

A flexible marginal modelling strategy for non-monotone missing data

Peer-reviewed author version

JANSEN, Ivy & MOLENBERGHS, Geert (2008) A flexible marginal modelling strategy for non-monotone missing data. In: JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY, 171(2). p. 347-373.

DOI: 10.1111/j.1467-985X.2007.00524.x

Handle: <http://hdl.handle.net/1942/8050>

A flexible marginal modeling strategy for non-monotone missing data

Ivy Jansen

Center for Statistics, Hasselt University, Agoralaan, B-3590 Diepenbeek, Belgium

Geert Molenberghs

Center for Statistics, Hasselt University, Agoralaan, B-3590 Diepenbeek, Belgium

Summary. A lot of research has been devoted to modeling strategies for longitudinal data with missingness, recently especially within the MNAR context. In this paper, the relatively unexplored but practically highly relevant domain of non-monotone missingness with multivariate ordinal responses will be broached. To this end, a dedicated version of the multivariate Dale model (Molenberghs and Lesaffre, 1994) will be formulated. Furthermore, we will also assess the sensitivity of these models to their assumptions, using the technique of global influence (Cook, 1979, 1986; Thijs et al., 2000).

Keywords: Contingency Table; Global Influence; Multivariate Dale Model; Non-random Missingness.

1. Introduction

Categorical data modeling has received a lot of attention during the past decades (Agresti, 2002). More recently, quite a bit of attention has been devoted to repeated categorical data (Diggle et al., 2002; Fahrmeir and Tutz, 1994; Molenberghs and Verbeke, 2005). One generally distinguishes between three families: (1) marginal models (Ashford and Sowden, 1970; Bahadur, 1961; 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; Liang and Zeger, 1989; Molenberghs and Ryan, 1999; Rosner, 1984); and (3) random-effects approaches (Breslow and Clayton, 1993; Stiratelli et al., 1984; Wolfinger and O'Connell, 1993).

Especially with longitudinally measured outcomes, measurement sequences are prone to incompleteness. When referring to the missing-value, or non-response, process we will use terminology of Little and Rubin (1987, Ch.6). Missing outcomes are 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 (or Bayesian) inference, and when the parameters describing the measurement process (θ) are functionally independent of the parameters describing the missingness process (ψ), MCAR and MAR are *ignorable*, while a nonrandom process is nonignorable.

A model for incomplete data starts from the joint distribution of the outcomes, \mathbf{Y} say, and the non-response process, \mathbf{R} say. This joint distribution $f(\mathbf{y}, \mathbf{r} | \theta, \psi)$ can be factorized in several ways. A selection model is based on the factorization $f(\mathbf{y} | \theta) f(\mathbf{r} | \mathbf{y}, \psi)$, whereas the reverse factorization is referred to as a pattern-mixture model (Little, 1993, 1994). When a common set of random-effects is thought to influence both the \mathbf{Y} and \mathbf{R} processes, conditional upon which they are independent, then the so introduced model is referred to as a shared-parameter model. For reviews, see Kenward and Molenberghs (1998); Little (1995); Wu and Carroll (1988).

The non-response process \mathbf{R} can either be monotone, also called dropout, or non-monotone when there are intermittent missing values. For each of these processes different modeling strategies are needed. Monotone missingness in continuous outcomes can be modeled, for example using a linear mixed model for the measurements, and a logistic regression for the dropout, depending on the previous, and possibly also the current, measurements (Diggle and Kenward, 1994). Monotone missingness in ordinal outcomes is discussed in Molenberghs, Kenward, and Lesaffre (1997). They use a Dale model for the measurements and a logistic regression for dropout. Also for non-monotone missingness several modeling strategies exist. For continuous outcomes, we refer to Troxel et al. (1998), who used a multivariate normal distribution for the measurements, together with the Markov Chain assumption, and a logistic regression, which only depends on the current measurement, for the missingness process. They encountered problems with the heavy computational load due to a multi-modal likelihood surface. For 2 binary outcomes Baker et al. (1992) already proposed a model for dealing with non-monotone missingness. Later, Jansen et al. (2003) reformulated this model such that its membership of the selection model family is unambiguously clear, extended it to accommodate, possibly continuous, covariates, turning the model into a regression tool for several categorical outcomes, and proposed a parameterization that avoids the risk of parameter space violations. The measurement model, as well as the missingness model are a form of multigroup logistic regression. Baker (1995) proposed a model for three binary outcomes with non-monotone missingness, based on marginal and association models for the measurements, and a logistic regression for the missingness mechanism, depending on the last observed and last unobserved measurement.

In this paper, we propose a flexible and elegant set of models that can be used for multivariate categorical outcomes with non-monotone missingness, based on the multivariate Dale model (Molenberghs and Lesaffre, 1994) for the measurements, as well as for the missingness mechanism. Since MNAR models are sensitive to the underlying model assumptions, generally performing a sensitivity analysis is strongly advisable. Many different routes might be followed, either at the level of the models, or at the level of the individuals (Draper, 1995; Glynn et al., 1986; Molenberghs et al., 2001; Rubin, 1977, 1994; Scharfstein et al., 1999). We will focus on global influence analysis, also known as the case-deletion method (Cook and Weisberg, 1982), to determine which subjects might be influential for the analysis.

The rest of the paper is organized as follows. Section 2 introduces data from the Belgian Health Interview Survey 1997 (HIS1997). In Section 3 we sketch the multivariate Dale model, and how two copies of it can be combined, describing the measurement and missingness part of the model, respectively. Its application to the HIS is presented in Section 4. In Section 5, the global influence methodology as introduced by Cook (1979, 1986) in linear regression, and by Molenberghs et al. (2003) and Thijs et al. (2000) in linear mixed models, is reviewed. Finally, the proposed methodology is applied to the Health Interview Survey data in Section 6.

2. Belgian Health Interview Survey

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

The main goal of the HIS1997 was to provide a description of the subjective health status of the entire Belgian population as well as of three regional subpopulations (Flemish, Walloon and Brussels region), and in addition of the Germanophone community. The idea was to obtain a reflection of how specific groups of people experience their health, to what extent they use health care facilities, and how they take care of their health by adopting a certain life style and/or by taking preventive or other health-related measures.

The target population was defined as all people residing in Belgium during 1997. The National Register was used as the sampling frame. The total number of successful interviews for the sample was set to 10000

Table 1. Data from the HIS (10786 subjects). ‘Mental health’ versus ‘fixed general practitioner’.

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

Table 2. Data from the HIS (10786 subjects). ‘Mental health’ versus ‘body mass index’.

	< 18	18–25	25–30	> 30	Missing
Good	100	2923	1685	581	298
Bad	81	1421	691	285	147
Missing	17	168	97	25	2267

(0.1% of the Belgian population), 3500 in both the Flemish and Walloon regions, and 3000 in the Brussels region. Sampling was based on a combination of systematic sampling, stratification, multistage sampling, clustering, and selection with differential probabilities. More details on the design of the study, together with a descriptive analysis of missingness, can be found in Burzykowski et al. (1999) and Renard et al. (1998).

In this paper, we will focus on three specific aspects of the HIS1997, i.e., the status of mental health (0 for ‘good’; 1 for ‘bad’; ‘.’ for ‘not interviewed’, ‘no valid information’, ‘error interviewer’, ‘not applicable’, ‘does not know’, ‘no answer’), having a fixed general practitioner (0 for ‘yes’; 1 for ‘no’; ‘.’ for ‘not interviewed’, ‘no valid information’, ‘error interviewer’, ‘not applicable’, ‘does not know’, ‘no answer’), and body mass index (1 for ‘< 18’; 2 for ‘18–25’; 3 for ‘25–30’; 4 for ‘> 30’; ‘.’ for ‘not interviewed’, ‘no valid information’, ‘error interviewer’, ‘not applicable’, ‘does not know’, ‘no answer’). In total, information is available for 10,786 subjects. Tabulated summary data are shown in Tables 1 and 2. Covariates of interest are sex (‘male’ and ‘female’) and educational level (‘none’, ‘primary’, ‘lower secondary’, ‘higher secondary’, ‘higher’). The distribution of the subjects over the different covariate levels of sex and education are shown in Tables 3 and 4, respectively, in which the structure is changed a bit to save space (blocks for a missing value for mental health are moved after other blocks).

3. Models for Non-Monotone Multivariate Categorical Outcomes

3.1. The Multivariate Dale Model

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

Table 3. Data from the HIS. ‘Mental health’ versus ‘fixed general practitioner’. Separate for males (5288) and females (5498), respectively.

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

Table 4. Data from the HIS. ‘Mental health’ versus ‘fixed general practitioner’. Separate for no education (267), primary education (1486), lower secondary education (1844), higher secondary education (3265), higher education (3843), and education information missing (81), respectively.

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

and interactions (described by cross-ratios of second and higher orders). A trivariate version of the model has been presented by McCullagh and Nelder (1989, Sec. 6.5.4). However, these authors did not provide a way to calculate the joint probabilities, needed in the likelihood equations. Several authors, including Molenberghs and Lesaffre (1994, 1999), Lang and Agresti (1994), and Glonek and McCullagh (1995) have provided solutions for the multivariate case. A detailed review is given in Molenberghs and Verbeke (2005).

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

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

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

$$Z_i(\mathbf{k}) = \sum_{\ell \leq \mathbf{k}} Z_i^*(\ell). \quad (3.1)$$

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

$$p_i(\mathbf{k}) = P(\mathbf{Y}_{iv} \leq \mathbf{k} | \mathbf{x}_i, \boldsymbol{\theta}) \quad (3.2)$$

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

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

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

$$\eta_{itj_t} = \text{logit}(p_{itj_t} | \mathbf{x}_i) = \boldsymbol{\theta}_{itj_t}^T \mathbf{x}_i, \quad (1 \leq t \leq T, 1 \leq j_t < c_t). \quad (3.3)$$

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

$$\varphi_{i,ts,j_tj_s} = \frac{(p_{i,ts,j_tj_s})(1 - p_{itj_t} - p_{isj_s} + p_{i,ts,j_tj_s})}{(p_{isj_s} - p_{i,ts,j_tj_s})(p_{itj_t} - p_{i,ts,j_tj_s})}. \quad (3.4)$$

They are usefully modeled on the log scale as

$$\eta_{i,ts,j_tj_s} = \ln \varphi_{i,ts,j_tj_s} = \boldsymbol{\theta}_{i,ts,j_tj_s}^T \mathbf{x}_i, \quad (1 \leq t, s \leq T, 1 \leq j_t < c_t, 1 \leq j_s < c_s).$$

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

3.2. The Bivariate Dale Model

Since, in the application, we only have two responses of interest (mental health and fixed general practitioner), the multivariate Dale model can be replaced by its bivariate version (Dale, 1986). Both responses are restricted to two possible outcomes only ('good' versus 'bad' and 'yes' versus 'no', respectively), obviating the need to construct tables with cumulative counts and probabilities, thence the $*$ is omitted from notation. The joint probabilities

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

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

$$\begin{aligned} h_1(p_{1+}(\mathbf{x})) &= \boldsymbol{\theta}_1^T \mathbf{x}, \\ h_2(p_{+1}(\mathbf{x})) &= \boldsymbol{\theta}_2^T \mathbf{x}, \\ h_3\left(\frac{p_{11}(\mathbf{x})p_{22}(\mathbf{x})}{p_{12}(\mathbf{x})p_{21}(\mathbf{x})}\right) &= \boldsymbol{\theta}_3^T \mathbf{x}, \end{aligned} \quad (3.5)$$

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

$$\ln \varphi = \ln \frac{p_{11}(\mathbf{x})p_{22}(\mathbf{x})}{p_{12}(\mathbf{x})p_{21}(\mathbf{x})}, \quad (3.6)$$

which is assumed linear in the covariates, although non-linear extensions are feasible but computationally more demanding. Solving equations (3.5) yields the probabilities $p_{j_1j_2}(\mathbf{x})$ (the dependence on \mathbf{x} is omitted

Table 5. Theoretical distribution of complete and observed cells of a bivariate binary outcome.

$\pi_{11,11}$	$\pi_{11,12}$	$\pi_{10,11}$	$\pi_{10,12}$	$\pi_{01,11}$	$\pi_{01,12}$	$\pi_{00,11}$	$\pi_{00,12}$
$\pi_{11,21}$	$\pi_{11,22}$	$\pi_{10,21}$	$\pi_{10,22}$	$\pi_{01,21}$	$\pi_{01,22}$	$\pi_{00,21}$	$\pi_{00,22}$

$\pi_{11,11}$	$\pi_{11,12}$	$\pi_{10,1+}$	$\pi_{01,11}$	$\pi_{01,12}$	$\pi_{00,++}$
$\pi_{11,21}$	$\pi_{11,22}$	$\pi_{10,2+}$	$\pi_{01,21}$	$\pi_{01,22}$	

for the ease of notation):

$$p_{11} = \begin{cases} \frac{1 + (p_{1+} + p_{+1})(\varphi - 1) - S(p_{1+}, p_{+1}, \varphi)}{2(\varphi - 1)} & \text{if } \varphi \neq 1, \\ p_{1+}p_{+1} & \text{if } \varphi = 1, \end{cases} \quad (3.7)$$

and

$$\begin{aligned} p_{12} &= p_{1+} - p_{11}, \\ p_{21} &= p_{+1} - p_{11}, \\ p_{22} &= 1 - p_{12} - p_{21} - p_{11}, \end{aligned} \quad (3.8)$$

with

$$S(\lambda_1, \lambda_2, \varphi) = \sqrt{[1 + (\lambda_1 + \lambda_2)(\varphi - 1)]^2 + 4\varphi(1 - \varphi)\lambda_1\lambda_2}. \quad (3.9)$$

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

3.3. Joint Model for Measurements and Missingness

In the case of two binary outcomes, the bivariate Dale model, as introduced in Section 3.2, will be used for the measurement model and for the missingness model given the measurements (Section 3.3.1). Also in the case that one or both outcomes have more than two outcome categories, the bivariate Dale model can be used, but a slightly different version than in Section 3.2 is needed. Therefore, we start from the multivariate Dale model, as introduced in Section 3.1, for the measurement model, combined with the bivariate Dale model for the missingness model given the measurements (Section 3.3.2). In both situations, again a selection model is obtained. Discrete as well as continuous covariates can be included in measurement and missingness models.

3.3.1. Two binary outcomes

Let $i = 1, \dots, N$ index distinct covariate levels. In this section, the index i will be suppressed from notation. Let $j_1, j_2 = 1, 2$ correspond to the outcome categories of the first and second measurement, respectively and let $r_1, r_2 = 0, 1$ correspond to the missingness indicators (1 for an observed and 0 for a missing measurement). This leads to a four-way classification as in Table 1. The complete data and observed data cell probabilities $\pi_{r_1 r_2, j_1 j_2}$ for this setting are presented in Table 5, and are factorized as:

$$\pi_{r_1 r_2, j_1 j_2} = p_{j_1 j_2} q_{r_1 r_2 | j_1 j_2}, \quad (3.10)$$

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

$$\begin{aligned} \eta_1 &= \text{logit } p_{1+} = X_1 \boldsymbol{\theta}, \\ \eta_2 &= \text{logit } p_{+1} = X_2 \boldsymbol{\theta}, \\ \eta_3 &= \ln \varphi_p = X_3 \boldsymbol{\theta}, \end{aligned} \quad (3.11)$$

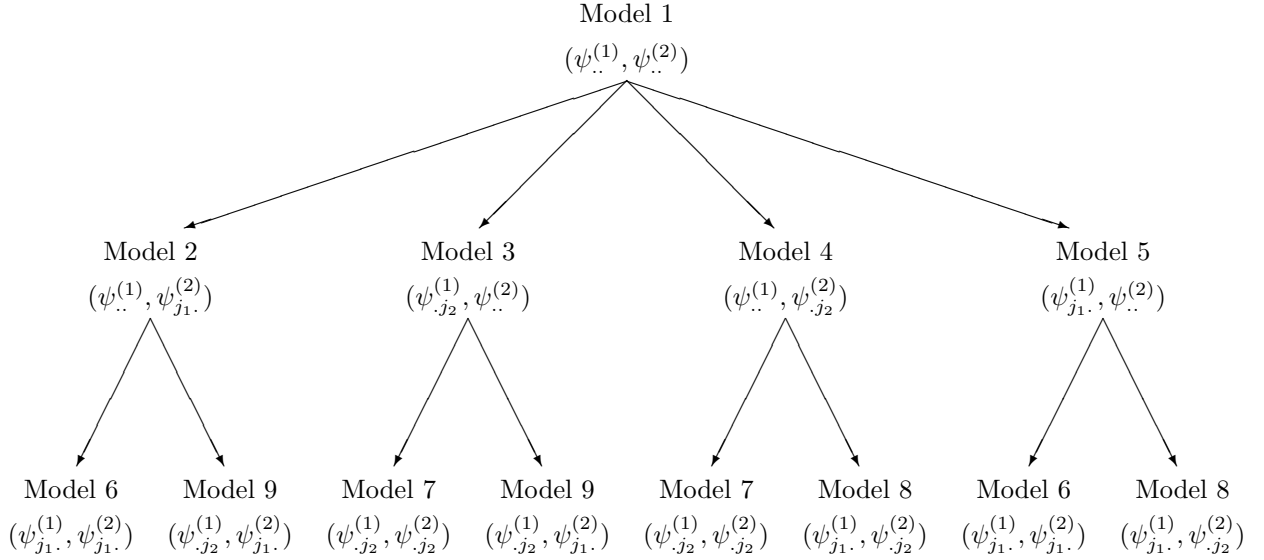


Fig. 1. Graphical representation of the model nesting structure.

$$\begin{aligned}
 \delta_{4|j_1 j_2} &= \text{logit } q_{1+|j_1 j_2} = Z_{1|j_1 j_2} \psi, \\
 \delta_{5|j_1 j_2} &= \text{logit } q_{+1|j_1 j_2} = Z_{2|j_1 j_2} \psi, \\
 \delta_{6|j_1 j_2} &= \ln \varphi_q = Z_{3|j_1 j_2} \psi.
 \end{aligned} \tag{3.12}$$

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

$$\begin{array}{lll}
 \text{Model 1} : & (\psi_{..}^{(1)}, \psi_{..}^{(2)}) & \text{Model 4} : & (\psi_{..}^{(1)}, \psi_{.j_2}^{(2)}) & \text{Model 7} : & (\psi_{.j_2}^{(1)}, \psi_{.j_2}^{(2)}) \\
 \text{Model 2} : & (\psi_{..}^{(1)}, \psi_{j_1.}^{(2)}) & \text{Model 5} : & (\psi_{j_1.}^{(1)}, \psi_{..}^{(2)}) & \text{Model 8} : & (\psi_{j_1.}^{(1)}, \psi_{.j_2}^{(2)}) \\
 \text{Model 3} : & (\psi_{.j_2}^{(1)}, \psi_{..}^{(2)}) & \text{Model 6} : & (\psi_{j_1.}^{(1)}, \psi_{j_1.}^{(2)}) & \text{Model 9} : & (\psi_{.j_2}^{(1)}, \psi_{j_1.}^{(2)}),
 \end{array}$$

where $\psi_{j_1 j_2}^{(1)}$ models the non-response for the first outcome and $\psi_{j_1 j_2}^{(2)}$ models the non-response for the second outcome. Index j_1 indicates the dependence on the first outcome, index j_2 on the second outcome. The nesting structure of these models is schematically represented in Figure 1. Interpretation is straightforward. For example, Model 1 is MCAR, in Model 4 missingness in the first variable is constant, while missingness in the second variable depends on its value. Note that this is the first time the Dale model is applied to a situation where missingness is non-monotone.

No a priori ordering is imposed on the outcomes. The advantage is that genuine multivariate settings (e.g., several questions in a survey) can be handled as well as longitudinal studies. When deemed necessary, the implications of ordering can be imposed by considering specific models and leaving out others. For example, one may want to avoid missingness on future observations. In the current bivariate case, the index j_2 would have to be removed from $\psi^{(1)}$ in the above model.

Models (3.11) and (3.12) can be combined into the model

$$\xi = W\phi, \tag{3.13}$$

where

$$\xi = (\eta_1, \eta_2, \eta_3, \delta_{4|11}, \delta_{5|11}, \delta_{6|11}, \delta_{4|12}, \delta_{5|12}, \delta_{6|12}, \delta_{4|21}, \delta_{5|21}, \delta_{6|21}, \delta_{4|22}, \delta_{5|22}, \delta_{6|22})^T,$$

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

and $\phi = (\theta^T, \psi^T)^T$. This new model will be referred to as the *Dale-Dale model*, separate Dale models being used for the marginal measurement process and for the process governing missingness, given the outcomes.

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

$$\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)^T,$$

$$\psi = (\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8)^T,$$

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

and

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

where $I_1 = 1$ in Models 3, 7 and 9, $I_2 = 1$ in Models 5, 6 and 8, $I_3 = 1$ if $I_1 = 1$ or $I_2 = 1$, $J_1 = 1$ in Models 4, 7 and 8, $J_2 = 1$ in Models 2, 6 and 9, $J_3 = 1$ if $J_1 = 1$ or $J_2 = 1$. These indicators are equal to 0 in all other models. Note that the interpretation of ψ_1 , ψ_2 , ψ_7 and ψ_8 depends on the restrictions made in a specific model. For example, model 2 assumes restrictions $\psi_{..}^{(1)}$ and $\psi_{j1.}^{(2)}$, which are equivalent to $\psi_{11}^{(1)} = \psi_{12}^{(1)} = \psi_{21}^{(1)} = \psi_{22}^{(1)}$, $\psi_{11}^{(2)} = \psi_{12}^{(2)}$ and $\psi_{21}^{(2)} = \psi_{22}^{(2)}$, such that ψ_1 models the non-response for the first variable, independent of any observed values, ψ_2 models the non-response for the second variable, when the first variable has a value equal to 1, and ψ_8 models the difference in non-response for the second variable between a value for the first variable equal to 1 and equal to 2.

To include individual-specific covariates, a subscript i has to be introduced to the vector ξ and the matrix W and hence to their constituent components. θ , ψ , X , and Z can easily be changed to increase or reduce the number of covariates.

3.3.2. One binary and one categorical outcome

Let $i = 1, \dots, N$ index distinct covariate levels. Also in this section, we suppress i from notation. Let $j_1 = 1, 2, j_2 = 1, 2, 3, 4$ 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 as in Section 3.3.1. This leads to a classification as in Table 2. The complete data and observed data cell probabilities $\pi_{r_1 r_2, j_1 j_2}$ for this setting are presented in Table 6. They can be factorized as in (3.10), but now (3.11) needs to be replaced by

Table 6. Theoretical distribution of complete and observed cells of a binary and a categorical (4 levels) outcome.

$\pi_{11,11}$	$\pi_{11,12}$	$\pi_{11,13}$	$\pi_{11,14}$	$\pi_{10,11}$	$\pi_{10,12}$	$\pi_{10,13}$	$\pi_{10,14}$	$\pi_{01,11}$	$\pi_{01,12}$	$\pi_{01,13}$	$\pi_{01,14}$	$\pi_{00,11}$	$\pi_{00,12}$	$\pi_{00,13}$	$\pi_{00,14}$	
$\pi_{11,21}$	$\pi_{11,22}$	$\pi_{11,23}$	$\pi_{11,24}$	$\pi_{10,21}$	$\pi_{10,22}$	$\pi_{10,23}$	$\pi_{10,24}$	$\pi_{01,21}$	$\pi_{01,22}$	$\pi_{01,23}$	$\pi_{01,24}$	$\pi_{00,21}$	$\pi_{00,22}$	$\pi_{00,23}$	$\pi_{00,24}$	
$\pi_{11,11}$	$\pi_{11,12}$	$\pi_{11,13}$	$\pi_{11,14}$					$\pi_{10,1+}$								
$\pi_{11,21}$	$\pi_{11,22}$	$\pi_{11,23}$	$\pi_{11,24}$					$\pi_{10,2+}$								
								$\pi_{01,+1}$	$\pi_{01,+2}$	$\pi_{01,+3}$	$\pi_{01,+4}$					$\pi_{00,++}$

$$\begin{aligned}
 \eta_1 &= \text{logit } p_{1+} = X_1 \boldsymbol{\theta}, & \eta_5 &= \ln \varphi_1 = X_5 \boldsymbol{\theta}, \\
 \eta_2 &= \text{logit } p_{+1} = X_2 \boldsymbol{\theta}, & \eta_6 &= \ln \varphi_2 = X_6 \boldsymbol{\theta}, \\
 \eta_3 &= \text{logit } p_{+2} = X_3 \boldsymbol{\theta}, & \eta_7 &= \ln \varphi_3 = X_7 \boldsymbol{\theta}, \\
 \eta_4 &= \text{logit } p_{+3} = X_4 \boldsymbol{\theta}, & & & &
 \end{aligned} \tag{3.14}$$

with

$$\varphi_1 = \frac{p_{11}(p_{22} + p_{23} + p_{24})}{(p_{12} + p_{13} + p_{14})p_{21}}, \quad \varphi_2 = \frac{(p_{11} + p_{12})(p_{23} + p_{24})}{(p_{13} + p_{14})(p_{21} + p_{22})} \quad \text{and} \quad \varphi_3 = \frac{(p_{11} + p_{12} + p_{13})p_{24}}{p_{14}(p_{21} + p_{22} + p_{23})},$$

the proportional odds ratios, rather than marginal global odds ratios as defined in (3.4). However, both parameterizations should give similar results. Equation (3.12) will be used here too. However, we now need it 8 times, once for each possible combination of j_1 and j_2 . Now,

$$\boldsymbol{\xi} = (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7, \delta_{4|11}, \delta_{5|11}, \delta_{6|11}, \dots, \delta_{4|14}, \delta_{5|14}, \delta_{6|14}, \delta_{4|21}, \delta_{5|21}, \delta_{6|21}, \dots, \delta_{4|24}, \delta_{5|24}, \delta_{6|24})^T.$$

For simplicity, we assume that no covariate effects will be included in this model, since the aim here is to show how the models should be adapted when an outcome has more than two categories. However, covariates can easily be included in the same way as before. Therefore,

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7)^T,$$

$$\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, \psi_4, \psi_5)^T,$$

where ψ_4 and ψ_5 do not appear in model 1, and ψ_5 does not appear in models 2–5. X is in this case a 7×7 identity matrix, while

$$Z = \begin{pmatrix} Z_{1|11} \\ Z_{2|11} \\ Z_{3|11} \\ Z_{1|12} \\ Z_{2|12} \\ Z_{3|12} \\ Z_{1|13} \\ Z_{2|13} \\ Z_{3|13} \\ Z_{1|14} \\ Z_{2|14} \\ Z_{3|14} \\ Z_{1|21} \\ Z_{2|21} \\ Z_{3|21} \\ Z_{1|22} \\ Z_{2|22} \\ Z_{3|22} \\ Z_{1|23} \\ Z_{2|23} \\ Z_{3|23} \\ Z_{1|24} \\ Z_{2|24} \\ Z_{3|24} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & I_1 & 0 \\ 0 & 1 & 0 & 0 & J_1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & I_2 & 0 \\ 0 & 1 & 0 & 0 & J_2 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & I_3 & 0 \\ 0 & 1 & 0 & 0 & J_3 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & I_4 & 0 \\ 0 & 1 & 0 & 0 & J_4 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & I_5 & 0 \\ 0 & 1 & 0 & 0 & J_5 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & I_6 & 0 \\ 0 & 1 & 0 & 0 & J_6 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & I_7 & 0 \\ 0 & 1 & 0 & 0 & J_7 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix},$$

where, in Models 3, 7 and 9 $I_1 = I_5 = 1$, $I_2 = I_6 = 2$ and $I_3 = I_7 = 3$, in Models 5, 6 and 8 $I_4 = I_5 = I_6 = I_7 = 1$, in Models 4, 7 and 8 $J_1 = J_5 = 1$, $J_2 = J_6 = 2$ and $J_3 = J_7 = 3$, and in Models 2, 6 and 9 $J_4 = J_5 = J_6 = J_7 = 1$. In all other models, these indicators are equal to 0.

4. Models Fitted to the Health Interview Survey Data

4.1. Mental Health versus Fixed General Practitioner

In the analysis, all patients with recorded covariate information are considered. The effect of the covariate on the marginal probabilities can either be constant, or different for both probabilities. We assume that this choice is identical for the measurement model and the missingness model. So either θ_4 and ψ_4 in (3.13) are included in the model, or θ_5 , θ_6 and ψ_5 , ψ_6 , for parsimony. The association between both outcomes is assumed to be constant in all settings.

We will consider in detail five different sets of Dale-Dale models. First, models are considered without covariates. Results (estimates, standard errors, and negative loglikelihoods) are presented in Table 7. Second, a constant effect of sex is added (Table 8), with distribution of the data over males and females shown in Table 3. In Table 9 the effect of sex is allowed to differ for both marginal probabilities. Next, education is included. Table 10 shows the results of the models with constant effect of education, while in Table 11 this effect changes over the marginal probabilities. Table 4 contains the distribution of the data over the different categories of education. Several other ways to include covariate information are possible (only including a covariate effect in the measurement or missingness model, a constant covariate effect in one part of the model and a varying effect in the other part, more than one covariate effect, ...), but will not be presented here for conciseness.

Table 7. Maximum likelihood estimates and standard errors of the Dale-Dale models. Mental health versus fixed general practitioner. All observations included. No covariates.

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

Sampling zeroes or small counts in some of the cells forces certain parameters to lie on the boundary of their corresponding parameter space which, due to the parameterization, is equal to ∞ . This should not be seen as a disadvantage of our model, since boundary solutions are a well known feature of MNAR models in general (Rubin, 1996). The advantage of our parameterization (see also Jansen et al., 2003) is that either an interior or a boundary solution is obtained, but never an invalid solution. Note that there are ramifications regarding hypothesis testing when parameters lie on the boundary of their space. For details, we refer to Molenberghs and Verbeke (2007).

From Table 7, the likelihood ratio tests reject Model 1 in favor of Model 3 ($LRT = 24.14$) and Model 4 ($LRT = 22.28$). Since these models make different assumptions about the missingness mechanism (in Model 3, a missing value for mental health depends on the (un)observed value for fixed general practitioner, in Model 4, a missing value for fixed general practitioner depends on the (un)observed value for fixed general practitioner), both have to be considered. The extension to Model 7, which combines the missingness mechanism assumptions from Model 3 and Model 4, is not significant (Models 3 and 7: $LRT = 0.00$, meaning they are equivalent at the observed data level, Models 4 and 7: $LRT = 1.86$). We also notice that in Model 4 the estimate for ψ_8 , the MNAR parameter, lies on the boundary of the parameter space, which makes Model 4 less favorable than Model 3. Nevertheless, a more careful study is necessary to get insight into why two models with completely different assumptions about the missingness mechanism are significant, and therefore a sensitivity analysis is performed in Section 6. But first we will discuss the results of the analyses where covariates are included into the model.

The effect of sex, assumed to be equal for both marginal probabilities, on the measurement and missingness model simultaneously, is highly significant, whichever of the models is chosen to conduct a likelihood ratio test (with 2 degrees of freedom). Further, within Table 8, Model 3 and Model 4 still provide the most adequate description. The likelihood ratio test statistics for comparing Models 1 and 3 and Models 1 and 4 equal 23.00 and 22.98, respectively, while those for extending Model 3 and Model 4 further along the edges of Figure 1, are all smaller than 1.50, and therefore not significant. Thus, from this set of models, one observes that sex improves the fit but the nature of the missingness mechanism does not change. In Jansen et al. (2003) a similar result was found. Without covariates, an MCAR model was found to be the most adequate choice. Including the effect of duration in the measurement model changed the nature of the missingness mechanism, by making it more complex. When further including duration in the missingness part, all evidence for non-MCAR missingness disappeared and again the same MCAR model came out as the best. It was also clear that the effect of duration on the missingness model could not be neglected.

Table 8. Maximum likelihood estimates and standard errors of the Dale-Dale models. Mental health versus fixed general practitioner. All observations included. Constant effect of sex.

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

Table 9. Maximum likelihood estimates and standard errors of the Dale-Dale models. Mental health versus fixed general practitioner. All observations included. Effect of sex different on both marginal probabilities.

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

Table 10. Maximum likelihood estimates and standard errors of the Dale-Dale models. Mental health versus fixed general practitioner. 81 observations not included. Constant effect of education.

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

Table 11. Maximum likelihood estimates and standard errors of the Dale-Dale models. Mental health versus fixed general practitioner. 81 observations not included. Effect of education different on both marginal probabilities.

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

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

For the covariate education, the same conclusions can be drawn concerning the model selection. Models 3 and 4, with a varying education effect, provide the most adequate description from all models in Tables 10–11, although the extension to Model 7 is again borderline non-significant, and will be considered as well. In the measurement model, the effect of education seems to be significant only for the marginal probability of fixed general practitioner, not for the marginal probability of mental health, with the interpretation that the

Table 12. Maximum likelihood estimates and standard errors of the Dale-Dale models. Mental health versus body mass index. All observations included. Equal odds ratios.

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
<i>Measurement model</i>									
θ_1	0.63 (0.02)	0.63 (0.02)	0.64 (0.02)	0.63 (0.02)	0.62 (0.02)	0.05 (0.02)	0.56 (0.02)	0.61 (0.02)	0.64 (0.02)
θ_2	-3.73 (0.07)	-3.73 (0.07)	-3.73 (0.07)	-3.76 (0.07)	-3.73 (0.07)	-3.69 (0.07)	-3.98 (0.07)	-3.76 (0.07)	-3.73 (0.07)
θ_3	0.34 (0.02)	0.34 (0.02)	0.34 (0.02)	0.32 (0.02)	0.34 (0.02)	0.33 (0.02)	-0.09 (0.06)	0.31 (0.02)	0.34 (0.02)
θ_4	2.07 (0.04)	2.07 (0.03)	2.07 (0.03)	2.04 (0.04)	2.07 (0.04)	2.06 (0.03)	1.33 (0.12)	2.03 (0.04)	2.07 (0.03)
θ_5	0.43 (0.02)	0.43 (0.02)	0.43 (0.02)	0.43 (0.02)	0.42 (0.02)	0.01 (0.01)	0.39 (0.02)	0.42 (0.02)	0.43 (0.02)
<i>Missingness model</i>									
ψ_1	1.16 (0.02)	1.16 (0.02)	1.08 (0.05)	1.16 (0.02)	1.18 (0.02)	$+\infty$	3.08 (0.38)	1.19 (0.03)	1.08 (0.05)
ψ_2	1.09 (0.02)	1.10 (0.02)	1.09 (0.02)	1.19 (0.05)	1.09 (0.02)	2.88 (0.06)	3.63 (0.54)	1.21 (0.06)	1.10 (0.02)
ψ_3	4.86 (0.07)	4.86 (0.07)	4.87 (0.07)	4.87 (0.07)	4.86 (0.07)	4.82 (0.10)	4.94 (0.08)	4.88 (0.08)	4.87 (0.07)
ψ_4		-0.02 (0.03)	0.05 (0.03)	-0.07 (0.03)	-0.07 (0.04)	$-\infty$	-1.03 (0.17)	-0.09 (0.04)	0.05 (0.03)
ψ_5						-2.75 (0.06)	-1.34 (0.23)	-0.08 (0.03)	-0.02 (0.03)
-loglik	22090.77	22090.56	22089.01	22088.81	22089.49	21994.64	22085.11	22086.93	22088.91

higher the education, the higher the probability of having a fixed general practitioner. In the missingness model, the influence of education is opposite for both marginal probabilities (negative for missingness in mental health, positive for missingness in fixed general practitioner).

Important modeling and data-analytic conclusions can be drawn. It is clear that covariate effects need not be the same on all marginal probabilities, in the measurement model as well as in the missingness model. Thus, a varying covariate effect can be investigated. It should also be noted that the parameter estimates of the covariates in the measurement part are remarkably stable among the 9 Dale-Dale models. This implies that, in case one is primarily interested in a specific covariate effect on mental health and fixed general practitioner, all models containing this effect provide very similar evidence. This need not be the case in general. However, attention is more and more devoted to the nature of the missingness mechanism. Therefore, one should still exercise caution, since non-random missingness models heavily rely on untestable assumptions (Verbeke and Molenberghs, 2000). In our analyses, more than one model, with completely different assumptions, turned out to describe the data well. So it is important to search for observations which may drive these conclusions (Verbeke et al., 2001). This naturally leads to sensitivity analysis. In Section 5, we will describe the global influence.

4.2. Mental Health versus Body Mass Index

In this analysis, we will not include covariate effects, since the way it should be done, is sufficiently described in the previous analysis of mental health versus fixed general practitioner. Thus, all 10,786 patients can be included in the analysis. Two different sets of Dale-Dale models will be considered. In a first model, we assume that the three proportional odds ratios are equal. Results are shown in Table 12. The second model is more complex in the sense that now the proportional odds ratios are allowed to be different (Table 13).

Similar to what was observed in Section 4.1, we notice that the estimates for the measurement model remain stable in all models in Table 12, except Model 6, probably caused by the boundary solution ($\pm\infty$). Therefore, when primary interest is not in the missingness process, but e.g. in estimating the association between both outcomes, all models (ignoring Model 6) can be used for this purpose. In this case, the odds ratio is estimated to be equal to 1.537. However, also interpreting the missingness process might be of interest, leading to the conclusion that Model 7 fits the data best ($LRT = 7.40$ when extending Model 4), meaning that the body mass index is responsible for a missing value for mental health as well as for body mass index.

Including three different proportional odds ratios in the model, improves the fit a lot (for all models, $LRT > 200$ with $2df$). However now, Model 3 is to be preferred while Model 4 is borderline non-significant ($LRT =$

Table 13. Maximum likelihood estimates and standard errors of the Dale-Dale models. Mental health versus body mass index. All observations included. Different odds ratios.

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
<i>Measurement model</i>									
θ_1	0.75 (0.02)	0.75 (0.02)	0.75 (0.02)	0.76 (0.02)	0.76 (0.02)	0.07 (0.02)	0.75 (0.02)	0.76 (0.02)	0.75 (0.02)
θ_2	-3.68 (0.06)	-3.68 (0.07)	-3.68 (0.07)	-3.71 (0.07)	-3.68 (0.07)	-3.59 (0.07)	-3.56 (0.45)	-3.71 (0.07)	-3.68 (0.07)
θ_3	0.34 (0.02)	0.34 (0.02)	0.34 (0.02)	0.31 (0.02)	0.34 (0.02)	0.36 (0.02)	0.43 (0.28)	0.31 (0.02)	0.34 (0.02)
θ_4	2.09 (0.03)	2.09 (0.03)	2.09 (0.03)	2.06 (0.04)	2.09 (0.03)	2.09 (0.03)	2.19 (0.31)	2.06 (0.04)	2.09 (0.03)
θ_5	-0.56 (0.14)	-0.56 (0.14)	-0.56 (0.14)	-0.56 (0.14)	-0.56 (0.14)	-0.64 (0.13)	-0.56 (0.14)	-0.56 (0.14)	-0.56 (0.14)
θ_6	-0.23 (0.05)	-0.23 (0.05)	-0.23 (0.05)	-0.23 (0.05)	-0.23 (0.05)	-0.15 (0.04)	-0.25 (0.08)	-0.23 (0.05)	-0.23 (0.05)
θ_7	0.59 (0.02)	0.59 (0.02)	0.59 (0.02)	0.60 (0.02)	0.59 (0.02)	0.06 (0.02)	0.56 (0.09)	0.60 (0.02)	0.59 (0.02)
<i>Missingness model</i>									
ψ_1	1.16 (0.02)	1.16 (0.02)	1.08 (0.04)	1.16 (0.02)	1.16 (0.02)	$+\infty$	0.74 (1.04)	1.15 (0.02)	1.08 (0.05)
ψ_2	1.09 (0.02)	1.10 (0.02)	1.09 (0.02)	1.19 (0.06)	1.09 (0.02)	2.88 (0.05)	0.68 (1.28)	1.19 (0.05)	1.10 (0.02)
ψ_3	4.86 (0.08)	4.86 (0.08)	4.87 (0.07)	4.88 (0.08)	4.86 (0.07)	4.82 (0.10)	4.85 (0.11)	4.88 (0.08)	4.87 (0.07)
ψ_4		-0.01 (0.03)	0.06 (0.03)	-0.07 (0.03)	0.02 (0.04)	$-\infty$	0.29 (0.77)	0.02 (0.04)	0.06 (0.03)
ψ_5						-2.75 (0.06)	0.28 (0.94)	-0.07 (0.03)	-0.02 (0.03)
-loglik	21981.37	21981.29	21979.42	21979.47	21981.31	21979.53	21979.36	21979.34	21979.30

3.80), but the extension to Model 7, which combines the assumptions regarding missingness from Models 3 and 4, is not significant anymore ($LRT = 0.12$ and 0.22 respectively). So body mass index is now only responsible for a missing value for mental health. Also in this case, for Model 6, problems were encountered with solutions on the boundary of the parameter space, and the estimates for the measurement model were not comparable with the estimates of other models. It can be concluded that, based in the 8 other models, the odds ratios $\varphi_1, \varphi_2, \varphi_3$ are equal to 0.571, 0.795 and 1.804. It is clear from those estimates that it is important to include all three proportional odds ratios into the model.

5. Global Influence

One of the tools to perform a sensitivity analysis is global influence, starting from case-deletion. This methodology is based on the difference in log-likelihood between the model fitted to the dataset as a whole on the one hand, and the dataset minus one subject on the other hand. Denote the log-likelihood function, corresponding to model (3.13), as

$$\ell(\phi) = \sum_{i=1}^N \ell_i(\phi), \quad (5.1)$$

in which $\ell_i(\phi)$ is the contribution of the i th individual to the log-likelihood, and where ϕ is the s -dimensional vector as defined in Section 3.3. Further, we denote by

$$\ell_{(-i)}(\phi), \quad (5.2)$$

the log-likelihood function, where the contribution of the i th subject has been removed. Cook's distances (CD) are based on measuring the discrepancy between either the maximized log-likelihoods (5.1) and (5.2) or (subsets of) the estimated parameter vectors $\hat{\phi}$ and $\hat{\phi}_{(-i)}$, with obvious notation. Precisely, we will consider both

$$CD_{1i} = 2(\hat{\ell}(\phi) - \hat{\ell}_{(-i)}(\phi)), \quad (5.3)$$

as well as

$$CD_{2i}(\phi) = 2(\hat{\phi} - \hat{\phi}_{(-i)})' \ddot{L}^{-1} (\hat{\phi} - \hat{\phi}_{(-i)}). \quad (5.4)$$

Formulation (5.4) easily allows for the consideration of global influence in a subvector of ϕ , such as the measurement model parameters θ , or missingness model parameters ψ . This will be indicated using notation of the form $CD_{2i}(\theta)$, $CD_{2i}(\psi)$, and so on.

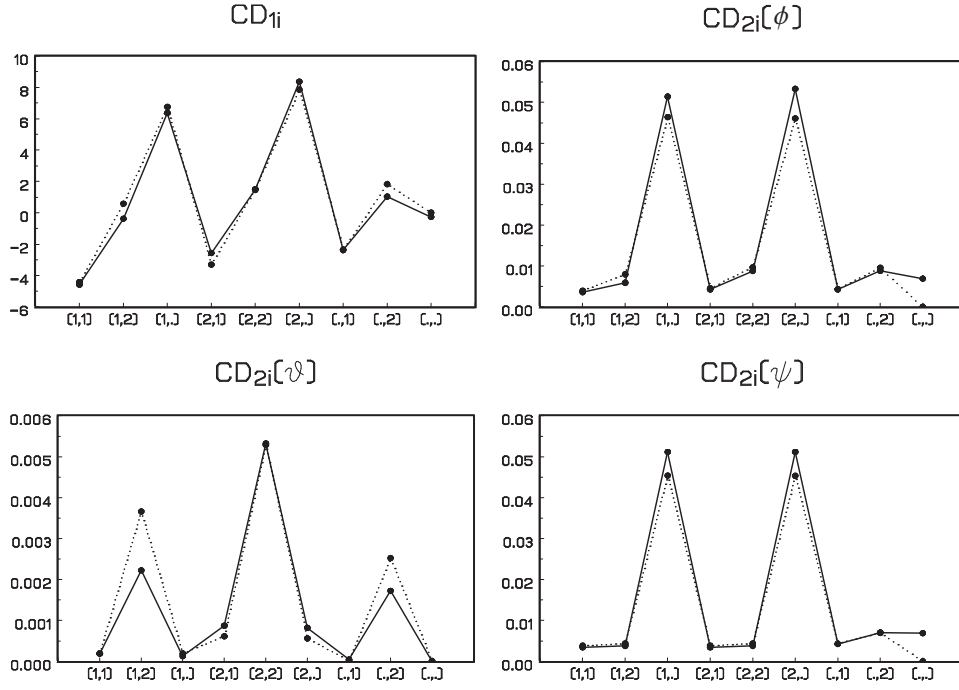


Fig. 2. Global influence for Model 3 with varying sex effect. Solid line for males, dotted for females.

6. Influence Analysis of the Health Interview Survey Data

We will apply the global influence ideas to the Dale-Dale models in order to explain the conclusions of Section 4.1. Whereas all models in Figure 1 can be used, we focus on Model 3 and Model 4, since those were the most adequate ones when, on the one hand, a varying sex effect and, on the other hand, a varying education effect is included in both parts of the model. In addition, we will consider Model 7, a supermodel of both Model 3 and 4, which was, in both settings, borderline non-significant.

Performing a global influence analysis on data with categorical outcomes is less time consuming than on data with continuous outcomes, since the data can be summarized in cells such as in Table 1. Thus, instead of removing every subject one by one, we only need to remove one subject per cell and per covariate level. So in our data set, when including sex, only 2×9 subjects need to be removed, and when including education, 5×9 subjects. Results are shown in Figures 2–7.

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

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

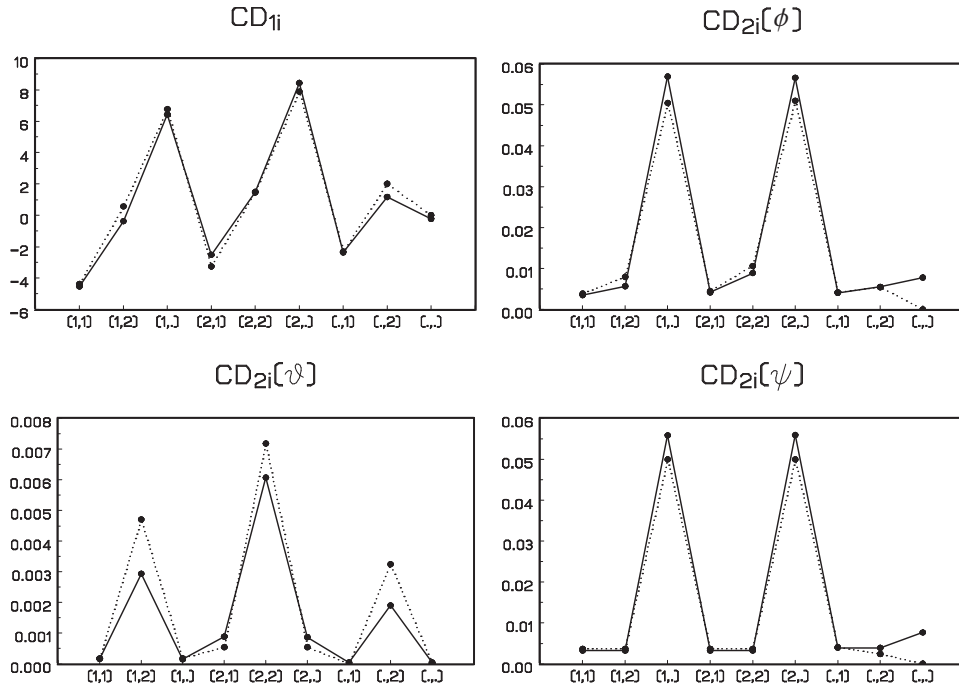


Fig. 3. Global influence for Model 4 with varying sex effect. Solid line for males, dotted for females.

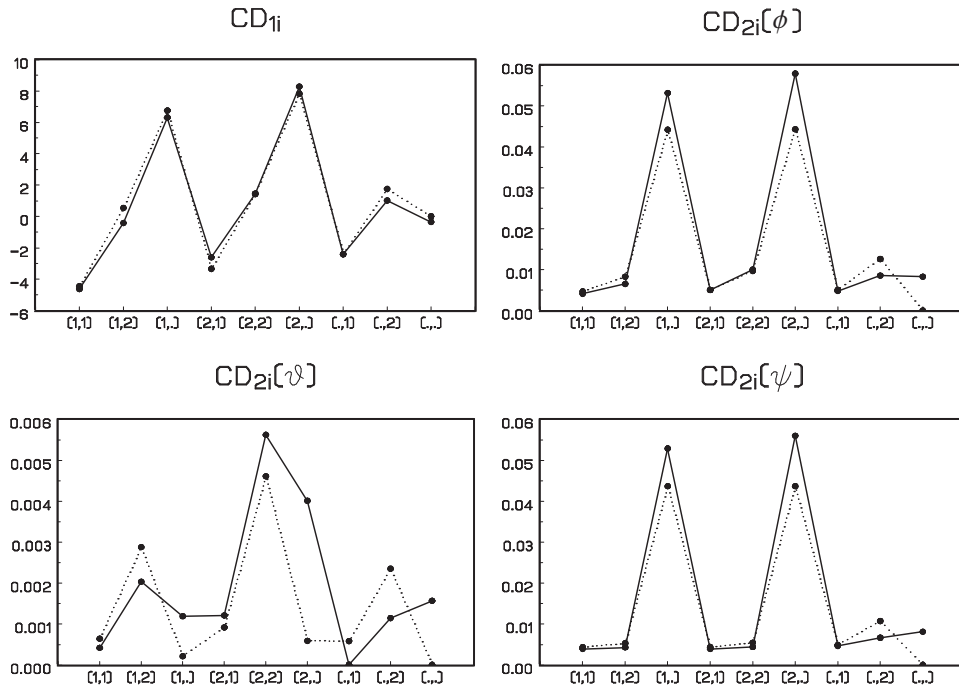


Fig. 4. Global influence for Model 7 with varying sex effect. Solid line for males, dotted for females.

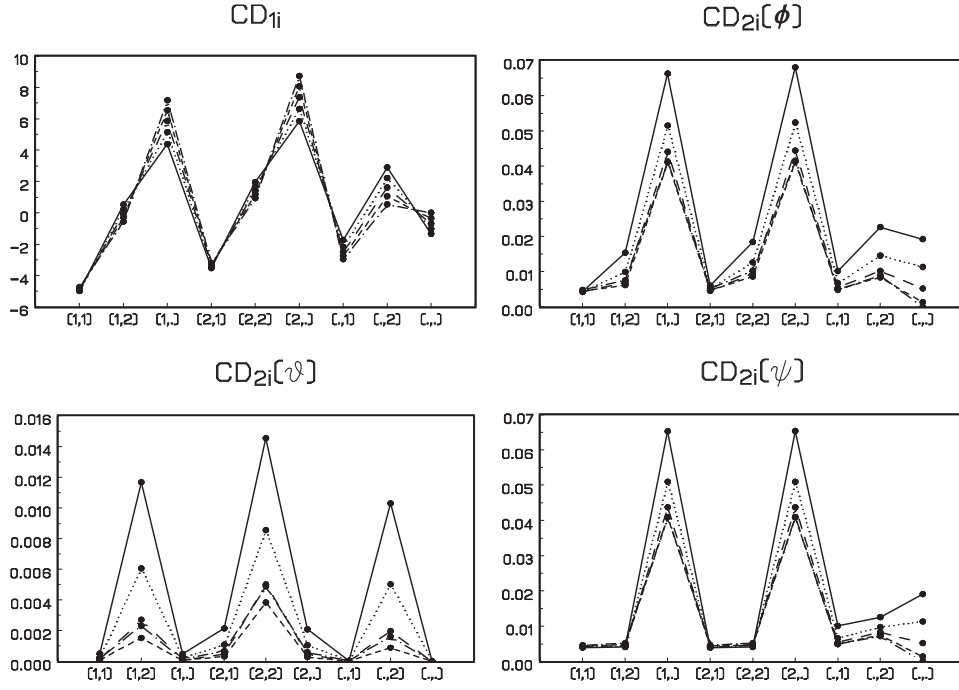


Fig. 5. Global influence for Model 3 with varying education effect. Solid line for no education, dotted for primary education, dashed for low secondary education, short dashes for high secondary education, dots and dashes for higher education.

Varying parameterization invariably produces the same result.

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

When education is included in the model, the influence graphs look similar for all models considered. For CD_{1i} , $CD_{2i}(\phi)$ and $CD_{2i}(\psi)$ the largest Cook's distances were measured for subjects with a missing value for fixed general practitioner. This result is similar to the one for sex, and can be explained in the same way. The Cook's distances $CD_{2i}(\theta)$ are largest for subjects without fixed general practitioner, for all levels of education, but this influence is negligible. Again, the Cook's distances for the measurement model parameters are much smaller than those for the missingness model parameters. Some more conclusions have to be drawn. Subjects having no measurements at all (for mental health and fixed general practitioner) have no influence on the measurement model parameters, but a non-negligible effect on the missingness model parameters, since, as in Jansen et al. (2003), the observations without any follow-up are explicitly modeled in the Dale-Dale models. Subjects without education (solid line) have the highest influence on all parameters of the model, possibly due to their low presence (267 out of 10705) in the data. In Jansen et al. (2006), it was stated that the local influence measure for a subject decreases linearly with the sample size of the dataset, and hence the total influence for a dataset is roughly constant. Our result may indicate that a similar result might be true for global influence, and when the influence is measured for a group of subjects with identical characteristics, this influence has to be rescaled according to the number of subjects in that group. Further investigation of this feature will be necessary.

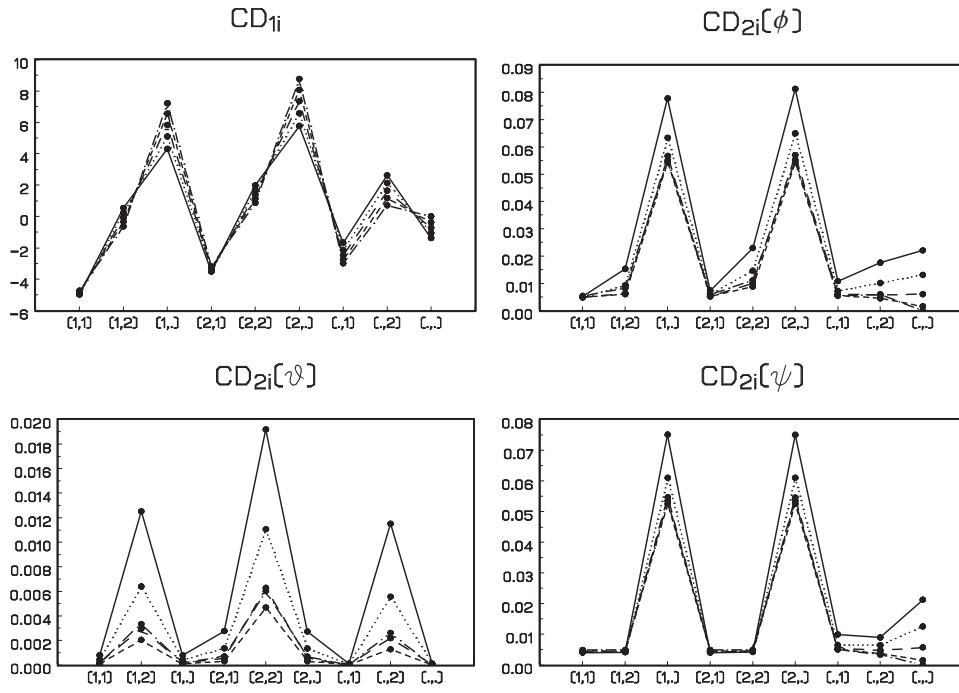


Fig. 6. Global influence for Model 4 with varying education effect. Solid line for no education, dotted for primary education, dashed for low secondary education, short dashes for high secondary education, dots and dashes for higher education.

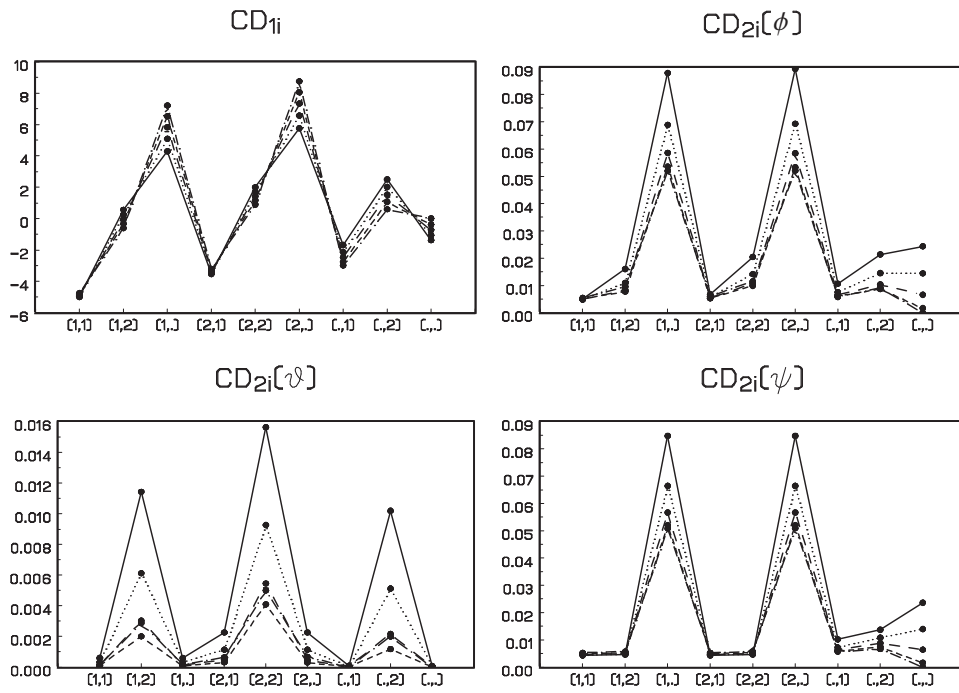


Fig. 7. Global influence for Model 7 with varying education effect. Solid line for no education, dotted for primary education, dashed for low secondary education, short dashes for high secondary education, dots and dashes for higher education.

7. Concluding Remarks

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

As is commonly known, to strengthen the data-analytic findings, it is best to conduct a sensitivity analysis, for example, a global influence analysis. In linear regression, global influence is conceptually simple, straightforward in computation and well studied. In more general settings, the calculation of the Cook's distances requires $N + 1$ model fits (once to fit the model for all observations, and once for each individual that has been excluded from the analysis). This is extremely time-consuming. Nevertheless, in the case of ordinal outcomes, the work is largely reduced. Only a small number of Cook's distances need to be calculated, since there are only a few outcome combinations possible. In the HIS data, only 9 model fits per covariate level were necessary. A disadvantage of global influence, is that it does not give any information on the reason why some individuals are more influential than others, since by deleting a subject, all types of influence stemming from it are lumped together. For this reason the so-called local influence tool (Cook, 1986; Jansen et al., 2003; Thijs et al., 2000; Van Steen et al., 2001; Verbeke et al., 2001) is a good alternative.

In our first data analysis, the estimates of the measurement model parameters are remarkably stable, no matter which assumptions are made regarding the reasons for non-response. When interest lies only in the effect of a specific covariate on the measurement model, all models containing this effect provide very similar evidence. The model selection (independent of the included covariate effects) pointed out that a missing value for either mental health or fixed general practitioner depends on the value of fixed general practitioner. The model that combines both assumptions is (borderline) non-significant. The influence analysis indicated that subjects with a missing value for fixed general practitioner are the most influential ones in these models. So they drive the conclusions towards these models. This is in contrast with the analyses done by Molenberghs et al. (2001) and Verbeke et al. (2001) who found that the influential observations are invariably completers, and with Jansen et al. (2003), where empty observations were influential. This illustrates that any type of subject, no matter how many of the outcomes are measured, can have an impact on the assumptions for the missingness model, and that the results of the data analysis should be carefully explored. The second data analysis shows how the new method can be applied on ordinal data, which strengthens the motivation why this research is needed in practice.

Acknowledgment

Ivy Jansen and Geert Molenberghs gratefully acknowledge support from *Fonds Wetenschappelijk Onderzoek-Vlaanderen* Research Project G.0002.98 "Sensitivity Analysis for Incomplete and Coarse Data" and from Belgian IUAP/PAI network "Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data".

References

- Agresti, A. (2002). *Categorical Data Analysis*. Hoboken, NJ: Wiley.
- Ashford, J. R. and R. R. Sowden (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 H. Solomon (Ed.), *Studies in Item Analysis and Prediction*, Stanford Mathematical Studies in the Social Sciences VI. Stanford, CA: Stanford University Press.
- Baker, S. G. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics* 51, 1042–1052.
- Baker, S. G., W. F. Rosenberger, and R. DerSimonian (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine* 11, 643–657.
- Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88, 9–25.
- Burzykowski, T., G. Molenberghs, J. Tafforeau, H. Van Oyen, and S. Demarest (1999). Missing data in the health interview survey 1997 in belgium. *Archives of Public Health* 57, 107–129.
- Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association* 74, 169–174.
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B* 48, 133–169.
- Cook, R. D. and S. Weisberg (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.
- Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* 42, 909–917.
- Diggle, P. D. and M. G. Kenward (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* 43, 49–93.
- Diggle, P. J., P. J. Heagerty, K.-Y. Liang, and S. L. Zeger (2002). *Analysis of Longitudinal Data (2nd ed.)*. Oxford Science Publications. Oxford: Clarendon Press.
- 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 G. Tutz (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Heidelberg: Springer-Verlag.
- Glonek, G. F. V. and P. McCullagh (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* 57, 477–482.
- Glynn, R. J., N. M. Laird, and D. B. Rubin (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples*, pp. 115–142. New York: Springer-Verlag.
- Jansen, I., N. Hens, G. Molenberghs, M. Aerts, G. Verbeke, and M. G. Kenward (2006). The nature of sensitivity in monotone missing not at random models. *Computational Statistics and Data Analysis* 50, 830–858.
- Jansen, I., G. Molenberghs, M. Aerts, H. Thijs, and K. Van Steen (2003). A local influence approach applied to binary data from a psychiatric study. *Biometrics* 59, 409–418.

- Kenward, M. G. and G. Molenberghs (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science* 12, 236–247.
- Lang, J. B. and A. Agresti (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association* 89, 625–632.
- Liang, K.-Y. and S. L. Zeger (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. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 88, 125–134.
- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* 81, 471–483.
- Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 90, 1112–1121.
- Little, R. J. A. and D. B. Rubin (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. London: Chapman & Hall.
- Molenberghs, G., M. G. Kenward, and E. Goetghebeur (2001). Sensitivity analysis for incomplete contingency tables: the slovenian plebiscite case. *Applied Statistics* 50, 15–29.
- Molenberghs, G., M. G. Kenward, and E. Lesaffre (1997). The analysis of longitudinal ordinal data with nonrandom dropout. *Biometrika* 84, 33–44.
- Molenberghs, G. and E. Lesaffre (1994). Marginal modelling of correlated ordinal data using a multivariate plackett distribution. *Journal of the American Statistical Association* 89, 633–644.
- Molenberghs, G. and E. Lesaffre (1999). Marginal modelling of multivariate categorical data. *Statistics in Medicine* 18, 2237–2255.
- Molenberghs, G. and L. M. Ryan (1999). Likelihood inference for clustered multivariate binary data. *Environmetrics* 10, 279–300.
- Molenberghs, G., H. Thijs, M. G. Kenward, and G. Verbeke (2003). Sensitivity analysis for continuous incomplete longitudinal outcomes. *Statistica Neerlandica* 57, 112–135.
- Molenberghs, G. and G. Verbeke (2005). *Discrete Longitudinal Data*. New York: Springer-Verlag.
- Molenberghs, G. and G. Verbeke (2007). Likelihood ratio, score, and wald tests in a constrained parameter space. *The American Statistician* 61, 1–6.
- Molenberghs, G., G. Verbeke, H. Thijs, E. Lesaffre, and M. G. Kenward (2001). Mastitis in dairy cattle: local influence to assess sensitivity of the dropout process. *Computational Statistics and Data Analysis* 37, 93–113.
- Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association* 60, 516–522.
- Renard, D., G. Molenberghs, H. Van Oyen, and J. Tafforeau (1998). Investigation of the clustering effect in the belgian health interview survey. *Archives of Public Health* 56, 345–361.
- 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., A. Rotnitzky, and J. M. Robins (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse model (with discussion). *Journal of the American Statistical Association* 94, 1096–1146.
- Stiratelli, R., N. Laird, and J. Ware (1984). Random effects models for serial observations with dichotomous response. *Biometrics* 40, 961–972.
- Thijs, H., G. Molenberghs, and G. Verbeke (2000). The milk protein trial: influence analysis of the dropout process. *Biometrical Journal* 42, 617–646.
- Troxel, A. B., D. P. Harrington, and S. R. Lipsitz (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics* 47, 425–438.
- Van Steen, K., G. Molenberghs, G. Verbeke, and H. Thijs (2001). A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling: An International Journal* 1, 125–142.
- Verbeke, G. and G. Molenberghs (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer Verlag.
- Verbeke, G., G. Molenberghs, H. Thijs, E. Lesaffre, and M. G. Kenward (2001). Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics* 57, 7–14.
- Wolfinger, R. and M. O’Connell (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 48, 233–243.
- Wu, M. C. and R. J. Carroll (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 44, 175–188.