

Sensitivity analysis for continuous incomplete longitudinal outcomes

Peer-reviewed author version

MOLENBERGHS, Geert; THIJS, Herbert; Kenward, Michael G. & VERBEKE, Geert
(2003) Sensitivity analysis for continuous incomplete longitudinal outcomes. In:
Statistica Neerlandica, 57(1). p. 112-135.

DOI: 10.1111/1467-9574.00224

Handle: <http://hdl.handle.net/1942/424>

Sensitivity Analysis for Continuous Incomplete Longitudinal Outcomes

GEERT MOLENBERGHS¹ HERBERT THIJS¹ MICHAEL G. KENWARD²
GEERT VERBEKE³

¹ Limburgs Universitair Centrum, transnationale Universiteit Limburg

² London School of Hygiene and Tropical Medicine

³ Katholieke Universiteit Leuven

geert.molenberghs@luc.ac.be, herbert.thijs@luc.ac.be,
mike.kenward@lshtm.ac.uk, geert.verbeke@med.kuleuven.ac.be

August 8, 2002

Abstract

Even though models for incomplete longitudinal data are in common use, they are surrounded with problems, largely due to the untestable nature of the assumptions one has to make regarding the missingness mechanism. Two extreme views on how to deal with this problem are (1) to avoid incomplete data altogether and (2) to construct ever more complicated joint models for the measurement and missingness processes. In this paper, it is argued that a more versatile approach is to embed the treatment of incomplete data within a sensitivity analysis. Several such sensitivity analysis routes are presented and applied to a case study, the milk protein trial analyzed before by Diggle and Kenward (1994). Apart from the use of local influence methods, some emphasis is put on pattern-mixture modeling. In the latter case, it is shown how multiple-imputation ideas can be used to define a practically feasible modeling strategy.

Some Key Words: Local Influence, Multiple Imputation, Missing Data, Pattern-Mixture Model, Selection Model.

1 Introduction

The problem of dealing with missing values is common throughout statistical work and is almost everpresent in the analysis of longitudinal, or repeated measurements data. Early work on missing values was largely concerned with algorithmic and computational solutions to the induced lack of balance or deviations from the intended study design. See for example the reviews by Afifi and Elashoff (1966) and Hartley and Hocking (1971). More recently general algorithms such as the expectation-maximisation (EM) (Dempster, Laird, and Rubin 1977), and data imputation and augmentation procedures (Rubin 1987), combined with powerful computing resources have largely

provided a solution to this aspect of the problem. There remains the very difficult and important question of assessing the impact of missing data on subsequent statistical inference. Conditions can be formulated, under which an analysis that proceeds as if the missing data are missing by design, that is, ignoring the missing value process, can provide valid answers to study questions. The difficulty in practice is that such conditions can rarely be assumed to hold. In this paper we review model-based approaches to the analysis of longitudinal data that explicitly allow departures from such conditions. We make a key point here, and return to it repeatedly: however we approach such analyses, assumptions will be required that cannot be assessed from the data under analysis. Hence in this setting there cannot be anything that could be termed a definitive analysis, and we shall argue that the appropriate statistical framework is one of sensitivity analysis.

Certain important concepts are now in common use in the missing value literature. We begin by introducing these. When referring to the missing-value, or non-response, process we will use terminology of Little and Rubin (1987, Chapter 6). A non-response process is said to be *missing completely at random* (MCAR) if the missingness is independent of both unobserved and observed data and *missing at random* (MAR) if, conditional on the observed data, the missingness is independent of the unobserved measurements. A process that is neither MCAR nor MAR is termed *non-random* (MNAR). In the context of likelihood inference, and when the parameters describing the measurement process are functionally independent of the parameters describing the missingness process, MCAR and MAR are *ignorable*, while a non-random process is non-ignorable.

Most methods are formulated within the selection modeling frame (Little and Rubin 1987) as opposed to pattern-mixture modeling (PMM; Little 1993, 1994). A selection model factors the joint distribution of the measurement and response mechanisms into the marginal measurement distribution and the response distribution, conditional on the measurements. This is intuitively appealing since the marginal measurement distribution would be of interest also with complete data. Further, Little and Rubin's taxonomy is most easily developed in the selection setting. Based on the model of Diggle and Kenward (1994), we show how local influence ideas (Cook 1986) can be used as a sensitivity analysis tool for incomplete longitudinal data, in a selection model context.

This will underscore that, especially in the context of non-random missingness models, selection models, although identifiable, should be approached with caution (Glynn, Laird, and Rubin 1986). Therefore, pattern-mixture models, employing the reverse factorization, have gained renewed interest in recent years (Little 1993, 1994, Hogan and Laird 1997). Note that a further framework consist of shared random-effects models (e.g., Wu and Carroll 1988).

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.

The milk protein content trial is introduced in Section 2. Selection models and sensitivity analysis tools based on local influence are described in Section 3. Section 4 is devoted to pattern-mixture models. The milk protein content data are analyzed, in several ways, in Section 5.

2 The Milk Protein Content Trial

The milk protein data were introduced by Verbyla and 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. 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. 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 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.

3 Selection Models and Local Influence

In modeling missing data one is interested in $f(\mathbf{y}_i, d_i | \boldsymbol{\theta}, \boldsymbol{\psi})$ which is the joint distribution of the measurements Y_i and the dropout indicators D_i defined by adding 1 to the time of the last measurement. A first and most popular approach is by using selection models based on the factorization $f(\mathbf{y}_i, d_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \boldsymbol{\theta})f(d_i | \mathbf{y}_i, \boldsymbol{\psi})$. In this framework standard missing data concepts such as MCAR, MAR, MNAR (Rubin 1976, Little and Rubin 1987) can be constructed. Recently more interest is put in the opposite factorization $f(\mathbf{y}_i, d_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | d_i, \boldsymbol{\theta})f(d_i | \boldsymbol{\psi})$ being the basis for pattern-mixture models. Molenberghs, Michiels, Kenward, and Diggle (1998) showed that pattern-mixture models allow for a natural analog of MAR, hence enabling a similar classification of missing data mechanisms. We will use a popular model for repeated measurements, incorporating random effects and serial correlation. A linear mixed-effects model with serial correlation can be written as

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \mathbf{W}_i + \boldsymbol{\varepsilon}_i, \quad (1)$$

(Verbeke and Molenberghs 2000) where \mathbf{Y}_i is the n dimensional response vector for subject i , $1 \leq i \leq N$, N is the number of subjects, X_i and Z_i are $(n \times p)$ and $(n \times q)$ known design matrices, $\boldsymbol{\beta}$ is the p dimensional vector containing the fixed effects, $\mathbf{b}_i \sim N(\mathbf{0}, D)$ is the q dimensional vector containing the random effects, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 I_{n_i})$ is a n dimensional vector of measurement error components, and $\mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N$ are assumed to be independent. The serial correlation is captured by the realization of a Gaussian stochastic process, \mathbf{W}_i , which is assumed to follow a $N(\mathbf{0}, \tau^2 H_i)$ law. The serial covariance matrix H_i only depends on i through the number n of observations and through the time points t_{ij} at which measurements are taken. The structure of the matrix H_i is determined through the autocorrelation function $\rho(t_{ij} - t_{ik})$. Two popular choices for $\rho(\cdot)$ are the exponential and Gaussian models defined respectively as $\rho(u) = \exp(-\phi u)$ and $\rho(u) = \exp(-\phi u^2)$, with $u = |t_{ij} - t_{ik}|$ and $\phi > 0$. Finally, D is a general $(q \times q)$ covariance matrix with (i, j) element $d_{ij} = d_{ji}$. Inference is based on the marginal distribution of the response \mathbf{Y}_i which, after integrating over random effects, can be expressed as

$$\mathbf{Y}_i \sim N(X_i \boldsymbol{\beta}, Z_i D Z_i' + \Sigma_i). \quad (2)$$

Here, $\Sigma_i = \sigma^2 I_{n_i} + \tau^2 H_i$ is a $(n \times n)$ covariance matrix grouping the measurement error and serial components. Random effects in model (1) stem from heterogeneity between subjects.

We assume that incompleteness is due to dropout only, and that the first measurement Y_{i1} is obtained for everyone. The model for the dropout process 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 $g(\mathbf{h}_{ij}, y_{ij})$ in which \mathbf{h}_{ij} is a vector containing all responses observed up to but not including occasion j , as well as relevant covariates. We then assume that $g(\mathbf{h}_{ij}, y_{ij})$ satisfies

$$\text{logit}[g(\mathbf{h}_{ij}, y_{ij})] = \text{logit}[\text{pr}(D_i = j | D_i \geq j, \mathbf{y}_i)] = \mathbf{h}_{ij}\boldsymbol{\psi} + \omega y_{ij}, \quad i = 1, \dots, N. \quad (3)$$

When ω equals zero, the dropout model is random, and all parameters can be estimated using standard software since the measurement model for which we use a linear mixed model and the dropout model, assumed to follow a logistic regression, can then be fitted separately. If $\omega \neq 0$, the dropout process is assumed to be non-random.

Model (3) is now used to construct the dropout process:

$$f(d_i | \mathbf{y}_i, \boldsymbol{\psi}) = \begin{cases} \prod_{j=2}^{n_i} [1 - g(\mathbf{h}_{ij}, y_{ij})] & \text{for a complete sequence } (d_i = n_i + 1), \\ \prod_{j=2}^{d-1} [1 - g(\mathbf{h}_{ij}, y_{ij})] g(\mathbf{h}_{id}, y_{id}) & \text{for a dropout } (d_i = d \leq n_i). \end{cases} \quad (4)$$

Let us now shift attention to sensitivity and influence analysis issues. Whereas a global influence approach is based on case-deletion, a local influence based sensitivity assessment of the relevant quantities, such as treatment effect or time evolution parameters, with respect to assumptions about the dropout model is based on the following perturbed version of (3):

$$\text{logit}(g(\mathbf{h}_{ij}, y_{ij})) = \text{logit}[\text{pr}(D_i = j | D_i \geq j, \mathbf{y}_i)] = \mathbf{h}_{ij}\boldsymbol{\psi} + \omega_i y_{ij}, \quad i = 1, \dots, N, \quad (5)$$

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, the model reduces to a MAR model. Hence (5) 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) and Rubin (1994).

3.1 Influence Diagnostics

Cook (1986) 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, Chatterjee and Hadi 1988) in which the effect is studied of completely removing

cases from the analysis. They were introduced by Cook (1977, 1979) for the linear regression context. Denote the log-likelihood function, corresponding to measurement model (2) and dropout model (3) by

$$\ell(\boldsymbol{\gamma}) = \sum_{i=1}^N \ell_i(\boldsymbol{\gamma}) \quad (6)$$

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

$$\ell_{(-i)}(\boldsymbol{\gamma}) \quad (7)$$

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 (6) and (7) or (subsets of) the estimated parameter vectors $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}_{(-i)}$, with obvious notation. Precisely, we will consider both

$$CD_{1i} = 2(\hat{\boldsymbol{\ell}} - \hat{\boldsymbol{\ell}}_{(-i)}) \quad (8)$$

as well as

$$CD_{2i}(\boldsymbol{\gamma}) = 2 (\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}_{(-i)})' \ddot{L}^{-1} (\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}_{(-i)}). \quad (9)$$

Formulation (9) 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.

In linear regression, global influence is conceptually simple, computationally straightforward and well studied. The latter two of these features do not carry over to more general settings. To overcome these limitations, *local* influence methods have been suggested. The principle is to investigate how the results of an analysis are changed under infinitesimal perturbations of the model. In the framework of the linear mixed model Beckman, Nachtsheim, and Cook (1987) used local influence to assess the effect of perturbing the error variances, the random-effects variances and the response vector. In the same context, Lesaffre and Verbeke (1998) have shown that the local influence approach is also useful for the detection of influential subjects in a longitudinal data analysis. Moreover, since the resulting influence diagnostics can be expressed analytically, they often can be decomposed in interpretable components, which yield additional insights in the reasons why some subjects are more influential than others.

Verbeke *et al* (2001) studied the influence the non-randomness of dropout exerts on the model parameters. Let us briefly sketch the principles of local influence and then apply them to our MNAR problem.

We denote the log-likelihood function corresponding to model (5) by

$$\ell(\boldsymbol{\gamma}|\boldsymbol{\omega}) = \sum_{i=1}^N \ell_i(\boldsymbol{\gamma}|\omega_i) \quad (10)$$

in which $\ell_i(\boldsymbol{\gamma}|\omega_i)$ 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 $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_N)'$ of weights defining the perturbation of the MAR model. Let $\hat{\boldsymbol{\gamma}}$ be the maximum likelihood estimator for $\boldsymbol{\gamma}$, obtained by maximizing $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega}_0)$, and let $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\omega}}$ denote the maximum likelihood estimator for $\boldsymbol{\gamma}$ under $\ell(\boldsymbol{\gamma}|\boldsymbol{\omega})$. Cook (1986) proposed to measure the distance between $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\gamma}}$ by the so-called likelihood displacement, defined by $LD(\boldsymbol{\omega}) = 2(\ell(\hat{\boldsymbol{\gamma}}|\boldsymbol{\omega}_0) - \ell(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\omega}}|\boldsymbol{\omega}))$. Since this quantity can only be depicted when $N = 2$, Cook (1986) proposed to look at local influence, i.e., at the normal curvatures $C_{\mathbf{h}}$ of $\boldsymbol{\xi}(\boldsymbol{\omega})$ in $\boldsymbol{\omega}_0$, in the direction of some N dimensional vector \mathbf{h} of unit length. It can be shown that a general form is given by

$$\begin{aligned} C_{\mathbf{h}}(\boldsymbol{\theta}) &= -2\mathbf{h}' \left[\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i} \Big|_{\omega_i=0} \right]' \ddot{L}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\theta} \partial \omega_i} \Big|_{\omega_i=0} \right] \mathbf{h} \\ C_{\mathbf{h}}(\boldsymbol{\psi}) &= -2\mathbf{h}' \left[\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i} \Big|_{\omega_i=0} \right]' \ddot{L}^{-1}(\boldsymbol{\psi}) \left[\frac{\partial^2 \ell_{i\omega}}{\partial \boldsymbol{\psi} \partial \omega_i} \Big|_{\omega_i=0} \right] \mathbf{h}, \end{aligned}$$

evaluated at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$, where indeed the influence for the measurement and dropout model parameters split, since the second derivative matrix of the log-likelihood, \ddot{L} is block-diagonal with blocks $\ddot{L}(\boldsymbol{\theta})$ and $\ddot{L}(\boldsymbol{\psi})$. Verbeke *et al* (2001) have decomposed local influence into meaningful and interpretable components.

4 Pattern-Mixture Modeling Approach

Fitting pattern-mixture models can be approached in several ways. It is important to decide whether pattern-mixture and selection modeling are to be contrasted with one another or rather the pattern-mixture modeling is the central focus.

In the latter case, it is natural to conduct an analysis, and preferably a sensitivity analysis, *within* the pattern-mixture family. 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. One of those, ACMV (available case missing values), is the natural counterpart of MAR in the PMM framework.

Strategy 2. As opposed to identifying restrictions, model simplification can be done to identify the parameters. Thijs *et al* (2001) discussed several sub-strategies in detail.

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, Beale and Dunne (1997), are also making untestable assumptions, just as in the selection model case. In the identifying restrictions setting on the other hand, the assumptions are clear from the start.

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. An example of such a marginalization is given in Section 5.

4.1 Identifying Restriction Strategies

In line with the results obtained by Molenberghs, Michiels, Kenward, and Diggle (1998), we restrict attention to monotone patterns. In general, let us assume we have $t = 1, \dots, T$ dropout patterns where the dropout indicator, introduced earlier, is $d = t + 1$. For pattern t , the complete data density is given by

$$f_t(y_1, \dots, y_T) = f_t(y_1, \dots, y_t) f_t(y_{t+1}, \dots, y_T | y_1, \dots, y_t). \quad (11)$$

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

$$f_t(y_s | y_1, \dots, y_{s-1}) = \sum_{j=s}^T \omega_{sj} f_j(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (12)$$

We will use ω_s as shorthand for the set of ω_{sj} ’s used. Every ω_s which sums to one provides a valid identification scheme.

Let us incorporate (12) into (11):

$$f_t(y_1, \dots, y_T) = f_t(y_1, \dots, y_t) \prod_{s=0}^{T-t-1} \left[\sum_{j=T-s}^T \omega_{T-s,j} f_j(y_{T-s} | y_1, \dots, y_{T-s-1}) \right]. \quad (13)$$

Let us consider three special but important cases. Little (1993) proposes CCMV (complete case missing values) which uses the following identification:

$$f_t(y_s | y_1, \dots, y_{s-1}) = f_T(y_s | y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (14)$$

In other words, information which is unavailable is always borrowed from the completers. Alternatively, the nearest identified pattern can be used:

$$f_t(y_s|y_1, \dots, y_{s-1}) = f_s(y_s|y_1, \dots, y_{s-1}), \quad s = t + 1, \dots, T. \quad (15)$$

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

The third special case of (12) will be ACMV. Thus, ACMV is reserved for the counterpart of MAR in the PMM context. The corresponding ω_s vectors can be shown to have components:

$$\omega_{sj} = \frac{\alpha_j f_j(y_1, \dots, y_{s-1})}{\sum_{\ell=s}^T \alpha_\ell f_\ell(y_1, \dots, y_{s-1})}, \quad (16)$$

where α_j is the fraction of observations in pattern j (Molenberghs, Michiels, Kenward and Diggle 1998).

This MAR-ACMV link connects the selection and pattern-mixture families. It is further of interest to consider specific sub-families of the MNAR family. In the selection model context, (3) restricts attention to a class of mechanisms where dropout may depend on the current, possibly unobserved, measurement, but not on future measurements. The entire class of such models will be termed non-future dependent (MNFD). While they are natural and easy to consider in a selection model context, there exist important examples of mechanisms that do not satisfy MNFD, such as shared-parameter models (Wu and Bailey 1989, Little 1995).

Kenward, Molenberghs, and Thijs (2001) have shown there is a counterpart to MNFD in the pattern-mixture context. The MNFD selection models obviously satisfy

$$f(r = t|y_1, \dots, y_T) = f(r = t|y_1, \dots, y_{t+1}). \quad (17)$$

Within the PMM framework, we define non-future dependent missing value restrictions (NFMV) as follows:

$$\forall t \geq 2, \forall j < t - 1 : f(y_t|y_1, \dots, y_{t-1}, r = j) = f(y_t|y_1, \dots, y_{t-1}, r \geq t - 1). \quad (18)$$

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

$$f(y_{t+1}|y_1, \dots, y_t, r = t). \quad (19)$$

In other words, the distribution of the “current” unobserved measurement, given the previous ones, is unconstrained. Note that (18) excludes such mechanisms as CCMV and NCMV. Kenward, Molenberghs, and Thijs (2001) have shown that, for longitudinal data with dropouts, MNFD and NFMV are equivalent.

For pattern t , the complete data density is given by

$$f_t(y_1, \dots, y_T) = f_t(y_1, \dots, y_t) f_t(y_{t+1}|y_1, \dots, y_t) f_t(y_{t+2}, \dots, y_T|y_1, \dots, y_{t+1}). \quad (20)$$

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 and third components. First, From the data, estimate $f_t(y_1, \dots, y_t)$. Second, the user has full freedom to choose

$$f_t(y_{t+1}|y_1, \dots, y_t). \quad (21)$$

Substantive considerations can be used to identify this density. Or a family of densities can be considered by way of sensitivity analysis. Third, using (18), the densities $f_t(y_j|y_1, \dots, y_{j-1})$, ($j \geq t+2$) 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.

Two obvious mechanisms, within the MNFD family but outside MAR, are FD1, i.e., choose (21) according to CCMV, and FD2, i.e., choose (21) according to NCMV. FD1 and FD2 are strictly different from CCMV and NCMV.

4.2 How to Use Restrictions ?

We will briefly outline a general strategy. Several points which require further specification will be discussed in what follows. (1) Fit a model to the pattern-specific identifiable densities: $f_t(y_1, \dots, y_t)$. This results in a parameter estimate, $\hat{\gamma}_t$. (2) Select an identification method of choice. (3) Using this identification method, determine the conditional distributions of the unobserved outcomes, given the observed ones:

$$f_t(y_{t+1}, \dots, y_T|y_1, \dots, y_t). \quad (22)$$

(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 (22). (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).

We have seen how general identifying restrictions (12), 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 ω , such as in (16) for ACMV. In addition, we need to draw imputations from the conditional densities.

Let us proceed by studying the special case of three measurements first. To this end, we consider an identification scheme 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: from pattern 3, $f_3(y_1, y_2, y_3)$; from pattern 2, $f_2(y_1, y_2)$; and from pattern 1 $f_1(y_1)$. (2) To properly account for the uncertainty with which the parameters are estimated, we need to draw

from them as is customarily done in multiple imputation. It will be assumed that in all densities from which we draw, this parameter vector is used. (3) **For pattern 2.** Given an observation in this pattern, with observed values (y_1, y_2) , calculate the conditional density $f_3(y_3|y_1, y_2)$ and draw from it. (4) **For pattern 1.** We now have to distinguish three substeps.

1. There is now only one ω involved: for pattern 1, in order to determine $f_1(y_2|y_1)$, as a combination of $f_2(y_2|y_1)$ and $f_3(y_2|y_1)$. Every ω in the unit interval is valid. Specific cases are: for NCMV, $\omega = 1$; for CCMV, $\omega = 0$; for ACMV, ω is calculated from (??). Note that, given y_1 , this is a constant, depending on α_2 and α_3 . For FD1 and FD2, the first unidentified conditional density can be chosen freely, thereafter a system of ω 's has to be chosen as well.

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.

2. If $U \leq \omega$, calculate $f_2(y_2|y_1)$ and draw from it. Otherwise, do the same based on $f_3(y_2|y_1)$.
3. 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).

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

In several cases, the conditional density is a mixture of normal densities. Then an additional and straightforward draw from the components of the mixture is necessary.

5 Analysis of Milk Protein Trial Data

In this section we will present several analyses of the milk protein trial, introduced in Section 2. We will start by briefly reconsidering the analyses done by Diggle and Kenward (1994) and Kenward (1998) where a first attempt was made towards sensitivity analysis. Thereafter, a sensitivity analysis will be presented using global influence, local influence, and pattern-mixture modeling.

Diggle and Kenward (1994) considered a linear mixed-effects measurement model, 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. Their model can be expressed as follows:

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

To acknowledge the fact that dropout starts from week 15 onwards, the product in (4) is over $j = 15, \dots, n_i$ instead of over $j = 2, \dots, n_i$.

Diggle and Kenward (1994) found that the dropout 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 date, 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. We will return to a more formal pattern-mixture analysis later, and 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:

$$\text{logit}[g(y_{i,j-1}, y_{ij})] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij}. \quad (23)$$

As in the Diggle and Kenward (1994) model, the product in (4) is over $j = 15, \dots, n_i$ instead of over $j = 2, \dots, n_i$.

Parameter estimates for this model under both MAR and MNAR, are listed in Tables 1 and 1. A number of additional analyses, based on deletion of certain subgroups, will be discussed in Section 5.4. 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 (23) is retained, while the product in (4) is over $j = 2, \dots, n_i$, 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 not shown. 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. As results from theory, the measurement model parameters do not change under the MAR model. 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 as well, 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 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.

Table 1: *Milk protein trial, maximum likelihood estimates (standard errors) of the MAR and MNAR models, 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.*

MAR				
Effect	Par.	all	set 1	set 2
<u>Measurement model:</u>				
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)
<u>Dropout model:</u>				
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
MNAR				
<u>Measurement model:</u>				
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)
<u>Dropout model:</u>				
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

5.1 Global Influence

Global influence results are shown in Figures 1–3. They are based on fitting a MNAR model for each cow deleted. The Cook’s distances for the first model are shown in Figure 2. The individual

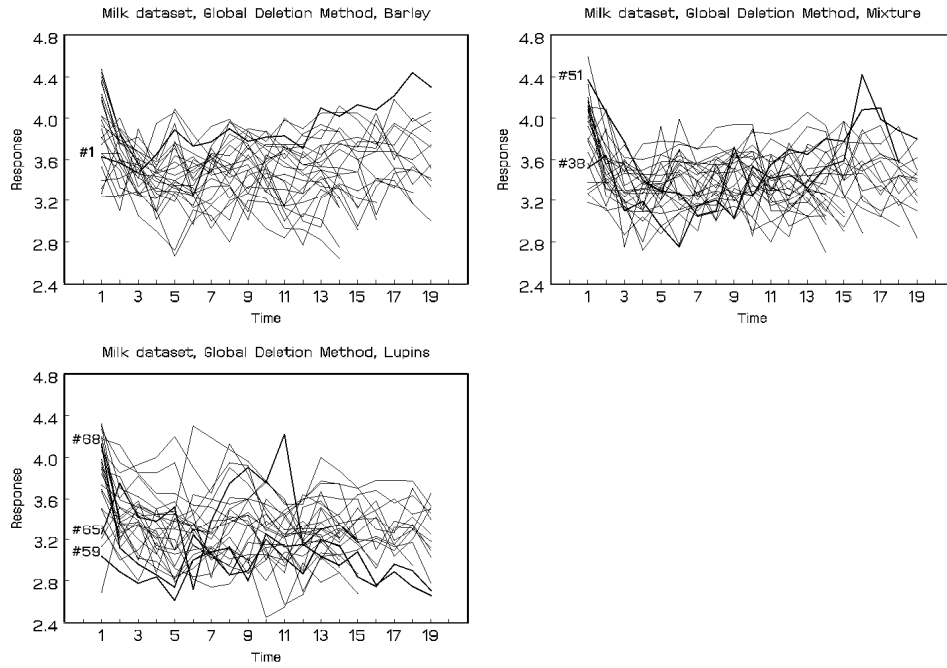


Figure 1: *Milk protein trial, individual profiles, with globally influential subjects highlighted, dropout modeled from week 15.*

curves with influential subjects highlighted are plotted in Figure 1 where subject #38 pertains to the first model only.

All of the Cook's distance plots are very similar. 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. Figure 3 considers the CD_{2i} measures for the diet group contrasts. While Figure 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 $CD_{2i}(\cdot)$, whether based on the entire parameter vector γ , 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

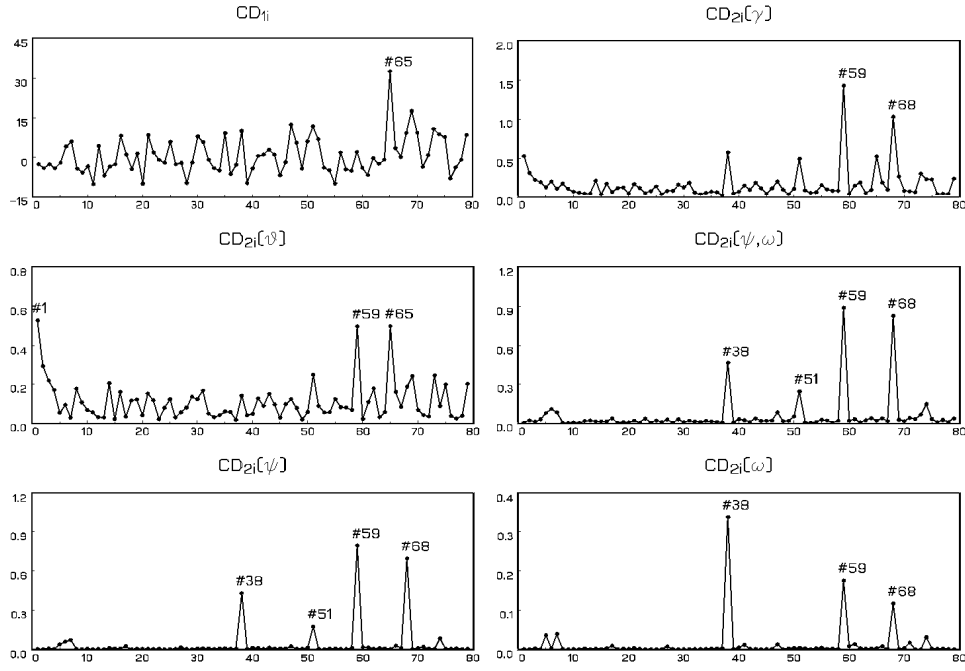


Figure 2: *Milk protein trial, index plots of CD_{1i} , $CD_{2i}(\gamma)$, $CD_{2i}(\theta)$, $CD_{2i}(\psi, \omega)$, $CD_{2i}(\psi)$, $CD_{2i}(\omega)$, dropout modeled from week 15.*

way. From a plot of the dropout parameter estimates for each deleted case (not shown here) it is very clear that upward peaks in $\hat{\psi}_{0(-i)}$ for subjects #51 and #59 are compensated with downward peaks in $\hat{\omega}_{(-i)}$. 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 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

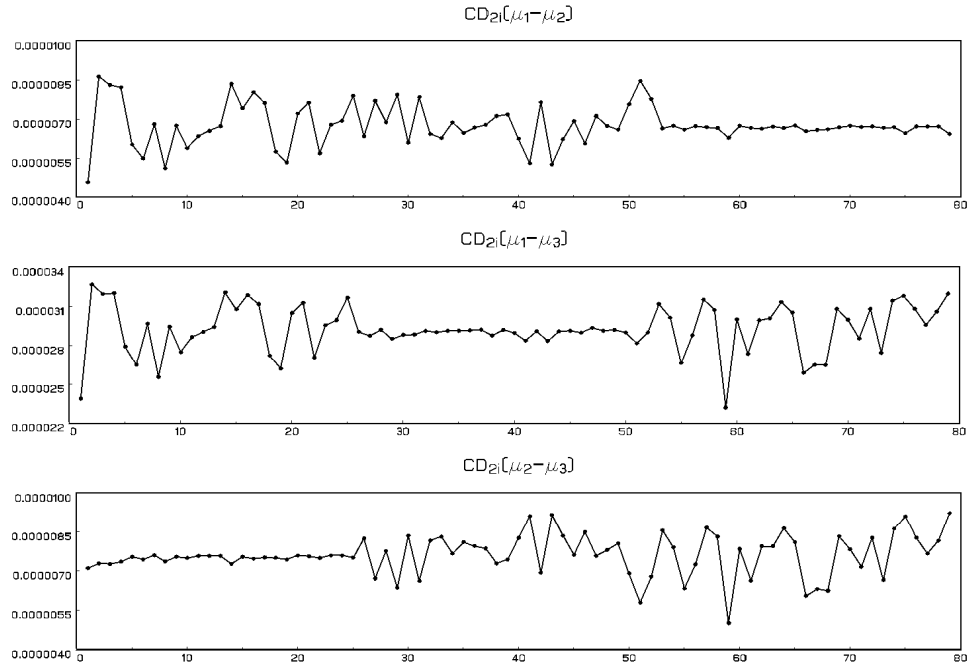


Figure 3: *Milk protein trial, index plots of $CD_{2i}(\mu_1 - \mu_2)$, $CD_{2i}(\mu_1 - \mu_3)$, $CD_{2i}(\mu_2 - \mu_3)$, dropout modeled from week 15.*

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.

5.2 Local Influence

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

$$\psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij} = \lambda_0 + \lambda_1 y_{i,j-1} + \lambda_2 (y_{ij} - y_{i,j-1}). \quad (24)$$

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

The plots for C_i and $C_i(\boldsymbol{\psi})$ are virtually identical (not shown). This is due to the relative magnitudes of the $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ components. Profiles #51, #59, and #66–#68 are influential when dropout is modeled from week 2 onwards. An explanation for the influence in $\boldsymbol{\psi}$ is found by studying the local influence expressions in some detail. Without going into detail, this expression is similar to

$$F(y) = y^2 g(1 - g), \quad (25)$$

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.

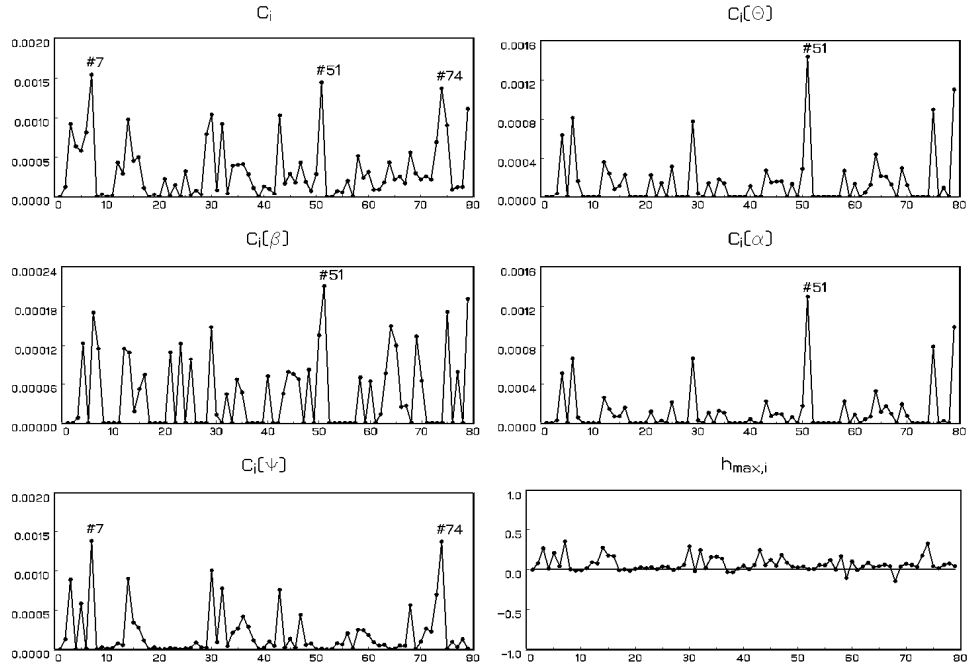


Figure 4: *Milk protein trial, index plots of C_i , $C_i(\theta)$, $C_i(\beta)$, $C_i(\alpha)$, $C_i(\psi)$, and of the components of the direction \mathbf{h}_{\max} of maximal curvature, dropout modeled from week 15, incremental analysis.*

Apparently, for ψ_0 and ψ_1 the maximum is obtained for $y = 2.51$, exactly as seen in the influential profiles, which are all in the lupins group. 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 (25) is the square of the response. This is a direct consequence of our raw parameterization of the dropout process, the logit of which is 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 (23) 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 (24) are equivalent for model fitting purposes, Molenberghs *et al* (1999) showed that they lead to different perturbation schemes of the form (23). 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.

Finally, we will compare both models. The direct-variable results found in the analysis from week 15 agree fairly well with those from week 2, 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 (25). 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. While the direction of maximal curvature still shows no unusual subjects, C_i shows somewhat different subjects to be 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.

5.3 Overview of Global and Local Influence

Table 2 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 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

Table 2: *Milk protein trial, summary of influential subjects.*

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

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.

5.4 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 2). 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 1 – 1. 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 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.

5.5 Pattern-Mixture Analysis

We will group the various occasions of dropout into three patterns ($t = 1, 2, 3$): (1) those dropping out prior to week 14, (2) those dropping out from week 15 until week 18, and (3) the completers, having data on all 19 weeks. To each pattern, we now fit a model equal to the selection model described at the start of Section 5. Since we have three sets of group differences rather than a single one, the marginal diet effects can be composed using:

$$\beta_\ell = \sum_{t=1}^3 \beta_{\ell t} \pi_t, \quad \ell = 1, 2, \quad (26)$$

where ℓ runs over both diet contrasts. The corresponding variance is

$$\text{Var}(\beta_1, \beta_2) = AVA'$$

where

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

and

$$A = \frac{\partial(\beta_1, \beta_2)}{\partial(\beta_{11}, \dots, \beta_{23}, \pi_1, \pi_2, \pi_3)}.$$

Given that marginal quantities have been determined using the pattern-mixture approach, it is useful to compare them to the corresponding selection model. This is done graphically in Figure 5. Let us now turn attention to the assessment of diet effect. Using a 2 degree of freedom selection model test, one obtains $F = 8.51$, whereas the marginalized pattern-mixture model test is $F = 17.82$ (also on 2 degrees of freedom). The test is larger since one corrects for pattern-specific differences, thereby rendering the test more efficient. Alternatively, a stratified test can be conducted in the pattern-mixture setting, i.e., a test on all 6 treatment differences simultaneously, providing $F = 6.05$.

The pattern-mixture model fitted belongs to Strategy 2, described in Section 4. Alternatively, Strategy 1 can be employed, in particular the CCMV, NCMV, ACMV, FD1, and FD2 mechanisms.

This way, we have 5 marginal diet effects tests. The results are displayed in Table 3. Clearly, there is very little difference among the test statistics, thus suggesting that the conclusion of a strong difference between diets is largely unaffected across a number of identifying restrictions. However, there is a substantial difference with the marginalized test of diet effect, reached earlier under Strategy 2, underscoring that simple models, while tempting to use, could be deceptive.

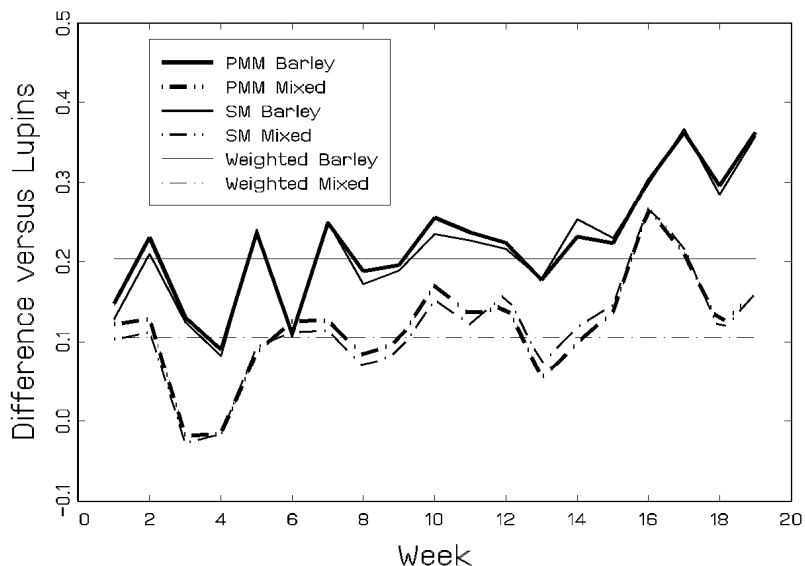


Figure 5: Graphical comparison of selection and pattern-mixture models. The lines correspond to the diet effects relative to the lupins diet. The weighted profile refers to a single-parameter pattern-mixture diet effect, as in (26).

Table 3: Marginal (2 d.f.) tests for diet effect (p values) for each of the 5 identifying restrictions considered.

Restrictions	F	p value
CCMV	6.08	< 0.0001
NCMV	5.56	< 0.0001
ACMV	5.70	< 0.0001
FD1	6.02	< 0.0001
FD2	5.77	< 0.0001

6 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* 2001) and global influence (Chatterjee and Hadi 1988) on the one hand, and pattern-mixture modeling on the other hand. These methods were applied to the milk protein data of Diggle and Kenward (1994).

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.

We 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. 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.

We believe that our approaches can play a useful role, as a member of a collection of sensitivity tools. Of course, a sensitivity analysis can be conducted within yet 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.

Acknowledgements

We gratefully acknowledge support from *Fonds Wetenschappelijk Onderzoek-Vlaanderen* Research Project G.0002.98 “Sensitivity Analysis for Incomplete and Coarse Data”, and from *Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in Industrie*.

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.
- Beckman, R.J., Nachtsheim, C.J. and Cook, R.D. (1987) Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413–426.
- Chatterjee, S. and Hadi, A.S. (1988) *Sensitivity Analysis in Linear Regression*. New York: Wiley.
- Cook, R.D. (1977) Detection of influential observations in linear regression. *Technometrics*, **19**, 15–18.
- 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 & Hall.
- 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. (1998) Milk protein trial: informative dropout versus random drop-in. *Technical Report*.
- 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. (1990) *Time Series: A Biostatistical Introduction*. Oxford: Oxford University Press.
- Diggle, P.J. and Kenward, M.G. (1994) Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.
- Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B*, **57**, 45–97.
- Glynn, R.J., Laird, N.M. and Rubin, D.B. (1986) Selection modeling versus mixture modeling with nonignorable nonresponse. In *Drawing Inferences from Self-Selected Samples*, Ed. H. Wainer, pp. 115–142. New York: Springer-Verlag.
- Hartley, H.O. and Hocking, R. (1971) The analysis of incomplete data. *Biometrics*, **27**, 7783–7808.
- 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.
- 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., Molenberghs, G. and Thijs, H. (2001) Pattern-mixture models with proper time dependence. *Submitted for publication*.
- Laird, N.M. (1994) Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 84.
- Lesaffre, E. and Verbeke, G. (1998) Local influence in linear mixed models. *Biometrics*, **54**, 570–582.
- Little, R.J.A. (1993) Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, **88**, 125–134.
- Little, R.J.A. (1994) A class of pattern-mixture models for normal incomplete data. *Biometrika*, **81**, 471–483.
- Little, R.J.A. (1995) Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.

- Little, R.J.A. and Rubin, D.B. (1987) *Statistical Analysis with Missing Data*. New York: Wiley.
- 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., Goetghebeur, E.J.T., 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., Kenward, M. G. and Lesaffre, E. (1997) The analysis of longitudinal ordinal data with non-random dropout. *Biometrika*, **84**, 33–44.
- Molenberghs, G., Michiels, B. and Lipsitz, S.R. (1999) Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics*, **55**, 978–983.
- Rubin, D.B. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.
- 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.
- Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.
- 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.
- 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.
- Thijs, H., **Molenberghs, G.**, Michiels, B., Verbeke, G., and Curran, D. (2001) Strategies to fit pattern-mixture models. *Biostatistics*, **3**, 245–265.
- Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E. and Kenward, M.G. (2001) Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics*, **57**, 7–14.
- Verbyla, A.P. and Cullis, B.R. (1990) Modelling in repeated measures experiments. *Applied Statistics*, **39**, 341–356.
- 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.