BIOMETRICS 57, 7–14
March 2001

Sensitivity Analysis for Nonrandom Dropout: A Local Influence Approach

Geert Verbeke,[1] Geert Molenberghs,[2,*] Herbert Thijs,[2] Emmanuel Lesaffre,[1] and Michael G. Kenward[3]

[1]Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium
[2]Biostatistics, Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium
[3]Institute of Mathematics and Statistics, The University of Kent, Canterbury, Kent CT2 7NF, U.K.
*email: geert.molenberghs@luc.ac.be

SUMMARY. Diggle and Kenward (1994, *Applied Statistics* **43**, 49–93) proposed a selection model for continuous longitudinal data subject to nonrandom dropout. It has provoked a large debate about the role for such models. The original enthusiasm was followed by skepticism about the strong but untestable assumptions on which this type of model invariably rests. Since then, the view has emerged that these models should ideally be made part of a sensitivity analysis. This paper presents a formal and flexible approach to such a sensitivity assessment based on local influence (Cook, 1986, *Journal of the Royal Statistical Society, Series B* **48**, 133–169). The influence of perturbing a missing-at-random dropout model in the direction of nonrandom dropout is explored. The method is applied to data from a randomized experiment on the inhibition of testosterone production in rats.

KEY WORDS: Compound symmetry; Global influence; Linear mixed model; Missing data; Normal curvature.

1. Introduction

In a longitudinal study, each unit is measured on several occasions. It is not unusual for some sequences of measurements to terminate early for reasons outside the control of the investigator, and any unit so affected is often called a dropout. Little and Rubin (1987, Chapter 6) make important distinctions between different missing values processes. A dropout process is said to be completely random (MCAR) if the dropout is independent of both unobserved and observed data and random (MAR) if, conditional on the observed data, the dropout is independent of the unobserved measurements; otherwise, the dropout process is termed nonrandom (MNAR) or nonignorable. One approach is to estimate from the available data the parameters of a model representing a nonrandom dropout mechanism (Diggle and Kenward, 1994; DK). (For reviews, see Little (1995) and Kenward and Molenberghs (1999).)

With the volume of literature on nonrandom missing data increasing, there has been growing concern about the fact that models often rest on strong assumptions and relatively little evidence from the data themselves (Glynn, Laird, and Rubin, 1986), especially for selection models. For example, formal tests for the null hypothesis of random missingness, while technically possible, should be approached with caution. Thus, there is a growing awareness of the need for methods that investigate the sensitivity of the results with respect to the model assumptions (Laird, 1994; Little, 1994; Rubin, 1994; Molenberghs et al., 1999). As a general rule, fitting a nonrandom dropout model should be subject to careful scrutiny. First, there is the impact of the assumed distributional form (Kenward, 1998; Scharfstein, Rotnitzky, and Robins, 1999). Second, one should consider the impact one or a few influential subjects may have on the model parameters. In this article, we argue that this problem can also be usefully approached by means of local influence methodology (Cook, 1986).

Section 2 sketches a general modeling framework for incomplete data and introduces the selection model of DK for continuous longitudinal data subject to dropout. In Section 3, the local influence approach of Cook (1986) is discussed. Afterwards, in Section 4, the method is extensively illustrated in the analysis of growth data from a randomized experiment designed to study the effect of the inhibition of testosterone production in rats.

2. A Selection Model for Nonrandom Dropout

We assume that, for subject $i = 1, \ldots, N$ in the study, a sequence of responses Y_{ij} is designed to be measured at occasions $j = 1, \ldots, n_i$. The outcomes are grouped into a vector $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$. Let D_i be the occasion at which dropout

occurs, and we split $\boldsymbol{Y}_i$ into observed ($\boldsymbol{Y}_i^{\mathrm{o}}$) and missing ($\boldsymbol{Y}_i^{\mathrm{m}}$) components, respectively. Let $\boldsymbol{\theta}$ parameterize the measurement model and $\boldsymbol{\psi}$ the dropout model. DK combine a linear mixed model (Laird and Ware, 1982) for the measurement process with a logistic regression model for the dropout process. The measurement model assumes that the vector $\boldsymbol{Y}_i$ of repeated measurements for the ith subject satisfies the linear regression model

$$\boldsymbol{Y}_i \sim \mathrm{N}(X_i\boldsymbol{\beta}, V_i), \qquad i = 1, \ldots, N, \tag{2.1}$$

in which $\boldsymbol{\beta}$ is a vector of population-averaged regression coefficients called fixed effects and where $V_i = Z_i G Z_i' + \Sigma_i$ (Verbeke and Molenberghs, 2000) for positive definite matrices G and Σ_i. The matrices X_i and Z_i contain covariate values, used for modeling the mean and the covariance structure, respectively. The parameters in $\boldsymbol{\beta}$, G, and Σ_i are assembled into $\boldsymbol{\theta}$.

Since no data would be observed otherwise, we assume that the first measurement Y_{i1} is obtained for every subject in the study. The model for the dropout process is based on a logistic regression for the probability of dropout at occasion j, given the subject was still in the study up to occasion j. We denote this probability by $g(\boldsymbol{h}_{ij}, y_{ij})$, in which $\boldsymbol{h}_{ij}$ is a subvector of the history $\overline{\boldsymbol{h}}_{ij}$ containing all responses observed up to but not including occasion j, as well as covariates. We assume

$$\begin{aligned} \mathrm{logit}[g(\boldsymbol{h}_{ij}, y_{ij})] &= \mathrm{logit}\left[\mathrm{pr}(D_i = j \mid D_i \geq j, \boldsymbol{y}_i)\right] \\ &= \boldsymbol{h}_{ij}\boldsymbol{\psi} + \omega y_{ij} \qquad i = 1, \ldots, N. \end{aligned} \tag{2.2}$$

In our case, $\boldsymbol{h}_{ij}$ will contain the previous measurement y_{ij}.

When ω equals zero and the model assumptions made are correct, the dropout model is random and all parameters can be estimated using standard software since the measurement model and dropout model parameters can then be fitted separately. If $\omega \neq 0$, the dropout process is assumed to be nonrandom. Earlier we pointed to the sensitivity of such an approach, and a dropout model may be found to be nonrandom solely because one or a few influential subjects have driven the analysis.

To investigate sensitivity of estimation of quantities of interest, such as treatment effect, growth parameters, or the dropout model parameters, with respect to assumptions about the dropout model, we consider the following perturbed version of (2.2):

$$\begin{aligned} \mathrm{logit}(g(\boldsymbol{h}_{ij}, y_{ij})) &= \mathrm{logit}\left[\mathrm{pr}(D_i = j \mid D_i \geq j, \boldsymbol{y}_i)\right] \\ &= \boldsymbol{h}_{ij}\boldsymbol{\psi} + \omega_i y_{ij} \qquad i = 1, \ldots, N. \end{aligned} \tag{2.3}$$

There is a fundamental difference with model (2.2) since the ω_i should not be viewed as parameters: they are local, individual-specific perturbations around a null model. In our case, the null model will be the MAR model, corresponding to setting $\omega = 0$ in (2.2). Thus, the ω_i are perturbations that will be used only to derive influence measures (Cook, 1986). Perturbation scheme (2.3) is not unique. For example, one might want to consider the posited MNAR model (2.2) as a null model. Also, within the posited family of models, alternative versions for the dropout model (e.g., using more than one previous observation) can be considered. Finally, we could envisage a sensitivity analysis directly in terms of (2.2), e.g., by studying how parameter estimates or functions thereof change with changing values of ω. A relevant range for ω could include both zero and its maximum likelihood estimate under (2.2), $\hat{\omega}$.

Our scheme enables studying the effect of how small perturbations in the MNAR direction can have a large impact on key features of the model. Practically, one way of doing this is to construct local influence measures (Cook, 1986). Clearly, not all possible forms of impact, resulting from sensitivity to dropout model assumptions, will be found in this way, and the method proposed here should be viewed as one component of a sensitivity analysis (e.g., Molenberghs, Goetghebeur, and Kenward, 2000).

When small perturbations in a specific ω_i lead to relatively large differences in the model parameters, it suggests that the subject is likely to drive the conclusions. For example, if such a subject would drive the model toward MNAR, then the conditional expectations of the unobserved measurements, given the observed ones, may deviate substantially from the ones under an MAR mechanism (Kenward, 1998). Such an observation is important also for our approach because then the impact (e.g., from influential subjects) on dropout model parameters extends to all functions that include these dropout parameters. One such function is the conditional expectation of the unobserved measurements, given the corresponding dropout pattern (as would be used in the E step of the EM algorithm), $\mathrm{E}(\boldsymbol{Y}_i^{\mathrm{m}} \mid y_i^{\mathrm{o}}, d_i, \boldsymbol{\theta}, \boldsymbol{\psi})$. As a consequence, the corresponding measurement model parameters will indirectly be affected as well. Therefore, influence on the measurement model parameters can arise, not only from incomplete observations but also from complete ones.

3. Local Influence

Cook (1986) suggests that more confidence can be put in a model that is relatively stable under small modifications. The best known perturbation schemes are based on case deletion (Cook and Weisberg, 1982). A quite different paradigm is the local influence approach, where one investigates how the results are changed under small perturbations of the model. Since the resulting influence diagnostics can be expressed analytically, they often can be decomposed in interpretable components, which yields additional insight.

We are interested in the influence the nonrandomness of dropout exerts on the parameters of interest. This can be done by considering (2.3) as the dropout model. When small perturbations in a specific ω_i lead to relatively large differences in the model parameters, then this suggests that these subjects may have a large impact on the final analysis. Therefore, even though we may be tempted to conclude that such subjects drop out nonrandomly, this conclusion is misguided since we are not aiming to detect (groups of) subjects that drop out nonrandomly but rather subjects that have a considerable impact on the dropout and measurement model parameters.

We denote the log-likelihood function corresponding to model (2.3) by $\ell(\boldsymbol{\gamma} \mid \boldsymbol{\omega}) = \Sigma_{i=1}^{N} \ell_i(\boldsymbol{\gamma} \mid \omega_i)$, in which $\ell_i(\boldsymbol{\gamma} \mid \omega_i)$ is the contribution of the ith individual to the log likelihood and where $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\psi})$ is the s-dimensional vector, grouping the parameters of the measurement model and the dropout model, not including the $N \times 1$ vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)'$ of weights defining the perturbation of the MAR model. It is assumed that $\boldsymbol{\omega}$ belongs to an open subset Ω of $\mathbb{R}^N$. For

ω equal to $\omega_0 = (0, 0, \ldots, 0)'$, $\ell(\gamma \mid \omega_0)$ is the log-likelihood function that corresponds to an MAR dropout model.

Let $\hat{\gamma}$ be the maximum likelihood estimator for γ, obtained by maximizing $\ell(\gamma \mid \omega_0)$, and let $\hat{\gamma}_\omega$ denote the maximum likelihood estimator for γ under $\ell(\gamma \mid \omega)$. The local influence approach now compares $\hat{\gamma}_\omega$ with $\hat{\gamma}$. Strongly different estimates suggest that the estimation procedure is highly sensitive to such perturbations. Cook (1986) proposed measuring the distance between $\hat{\gamma}_\omega$ and $\hat{\gamma}$ by the so-called likelihood displacement, defined by $LD(\omega) = 2[\ell(\hat{\gamma} \mid \omega_0) - \ell(\hat{\gamma}_\omega \mid \omega)]$. This takes into account the variability of $\hat{\gamma}$. Indeed, $LD(\omega)$ will be large if $\ell(\gamma \mid \omega_0)$ is strongly curved at $\hat{\gamma}$, which means that γ is estimated with high precision, and is small otherwise. Therefore, a graph of $LD(\omega)$ versus ω contains essential information on the influence of perturbations. It is useful to view this graph as the geometric surface formed by the values of the $N + 1$ dimensional vector $\xi(\omega) = (\omega', LD(\omega))'$ as ω varies throughout Ω. Since this influence graph can only be depicted when $N = 2$, Cook (1986) proposed looking at local influences, i.e., at the normal curvatures C_h of $\xi(\omega)$ in ω_0, in the direction of some N-dimensional vector h of unit length. Let Δ_i be the s-dimensional vector defined by

$$\Delta_i = \left.\frac{\partial^2 \ell_i(\gamma \mid \omega_i)}{\partial \omega_i \partial \gamma}\right|_{\gamma=\hat{\gamma}, \omega_i=0}$$

and define Δ as the $(s \times N)$ matrix with Δ_i as its ith column. Further, let $\ddot{L}$ denote the $(s \times s)$ matrix of second-order derivatives of $\ell(\gamma \mid \omega_0)$ with respect to γ, also evaluated at $\gamma = \hat{\gamma}$. Cook (1986) has then shown that C_h can be easily calculated by $C_h = 2|h'\Delta'\ddot{L}^{-1}\Delta h|$. Obviously, C_h can be calculated for any direction h. One evident choice is the vector h_i containing one in the ith position and zero elsewhere, corresponding to the perturbation of the ith weight only. This reflects the influence of allowing the ith subject to drop out nonrandomly while the others can only drop out at random. The corresponding local influence measure, denoted by C_i, then becomes $C_i = 2|\Delta_i'\ddot{L}^{-1}\Delta_i|$. Another important direction is the direction $h_{\max}$ of maximal normal curvature $C_{\max}$. It shows how to perturb the MAR model to obtain the largest local changes in the likelihood displacement. $C_{\max}$ is the largest eigenvalue of $-2\Delta'\ddot{L}^{-1}\Delta$ and $h_{\max}$ is the corresponding eigenvector.

When a subset γ_1 of $\gamma = (\gamma_1', \gamma_2')'$ is of special interest, a similar approach can be used, replacing the log likelihood by the profile log likelihood for γ_1, and the methods discussed above for the full parameter vector directly carry over (Lesaffre and Verbeke, 1998).

3.1 *Applied to the Model of DK*

We need expressions for Δ and $\ddot{L}$. Expressions for the elements of $\ddot{L}$ in case $\Sigma_i = \sigma^2 I$ are given by Lesaffre and Verbeke (1998) and can easily be extended to the case considered here. Straightforward derivation shows that the columns Δ_i of Δ are given by

$$\left.\frac{\partial^2 \ell_{i\omega}}{\partial \theta \partial \omega_i}\right|_{\omega_i=0} = 0, \tag{3.1}$$

$$\left.\frac{\partial^2 \ell_{i\omega}}{\partial \psi \partial \omega_i}\right|_{\omega_i=0} = -\sum_{j=2}^{n_i} h_{ij} y_{ij} g(h_{ij})[1 - g(h_{ij})] \tag{3.2}$$

for complete sequences (no drop out) and by

$$\left.\frac{\partial^2 \ell_{i\omega}}{\partial \theta \partial \omega_i}\right|_{\omega_i=0} = [1 - g(h_{id})]\frac{\partial \lambda(y_{id} \mid h_{id})}{\partial \theta} \tag{3.3}$$

$$\left.\frac{\partial^2 \ell_{i\omega}}{\partial \psi \partial \omega_i}\right|_{\omega_i=0} = -\sum_{j=2}^{d-1} h_{ij} y_{ij} g(h_{ij})[1 - g(h_{ij})] - h_{id}\lambda(y_{id} \mid h_{id}) g(h_{id})[1 - g(h_{id})] \tag{3.4}$$

for incomplete sequences. All above expressions are evaluated at $\hat{\gamma}$, and $g(h_{ij}) = g(h_{ij}, y_{ij})\ |_{\omega_i=0}$ is the MAR version of the dropout model. In (3.3), we make use of the conditional mean

$$\lambda(y_{id} \mid h_{id}) = \lambda(y_{id}) + V_{i,21}V_{i,11}^{-1}[h_{id} - \lambda(h_{id})]. \tag{3.5}$$

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

The derivatives of (3.5) with respect to the measurement model parameters are

$$\frac{\partial \lambda(y_{id} \mid h_{id})}{\partial \beta} = x_{id} - V_{i,21}V_{i,11}^{-1}X_{i,(d-1)},$$

$$\frac{\partial \lambda(y_{id} \mid h_{id})}{\partial \alpha} = \left[\frac{\partial V_{i,21}}{\partial \alpha} - V_{i,21}V_{i,11}^{-1}\frac{\partial V_{i,11}}{\partial \alpha}\right] \times V_{i,11}^{-1}[h_{id} - \lambda(h_{id})],$$

where x_{id}' is the dth row of X_i and where $X_{i,(d-1)}$ indicates the first $(d-1)$ rows X_i. Further, α indicates the subvector of covariance parameters within the vector θ.

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

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

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

evaluated at $\gamma = \hat{\gamma}$. It now immediately follows from (3.1) and (3.3) that direct influence on θ only arises from those measurement occasions at which dropout occurs. In particular, from expression (3.3), it is clear that the corresponding contribution is large only if (1) the dropout probability is small but the subject disappears nevertheless and (2) the conditional mean strongly depends on the parameter of interest. This implies that complete sequences cannot be influential in the strict sense ($C_i(\theta) = 0$) and that incomplete sequences only contribute, in a direct fashion, at the actual dropout time. However, in Section 2, we made an important distinction between direct and indirect influence. It was shown that complete sequences can have an impact by changing the conditional expectation of the unobserved measurements given the observed ones and given the dropout mechanism. Thus, a complete observation that has a strong impact on the dropout model parameters can still drastically change the measurement model parameters and functions thereof.

3.2 *Special Case: Compound Symmetry*

A special but enlightening case is the compound symmetry model, $Z_i = \mathbf{1}_{n_i}$, a vector of ones. The matrix G then reduces to τ^2. Assuming further that $\Sigma_i = \sigma^2 I_{n_i}$, the covariance matrix becomes $V_i = \sigma^2 I_{n_i} + \tau^2 J_{n_i}$, where J_{n_i} is an $(n_i \times n_i)$ matrix of ones. Since $\ddot{L}$ is block diagonal,

$$C_i(\boldsymbol{\theta}) = 2[1 - g(\boldsymbol{h}_{id})]^2 \frac{\partial \lambda(y_{id} \mid \boldsymbol{h}_{id})}{\partial \boldsymbol{\theta}}' \ddot{L}^{-1}(\boldsymbol{\theta}) \frac{\partial \lambda(y_{id} \mid \boldsymbol{h}_{id})}{\partial \boldsymbol{\theta}}, \tag{3.6}$$

in which the first factor is large for a small dropout probability at the time of dropout, i.e., for an unlikely event. This is appealing since g_{id} then has the potential of being improved by including dependence on y_{id}. For such a subject, apparent nonrandomness would help. The second factor of (3.6) involves $\ddot{L}(\boldsymbol{\theta})$, and it is therefore harder to derive closed-form expressions. We will derive an insightful approximation. The off-diagonal block of the observed information matrix $\ddot{L}(\boldsymbol{\theta})$ pertaining to the mixed derivatives with respect to $\boldsymbol{\beta}$ and α is not equal to zero (Kenward and Molenberghs, 1998). But these authors also show that, in many practical settings, the difference is negligible, i.e., $C_i(\boldsymbol{\theta}) \simeq C_i^{\text{ap}}(\boldsymbol{\beta}) + C_i^{\text{ap}}(\sigma^2, \tau^2)$.

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

$$\frac{\partial \lambda(y_{id} \mid \boldsymbol{h}_{id})}{\partial \boldsymbol{\beta}} = \xi_{id} \boldsymbol{x}_{id} + (1 - \xi_{id}) \boldsymbol{\rho}_{id}, \tag{3.7}$$

with $\xi_{id} = \sigma^2/[\sigma^2 + (d-1)\tau^2]$ and $\boldsymbol{\rho}_{id} = \boldsymbol{x}_{id} - X'_{i(d-1)} \mathbf{1}_{d-1}/(d-1)$. Note that (3.7) is a weighted average of the covariate $\boldsymbol{x}_{id}$ and the within-series residual covariate $\boldsymbol{\rho}_{id}$ at time d. Further, the matrix of second derivates $\ddot{L}^{-1}(\boldsymbol{\beta})$ is

$$\ddot{L}^{-1}(\boldsymbol{\beta}) = \sum_{i=1}^{N} X_{i(d-1)}' \left(I_{d-1} + \frac{\tau^2}{\sigma^2 + (d-1)\tau^2} J_{d-1} \right) X_{i(d-1)},$$

from which it follows that

$$\begin{aligned} C_i^{\text{ap}}(\boldsymbol{\beta}) = {} & 2[1 - g(\boldsymbol{h}_{id})]^2 (\xi_{id} \boldsymbol{x}_{id} + (1 - \xi_{id}) \boldsymbol{\rho}_{id})' \\ & \times \sigma^2 \left[\sum_{i=1}^{N} \left(\xi_{id} X_{i(d-1)}' X_{i(d-1)} \right.\right. \\ & \left.\left. + (1 - \xi_{id}) R_{i(d-1)}' R_{i(d-1)} \right) \right]^{-1} \\ & \times (\xi_{id} \boldsymbol{x}_{id} + (1 - \xi_{id}) \boldsymbol{\rho}_{id}), \end{aligned} \tag{3.8}$$

where $R_{i,d-1} = X_{i(d-1)} - \mathbf{1}_{d-1} \overline{X_{i(d-1)}}$. Here $\overline{X_{i(d-1)}} = \mathbf{1}'_{d-1} X_{i(d-1)}/(d-1)$. Expression (3.8) is the product of the factor that purely depends on the dropout probability and a factor that has the structure of a leverage. This motivates calling the second factor of $C_i^{\text{ap}}(\boldsymbol{\beta})$ a generalized leverage, not only for compound symmetry but also for general covariances.

Similar calculations can be performed for the variance components (σ^2, τ^2), yielding

$$\begin{aligned} C_i^{\text{ap}}(\sigma^2, \tau^2) = {} & 2[1 - g(\boldsymbol{h}_{id})]^2 \xi_{id}^2 (1 - \xi_{id})^2 \overline{[\boldsymbol{h}_{id} - \lambda(\boldsymbol{h}_{id})]}^2 \\ & \times \left(-1, \frac{1}{\tau^2} \right) \ddot{L}^{-1}(\sigma^2, \tau^2) \begin{pmatrix} -1 \\ \frac{1}{\tau^2} \end{pmatrix}, \end{aligned} \tag{3.9}$$

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

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

Even though $\ddot{L}^{-1}(\sigma^2, \tau^2)$ has a somewhat complicated form, it occurs in (3.9) only through a scalar. Thus, $C_i^{\text{ap}}(\sigma^2, \tau^2)$ can in practice be decomposed into three interpretable components. The first factor is shared with $C_i^{\text{ap}}(\boldsymbol{\beta})$ and has the same interpretation. The second factor disappears when either the measurement error variance or the variance of the random intercept is reduced to zero. It is maximal when there is balance between both components of variability ($\xi_{id} = 0.5$). The third factor is large when the squared average residual of the history at the time of dropout is large.

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

$$\begin{aligned} C_i(\boldsymbol{\psi}) = {} & 2 \left(\sum_{j=2}^{d} \boldsymbol{h}_{ij} y_{ij} v_{ij} \right)' \left(\sum_{i=1}^{N} \sum_{j=2}^{d} v_{ij} \boldsymbol{h}_{ij} \boldsymbol{h}_{ij}' \right)^{-1} \\ & \times \left(\sum_{j=2}^{d} \boldsymbol{h}_{ij} y_{ij} v_{ij} \right), \end{aligned} \tag{3.10}$$

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

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

for incomplete sequences and $v_{ij} = g(h_{ij})[1 - g(h_{ij})]$. Expression (3.10) bears some resemblance with the hat-matrix diagonal from logistic regression. The contributions from a single individual are summed in the first and third factor of (3.10), even though they contribute independent pieces of information to the logistic regression. This is because each individual is given a single weight, ω_i. Alternatively, one can consider measurement-specific weights, ω_{ij}. The calculations for this case are straightforward.

4. Analysis of Rat Data

The data come from a randomized experiment designed to study the effect of the inhibition of testosterone production in rats (Department of Orthodontics of the Catholic University of Leuven [K.U.L.] in Belgium; Verdonck et al., 1998). A total of 50 male Wistar rats were randomized to either control or one of two treatment groups (low or high dose of the drug cecapeptyl, an inhibitor of testosterone production). The treatment started at the age of 45 days, and measurements were taken every 10 days, with the first observation taken at the age of 50 days. Our response is a characterization of the height of the skull, taken under anesthesia. Many rats do not survive anesthesia, resulting in only 22 (44%) rats having all seven designed measurements taken. The investigators' impression was that dropout was independent of the measurements.

The individual profiles are shown in Figure 1. To linearize, we use the logarithmic transformation $t = \ln(1 + (\text{age} - 45)/10)$ for the time scale. Let y_{ij} denote the jth measurement for the ith rat, taken at $t = t_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, N$.

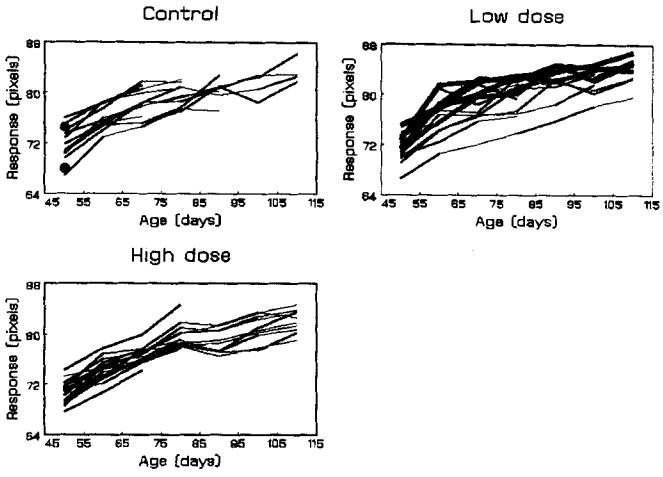

Figure 1. Individual growth curves for the three treatment groups separately. Influential subjects are highlighted.

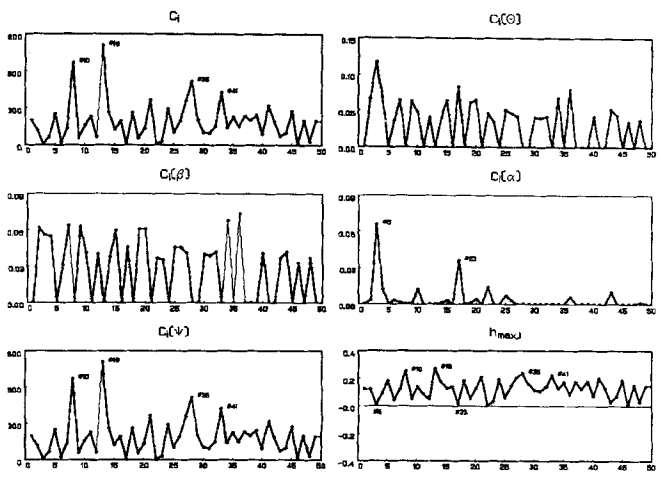

Figure 2. Index plots of C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature.

A simple statistical model then assumes that y_{ij} satisfies a model of the form (2.1) with common average intercept β_0 for all three groups; average slopes β_1, β_2, and β_3 for the three treatment groups, respectively; and assuming compound symmetry covariance structure, with common variance $\sigma^2+\tau^2$ and common covariance τ^2. These models are estimated under MCAR, MAR, and MNAR processes. Following these models, there is little evidence of MAR and no evidence for MNAR.

Figure 2 displays overall C_i and influences for subvectors $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\psi}$. In addition, the direction $\boldsymbol{h}_{\max}$ corresponding to maximal local influence is given. Apart from the last one of these graphs, the scales are not unitless and therefore it would be hard to use a common one for all of the panels. This implies that the main emphasis should be on relative magnitudes. For example, one could compare the original analysis with one where the most influential ones are deleted. We observe large absolute scale differences for different influence graphs.

The largest C_i are observed for rats 10, 16, 35, and 41, and virtually the same picture holds for $C_i(\boldsymbol{\psi})$. They are highlighted in Figure 1. All four belong to the low-dose group. Arguably, their relatively large influence is caused by an interplay of three facts. First, the profiles are relatively high and hence y_{ij} and h_{ij} in (3.10) are large. Second, since all four profiles are complete, the first factor in (3.10) contains a maximal number of large terms. Third, the computed v_{ij} are relatively large, which is implied by the MAR dropout model parameter estimates in Table 1, which are based on model (1).

Turning attention to $C_i(\boldsymbol{\alpha})$ reveals peaks for rats 5 and 23. Both belong to the control group and drop out after a single measurement occasion. They are highlighted in the first panel of Figure 1. To explain this, observe that the relative magnitude of $C_i(\boldsymbol{\alpha})$, approximately given by (3.9), is determined by $1-g(h_{id})$ and $h_{id}-\lambda(h_{id})$. The first term is large when the probability of dropout is small. Now, when dropout occurs early in the sequence, the measurements are still relatively low, implying that the dropout probability is rather small (cf., Table 1). This feature is built into the model by writing the dropout probability in terms of the raw measurements with time-independent coefficients rather than, e.g., in terms of residuals. Further, the residual $h_{id}-\lambda(h_{id})$ is large since these two rats are somewhat distant from the group by time mean. A practical implication of this is that the time-constant nature of the dropout model may be unlikely to hold. Therefore, a time-varying version was considered, where the logit of the dropout model takes the form $\psi_0+\psi_1 y_{i,j-1}+\nu_0 t_{ij}+\nu_0 t_{ij} y_{i,j-1}$. There is overwhelming evidence in favor of such a more elaborate MAR model (likelihood ratio statistic of 167.4 on 2 d.f.). Thus, local influence can be used to call into question the posited MAR (and MNAR) models and to guide further selection of more elaborate, perhaps MAR, models.

Since all deviations are rather moderate, we further explore our approach by considering a second analysis where all responses for rats 10, 16, 35, and 41 have been increased with 20 units. The effect of this distortion will primarily be seen in the variance structure. Precisely, such a change is likely to inflate the random intercept variance at the expense of the other variance components. In doing so, we will illustrate that (1) such a change is likely to show up in the assessment of the dropout model, underscoring the sensitivity, and that (2) the local influence approach is able to detect such an effect. The parameter estimates for all three models are also shown in Table 1. Clearly, while the fixed-effect parameters remain virtually unchanged, the random intercept parameter has, of course, drastically increased. Likewise, the dropout parameters are affected. In addition, the likelihood ratio statistic for MAR versus MCAR changes from 2.8 to 0.3 and for MNAR versus MAR changes from 0.1 to 2.9. Thus, the evidence has shifted from the first to the second test. While all of these statistics are nonsignificant, there is an important qualitative effect. In order to check whether these findings are recovered by the local influence approach, let us study Figure 3. In line with the changes in parameter estimates, $C_i(\boldsymbol{\beta})$ shows no peaks in these observations, but peaks in $C_i(\boldsymbol{\alpha})$ and $C_i(\boldsymbol{\psi})$ indicate a relatively strong influence from the four extreme profiles.

Graphical representations such as Figure 3 are sometimes judged misleading since the apparent magnitude of a subject

Table 1
Maximum likelihood estimates (standard errors) of completely random, random, and nonrandom dropout models fitted to the rat data set with and without modification

Effect	Parameter	MCAR	MAR	MNAR
		Original Data		
Measurement model				
Intercept	β_0	68.61 (0.33)	68.61 (0.33)	68.60 (0.33)
Slope control	β_1	7.51 (0.22)	7.51 (0.22)	7.53 (0.24)
Slope low dose	β_2	6.87 (0.23)	6.87 (0.23)	6.89 (0.23)
Slope high dose	β_3	7.31 (0.28)	7.31 (0.28)	7.35 (0.30)
Random intercept	τ^2	3.44 (0.77)	3.44 (0.77)	3.43 (0.77)
Measurement error	σ^2	1.43 (0.14)	1.43 (0.14)	1.43 (0.14)
Dropout model				
Intercept	ψ_0	−1.98 (0.20)	−8.48 (4.00)	−10.30 (6.88)
Previous measurement	ψ_1		0.08 (0.05)	0.03 (0.16)
Current measurement	$\omega = \psi_2$			0.07 (0.22)
−2 log likelihood		1100.4	1097.6	1097.5
		Modified Data		
Measurement model				
Intercept	β_0	70.20 (0.92)	70.20 (0.92)	70.25 (0.92)
Slope control	β_1	7.52 (0.25)	7.52 (0.25)	7.42 (0.26)
Slope low dose	β_2	6.97 (0.25)	6.97 (0.25)	6.90 (0.25)
Slope high dose	β_3	7.21 (0.31)	7.21 (0.31)	7.04 (0.33)
Random intercept	τ^2	40.38 (0.18)	40.38 (0.18)	40.71 (8.25)
Measurement error	σ^2	1.42 (0.14)	1.42 (0.14)	1.44 (0.15)
Dropout model				
Intercept	ψ_0	−1.98 (0.20)	−0.79 (1.99)	2.08 (3.08)
Previous measurement	ψ_1		−0.015 (0.03)	0.23 (0.15)
Current measurement	$\omega = \psi_2$			−0.28 (0.17)
−2 log likelihood		1218.0	1217.7	1214.8

is influenced by its neighbors. On the other hand, it preserves the order across all six index plots. One way to overcome this problem is by ordering one plot (e.g., according to C_i) and keeping this order across all six panels. Alternatively, scatter plots of (1) the measurement versus dropout components and (2) fixed effects versus variance component elements can be used. An example of the latter is presented in Figure 4.

5. Concluding Remarks

In this article, we have underscored the importance of assessing sensitivity when fitting parametric selection models, such as the one proposed by DK. One possible route of assessing sensitivity is by considering local influence methods (Cook, 1986). This approach, while only one way of studying sensitivity, is very broad and can potentially take many forms. Our method is based on the concept of individual-specific infinitesimal perturbations around the MAR model. Technically, our method assigns a perturbation, within the linear predictor of the dropout model, to the so-called current, potentially unobserved measurement. The advantage of this choice is that, after fitting the MAR model, relatively simple, noniterative calculations suffice to obtain the influence measures. In particular, the need to fit nonrandom dropout models is avoided. Moreover, in the special case of a compound-symmetry model, the influence measures are approximately decomposable into intuitive and interpretable components. In any case, influence decomposes in a measurement and dropout part, the first of which is zero in the case of a complete observation. The lat-

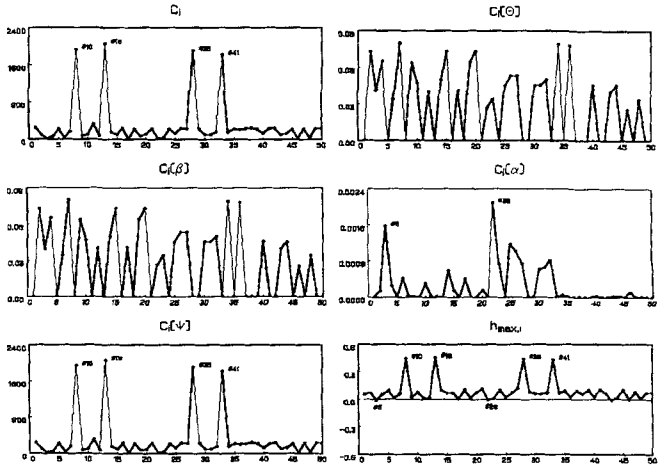

Figure 3. Index plots of C_i, $C_i(\boldsymbol{\theta})$, $C_i(\boldsymbol{\beta})$, $C_i(\boldsymbol{\alpha})$, $C_i(\boldsymbol{\psi})$, and of the components of the direction $\boldsymbol{h}_{\max}$ of maximal curvature, where four profiles have been shifted upward.

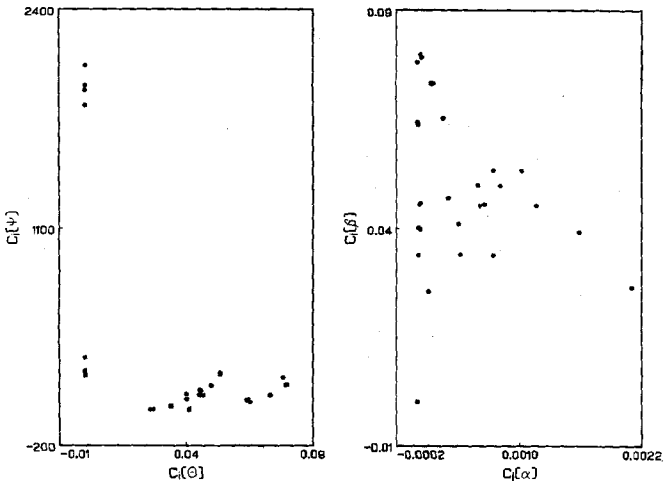

Figure 4. Scatter plots of (1) $C_i(\boldsymbol{\theta})$ versus $C_i(\boldsymbol{\psi})$ and (2) $C_i(\boldsymbol{\beta})$ versus $C_i(\boldsymbol{\alpha})$, where four profiles have been shifted upward.

ter comment needs careful qualification because it may seem counterintuitive at first sight. However, as indicated in Section 2 and further developed in Section 3, influence on the dropout parameters translates, through the conditional expectation of unobserved measurements given dropout, into influence on the measurement model parameters and functions thereof. Therefore, the study of local influence, together with its indirect implications, can provide valuable insight into which observations may lead to a seemingly nonrandom dropout model.

It has to be emphasized that, within the concept of perturbation around the MAR model, the proposed method is by no means the only one possible. Further, it is possible, although this falls outside the scope of this article, to consider perturbations around particular forms of MNAR models within or outside the DK framework. For example, when nonmonotone missingness is an issue, then a local influence approach can be developed for a suitable model, such as the ones proposed by Baker (1995) and Troxel, Harrington, and Lipsitz (1998). Finally, in a practical data analysis setting, various forms of sensitivity analysis should supplement each other.

The analysis of the rat data set (Verdonck et al., 1998) supports the claim that the influence measures are easy to interpret. In addition to the advantages quoted earlier, we claim that a careful study of the conditions under which the diagnostics become large can shed some light on the adequacy of the model formulation. For example, the DK model usually writes the logit of the dropout probability as a function of the raw measurements with time-independent coefficients. This implies that an expression such as (3.10) depends directly on the magnitude of the responses. An alternative parameterization of the dropout probability in terms of residuals $(Y_{ij} - \mu_{ij})/\sigma_{ij}$ would obviously yield a different picture. The authors have developed GAUSS code, which is available upon request.

Acknowledgements

We thank Rod Little and Don Rubin for helpful comments. We gratefully acknowledge support from FWO-Vlaanderen Research Project "Sensitivity Analysis for Incomplete and Coarse Data." The third author gratefully acknowledges support from a research grant of Vlaams Instituut voor de Bevordering van het Wetenschappelijk-Technologisch Onderzoek in de Industrie. Finally, we would like to thank an anonymous associate editor and three anonymous referees for very thorough and constructive comments.

Résumé

Diggle et Kenward (*Applied Statistics*, 1994) proposèrent un modèle de sélection pour des données continues longitudinales soumises à un processus de censure non aléatoire. Cet article a provoqué un large débat sur le rôle de tels modèles. L'enthousiasme initial fit place au scepticisme sur les hypothèses fortes mais non testables sur lesquelles ce type de modèle repose. Depuis lors, s'est dégagé l'opinion selon laquelle ces modèles devraient idéalement faire partie d'une analyse de sensibilité. Ce papier présente une approche formelle et flexible de cette analyse de sensibilité reposant sur la notion d'influence locale (Cook, *JRSS-B* 1986). L'influence de la perturbation d'un modèle MAR dans le sens d'un processus non-aléatoire est explorée. Cette méthode est appliquée aux données d'un essai randomisé de l'inhibition de la production de testostérone chez le rat.

References

Baker, S. G. (1995). Marginal regression for repeated binary data with outcomes subject to nonignorable nonresponse. *Biometrics* **51,** 1042–1052.

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.

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

Glynn, Laird, and Rubin. (1986). Selection modelling versus mixture modelling with nonignorable nonresponse. In *Drawing Inferences from Self-Selected Samples,* H. Wainer (ed), 115–142. New York: Springer Verlag.

Kenward, M. G. (1998). Selection and models for repeated measurements with nonrandom dropout: An illustration of sensitivity. *Statistics in Medicine* **17,** 2723–2732.

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

Kenward, M. G. and Molenberghs, G. (1999). Parametric models for incomplete continuous and categorical longitudinal studies data. *Statistical Methods in Medical Research* **8,** 51–83.

Laird, N. M. (1994). Discussion to Diggle, P. J. and Kenward, M. G.: Informative dropout in longitudinal data analysis. *Applied Statistics* **43,** 84.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38,** 963–974.

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

Little, R. J. A. (1994). Discussion to Diggle, P. J. and Kenward, M. G.: Informative dropout in longitudinal data analysis. *Applied Statistics* **43,** 78.

Little, R. J. A. (1995). Modelling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90,** 1112–1121.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Molenberghs, G., Goetghebeur, E., Lipsitz, S. R., and Kenward, M. G. (1999). Non-random missingness in categorical data: Strengths and limitations. *The American Statistician* **53,** 110–118.

Molenberghs, G., Goetghebeur, E., and Kenward, M. G. (2000). Sensitivity analysis for incomplete contingency tables. *Applied Statistics*, in press.

Rubin, D. B. (1994). Discussion to Diggle, P. J. and Kenward, M. G.: Informative dropout in longitudinal data analysis. *Applied Statistics* **43,** 80–82.

Scharfstein, D. O., Rotnizky, A., and Robins, J. M. (1999). Adjusting for non-ignorable drop-out using semiparametric nonresponds models (with discussion). *Journal of the American Statistical Association* **94,** 1096–1146.

Troxel, A. B., Harrington, D. P., and Lipsitz, S. R. (1998). Analysis of longitudinal data with non-ignorable non-monotone values. *Applied Statistics* **47,** 425–438.

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

Verdonck, A., De Ridder, L., Verbeke, G., Bourguignon, J. P., Carels, C., Kuhn, E. R., Darras, V., and de Zegher, F. (1998). Comparative effects of neonatal and prepubertal castration on craniofacial growth in rats. *Archives of Oral Biology* **43,** 861–871.

Received January 1999. *Revised June* 2000.
Accepted July 2000.