from incomplete multivariate data. *Biometrika*, **63**, 593–604.

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

Little, R.J.A. (1994b) Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 78.

Little, R.J.A. (1995) Modelling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.

Little, R.J.A. and Wang, Y. (1996) Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, **52**, 98–111.

Little, R.J.A. and Rubin, D.B. (1987) *Statistical Analysis with Missing Data.* New York: Wiley.

Louis, T.A. (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society - Series B*, **44**, 226-233.

Longford, N.T. (1993) *Random Coefficient Models.* Oxford: Oxford University Press.

Mardia, K.V. (1970). *Families of Bivariate Distributions*, London: Griffin.

McArdle, J.J. and Hamagami, F. (1992) Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research*, **18**, 145–166.

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

Meilijson, I. (1989) A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society - Series B*, **51**, 127–138.

Meng, X.-L. and Rubin, D.B. (1991) Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association*, **86**, 899–909.

Michiels, B. and Molenberghs, G. (1995) Protective estimation of longitudinal categorical data with nonrandom dropout. *Lecture Notes in Statistics*, **104**, 177–184.

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 Modenls and Pattern Mixture Models to Analyze Longitudinal Quality of Life Data Subject to Dropout. *Submitted to Statistics in Medicine*

Michiels, B., Molenberghs, G. and Lipsitz, S.R. (1999) A Pattern-Mixture Odds Ratio Model for Incomplete Categorical Data. *Communications in Statistics - Theory and Methods*, **28**, 2843–2870

Molenberghs, G. and Goetghebeur, E. (1997) Simple fitting algorithms for incomplete categorical data. *Journal of the Royal Statistical Society, Series B*, **59**, 401–414.

Molenberghs, G., Goetghebeur, E. and Kenward, M.G. (2001) Sensitivity analysis for incomplete contingency tables. *Applied Statistics*, **50**, 15–30.

Molenberghs, G., Goetghebeur, E., Lipsitz, S.R. and Kenward, M.G. (1999) Non-Random 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. (1999) Missing data perspectives of the fluvoxamine data set: a review. *Statistics in Medicine*, **18**, 2449–2464.

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**, 2237-2256.

Molenberghs, G., Michiels, B. and Kenward, M.G. (1998) Pseudo-likelihood for combined selection and pattern-mixture models for missing data problems. *Biometrical Journal*, **40**, 557–572.

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

Molenberghs, G., Michiels, B. and Lipsitz, S.R. (1999) Selection Models and Pattern-Mixture Models for Incomplete Data with Covariates. *Biometrics*, **55**, 978-983.

Molenberghs, G., Verbeke, G., Thijs, H., Lesaffre, E. and Kenward, M.G. (2001) Influence Analysis to Assess Sensitivity of the Dropout Process. *Computational Statistics and Data Analysis*, **37**, 93–113.

Mori, M., Woolson, R. F. and Woodsworth, G. G. (1992) Application of Empirical Bayes Inference to Estimation of Rate of Change in the Presence of Informative Right Censoring. *Statistics in Medicine*, **11**, 621–631.

Mori, M., Woolson, R. F. and Woodsworth, G. G. (1994) Slope Estimation in the Presence of Informative Censoring: Modeling the Number of Observations as a Geometric Random Variable. *Biometrics*, **50**, 39–50.

Muthén, B., Kaplan, D. and Hollis, M (1987) On structural equation modeling with data that are not missing completely at random. *Psychometrika*, **52**, 431–462.

Nelder, J.A. and Mead, R. (1965) A simplex method for function minimisation. *The Computer Journal*, **7**, 303–313.

Nordheim, E.V. (1984) Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *Journal of the American Statistical Association*, **79**, 772–780.

Park, T. and Brown, M.B. (1994) Models for categorical data with nonignorable nonresponse. *Journal of the American Statistical Association*, **89**, 44–52.

Piantadosi, S. (1997) *Clinical Trials: A Methodologic Perspective.* New York: John Wiley & Sons.

Plackett, R.L. (1965) A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516–522.

Reisberg, B., Borenstein, J., Salob, S.P., Ferris, S.H., Franssen, E. and Georgotas, A. (1987) Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *Journal of Clinical Psychiatry*, **48**, 9–13.

Ripley, B.D. (1981) *Spatial Statistics.* New York: John Wiley and Sons.

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 semi-parametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**, 106–121.

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

Rubin, D.B. (1977) Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys. *Journal of the American Statistical Association*, **72**, 538–543.

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

Rubin, D.B. (1977) Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys. *Journal of the American Statistical Association*, **72**, 538–543.

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

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. and Schenker, N. (1986) Multiple imputation for interval estimation from random samples. *Journal of the American Statistical Association*, **81**, 366–374.

Rubin, D.B., Stern, H.S. and Vehovar, V. (1995) Handling "Don't Know" Survey Responses: The Case of the Slovenian Plebiscite. *Journal of the American Statistical Association*, **90**, 822–828.

Schafer, J.L. (1997) *Analysis of incomplete multivariate data.* London: Chapman and Hall.

Schafer J.L., Khare M. and Ezatti-Rice T.M. (1993) Multiple imputation of missing data in NHANES III. In: *Proceedings of the Annual Research Conference* Washington, DC: Bureau of the Census. pp. 459–487.

Scharfstein, D.O., Rotnizky, A. and Robins, J.M. (1999) Adjusting for non-ignorable drop-out using semiparametric nonresponde models (with discussion). *Journal of the American Statistical Association*, **94**, 1096–1146.

Schipper, H., Clinch, J. and McMurray, A. (1984) Measuring the quality of life of cancer patients: the Functional-Living Index-Cancer: development and validation. *Journal of Clinical Oncology*, **2**, 472–483.

Schluchter, M. (1988) Analysis of incomplete multivariate data using linear models with structured covariance matrices. *Statistics in Medicine*, **7**, 317–324.

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.

Selvin, S. (1996) *Statistical Analysis of Epidemiologic Data.* New York: Oxford University Press.

Smith, D.M., Robertson, B. and Diggle, P.J. (1996) *Object-oriented Software for the Analysis of Longitudinal Data in S.* Technical Report MA 96/192. Department of Mathematics and Statistics, University of Lancaster, LA1 4YF, United Kingdom.

Stasny, E.A. (1986) Estimating gross flows using panel data with nonresponse: an example from the Canadian Labour Force Survey. *Journal of the American Statistical Association*, **81**, 42–47.

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

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

Thijs, H., Molenberghs, G., Verbeke, G. Michiels, B. and Curran, D. (2002) Strategies to fit pattern-mixture models. *Biostatistics*, **0**, 000–000.

Van Steen, K., Molenberghs, G., Verbeke, G. and Thijs, H. (2002) A Local Influence Approach applied to Ordinal Data from a Longitudinal Study. *Biometrics*, **0**, 000–000.

Verbeke, G. and Lesaffre, E. (1999) The effect of drop-out on the efficiency of longitudinal experiments. *Applied Statistics*, **48**, 363–376

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

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

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

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E. and Kenward, M.G. (1998) Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics*, **57**, 7–14.

Verbyla, A.P. and Cullis, B.R. (1990) Modelling in repeated measures experiments. *Applied Statistics*, **39**, 341–356.

Verdonck, A., De Ridder, L., Verbeke, G., Bourguignon, J.P., Carels, C., Kuhn, E.R., Darras, V. and de Zegher, F. (1998) Craniofacial growth in the male rat: Effect of endogeneous testosterone. *Archives of Oral Biology*, **43**, 861–872

Vonesh, E.F. and Chinchilli, V.M. (1997) *Linear and non-linear models for the analysis of repeated measurements*. Basel: Marcel Dekker.

Wainer, H. (1989) Eelworms, Bullet Holes and Geraldine Ferraro: Some Problems with Statistical Adjustment and Some Solutions. *Journal of Educational Statistics*, **14**, 121–140.

Welsh, A.H. (1996) *Aspects of Statistical Inference*. New York: Wiley.

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

Wu, M.C. and Bailey, K.R. (1988) Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, **7**, 337–346.

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.

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

Yates, F. (1933) The analysis of replicated experiments when the field results are incomplete. *Empirical Journal of Experimental Agriculture*, **1**, 129–142.

Zhao, L.P., Lipsitz, S. and Lew, D. (1996) Regression analysis with missing covariate data using estimating equations. *Biometrics*, **52**, 1165–1182.