A Local Influence Approach applied to Binary Data from a Psychiatric Study
Peer-reviewed author version

A Local Influence Approach applied to Binary Data from a Psychiatric Study

Ivy Jansen, Geert Molenberghs, Marc Aerts,

Herbert Thijs, Kristel Van Steen

Biostatistics, Center for Statistics, Limburgs Universitair Centrum,
Universitaire Campus, B-3590 Diepenbeek, Belgium

SUMMARY

Recently, a lot of concern has been raised about assumptions needed in order to fit
statistical models to incomplete multivariate and longitudinal data. In response,
research efforts are being devoted to the development of tools that assess the
sensitivity of such models to often strong but always, at least in part, unverifiable
assumptions. Many efforts have been devoted to longitudinal data, primarily in
the selection model context, although some researchers have expressed interest
in the pattern-mixture setting as well.

A promising tool, proposed by Verbeke et al. (2001), is based on local influence
(Cook, 1986). These authors considered the Diggle and Kenward (1994) model,
which is based on a selection model, integrating a linear mixed model for contin-
uous outcomes with logistic regression for dropout. In this paper, we show that
a similar idea can be developed for multivariate and longitudinal binary data,
subject to non-monotone missingness. We focus on the model proposed by Baker,
Rosenberger, and DerSimonian (1992). The original model is first extended to
allow for, possibly continuous, covariates, whereafter a local influence strategy is
developed to support the model building process. The model is able to deal with
non-monotone missingness but has some limitations as well, stemming from the
conditional of the model parameters. Some analytical insight is provided into the
behavior of the local influence graphs.

Key words: Contingency Table; Influence Analysis; Longitudinal Binary Data;
Non-random Missingness; Sensitivity Parameter.

1

1   INTRODUCTION

Categorical data modeling has received a lot of attention during the past decades (Agresti, 1990). One distinguishes between three modeling families: (1) marginal models (Bahadur, 1961; Ashford and Sowden, 1970; Molenberghs and Lesaffre, 1994, 1999); (2) conditional models where parameters associated with a particular set of outcomes are interpreted relative to values for (a subset of) the other outcomes (Cox, 1972; Rosner, 1984; Liang and Zeger, 1989; Molenberghs and Ryan, 1999); (3) random-effects approaches (Stiratelli, Laird, and Ware, 1984; Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993). For reviews see Pendergast et al. (1996) and Fahrmeir and Tutz (1994).

When referring to the missing-value, or nonresponse, process we will use terminology of Little and Rubin (1987, Ch. 6). A nonresponse 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 *nonrandom* (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 nonrandom process is nonignorable.

A model for incomplete data starts from the joint distribution of the outcomes, Y say, and the non-response process, R say. Model families can be distinguished based on the way this joint distribution $f(y, r)$ is factorized (parameters are suppressed from notation). A selection model is based on the factorization $f(y)f(r|y)$, whereas the reverse factorization is referred to as a pattern-mixture model. When a common set of random-effects is thought to influence both the Y and R processes, conditional upon which these processes are independent, then

the so introduced model is referred to as a shared-parameter model. For reviews, see Little (1995), and Kenward and Molenberghs (1998). Baker, Rosenberger, and DerSimonian (1992, henceforth referred to as BRD) proposed a model for bivariate binary data subject to non-random non-response.

There has been a growing awareness of the sensitivity of such models with respect to the underlying model assumptions. It turns out in practice that there are numerous subtle issues not encountered with complete tables. Some of these issues are purely technical (nonunique, invalid, or boundary estimates), others are of a more interpretational and philosophical nature (e.g., models that yield the same or similar fits to the observed data can produce qualitatively different predictions for the unobserved data; Molenberghs et al., 1999a). In response, several authors have illustrated the need of and/or proposed tools for sensitivity analysis (Rubin, 1977; discussion of Diggle and Kenward, 1994; Draper, 1995; Glynn, Laird, and Rubin, 1986; Scharfstein, Robins, and Rotnitzky, 1999; Molenberghs, Kenward, and Goetghebeur, 2001).

Verbeke et al. (2001) and Thijs, Molenberghs, and Verbeke (2000) developed a local-influence based approach for the detection of subjects that strongly influence the conclusions. These authors focused on the Diggle and Kenward (1994) model for continuous outcomes. Van Steen et al. (2001) adapted these ideas to the model of Molenberghs, Kenward, and Lesaffre (1997), for monotone repeated ordinal data.

In this paper, we focus on the model family proposed by BRD. A number of contributions are made. First, the model is reformulated such that its membership of the selection model family is unambiguously clear. Second, the original model is extended to accommodate for, possibly continuous, covariates, turning the model into a regression tool for several categorical outcomes. Third, a parameterization is proposed that avoids the risk of invalid

solutions. In other words, all combinations of the natural parameters produce probabilities between 0 and 1. As a consequence, the closed-form solutions of BRD no longer apply; given the focus on continuous covariates, the derivation of closed-form solutions should not be of primary concern. Fourth, a local influence approach is coupled with the model strategy, to assess which observations have a strong impact on the comparison of two nested models within the BRD family. Finally, some insight is provided into the behavior of the derived local influence measures. In spite of the advantages, there are some limitations, stemming from the conditional nature of parameter interpretation. This is especially the case when planned sequences are of unequal length and/or time-varying covariates are present.

The rest of the paper is organized as follows. Section 2 introduces the data from a psychiatric study. In Section 3 we sketch the BRD models, extended to incorporate covariate effects, and its application to the psychiatric study is presented in Section 4. In Section 5, the local influence methodology as introduced by Cook (1986) is reviewed. Finally, this method is exemplified on the psychiatric data in Section 6.

2 THE PSYCHIATRIC STUDY

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

After enrollment into the study, a number of baseline characteristics was scored, and the patient was assessed at four follow-up visits. The therapeutic effect and the extent of worsening side effects were scored at each visit on an ordinal scale. A side effect occurs if new

symptoms appear while there is therapeutic effect if old symptoms disappear. We will focus on a dichotomized version (present/absent) of side effects at the first and the last visit. While this may appear to be a limitation, it will allow us to highlight all features of the model and its associated influence analysis (including the effect of covariates, non-monotone missingness, boundary solutions, and several qualitative features of the influence analysis).

Accumulated experience with fluvoxamine in controlled clinical trials has shown that it is effective as a conventional tricyclic antidepressant (Burton, 1991). However, many patients who suffer from depression have concomitant morbidity with conditions such as obsessive-compulsive disorder, anxiety disorders and, to some extent, panic disorders. In most trials, patients with comorbidity are excluded and therefore, it is of interest to gather evidence as to the importance of such factors, with a view on improved diagnosis and treatment. A useful, easy to obtain and quantitative covariate, strongly related to the history of comorbidity, is duration of the mental illness, prior to inclusion in the trial (*prior duration*). Its effect on the clinical outcomes of the study is therefore of scientific importance (Lesaffre, Molenberghs, and Dewulf, 1996) and will be studied here.

The observed data are given in Table 1, both without stratification for covariate levels and broken up by binary duration category (below and above 4 years). A small subgroup has missing information on the duration covariate. The entire set of data, with duration in continuous format, is available from the authors upon request. There are two patients with a non-monotone pattern of follow-up while 14 subjects have no follow-up data at all. Therefore we will need models which can handle non-monotone missingness patterns, as will be described in Section 3.

3 MODELS FOR NON-MONOTONE PATTERNS

Baker, Rosenberger, and DerSimonian (1992) considered a log-linear type of model for two possibly binary outcomes, subject to non-monotone missingness. Let $i = 1, \ldots, n$ index distinct covariate levels. In this section, the index i will be suppressed from notation. Let $j, k = 1, 2$ correspond to the outcome categories of the first and second measurement, respectively and let $r_1, r_2 = 0, 1$ correspond to the missingness indicators (1 for an observed and 0 for a missing measurement). This leads to a four-way classification as in Table 1. The complete data and observed data cell probabilities $\pi_{r_1 r_2, jk}$ for this setting are presented in Table 2. To accommodate (possibly continuous) covariates, we will use a selection model parameterization, differing from and extending the original one:

$$\pi_{r_1 r_2, jk} = p_{jk} q_{r_1 r_2 | jk}, \tag{3.1}$$

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

$$p_{jk} = \frac{\exp(\boldsymbol{\theta}_{jk})}{\sum_{j,k-1}^{2} \exp(\boldsymbol{\theta}_{jk})}, \tag{3.2}$$

$$q_{r_1 r_2 | jk} = \frac{\exp[\beta_{jk}(1 - r_2) + \alpha_{jk}(1 - r_1) + \gamma(1 - r_1)(1 - r_2)]}{1 + \exp(\beta_{jk}) + \exp(\alpha_{jk}) + \exp(\beta_{jk} + \alpha_{jk} + \gamma)}. \tag{3.3}$$

No a priori ordering is imposed on the outcomes. The advantage is that genuine multivariate settings (e.g., several questions in a survey) can be handled as well. When deemed necessary, the implications of ordering can be imposed by considering specific models and leaving out others. For example, one may want to avoid missingness on future observations. In the current bivariate case, the index k would have to be removed from α in the above model. To identify the model, we set $\boldsymbol{\theta}_{22} = \mathbf{0}$ and further $\boldsymbol{\theta}_{jk} = X_{jk}\boldsymbol{\eta}$. This allows inclusion of covariate effects which, together with (3.2), is related to the multigroup logistic model (Albert and Lesaffre, 1986). Even though the parameters $\boldsymbol{\eta}$ are conditional in nature and

therefore somewhat difficult to directly interpret in case planned sequences are of unequal length, (3.2) allows easy calculation of the joint probabilities. Note that observed sequences could be of unequal length. The conditional interpretation of the parameters will be a major obstacle in the presence of time-varying covariates, as is often the case in longitudinal studies with longer measurement sequences. Arguably, in such a case, a different model may be more suitable. Generally, computational advantages become increasingly important as the length of the response vector grows. If necessary, specific functions of interest, such as a marginal treatment effect, can be derived. They will typically take the form of non-linear functions. Arguably, a model of the type here can be most useful as a component of a sensitivity analysis, in conjunction with the use of different (e.g., marginal) models.

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

$$W = \begin{pmatrix} X & 0 \\ 0 & Z \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\phi} = (\boldsymbol{\eta}^T, \boldsymbol{\psi}^T)^T. \tag{3.4}$$

The corresponding log-likelihood function can be written as:

$$
\begin{aligned}
\ell &= \sum_{j,k=1}^{2} y_{11jk} \ln \pi_{11jk} + \sum_{j=1}^{2} y_{10j+} \ln(\pi_{10j1} + \pi_{10j2}) \\
&+ \sum_{k=1}^{2} y_{01+k} \ln(\pi_{011k} + \pi_{012k}) \\
&+ y_{00++} \ln(\pi_{0011} + \pi_{0012} + \pi_{0021} + \pi_{0022}) \\
&= \sum_{j,k=1}^{2} \sum_{s=1}^{y_{11jk}} \ln \pi_{11jk} + \sum_{j=1}^{2} \sum_{s=1}^{y_{10j+}} \ln \pi_{10j+} + \sum_{k=1}^{2} \sum_{s=1}^{y_{01+k}} \ln \pi_{01+k} + \sum_{s=1}^{y_{00++}} \ln \pi_{00++}.
\end{aligned}
$$

Computation of derivatives is straightforward. A technical report can be obtained from the authors upon request. To include individual-specific covariates, a subscript i has to be introduced to the vector $\boldsymbol{\xi}_i$ and the matrix W_i and hence to their constituent components.

BRD consider nine identifiable models, setting α_{jk} and β_{jk} constant in one or more indices:

$$\text{BRD1} \;:\; (\alpha_{..}, \beta_{..}) \qquad \text{BRD4} \;:\; (\alpha_{..}, \beta_{.k}) \qquad \text{BRD7} \;:\; (\alpha_{.k}, \beta_{.k})$$
$$\text{BRD2} \;:\; (\alpha_{..}, \beta_{j.}) \qquad \text{BRD5} \;:\; (\alpha_{j.}, \beta_{..}) \qquad \text{BRD8} \;:\; (\alpha_{j.}, \beta_{.k})$$
$$\text{BRD3} \;:\; (\alpha_{.k}, \beta_{..}) \qquad \text{BRD6} \;:\; (\alpha_{j.}, \beta_{j.}) \qquad \text{BRD9} \;:\; (\alpha_{.k}, \beta_{j.}).$$

The nesting structure of these models is schematically represented in Figure 1. BRD considered these using the original parameterization, but carry over to parameterization (3.3) is immediate. Interpretation is straightforward. For example, BRD1 is MCAR, in BRD4 missingness in the first variable is constant, while missingness in the second variable depends on its value.

4   Models Fitted to the Psychiatric Study

In the analysis all patients with known duration level are considered, leaving a total of 310 out of 315 subjects in the study. In the measurement model, the effect of duration is held constant over both visits. Regarding the missingness model, an effect of duration is assumed in both the α and the β parameters. Each of the 9 models is represented by a specific choice for the design and the corresponding parameter vector.

We will consider three sets of BRD models in detail. Table 3 presents models (estimates, s.e., negative loglikelihoods) without duration. In Table 4, duration is added as a covariate to the measurement model only, whereas in the final set (Table 5) the effect of duration is included in both measurement and missingness parts. Sampling zeroes in some of the cells forces certain parameters to lie on the boundary of their corresponding parameter space which, due to the parameterization, is equal to ∞. This should not be seen as a disadvantage of our model, since boundary solutions are a well known feature of MNAR models (Rubin, 1996). The advantage of our parameterization is that either an interior or a boundary solution is obtained, and never an invalid solution.

From Table 3, the likelihood ratio tests fails to reject BRD1 in favor of a more complex model, implying MCAR would be adequate. However, this conclusion changes when duration is included in the measurement model (Table 4). The effect of duration is highly significant, whichever of the BRD models is chosen to conduct a likelihood ratio test. Further, within Table 4, not BRD1 but rather BRD4 provides the most adequate description. The likelihood ratio test statistic for comparing BRD1–4 equals 7.10, while those for BRD4–7 and BRD4–8 are 2.10 and 1.52, respectively. Thus, from this set of models, one observes that duration improves the fit and apparently duration, included in the measurement model, has the effect of changing the nature of the missingness mechanism, by making it more complex, even though it is often believed that including explanatory variables may help explaining structure in the missingness mechanism. BRD4 states that missingness at the second occasion depends on the (possibly unobserved) value at that same occasion, a so-called type I model, in the typology of Baker (2000), in contrast to type II models, where missingness in a variable depends at least also on other, possibly incomplete, assessments. Obviously, such models are particularly vulnerable to assumptions. Up to this point, no covariate effects have been considered on the missingness parameters. When switching to Table 5, including duration in the missingness part, the conclusions change drastically. First, all evidence for non-MCAR missingness disappears and BRD1 comes out as the most adequate description. Second, comparing corresponding BRD models between Tables 4 and 5 (p-values in bottom line of Table 5), it is clear that the effect of duration on the missingness model cannot be neglected.

Important modeling and data analytic conclusions can be drawn. First, it clearly does not suffice to consider covariate effects on the measurement model, but one has to carefully contemplate such effects on the missingness model as well. Therefore, the models in Table 5, should be regarded as the ones of primary interest. Second, it is found that a longer duration implies a less favorable side effects outcome, as well as an increased change of missing visits.

Obviously, duration acts as a confounding variable which, unless included in both parts of the model, may suggest a relationship between the measurement and missingness models and thus one may erroneously be led to believe that the missing data are MNAR. Third, it should be noted that the parameter estimates of duration in the measurement part are remarkably stable. This implies that, in case one is primarily interested in the effect of duration on the occurrence of side effects all 18 models containing this effect provide very similar evidence. While this need not be the case in general, it is a comforting aspect of this particular data analysis. However, while we have reached plausible conclusions, one should still exercise caution, since non-random missingness models heavily rely on untestable assumptions (Verbeke and Molenberghs, 2000). Therefore, it is important to search for observations which may drive these conclusions (Verbeke et al., 2001). This naturally leads to sensitivity analysis, a form of which will be undertaken in the next section.

5 LOCAL INFLUENCE

Due to the sensitivity of parametric MNAR models (Kenward, 1998; Molenberghs et al., 1999a; Scharfstein, Rotnitzky, and Robins, 1999) it is imperative to study this phenomenon, for example by means of local influence, as proposed by Verbeke et al. (2001) and Molenberghs et al. (2001). Local influence studies the effect of infinitesimally small model perturbations around a given null model. It contrasts with the more familiar case-deletion schemes (Cook and Weisberg, 1982).

Verbeke et al. (2001) studied local influence in the context of the Diggle and Kenward (1994) model where a linear mixed measurement model is combined with logistic models for dropout. They studied the effect of perturbing a MAR dropout model in the direction of MNAR. We will consider perturbations of a given BRD model in the direction of a model with one more parameter in which the original model is nested, implying that perturbations lie along the

edges of Figure 1: for each of the nested pairs in Figure 1, the simpler of the two models equates two parameters from the more complex one. For example, BRD4 includes $\beta_{.k}$, $(k = 1, 2)$, whereas in BRD1 only $\beta_{..}$ is included. For the influence analysis, ω_i is then included as a contrast between two such parameters; for the perturbation of BRD1 in the direction of BRD4, one considers $\beta_{..}$ and $\beta_{..} + \omega_i$. Such an ω_i is not a subject-specific parameter, but rather an infinitesimal perturbation. The vector of all ω_i's defines the direction in which such a perturbation is considered. Clearly, other perturbation schemes are possible as well, or one could consider a different route of sensitivity analysis altogether. Ideally, several could be considered within an integrated sensitivity analysis. The BRD family provides a versatile environment for sensitivity analysis, as opposed to the Diggle and Kenward model where, in its basic form, only a few missingness parameters are present. This is due in part to the ability to handle non-monotone missingness. Note that the influence analysis focuses on the missingness model, rather than on the measurement model parameters. This may be seen as slightly odd since often scientific interest focuses on the measurement model parameters. However, it has been documented (discussion to Diggle and Kenward, 1994; Kenward, 1998; Verbeke et al., 2001) that the missingness model parameters are often the most sensitive ones to take up all kinds of misspecification and influential features. These may then, in turn, impact conclusions coming from the measurement model parameters (e.g., time evolution) or combinations from both (e.g., covariate effects for certain groups of responders).

We will give a brief summary of local influence; see also Verbeke et al. (2001). Denote the log-likelihood corresponding to model (3.2)–(3.3) by $\ell(\boldsymbol{\phi}|\boldsymbol{\omega}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\phi}|\omega_i)$, in which $\ell_i(\boldsymbol{\phi}|\omega_i)$ is the contribution of the ith individual. The parameter $\boldsymbol{\phi}$ is as in (3.4). Assume that $\boldsymbol{\omega}$ belongs to an open subset Ω of $I\!\!R^N$. For $\boldsymbol{\omega}$ equal to $\boldsymbol{\omega_0} = (0, 0, \ldots, 0)'$, $\ell(\boldsymbol{\phi}|\boldsymbol{\omega_0})$ is the log-likelihood corresponding to the simpler of the two BRD models.

Let $\widehat{\boldsymbol{\phi}}$ be the maximum likelihood estimator for $\boldsymbol{\phi}$, obtained by maximizing $\ell(\boldsymbol{\phi}|\boldsymbol{\omega_0})$, and

let $\widehat{\boldsymbol{\phi}}_{\omega}$ denote the maximum likelihood estimator for $\boldsymbol{\phi}$ under $\ell(\boldsymbol{\phi}|\boldsymbol{\omega})$. The local influence approach compares $\widehat{\boldsymbol{\phi}}_{\omega}$ with $\widehat{\boldsymbol{\phi}}$. Similar estimates indicate that the parameter estimates are robust w.r.t. perturbations in the direction of the extended model. Cook (1986) proposed to measure the distance between $\widehat{\boldsymbol{\phi}}_{\omega}$ and $\widehat{\boldsymbol{\phi}}$ by the likelihood displacement: $LD(\boldsymbol{\omega}) = 2[\ell(\widehat{\boldsymbol{\phi}}|\boldsymbol{\omega_0}) - \ell(\widehat{\boldsymbol{\phi}}_{\omega}|\boldsymbol{\omega_0})]$. $LD(\boldsymbol{\omega})$ will be large if $\ell(\boldsymbol{\phi}|\boldsymbol{\omega_0})$ is strongly curved at $\widehat{\boldsymbol{\phi}}$, which means that $\boldsymbol{\phi}$ is estimated with high precision, and small otherwise. Therefore, a graph of $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ contains essential information on the influence of perturbations. Since this so-called influence graph can only be depicted when $N = 2$, Cook (1986) proposed to consider local influence, i.e., at the normal curvatures $C_{\boldsymbol{h}}$ of $\boldsymbol{\xi}(\boldsymbol{\omega})$ in $\boldsymbol{\omega_0}$, in the direction of some N dimensional vector $\boldsymbol{h}$ of unit length. Let $\boldsymbol{\Delta}_i$ be the s dimensional vector defined by

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

and define Δ as the $(s \times N)$ matrix with $\boldsymbol{\Delta}_i$ as its ith column. Let $\ddot{L}$ denote the $(s \times s)$ matrix of second order derivatives of $\ell(\boldsymbol{\phi}|\boldsymbol{\omega_0})$ with respect to $\boldsymbol{\phi}$, also evaluated at $\boldsymbol{\phi} = \widehat{\boldsymbol{\phi}}$. Cook (1986) has then shown that $C_{\boldsymbol{h}}$ can be easily calculated by $C_{\boldsymbol{h}} = 2|\boldsymbol{h}'\Delta'\ddot{L}^{-1}\Delta\boldsymbol{h}|$.

$C_{\boldsymbol{h}}$ can be calculated for any direction $\boldsymbol{h}$. One choice is the vector $\boldsymbol{h_i}$ containing one in the ith position and zero elsewhere, corresponding to the perturbation of the ith subject only, reflecting the influence of allowing the ith subject to drop out in a more general fashion than the others. The corresponding local influence measure is $C_i = 2|\boldsymbol{\Delta}'_i\ddot{L}^{-1}\boldsymbol{\Delta}_i|$. Another important direction is the direction $\boldsymbol{h}_{\text{max}}$ of maximal normal curvature C_{max}. It shows how to perturb the model to obtain the largest local changes in the likelihood displacement. It is readily seen that C_{max} is the largest eigenvalue of $-2\,\Delta'\,\ddot{L}^{-1}\,\Delta$, with $\boldsymbol{h}_{\text{max}}$ the corresponding eigenvector. Calculation of local influence measures reduces to evaluation of Δ and $\ddot{L}$ and a convenient computational scheme can be used whenever a program is available to fit the full alternative model, i.e., the model at the end of edge in Figure 1 since it then suffices to

compute the second derivative at $(\widehat{\boldsymbol{\phi}}, \omega_i = 0)$, for each observation separately, from which the $\boldsymbol{\Delta}_i = (\boldsymbol{\phi}, \omega)$ subvector is selected.

6   INFLUENCE ANALYSIS OF THE PSYCHIATRIC STUDY

We will apply the local influence ideas to the BRD models in order to contradict or strengthen the conclusions of Section 4. Whereas all comparisons along the edges of Figure 1 are possible, we focus on the comparison BRD1–4 (Figure 2), since the first one was the most adequate model when no duration effect is included and when duration is included in both parts of the model, while the second one was the model of choice when duration is included in the measurement model only. In addition, we will consider the comparisons BRD4–7 (Figure 3) and BRD4–8 (plot not shown), the supermodels of BRD4. The symbols used in these Figures are: $+$: both observations are available, (1,1) type; ▲: only the first observation is available, (1,0) type; ■: only the second observation is available, (0,1) type; •: both measurements are missing, (0,0) type.

We consider C_i and $\boldsymbol{h}_{\max}$. The top right panel in Figure 2 essentially shows no structure, while in the top left there are two important observations. First, a layering effect is present. This is not surprising, since there are quite a number of discrete features to the model: the responses and the missingness patterns. On the other hand, the continuous covariate duration is included in the measurement model. In this case, mainly the missingness patterns are noticeable, although the top layer shows a good deal of variability.

Two views can be taken. Either, focus is on two observations, #184 and #185, that stand out. These subjects have no measurements at all for side effects. Alternatively, the entire pattern without follow up measurements can be studied. We will return to this issue later in this section. This phenomenon is in contrast to the analyses made by Verbeke et al.

(2001) and Molenberghs et al. (2001) who found that the influential observations are invariably completers. In this case, the situation is different since the "empty" observations are explicitly modeled in the BRD models. Therefore, assumptions about the perturbations in the direction of such observations have an impact on the values such an individual *would have had* had the measurements been made; hence a strong sensitivity. This illustrates that studying influence by means of perturbations in the missingness model may lead to important conclusions regarding the measurement model parameters. Indeed, the measurement model conclusions depend, not only on the observations actually made, but also on the expectation of the missing measurements. In an MNAR model, such expectations depend on the missingness model as well, since they are made *conditional on an observation being missing.* A high level of sensitivity means that the expectations of the missing outcomes and the resulting measurement model parameters strongly depend on the missingness model. Based on this consideration, Verbeke et al. (2001) showed that, in spite of the fact that completers cannot have a direct influence on the measurement model parameters, they still can do so implicitly. Given the strong level of dependence of missingness models on assumptions, it is crucial to investigate the sensitivity of the measurement model conclusions, using local influence that targets the missingness model. As stated earlier, the only continuous characteristics of the observations are the levels for duration. These are 38 and 41, respectively, the largest values within the group without observations and the 91st and 92nd percentile values within the entire sample. Thus, the conclusions are driven by a very high value of duration.

Consider the bottom panels of Figure 2. The right hand panel still shows little or no structure. On the left hand side, the layering has been blurred due to the occurrence of duration as a continuous feature into the missingness model. The fact that no sets of observations stand out as such, confirms the impression that a good fit has been obtained by including duration in both parts of the model. Consider Figure 3. A qualitative difference with Fig-

ure 2 (top left panels) is that the entire group with no follow-up measurements shows more influential than all other subjects. In this case, $h_{\max}$ displays the same group of subjects with no follow-up. However, all of this disappears when one turns to the bottom panels, again underscoring the importance of duration in the missingness model. The consequence of these findings is that, as soon as duration is included in the missingness model, reasonable confidence can be put into the conclusions. Nevertheless, based on the comparison BRD1–4, it seems wise to further study the effect of subjects #184 and #185, as well as from the group without follow up. To this effect, three additional analyses are considered: two sets pertain to removal of subjects #184 and #185: without (I) and with (II) duration as a covariate in the measurement model. We do not consider removal in case duration is included in the missingness model since, in this case, these two subjects did not show up as locally influential. Finally, removing all subjects without follow-up measurements and using duration as covariate in the measurement model is reported as family III.

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

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

7 CONCLUDING REMARKS

We have presented a set of analyses for incomplete binary data. Apart from considering several plausible model strategies (depending on the inclusion of duration as a covariate in the measurement and/or missingness models), a number of sensitivity tools have been proposed and used to strengthen the findings. We have proposed a possible method to assess influence, using local influence (Cook, 1986) in the case of selection models for a bivariate binary outcome subject to non-monotone missingness. To this end, a joint model for outcomes and non-response has been proposed in which (possibly continuous) covariates are allowed. The model is based on an extension of Baker, Rosenberger, and DerSimonian (1992) towards the inclusion of covariates. While we focus on bivariate binary outcomes, the model can be extended to more than two assessments. To this end, extensions of (3.2) and (3.3) would have to be entertained. For both, log-linear type as well as marginal models can be considered. It should be emphasized that similar sensitivity analysis ideas could be developed within other, very interesting, model families, such as the one proposed by Baker (1996). This model would be particularly attractive in the case of ordered missingness. This is an important feature since the particular model considered in this paper suffers from interpretational problems when planned sequences are of unequal length and/or time-varying covariates are included.

In our case study, it turned out that the inclusion of a key covariate in both the measurement model and the missingness model, has to ability to substantially improve the fit of the model

and to explain missingness in the sense that an otherwise seemingly MNAR mechanism is brought back to MCAR. The latter implies a number of methodological and interpretational advantages. A sensitivity analysis, conducted to challenge these findings, has led us to an increased confidence in the data analytic findings.

Similar concepts, in the context of a selection model for continuous longitudinal data subject to dropout, have been used by Verbeke et al. (2001) and Verbeke and Molenberghs (2000) and by Van Steen et al. (2001) for discrete longitudinal data, also subject to dropout. These authors place the perturbation within the logistic regression for dropout, whereas in our case, a family of identified models is considered. These models have been put into an hierarchy (Figure 1) and then infinitesimal perturbations are studied for each model in the direction of all models that have exactly one more parameter in the missingness models. This approach allows detection of several possible patterns. First, different subsets of patients may be influential for different model extensions or, as was the case here, several comparisons may point to the same pair of influential observations. Second, in an influence graph, some subgroups of patients may almost lie on a straight line, while others tend to form a cloud. This is due to the combination of categorical aspects (outcomes, non-response patterns) with continuous aspects (covariates). These can lead to different actions, ranging from design and protocol changes in future studies to removal of observations or groups of observations from the analysis of the current study. However, the latter seemed unnecessary in our analysis, given that including duration as a covariate in the missingness model led to apparently very stable conclusions.

The authors have developed GAUSS code which is available upon request.

ACKNOWLEDGMENT

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

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

Ashford, J.R. and Sowden, R.R. (1970). Multivariate probit analysis. *Biometrics* **26**, 535–546.

Bahadur, R.R. (1961). A representation of the joint distribution of responses to n dichotomous items. In: *Studies in Item Analysis and Prediction,* H. Solomon (Ed.). Stanford Mathematical Studies in the Social Sciences VI. Stanford, CA: Stanford University Press.

Baker, S.G. (1996). The analysis of categorical case-control data subject to nonignorable nonresponse. *Biometrics* **52**, 362–369.

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

Baker, S.G. (2000). Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-none compliance. *Journal of the American Statistical Association* **95**, 43–50.

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

Burton, S.W. (1991). Review of fluvoxamine and its uses in depression. *International Clinical Psychopharmacology* **6** (Supplement 3), 1–17.

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.

Diggle, P.D. and Kenward, M.G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–93.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* **57**, 45–97.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Heidelberg: Springer-Verlag.

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.

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 estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* **50**, 945–953.

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

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.

Liang, K.-Y. and Zeger, S.L. (1989). A class of logistic regression models for multivariate binary time series. *Journal of the American Statistical Association* **84**, 447–451.

Little, R.J.A. (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.

Michiels, B. and Molenberghs, G. (1997). Protective estimation of longitudinal categorical data with nonrandom dropout. *Communications in Statistics, Theory and Methods* **26**, 65–94.

Molenberghs, G., Goetghebeur, E., Lipsitz, S.R., and Kenward, M.G. (1999a). 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. (1999b). Missing data perspectives of the fluvoxamine data set: a review. *Statistics in Medicine* **18**, 2449–2464.

Molenberghs, G., Kenward, M.G., and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics* **50**, 15–29.

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

Molenberghs, G. and Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* **89**, 633–644.

Molenberghs, G., and Lesaffre, E. (1999). Marginal modelling of multivariate categorical data. *Statistics in Medicine* **18**, 2237–2255.

Molenberghs, G. and Ryan, L.M. (1999). Likelihood inference for clustered multivariate binary data. *Environmetrics* **10**, 279–300.

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

Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M., and Fisher, M.R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review* **64**, 89–118.

Rosner, B. (1984). Multivariate methods in ophtalmology with applications to other paired-data situations. *Biometrics* **40**, 1025–1035.

Rubin, D.B. (1977). Formalizing subjective notions about the effect of nonresponse in sample surveys. *Journal of the American Statistical Association* **72**, 538–543.

Rubin, D.B. (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.

Scharfstein, D.O., Rotnitzky, 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.

Stiratelli, R., Laird, N., and Ware, J. (1984). Random effects models for serial observations with dichotomous response. *Biometrics* **40**, 961–972.

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

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

Verbeke, G. and 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**, 43–50.

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

Table 1: *Data from a psychiatric study. Side effects (yes/no) at two occasions.*

yes	89	13
no	57	65

26
49

2	0

14

Table 2: *Theoretical distribution over complete and observed cells of a bivariate binary outcome. Tables correspond to completely observed subjects and subjects with the second, the first, and both measurements missing, respectively.*

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

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

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

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

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

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

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

$\pi_{00,++}$

Table 3: *Maximum likelihood estimates and standard errors of BRD models. All observations included. No covariates.*

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

Table 4: *Maximum likelihood estimates and standard errors of BRD models. All observations included. Duration as covariate in the measurement model.*

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

Table 5: *Maximum likelihood estimates and standard errors of BRD models. All observations included. Duration as covariate in both measurement and missingness model.*

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

† p-value for the comparison with the corresponding BRD model in Table 4, to test the null hypothesis of no effect of duration in the missingness model.

Table 6: *Negative loglikelihood values for three additional sets of analysis. I: #184 and #185 removed, no covariates; II: #184 and #185 removed, duration as covariate in the measurement model; III: all observations in the (0,0) group removed, duration as covariate in the measurement model.*

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

Figure Captions

Figure 1. Graphical representation of the BRD model nesting structure.

Figure 2. Index plots of C_i (left panels) and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature (right panels) for comparison BRD1–4, without (top panels) or with (bottom panels) duration as a covariate in the missingness models.

Figure 3. Index plots of C_i (left panels) and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature (right panels) for comparison BRD4–7, without (top panels) or with (bottom panels) duration as a covariate in the missingness models.
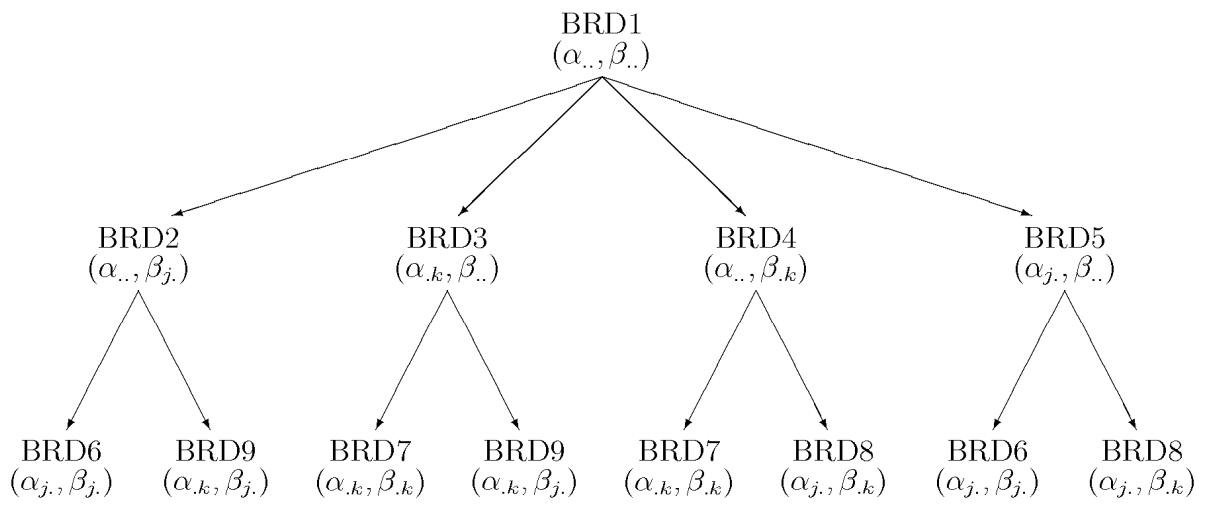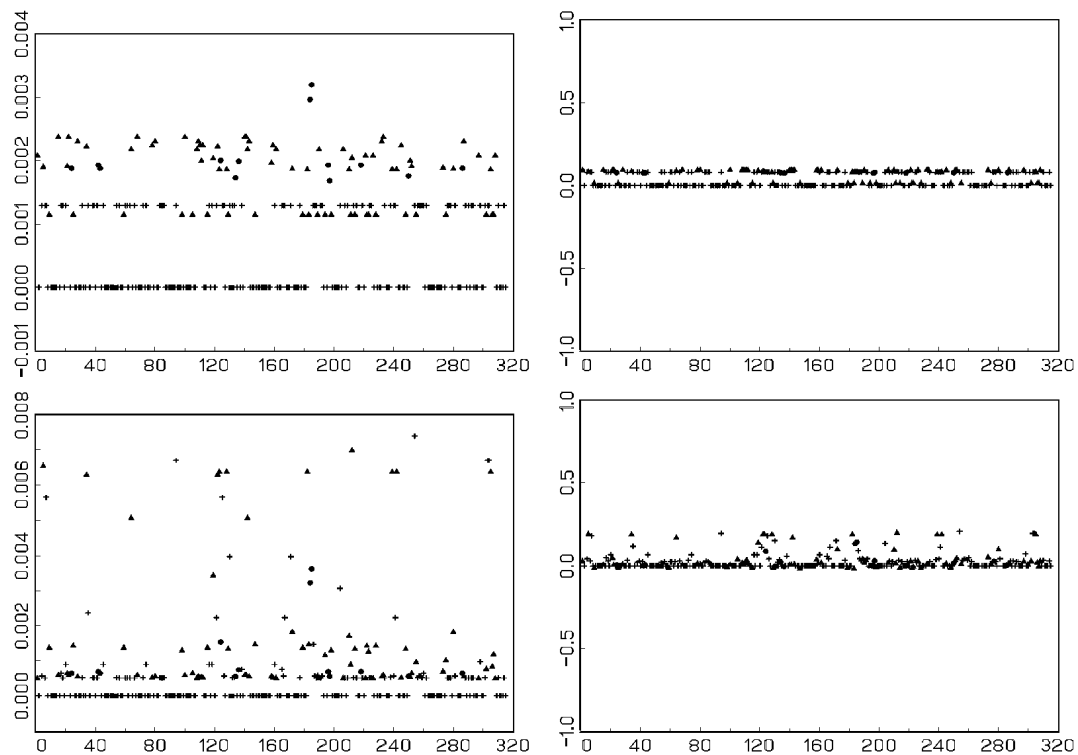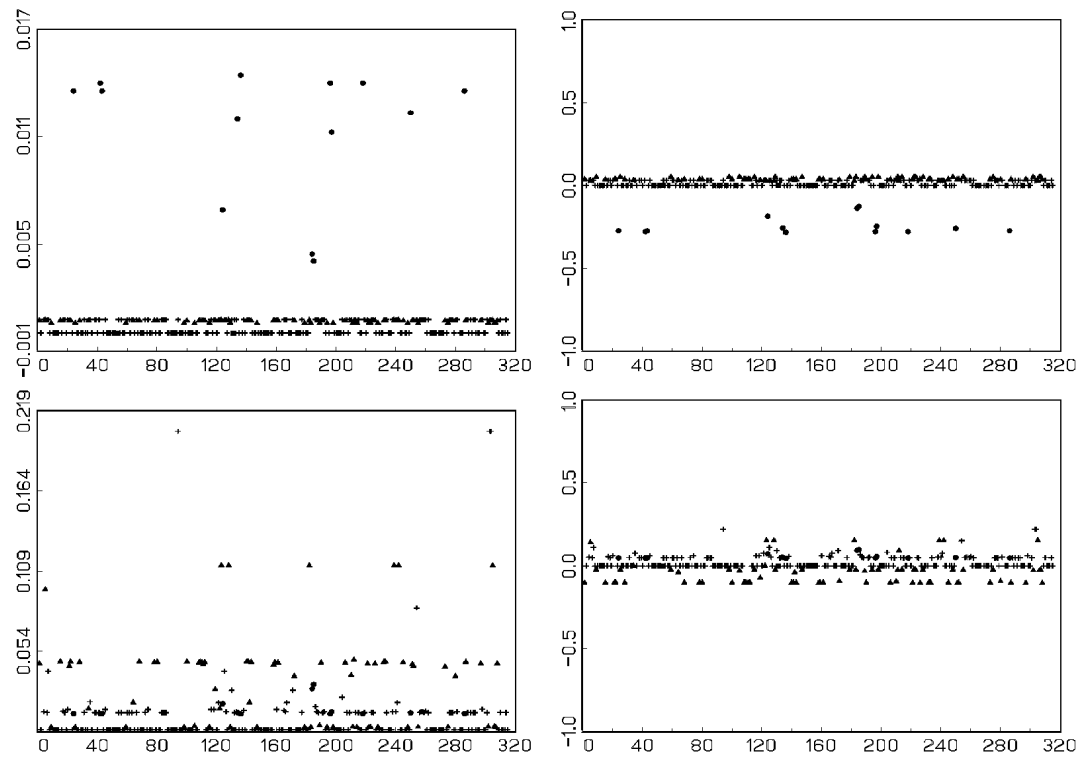
Figure 1:

Figure 2:

Figure 3: