STATISTICS IN MEDICINE
Statist. Med. **18**, 2449–2464 (1999)

MISSING DATA PERSPECTIVES OF THE FLUVOXAMINE DATA SET: A REVIEW

GEERT MOLENBERGHS[1]*, ELS J. T. GOETGHEBEUR[2], STUART R. LIPSITZ[3], MICHAEL G. KENWARD[4], EMMANUEL LESAFFRE[5] AND BART MICHIELS[1]

[1]*Biostatistics, Limburgs Universitair Centrum, B3590 Diepenbeek, Belgium*
[2]*Department of Applied Mathematics, Universiteit Gent, Gent, Belgium*
[3]*Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts, U.S.A.*
[4]*London School of Hygiene and Tropical Medicine, London, U.K.*
[5]*Biostatistical Center, Katholieke Universiteit Leuven, Leuven, Belgium*

SUMMARY

Fitting models to incomplete categorical data requires more care than fitting models to the complete data counterparts, not only in the setting of missing data that are non-randomly missing, but even in the familiar missing at random setting. Various aspects of this point of view have been considered in the literature. We review it using data from a multi-centre trial on the relief of psychiatric symptoms. First, it is shown how the usual expected information matrix (referred to as *naive information*) is biased even under a missing at random mechanism. Second, issues that arise under non-random missingness assumptions are illustrated. It is argued that at least some of these problems can be avoided using contextual information. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

Missing data occur often and for a variety of reasons. Many methods have become available to analyse incomplete data. Although most of the literature focuses on continuous outcomes, incomplete categorical data are also well studied. For categorical outcomes, incomplete data imply that a subject is not always classified into a single outcome category but rather into a set of categories.

When referring to the missing data mechanism we will use terminology of Little and Rubin[1] (Chapter 6). A non-response process is said to be *missing completely at random* (MCAR) if the missingness is independent of both unobserved and observed data and *missing at random* (MAR) if, conditional on the observed data, the missingness is independent of the unobserved measurements. A process that is neither missing completely at random nor missing at random is termed *non-random* (MNAR). In the context of likelihood inference, and when the parameters describing the measurement process are functionally independent of the parameters describing the missingness process, MCAR and MAR are *ignorable*, while a non-random process is non-ignorable.

Work on incomplete categorical data has largely been in the context of partially classified contingency tables.[2] Molenberghs *et al.*[3] introduced a method for the analysis of longitudinal ordinal data with non-random drop-out. The latter approach is based on Diggle and Kenward,[4]

* Correspondence to: Geert Molenberghs, Biostatistics, Center for Statistics, Limburgs Universitair Centrum, B3590 Diepenbeek, Belgium. E-mail: geert.molenberghs@luc.ac.be

CCC 0277–6715/99/182449–16$17.50

who treat non-random drop-out in continuous longitudinal data. The EM algorithm[5] is extensively used as a maximization tool, but other proposals have been made as well.[6]

This paper reviews several issues that require the modeller's attention, arising in the context of incomplete categorical data. In the literature, several illustrations have been given using data from a multi-centre trial, where the outcomes of interest (therapeutic effect and occurrence of side-effects) are scored repeatedly on an ordinal scale.[3,7–12] The data are introduced in Section 2.

Following the original work of Rubin and Little, there has evolved a general view that 'likelihood methods' that ignore the missing value mechanism are valid under an MAR process, where likelihood is interpreted in a frequentist sense. This statement needs careful qualification, however, which is the goal of Section 3. Indeed, Rubin[13] has shown that MAR and parameter distinctness (that is, the parameter space of the measurement and missing data processes is the product of the individual parameter spaces) is necessary and sufficient to ensure validity of *direct-likelihood* inference when ignoring the process that causes missing data, where direct-likelihood inference is defined as resulting solely from ratios of the likelihood function for various values of the parameter. However, the use of likelihood inference is often surrounded with references to frequentist concepts, such as identifying and using the appropriate sampling distribution. This is obviously relevant for determining distributions of test statistics, expected values of the information matrix, and measures of precision. Apart from the categorical case, Kenward and Molenberghs[10] provide illustrations for normally distributed data and when sampling is done subject to a stopping rule.

While the treatment of missing data that are missing at random requires some caution, one needs to be even more careful with non-randomly missing data. This contradicts a common belief that, with the availability of methods for incomplete data, fitting models is of the same level of complexity as any other statistical model building exercise and that in fact routine testing for the non-randomness of the non-response process is possible. However, many instances of the contrary have been reported. A classical example is found in Little and Rubin[1] (Section 11.6). Several issues are discussed in Section 4. It is illustrated how models are identifiable by virtue of model assumptions, which are usually impossible to verify merely on statistical grounds. In addition to the potential occurrence of non-unique, boundary solutions, and solutions that violate constraints, we show that models often yield the same or similar fits to the observed data, but produce qualitatively different predictions for the unobserved data. Other issues are presented in Molenberghs *et al.*[11]

2. THE FLUVOXAMINE TRIAL

The data come from a multi-centre study involving 315 patients that are treated by fluvoxamine for psychiatric symptoms described as possibly resulting from a dysregulation of serotonin in the brain. The data are discussed in Molenberghs and Lesaffre,[7] Kenward *et al.*,[8] Molenberghs *et al.*[3] and Michiels and Molenberghs.[12] After recruitment to the study, the patient was assessed at four visits. The therapeutic effect and the extent of side-effects were scored at each visit on an ordinal scale. The side-effect response is coded as: 1, none; 2, not interfering with functionality; 3, interfering significantly with functionality; 4, side-effects surpass the therapeutic effect. Similarly, the effect of therapy is recorded on a four point ordinal scale: (1), no improvement or worsening; 2, minimal improvement; 3, moderate improvement; 4, important improvement. Thus, a side-effect occurs if new symptoms occur while there is therapeutic effect if old symptoms disappear. A total of 299 patients have at least one measurement, including 242 completers. There is also

baseline covariate information on each subject: sex; age; initial severity (scale 1 to 7), and duration of actual mental illness.

3. LIKELIHOOD-BASED FREQUENTIST INFERENCE

It is generally believed that when the missing data are missing at random in the sense of Rubin,[13] the statistical analysis is of the same complexity as the corresponding complete data problem. In this section, we will illustrate that some caution is required.

Let the vector random variable $\boldsymbol{Y}$ correspond to the complete set of measurements on an individual and $\boldsymbol{R}$ the associated missing value indicator. For a particular realization of this pair $(\boldsymbol{y}, \boldsymbol{r})$ the elements of $\boldsymbol{r}$ take the values 1 and 0 indicating, respectively, whether the corresponding values of $\boldsymbol{y}$ are observed or not. Let $(\boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}})$ denote the partition of $\boldsymbol{y}$ into the respective sets of observed and missing data. We assume that the joint distribution of $(\boldsymbol{Y}, \boldsymbol{R})$ is regular.

We are concerned here with the sampling distributions of certain statistics under MCAR and MAR mechanisms. These mechanisms were described in the introduction and can be defined more formally as follows (Little and Rubin[1]). Under an MCAR mechanism $P(\boldsymbol{R} = \boldsymbol{r}|\boldsymbol{y}) = P(\boldsymbol{R} = \boldsymbol{r})$ and the joint distribution of the *observed* data partitions as follows: $f(\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}) = f(\boldsymbol{y}_{\text{obs}}) f(r)$. Under an MAR mechanism $\Pr(\boldsymbol{R} = \boldsymbol{r}|\boldsymbol{y}) = \Pr(\boldsymbol{R} = \boldsymbol{r}|\boldsymbol{y}_{\text{obs}})$ and again the joint distribution of the observed data can be partitioned, $f(\boldsymbol{y}_{\text{obs}}, \boldsymbol{r}) = f(\boldsymbol{y}_{\text{obs}}; \boldsymbol{\theta}) f(\boldsymbol{r}|(\boldsymbol{y}_{\text{obs}}; \boldsymbol{\beta})$ for parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. The corresponding log-likelihood function factors as

$$l(\boldsymbol{\theta}, \boldsymbol{\beta}; \boldsymbol{y}_{\text{obs}}, \boldsymbol{r}) = l_1(\boldsymbol{\theta}; \boldsymbol{y}_{\text{obs}}) + l_2(\boldsymbol{\beta}; \boldsymbol{r}). \tag{1}$$

It is assumed that $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are distinct (the assumption of separability). This partition of the likelihood has usually been taken for granted to mean that, under an MAR mechanism, likelihood methods based on l_1 alone are valid for inferences about $\boldsymbol{\theta}$ *even when interpreted in the broad frequentist sense*. We now consider more precisely the sense in which the different elements of the frequentist likelihood methodology can be regarded as valid in general under the MAR mechanism.

First we note that under the MAR mechanism $\boldsymbol{r}$ is *not* an ancillary statistic for $\boldsymbol{\theta}$. Hence we are not justified in restricting the sample space from that associated with the pair $(\boldsymbol{Y}, \boldsymbol{R})$. In considering the properties of frequentist procedures below we therefore define the appropriate sampling distributions to be that determined by this pair. We call this the *unconditional* sampling framework. By working within this framework we do need to consider the missing value mechanism. We shall be comparing this with the sampling distribution that would apply if $\boldsymbol{r}$ were fixed by design, that is if we repeatedly sampled using the distribution $f(\boldsymbol{y}_{\text{obs}}; \boldsymbol{\theta})$. If this sampling distribution were appropriate, this would lead directly to the use of $l_1(\cdot)$ as a basis for inference. We call this the *naive* sampling framework.

Certain elements of the frequentist methodology can be justified immediately from (1). The maximum likelihood estimator obtained from maximizing $l_1(\boldsymbol{\theta}; \boldsymbol{y}_{\text{obs}})$ alone is identical to that obtained from maximizing the complete log-likelihood function. Similarly the maximum likelihood estimator of $\boldsymbol{\beta}$ is functionally independent of $\boldsymbol{\theta}$ and so any maximum likelihood ratio concerning $\boldsymbol{\theta}$, with common $\boldsymbol{\beta}$, will involve $l_1(\cdot)$ only. Because these statistics are identical whether derived from $l_1(\cdot)$ or the complete log-likelihood, it follows at once that they have the required properties under the naive sampling framework.[1,13,14]

An important element of likelihood-based frequentist inference is the derivation of measures of precision of the maximum likelihood estimators from the information. For this either the

observed information, i_O, can be used where $i_O(\theta_j, \theta_k) = -\partial^2 l(\cdot)/\partial\theta_j\partial\theta_k$ or the expected information, i_E, where

$$i_E(\theta_j, \theta_k) = E\{i_O(\theta_j, \theta_k)\}. \tag{2}$$

The use of the expected information matrix is more problematical. The expectation in (2) needs to be taken over the *unconditional* sampling distribution (the *unconditional information* i_U) and consequently the use of the naive sampling framework (producing the *naive information* i_N) can lead to inconsistent estimates of precision. It is possible to calculate the unconditional information by taking expectations over the appropriate distribution and so correct this bias. Although this added complication is unnecessary in practice it does allow a direct examination of the effect of ignoring the missing value mechanism on the expected information.

3.1. Bivariate binary data

Suppose that each member of the pair of observations (Y_{i1}, Y_{i2}), from unit i, $i = 1, \ldots, n$, is a binary random variable, with associated probabilities $P(Y_{i1} = 1) = \lambda$ and $P(Y_{i2} = 1) = \theta$. It is assumed that an MAR mechanism is operating with respect to the second observation, that is, the probability of Y_{i2} being missing depends on Y_{i1} alone. It follows that Y_{i1} is always observed. We want to compare the naive information i_N with the unconditional information i_U for this set-up.

We express dependence between Y_{i1} and Y_{i2} through the conditional success probabilities of Y_{i2}: $\theta_1 = P(Y_{i2} = 1 \,|\, y_{i1} = 1)$ and $\theta_0 = P(Y_{i2} = 1 \,|\, y_{i1} = 0)$.

The off-diagonal elements of the observed information matrix are zero, so we need consider only the information for one of θ_0 and θ_1 to contrast the naive and unconditional forms of the expected information. Kenward and Molenberghs[10] have shown that the naive information equals

$$i_N(\theta_1, \theta_1) = \frac{m\lambda}{\theta_1(1-\theta_1)} \tag{3}$$

with $0 \leqslant m \leqslant n$ the number of complete observations. For the unconditional information they obtained

$$i_U(\theta_1, \theta_1) = \frac{n\lambda\eta_1}{\theta_1(1-\theta_1)}. \tag{4}$$

From (3) and (4), it can be seen that conditions for $E_R(i_N(\theta_1, \theta_1)) = i_U(\theta_1, \theta_1)$ and $E_R(i_N(\theta_0, \theta_0)) = i_U(\theta_0, \theta_0)$ are $E_R(m/n) = \eta_1 = \eta_0$ and hence $\eta = \eta_1 = \eta_0$, the requirement for an MCAR mechanism to operate. It follows that the MCAR mechanism is both a necessary and sufficient condition for the equivalence of the two forms of information.

3.2. The Fluvoxmine Trial

We will first study two dichotomized versions (category 1 versus higher categories 2, 3 and 4; and categories 1 and 2 versus 3 and 4) of side-effects and therapeutic effects at the first and the last measurement occasions. The data are shown in Table I. The model of the previous section is fitted to all four tables, which is particularly illustrative because naive and unconditional standard error estimates for λ coincide, concentrating potential differences between both estimators in the parameters θ_0 and θ_1. For the first analysis of side-effects, there are only small differences and inference at a common significance level is unaffected. This is different in setting 2. Indeed, the naive significance probability for $H_0 : \theta_0 = 0{\cdot}5$ is 0·0319 while the unconditional version is 0·1306.

Table I. Fluvoxamine trial: dichotomized outcome at first and last measurement occasions

Setting	Outcome	Dichotomized	(0, 0)	(0, 1)	(1, 0)	(1, 1)	(0, ∗)	(1, ∗)
1	side-effect	1/234	89	13	57	65	26	49
2	side-effect	12/34	203	5	14	2	48	27
3	therapeutic	1/234	11	1	124	88	7	68
4	therapeutic	12/34	77	9	119	19	28	47

Table II. Fluvoxamine trial: analysis of the data in Table I. Parameter estimates (naive standard errors; unconditional standard errors) are shown

Parameter	Side-effect 1/234	Side-effect 12/234	Therapeutic 1/234	Therapeutic 12/34
λ	0·572 (0·029; 0·029)	0·144 (0·020; 0·020)	0·937 (0·014; 0·014)	0·619 (0·028; 0·028)
θ_1	0·533 (0·044; 0·045)	0·125 (0·058; 0·083)	0·415 (0·034; 0·034)	0·138 (0·029; 0·029)
θ_0	0·128 (0·034; 0·033)	0·024 (0·011; 0·011)	0·083 (0·073; 0·080)	0·105 (0·033; 0·033)
η_1	0·714	0·372	0·757	0·746
η_0	0·797	0·813	0·632	0·754

Note that θ_1 is substantially different from θ_0, and, more importantly, that the missingness probabilities η_1 and η_0 are very difficult. For therapeutic effect, neither of the two settings leads to differences in standard errors of any importance (see Table II).

The analysis considered above is based on a simple Markov type model. It concentrates the discrepancy between the naive and robust frameworks in the conditional probabilities θ_j ($j = 0, 1$). Other parameterizations are less sensitive to the (mis)use of the naive framework. As an illustration, we analyse side-effects at the first, the second, and the fourth measurement occasion, on a three category scale (with original categories 3 and 4 combined). A trivariate odds ratio model[7] is adopted. Marginal cumulative logits for each outcome are combined with global marginal log-odds ratios for the pairwise and third-order interactions in order to specify the joint distribution. The marginal logits are assumed to depend on *duration*, whereas the log-odds ratios are assumed constant. Molenberghs *et al.*[3] observed that drop-out in the side-effects outcome depends both on the previous measurement, as well as on the value of *duration*. We analysed the set of 222 complete cases as well as all available data. Table III reports on the value of the (naive and unconditional) Wald statistic for a number of hypotheses. Although not spectacular, the differences between naive expected and observed information based tests is larger for the MAR analysis than for the complete case analysis. In particular, the *P*-value for the hypothesis of no *duration* effect (MAR) changes from 0·0049 with the naive information to 0·0110 with the observed information. In this example it was seen consistently that in MAR analyses the observed information yielded smaller test statistics than the naive information. For completers only analyses, this was not always the case.

4. MODELLING THE MISSING DATA MECHANISM

Consider a two-way contingency table where a subset of subjects only has margins observed, as discussed in Little.[15] Several non-response models are displayed in Table IV. The data are

Table III. Side-effects at times 1, 2 and 4. Wald test statistics for the completers only and for an MAR analysis

Hypothesis		Completer cases		MAR	
	d.f.	expected	observed	expected	observed
Common duration effect	2	1·36	1·19	2·54	2·44
No duration effect	3	2·98	2·54	12·90	11·13
Common two-way association	2	10·70	9·99	11·48	9·13
Intercepts equal across times	4	28·73	28·83	34·96	33·44
Common difference between intercepts	2	0·16	0·16	2·07	1·48
Linear trend in first intercept	1	0·0099	0·0099	0·16	0·18
Linear trend in second intercept	1	0·020	0·018	1·15	0·85
Linear trend in both intercepts	2	0·034	0·033	1·18	0·91

reproduced in Table V. Decompose the cell probabilities as $\mu_{jk}\phi_{r_1r_2|jk}$, where $j, k = 1, 2$ index the categorical outcome levels and $r_1, r_2 = 0, 1$ index the response pattern (1 indicating observed). For the data in Table V, the pattern $(r_1, r_2) = (0, 0)$ is absent. The problem of non-response patterns for which there are no observations is a very common one and requires careful attention, explaining why we will consider the pattern exhibited in Table V in detail.

Assuming that the missing data mechanism is MCAR (model I in Table IV), the missingness probabilities reduce to $\phi_{r_1r_2}$ $(r_1, r_2 = 0, 1)$. Note that the cell probabilities μ_{jk} are the same under MCAR and MAR, but of course the missingness probabilities would be different under MAR. One may assume that $\phi_{00} \equiv 0$ and hence drop it from the model. Alternatively, one can include this parameter and estimate it. Then, $\hat{\phi}_{00} = 0$ will follow. Both approaches are equivalent in terms of model fit, but might lead to differences in inference. Assuming ϕ_{00} is dropped, a set of non-redundant parameters is given by $(\mu_{11}, \mu_{12}, \mu_{21}, \phi_{11}, \phi_{10})'$. Parameter estimates and standard errors are presented in Table VI. The estimated complete data counts are shown in Table VII (model I). The fourth pattern in this table corresponds to the subjects without a single measurement; the observed count corresponding to this pattern is zero. Observe that the standard errors for the $\boldsymbol{\mu}$ parameters are slightly bigger than their complete data counterparts, reflecting the additional uncertainty.

Next we will consider several non-random missingness processes. Confining attention to the three patterns observed, there are 7 degrees of freedom in the data, suggesting that we can use at most four parameters for the missingness model, when the measurement model is left fully general. Family II in Table IV belongs to the recursive models proposed by Fay,[16] where: $p_1(j, k)$ is the probability of being observed at the second occasion, given outcomes j and k; $p_2(j, k)$ is the probability of being observed at the second occasion, given outcomes j and k, and given that a measurement was obtained at the first occasion; and $p_3(j, k)$ is the probability of being observed at the second occasion, given outcomes j and k, and given that the measurement at the first occasion was missing. When missingness at one occasion does not depend on missingness at the other occasion, $p_2(j, k) \equiv p_3(j, k)$, and family III is obtained. In family II, similarly to family I, the fact that there are only three out of four patterns observed is taken into account by setting $p_3(j, k) = 1$. When family III is seriously considered for candidate models, one must explicitly address the observations with pattern (0, 0). Below we discuss families II and III in turn.

Table IV. Non-response models for binary two-way tables with supplemental margins

I	$\phi_{11\mid jk} = p_1$ $\phi_{10\mid jk} = p_2$ $\phi_{01\mid jk} = p_3$ $\phi_{00\mid jk} = 1 - p_1 - p_2 - p_3$
II	$\phi_{11\mid jk} = p_1(j,k)p_2(j,k)$ $\phi_{10\mid jk} = p_1(j,k)[1 - p_2(j,k)]$ $\phi_{01\mid jk} = [1 - p_1(j,k)p_3(j,k)$ $\phi_{00\mid jk} = [1 - p_1(j,k)][1 - p_3(j,k)]$
A	$\text{logit}\, p_1(j,k) = \alpha_1 + \alpha_J I(j=1) + \alpha_K I(k=1)$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_J I(j=1) + \alpha_K I(k=1)$ $\text{logit}\, p_3(j,k) = \alpha_3$
B	$\text{logit}\, p_1(j,k) = \alpha_1 + \alpha_J I(j=1)$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_K I(k=1)$ $\text{logit}\, p_3(j,k) = \alpha_3 + \alpha_K I(k=1)$
C	$\text{logit}\, p_1(j,k) = \alpha_1 + \alpha_J I(j=1) + \alpha_K I(k=1)$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_J I(j=1) + \alpha_K I(k=1)$ $\text{logit}\, p_3(j,k) = \alpha_3 + \alpha_J I(j=1) + \alpha_K I(k=1)$
D	$\text{logit}\, p_1(j,k) = \alpha_1$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_J I(j=1)$ $\text{logit}\, p_3(j,k) = \alpha_3 + \alpha_J I(j=1)$
E	$\text{logit}\, p_1(j,k) = \alpha_1$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_K I(k=1)$ $\text{logit}\, p_3(j,k) = \alpha_3 + \alpha_K I(k=1)$
F	$\text{logit}\, p_1(j,k) = \alpha_1$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_{JK}[I(j=1) - I(k=1)]$ $\text{logit}\, p_3(j,k) = \alpha_3 + \alpha_{JK}[I(j=1) - I(k=1)]$
G	$\text{logit}\, p_1(j,k) = \alpha_1$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_{JK}[I(j=1) - I(k=1)]/2$ $\text{logit}\, p_3(j,k) = \alpha_3 + \alpha_{JK}[I(j=1) - I(k=1)]/2$
III	$\phi_{11\mid jk} = p_1(j,k)p_2(j,k)$ $\phi_{10\mid jk} = p_1(j,k)[1 - p_2(j,\text{k})]$ $\phi_{01\mid jk} = [1 - p_1(j,k)]p_2(j,k)$ $\phi_{00\mid jk} = [1 - p_1(j,k)][1 - p_2(j,k)]$
A	$\text{logit}\, p_1(j,k) = \alpha_1 + \alpha_J I(j=1)$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_K I(k=1)$
B	$\text{logit}\, p_1(j,k) = \alpha_0 + \alpha_1 I(j=1)$ $\text{logit}\, p_2(j,k) = \alpha_0 + \alpha_2 I(k=1)$
C	$\text{logit}\, p_1(j,k) = \alpha_1$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_J I(j=1)$
D	$\text{logit}\, p_1(j,k) = \alpha_1$ $\text{logit}\, p_2(j,k) = \alpha_2 + \alpha_K I(k=1)$
IV	$\text{logit}\, \phi_{11\mid jk} = \alpha_1 + \alpha_J I(j=1) + \alpha_K I(k=1)$ $\text{logit}\, \phi_{10\mid jk} = \alpha_2 + \alpha_J I(j=1) + \alpha_K I(k=1)$ $\phi_{01\mid jk} = 1 - \phi_{11\mid jk} - \phi_{10\mid jk}$ $\phi_{00\mid jk} = 0$

Table V. Two-way contingency table with two supplemental margins (Little[15])

100	50	30	28	60
75	75	60		

Table VI. Parameter estimates (standard errors) for models fitted to the data in Table V

Parameter	I(MCAR)	IIA	IIB	IIIA	IV	IV(EM)
μ_{11}	0·280 (0·023)	0·263 (0·022)	0·209 (0·019)	0·236 (0·028)	0·362 (0·029)	0·312
μ_{12}	0·174 (0·021)	0·168 (0·022)	0·167 (0·017)	0·141 (0·026)	0·253 (0·031)	0·200
μ_{21}	0·239 (0·023)	0·231 (0·023)	0·216 (0·019)	0·243 (0·034)	0·181 (0·025)	0·227
μ_{22}	0·308 (0·024)	0·338 (0·025)	0·408 (0·023)	0·380 (0·035)	0·204 (0·031)	0·262
ϕ_{11}	0·628 (0·022)					
ϕ_{10}	0·188 (0·018)					
ϕ_{01}	0·184 (0·018)					
α_1		0·942 (0·152)	0·870 (0·127)	0·870 (0·127)	1·198 (0·376)	0·543
α_2		0·596 (0·165)	0·329 (0·138)	1·054 (0·374)	−1·059 (0·297)	−1·521
α_J		0·559 (0·264)	$+\infty$ (−)	$+\infty$ (−)	−1·546 (0·480)	−0·515
α_K		0·795 (0·219)	$+\infty$ (−)	1·019 (1·245)	0·664 (0·144)	0·489
Odds ratio	2·071 (0·388)	2·295 (0·433)	2·367 (0·457)	2·621 (0·510)	1·613 (0·305)	1·809
Log-likelihood	−971·872	−958·674	−959·384	−986·506	−958·674	−960·747
Model d.f.	5	7	7	7	7	7

In model IIA, missingness is allowed to depend on the outcome at both measurements. The dependence is the same at both occasions, but the overall rate (see intercepts α_1 and α_2) is allowed to differ.

Parameter estimates are shown in Table VI. The model is saturated in the sense that the predicted and observed counts coincide and thus the likelihood ratio statistic $G^2 = 0$. The predicted probabilities for the hypothetical complete data all lie in the interior of the parameter space. Further, estimated complete data counts are all positive, as shown in Table VII (IIA). These add up to the observed counts in Table V. These properties are desirable, but will not always obtain. Furthermore, they do not yield conclusive evidence for the plausibility of the model. We will illustrate these points by changing the non-response model.

For model IIB, the probability of missingness in each outcome depends on its own values only, and these probabilities are allowed to differ at the two measurement occasions. Keeping $p_3 \equiv 1$, the number of parameters in IIA and IIB is the same. Model IIB clearly saturates the degrees of freedom and at the same time yields a non-zero deviance. This is also seen by inspecting the imputed cell counts (Table VII (IIB)); expected counts for the first pattern are different from the observed ones, as are the relevant margins for the second and the third pattern.

The zero cells in Table VII (IIB) are a consequence of high values found for the parameters α_J and α_K, which are diverging to infinity, implying that $p_1(1, k) = p_2(j, 1) = 1$, $\phi_{10}(j, 1) = \phi_{01}(1, k) = 0$. In other words, a boundary solution is found. Should the same model be fitted without constraints on the parameters (for example, by directly modelling missingness

Table VII. Complete data counts for models fitted to Table V

	(1, 1)		(1, 0)		(0, 1)		(0, 0)	
I(MCAR)	83·84	52·21	25·15	15·66	24·59	15·31	0	0
	71·62	92·33	21·49	27·70	21·00	27·08	0	0
IIA	100·00	50·00	14·24	15·76	11·51	14·66	0	0
	75·00	75·00	18·67	41·33	16·49	45·34	0	0
IIB	100·00	46·51	0·00	33·49	0·00	0·00	0	0
	72·58	79·89	0·00	57·52	30·42	57·58	0	0
IIIA	100·00	50·00	12·58	17·42	0·00	0·00	0·00	0·00
	72·58	95·13	9·13	33·15	30·42	39·87	3·82	13·89
IV	100·00	50·00	21·69	8·31	51·22	62·50	0	0
	75·00	75·00	34·87	25·13	−23·22	−2·50	0	0
IV (EM)	100·00	50·00	21·13	8·87	28·00	36·23	0	0
	75·00	75·00	33·57	26·43	0·00	23·77	0	0

probabilities), negative cell counts would be predicted. This phenomenon can be seen as evidence against the model, a point also raised by Baker *et al.*[17]

Since model IIB saturates the degrees of freedom and yet yields a non-zero deviance, the question is raised whether the model can be extended. Going one step further, one might include *two* additional parameters in the model, by extending IIB to

$$\text{logit } p_1(j,k) = \alpha_1 + \alpha_{J1} I(j=1) + \alpha_{K1} I(k=1)$$

$$\text{logit } p_2(j,k) = \alpha_2 + \alpha_{J2} I(j=1) + \alpha_{K2} I(k=1).$$

This model is clearly overparameterized. For different starting values, the maximization routine will lead to different solutions. The range of solutions thus obtained will reproduce the observed data counts exactly. Of course, the corresponding information matrix is singular.

Family III will always assign mass to all four patterns. Thus, it differs from the previous families in that the zero count in pattern (0, 0) has to be treated as a sampling zero. Model IIIA is similar in spirit to model IIB, but family III assumes missingness at both occasions to be independent. Complete data cell counts are displayed in Table VII (IIIA). Clearly, the fit of IIIA is inferior to IIB. This shows by calculating the deviance, but also by considering the prediction for the observed data counts. Note that this model predicts non-zero counts for pattern (0, 0), in spite of the zero count observed for this pattern. Furthermore, IIIA shows a boundary solution as well, albeit in one table only. This indicates that the assumption of independence is unrealistic.

The difference in fit between IIB and IIIA is expected from the difference in the observed data log-likelihood. However, log-likelihoods for IIA and IIB are fairly close, but the predicted complete data cell counts are radically different as well. This fact points towards a general problem with non-random missing data mechanisms. Indeed, we could decompose the full data into two parts: the observed counts; the distribution of the observations over the missing cells, given their observed margin. Models IIA and IIB are in good agreement on the first part, but very different on the second one (a reasonably balanced IIA versus a boundary IIB solution). This follows from the fact that missingness in IIA depends on a combination of influences of both

measurements, while in IIB missingness in a given outcome depends on its own realization only. These are indeed radically different assumptions. Some criticism applies to model IIA. Parameter identification is borrowed from equating the J and K effects at both times. Given the difference in interpretation of p_1 (unconditional) and p_2 (conditional on the status of the first outcome), this may be questionable.

Arbitrariness in distributing the observed counts over the missing cells is illustrated further by considering the somewhat peculiar model IV, which is a special case of the model considered by Baker.[18] Parameter estimate are shown in Table VI. It saturates the degrees of freedom and has a deviance of $G^2 = 0$, properties shared with IIA. However, the imputed cell counts are radically different; the non-response model does not constrain the probabilities to lie in the unit interval, and negative cell counts rather than a boundary solution are obtained under unconstrained maximum likelihood estimation. Although models IIA and IV describe the *observed* data equally well, there are large differences between both, exhibited by the (impossible) imputed values for the complete cell counts (Table VII (IV)). It has to be noted that the negative counts are not the true maximum likelihood estimates, which would be found by constraining the counts to be non-negative. Using the EM algorithm overcomes this problem. This solution is also displayed in Table VI, while the counts are given in Table VII (IV-EM). The fit for the completers does not change and the fit for the second pattern is very similar. However, the complete counts for the third pattern are drastically different. The model is saturated in terms of degrees of freedom, but the deviance is now positive. Baker *et al.*[13] argue that, particularly in large samples, a negative solution (and its corresponding boundary solution) can be viewed as evidence against the model and hence it is not necessary to compute boundary solutions. Apart from these problems, another point of criticism for model IV is that the missingness model treats the third pattern entirely differently from the others; whereas the effect on the first and second pattern is linear on the logit scale, the effect on the third pattern is highly non-linear, and not constrained to be non-negative. Arguably, one has to think harder about formulating a missingness model such that (i) undesirable asymmetries are avoided and (ii) non-negative solutions are ensured (with the possibility of having a boundary solution).

The fact that the predicted complete data counts can change dramatically with the non-response mechanism does not imply that all quantities of interest will change accordingly. It was noted for the models considered by Baker *et al.*[17] that the odds ratio in the 2×2 table, collapsed over all response patterns, is very stable and in fact, for many models, equal to the one in the completers' table. Thus, it is interesting to compute the marginal odds ratio. Estimates for this quantity (and standard errors, obtained with the delta method), have been calculated and are displayed in Table VI. Knowing that the odds ratio for the completers' table equals 2·000 (0·476), it is clear that the estimates for the models fitted are reasonably close, although the one for model IV, obtained with the Newton–Raphson method, is on the opposite side of the value for the completers than the other models. The value obtained for the boundary solution is again closer to the completers' value.

4.1. The Fluvoxamine Data

In the previous section it was argued that several models can look equally plausible, when the fit of the model to the *observed* data is considered as the sole criterion, even though the implications for the complete (partly unobserved) data can be radically different. It was argued that a real life application often has the benefit of subject matter background on the one hand, and the

Table VIII. Data from psychiatric study

(a) Side-effects

89	13	26	2	0	14		
57	65	49					

(b) Therapeutic effect

11	1	7	0	2	14
124	88	68			

(c) Side-effects, by sex

Time 1	Time 2		
	Completers		Drop-outs
	1	2	*
		Males	
1	34	6	10
2	12	19	23
		Females	
1	55	7	15
2	44	42	26

knowledge of a set of covariates on the other. In this section, several analyses of a real set of data will try to illustrate how progress can be made using this extra information. We will illustrate this point using the fluvoxamine study.

We study the occurrence of side-effects (no/yes) and the presence of therapeutic effect (no/yes), outcomes which were to be evaluated at two doctor's visits. All four non-response patterns are observed. The data are shown in Table VIII parts (a) and (b). Note that the data for side-effects agree with those in Table I (setting 1), but now the additional patterns are used as well. Also, the effect of sex and age on side-effects will be studied on the completers and on the patients that drop out (excluding 2 patients that are observed at the second occasion only, as well as 14 patients without measurements). The raw data, collapsed over age, are shown in Table VIII(c). The slight discrepancy between the counts in Table VIII(a) and the summed counts in Table VIII(c) is due to missing baseline information for a few patients.

To both Tables VIII(a) and (b), all models listed in Table IV are fitted (except for model IV). This means that a few models have been added (IID–G and IIIC–D). These models reflect *a priori* information: (i) the data are collected in a time-ordered fashion and hence missingness at the second time could possibly depend on the measurement at the first occasion while the reverse is unlikely; (ii) missingness at the second occasion is much more frequent than at the first occasion. Therefore, missingness at the first occasion could be considered purely accidental, while missingness at the second occasion is likely to be data dependent. All models hold $p_1(j, k)$ constant. *A priori* the family II models are considered more likely than the family III models since in a longitudinal study, missingness is often dominated by drop-out, forcing dependence between

Table IX. Model fit for side-effects (d.f. degrees of freedom; G^2, likelihood ratio test statistic, P-value and marginal odds ratio)

Model	d.f.	G^2	P	Odds ratio
I	6	4·52	0·1044	7·80 (2·39)
IIA	8	0·00	—	5·07 (1·71)
IIC	8	0·00	—	5·07 (1·71)
IID	7	1·52	0·2176	7·84 (2·35)
IIE	7	0·96	0·3272	7·70 (2·14)
IIF	7	2·04	0·1532	7·26 (2·25)
IIG	7	1·32	0·2506	7·98 (2·34)
IIIB	5	70·04	$<0·0001$	6·18 (2·04)
IIIC	6	27·88	$<0·0001$	7·81 (2·34)
IIID	6	27·88	$<0·0001$	7·81 (2·18)

non-response at the various occasions. This is reflected by the association in the marginal probabilities of falling in one of the four response patterns (0·71, 0·24, 0·0063 and 0·045, respectively, yielding an odds ratio of 21·1). Let us describe the models in terms of the effect of the measurements on missingness at the second occasion they assume. In models IIIC and IID, non-response depends on the outcome at the first occasion only; in models IIID and IIE it depends on the second occasion only; in model IIF it depends on the increment between both measurements; and in model IIG it depends on the average of both measurements.

Let us discuss side-effects first. Results are shown in Table IX. First, some models are not considered further since unconstrained maximization would yield negative expected complete data counts, if a boundary solution is not ensured. This additional phrase is important since, due to the sampling 0 in pattern (0, 1), some models (including several models that saturate the degrees of freedom) yield a boundary solution even with unconstrained maximization. Models that are not considered are IIB and IIIA. These two models are similar in the sense that they all assume drop-out in an outcome to depend only on the realization of that outcome. Among the remaining models, the ones belonging to family III are strongly rejected, which was anticipated due to the dependence between non-response at both occasions. The other models would be acceptable if goodness-of-fit were the only criterion considered. This includes MCAR (models I). The best fit, among the non-saturated models, is given by IIE, but models IID and IIG are very similar. These models assume constant non-response at the first occasion, and non-response at the second occasion that depends on either the first outcome, or on the second outcome, or on the average of both. Inspecting the complete data counts of these models, it is reassuring that all yield similar conclusions (with a slightly inferior fit for MCAR). They are displayed in Table X. Thus, the conclusion of our sensitivity analysis might be that missingness at the first occasion of side-effects is constant, whereas missingness at the second occasion depends on side-effects itself (immaterial whether measured at the first occasion, the second occasion, or both). Further, the association between both side-effect measures is considerable, with an odds ratio around 7·8 (standard error around 2·2).

The results for therapeutic effect are shown in Table XI. The same models are excluded on the basis of boundary values. Again, the fit of family III models is very poor. Among the remaining non-saturated models, the only convincing fit is for IIE. In IIE, non-response at time 2 depends

Table X. Complete data counts for models fitted to side-effects data

	(1, 1)		(1, 0)		(0, 1)		(0, 0)	
I(MCAR)	84·00	12·12	28·13	4·06	0·74	0·11	5·26	0·76
	60·21	67·67	20·16	22·66	0·53	0·60	3·77	4·23
IID	89·60	12·89	22·56	3·24	0·92	0·13	5·08	0·73
	57·12	64·41	23·11	26·06	0·44	0·49	3·86	4·35
IIE	89·69	13·12	17·55	7·84	0·95	0·07	4·79	1·05
	57·04	64·24	11·16	38·37	0·60	0·33	3·05	5·16
IIG	89·68	12·95	21·30	4·33	0·94	0·11	5·00	0·82
	57·06	64·35	19·07	30·26	0·48	0·44	3·59	4·62

Table XI. Model fit for therapeutic effect (d.f., degrees of freedom; G^2, likelihood ratio test statistic, P-value and marginal odds ratio)

Model	d.f.	G^2	P	Odds ratio
I	6	5·08	0·0789	7·77 (6·44)
IIA	8	0·00	—	1·13 (0·49)
IIC	8	0·00	—	1·13 (0·49)
IID	7	3·62	0·0571	7·86 (6·39)
IIE	7	0·08	0·7773	8·25 (8·32)
IIF	7	4·74	0·0295	7·10 (5·31)
IIG	7	2·90	0·0886	8·20 (7·18)
IIIB	5	27·56	$<0·0001$	7·67 (5·98)
IIIC	6	29·84	$<0·0001$	7·81 (6·27)
IIID	6	29·84	$<0·0001$	8·25 (8·44)

on the time 2 measurement. Model IIG would still be acceptable, but far less so than the others mentioned. Thus, the picture is much less clear than with side-effects. Even for the marginal odds ratio, two of the saturated models show a relatively small value while the others are higher. Of course, this effect is less severe than it appears due to the large standard errors. These are undoubtfully influenced by the count 1 in the completers' table.

Again, inspecting complete data counts sheds some light on these findings (Table XII). In fact, both models IIC and IIE yield boundary solutions for pattern (1, 0) while this does not need to be the case, not even for a saturated model. Indeed, several model parameters are estimated to be infinity. In addition, the way in which this pattern is filled in depends crucially on the model assumptions. In model IIC, all drop-outs are assumed to have arisen in spite of a therapeutic effect at the second occasion. In model IIE, the situation is exactly reversed. The conclusions for pattern (0, 0) are similar. Clearly, imputation in model IIC is driven by the zero count in the observed data of pattern (0, 1). This feature is less desirable and model IIC should be discarded. Model IIE, on the contrary, is able to reverse the zero columns in patterns (1, 0) and (0, 1), through two parameters at infinity (α_2 and α_K, with opposite signs). Model IIG is similar to but less extreme than IIE, with some differences in pattern (0, 0). Retaining the picture behind IIE and IIG, we might conclude that non-response at the second time is caused by a less favourable evolution and/or situation of therapeutic effect.

Table XII. Complete data counts for models fitted to therapeutic data

	(1, 1)		(1, 0)		(0, 1)		(0, 0)	
IIC	11·00	1·00	0·00	7·00	0·00	0·11	0·00	4·96
	124·00	88·00	0·00	68·00	0·00	1·89	0·00	9·04
IIE	11·57	1·00	6·43	0·00	0·00	0·02	0·96	0·03
	123·43	88·00	68·57	0·00	0·00	1·98	10·27	2·73
IIG	10·42	0·98	7·17	0·38	0·06	0·01	0·89	0·07
	123·51	89·19	47·87	19·48	0·91	0·96	8·26	4·86

Table XIII. Estimates (standard errors) for side-effects (covariates)

Parameter	MCAR	MAR	Non-R (i)	Non-R (ii)
Measurements model				
First time				
Intercept	0·640 (0·402)	0·640 (0·402)	0·642 (0·402)	0·639 (0·402)
Age effect	−0·022 (0·009)	−0·022 (0·009)	−0·022 (0·009)	−0·022 (0·009)
Second time				
Intercept	1.598 (0·489)	1·598 (0·489)	1·745 (0·597)	1·403 (0·489)
Age effect	−0·023 (0·011)	−0·023 (0·011)	−0·021 (0·011)	−0·025 (0·010)
Log-odds ratio	1·955 (0·357)	1·955 (0·357)	1·808 (0·515)	1·935 (0·347)
Drop-out model				
Intercept	0·766 (0·211)	1·085 (0·275)	0·951 (0·315)	1·382 (0·435)
First measurement		−0·584 (0·284)	−0·963 (0·623)	
Second measurement			1·237 (2·660)	−1·231 (0·627)
Sex effect	0·518 (0·275)	0·568 (0·277)	0·636 (0·294)	0·493 (0·289)
Log-likelihood	−480·485	−478·302	−478·149	−478·600
Model d.f.	7	8	9	8

From this model building it is clear that selecting a model which fits the observed data well is not sufficient when non-ignorable models are considered. First, models which produce boundary or invalid solutions should be treated with caution. Arguably, such models should be discarded. Second, one should question the plausibility of non-response mechanisms in terms of design information (for example, time ordering of measurements) and subject matter knowledge (for example, prior knowledge about the directionality of treatment effect).

Next, let us look at the progress that can be made through accounting for the effect of covariates. A marginal odds ratio model was fitted;[6] a logit link is assumed to link the outcome at each measurement occasion to covariates and the association between outcomes is modelled in terms of log-odds ratios. Results are presented in Table XIII.

We selected age as a predictor for the marginal measurement models. For both measurements, age increased the risk of side-effects. Although we allowed the association between both outcomes to be dependent on covariates, the log-odds ratio was found to be constant. In the drop-out model, sex was the only sufficiently significant predictor to be kept in the model. We consider four

drop-out models: (i) MCAR; (ii) MAR; (iii) Non-R(i), drop-out depends on both measurements; (iv) Non-R(ii), drop-out depends only on the value of the second measurement. The most general drop-out model we retained assumes the form

$$\text{logit}(\phi_{ijk}) = \text{logit}(P(\text{non-drop-out} \mid Y_1 = j,\ Y_2 = k, x_i)) = \alpha_0 + \alpha_1 I(j = 1) + \alpha_2 I(k = 1) + \alpha_3 x_i$$

where x_i is the sex of subject i. From the MAR model we conclude that drop-out chances increase for subjects with side-effects at the first visits and are higher for men than for women.

Observe that the MAR model and the Non-R models lead to approximately the same fit. This is in line with our findings in Table IX. First, the association between both measurements is considerable, given a log-odds ratio of about 2, found in all models. MAR and Non-R(ii) both indicate a strong dependence of drop-out on the level of side-effect; the regression coefficients have the same sign. In conclusion, all three models show a strong dependence of drop-out on the occurrence of side-effects, irrespective of whether the first, the second, or both measurements are used, that is, the same conclusion as in the analysis without covariates.

Diggle and Kenward[4] and Molenberghs *et al.*[3] considered several examples where a non-random drop-out model showed a markedly better fit than a random drop-out model. These authors then reparameterized the drop-out model in terms of *size* and *increment* (average of and difference between previous and current measurements). It turned out that in such cases drop-out usually depends on increment rather than on size.

5. CONCLUDING REMARKS

Fitting models to incomplete data by means of maximum likelihood, even when the missing data mechanism is MAR, requires care since likelihood inference is surrounded with references to the sampling distribution. In particular, the classical expected information matrix, while used by many authors, is inconsistent and should be replaced with the observed information matrix.

Non-random (selection) models necessarily encompass a part that cannot be tested by the data. Therefore, whether or not a drop-out model is acceptable cannot be determined solely by mechanical model building exercises. Arbitrariness can be removed partly by careful consideration of the plausibility of a model. One should use as much context derived information as possible. Prior knowledge can given an idea of which models are more plausible. Covariate information can be explicitly included in the models to increase the range of plausible models which can be fit.[2,19] Moreover, covariates can help explaining the dependence between response mechanism and outcomes. Good non-response models in a longitudinal setting should make use of the temporal and association structures among the repeated measures.

ACKNOWLEDGEMENTS

The authors wish to thank Solvay Duphar N.V. for the kind permission to use their data.

REFERENCES

1. Little, R. J. A. and Rubin, D. B. *Statistical Analysis with Missing Data*, Wiley, New York, 1987.
2. Baker, S. G. and Laird, N. M. 'Regression analysis for categorical variables with outcome subject to nonignorable non-response', *Journal of the American Statistical Association*, **83**, 62–69 (1988).
3. Molenberghs, G., Kenward, M. G. and Lesaffre, E. 'The analysis of longitudinal ordinal data with non-random dropout', *Biometrika*, **84**, 33–44 (1997).

4. Diggle, P. D. and Kenward, M. G. 'Informative dropout in longitudinal data analysis (with discussion)', *Applied Statistics*, **43,** 49–93 (1994).
5. Dempster, A. P., Laird, N. M. and Rubin, D. B. 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society*, *Series B*, **39**, 1–38 (1977).
6. Molenberghs, G. and Goetghebeur, E. 'Simple fitting algorithms for incomplete categorical data', *Journal of the Royal Statistical Society*, *Series B*, **59**, 401–414 (1997).
7. Molenberghs, G. and Lesaffre, E. 'Marginal modelling of correlated ordinal data using a multivariate Plackett distribution', *Journal of the American Statistical Association*, **89**, 633–644 (1994).
8. Kenward, M. G., Lesaffre, E. and Molenberghs, G. 'An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random', *Biometrics*, **50**, 945–953 (1994).
9. Lesaffre, E., Molenberghs, G. and Dewulf, L. 'Effect of dropouts in a longitudinal study: an application of a repeated ordinal model', *Statistics in Medicine*, **15**, 1123–1141 (1996).
10. Kenward, M. G. and Molenberghs, G. 'Likelihood based frequentist inference when data are missing at random', *Statistical Science*, **12**, 236–247 (1996).
11. Molenberghs, G., Goetghebeur, E., Lipsitz S. R. and Kenward, M. G. 'Non-Random Missingness in Categorical Data: Strengths and Limitations', *The American Statistician* **53**, 110–118. (1999).
12. Michiels, B. and Molenberghs, G. 'Protective estimation of longitudinal categorical data with nonrandom dropout', *Communications in Statistics – Theory and Methods*, **26**, 65–94 (1997).
13. Rubin, D. B. 'Inference and missing data', *Biometrika*, **63**, 581–592 (1976).
14. Little, R. J. A. 'Inference about means for incomplete multivariate data', *Biometrika*, **63**, 593–604 (1976).
15. Little, R. J. A. 'Pattern-mixture models for multivariate incomplete data', *Journal of the American Statistical Association*, **88**, 125–134 (1993).
16. Fay, R. E. 'Causal models for patterns of nonresponse', *Journal of the American Statistical Association*, **81**, 354–365 (1986).
17. Baker, S. G., Rosenberger, W. F. and DerSimonian, R. 'Closed-form estimates for missing counts in two-way contingency tables', *Statistics in Medicine*, **11**, 643–657 (1992).
18. Baker, S. G. 'Composite linear models for incomplete multinomial data', *Statistics in Medicine*, **13**, 609–622 (1994).
19. Baker, S. G. 'Marginal regression for repeated binary data with outcome subject to nonignorable nonresponse', *Biometrics*, **51**, 1042–1052 (1995).

