Marginalizing pattern-mixture models for categorical data subject to monotone missingness

DOI: 10.1007/s00184-008-0219-y
Handle: http://hdl.handle.net/1942/9265

Noname manuscript No.
(will be inserted by the editor)

Marginalizing pattern-mixture models for categorical data subject to monotone missingness

Cristina Sotto · Caroline Beunckens · Geert Molenberghs · Ivy Jansen · Geert Verbeke

Received: date / Accepted: date

Abstract Many models to analyze incomplete data that allow the missingness to be non-random have been developed. Since such models necessarily rely on unverifiable assumptions, considerable research nowadays is devoted to assess the sensitivity of resulting inferences. A popular sensitivity route, next to local influence (Cook, 1986; Jansen *et al.*, 2003) and so-called intervals of ignorance (Molenberghs, Kenward, and Goetghebeur, 2001), is based on contrasting more conventional selection models with members from the pattern-mixture model family. In the first family, the outcome of interest is modeled directly, while in the second family the natural parameter describes the measurement process, conditional on the missingness pattern. This implies that a direct comparison ought not to be done in terms of parameter estimates, but rather should pass by marginalizing the pattern-mixture model over the patterns. While this is relatively straightforward for linear models, the picture is less clear for the nevertheless important setting of categorical outcomes, since models ordinarily exhibit a certain amount of non-linearity. Following ideas laid out in Jansen and Molenberghs (2007), we offer ways to marginalize pattern-mixture-model-based parameter estimates, and supplement these with asymptotic variance formulas. The modeling context is provided by the multivariate Dale model. The performance of the method and its usefulness for sensitivity analysis is scrutinized using simulations.

Keywords Categorical Data · Multivariate Dale Model · Multiple Imputation

Mathematics Subject Classification (2000) 92B15 · 62P10

The authors gratefully acknowledge financial support from the Interuniversity Attraction Pole Research Network P6/03 of the Belgian Government (Belgian Science Policy).

Cristina Sotto, Caroline Beunckens, Geert Molenberghs, and Ivy Jansen
Center for Statistics, Hasselt University, Agoralaan, Building D, B3590 Diepenbeek, Belgium.
Tel: +32-11-268205; Fax: +32-11-268299. E-mail: cristina.sotto@uhasselt.be

Cristina Sotto
School of Statistics, University of the Philippines, Diliman, Quezon City, Philippines.

Geert Verbeke
Biostatistical Centre, Katholieke Universiteit Leuven, Belgium.

1 Introduction

Clinical studies are oftentimes of a longitudinal nature, and as such, have a tendency to be incomplete. The nature of such incompleteness can have implications on the conclusions arising, making it thus essential that appropriate techniques be used in the analysis of data with missing values. Historically, missing data were ignored and the analyses carried out on the completers only. With the recent developments in the area of missing data, however, it has been shown that such an approach would only be valid under the most restrictive of assumptions (Rubin, 1976). A more realistic approach takes into account the mechanism driving the missingness, in addition to the process governing the outcomes, rather than ignore it. Models for incomplete data, therefore, simultaneously consider both the outcome and non-response processes, $\boldsymbol{Y}$ and $\boldsymbol{R}$, respectively, thereby entailing working on the joint distribution, $f(\boldsymbol{y}, \boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\psi})$.

For the non-response process, $\boldsymbol{R}$, one can speak of *monotone missingness* (or dropout), in which the unobserved measurements within a longitudinal series all occur after a particular measurement occasion, and *non-monotone missingness*, for which missing values arise intermittently within the series. In general, the missingness process, be it monotone or not, can be classified using the taxonomy introduced by Rubin (1976). A mechanism is said to be *missing completely at random* (MCAR) if the processes governing the missingness and the outcomes are independent, possibly conditionally on covariates. Under such a case, any analysis ignoring the missing data would be valid. However, such an assumption is unrealistic and often not applicable in real-life contexts. A more relaxed assumption would be one of *missing at random* (MAR), for which the missingness may depend on the observed outcomes and on covariates but, given these, not further on the unobserved outcomes. When, in addition to such dependencies, the unobserved data provide further information about the missing data mechanism, then the mechanism is referred to as being *missing not at random* (MNAR).

Various routes can be taken in modeling incomplete data, and these depend on the chosen factorization of the joint distribution, $f(\boldsymbol{y}, \boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\psi})$, of the response and missingness processes. *Selection models* (Rubin, 1976; Little and Rubin, 2002) use the factorization $f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\psi})$, while the reverse factorization $f(\boldsymbol{y}|\boldsymbol{r}, \boldsymbol{\theta})f(\boldsymbol{r}|\boldsymbol{\psi})$ is the basis of *pattern-mixture models* (Little, 1993, 1994). Alternatively, if $\boldsymbol{Y}$ and $\boldsymbol{R}$ are taken to be independent, conditional on a common set of latent variables or random effects, via the factorization $f(\boldsymbol{y}|\boldsymbol{b}, \boldsymbol{\theta})f(\boldsymbol{r}|\boldsymbol{b}, \boldsymbol{\psi})$, one may use *shared-parameter models* (Wu and Carrol, 1988; Wu and Bailey, 1989). An obvious consequence of these different factorizations is a disparity in the interpretations of the parameters therein, making it difficult to allow comparison across the three model frameworks. Whereas $\boldsymbol{\theta}$ in a selection model would represent the marginal effects of the dependent variables on the response, it would denote pattern-specific effects in a pattern-mixture model; when turning to the shared-parameter model family, $\boldsymbol{\theta}$ would denote conditional effects, conditional on the missingness patterns and on the random effects.

References to pattern-mixture models include Rubin (1977); Glynn, Laird, and Rubin (1986); Little and Rubin (2002); Hogan and Laird (1997); Molenberghs and Kenward (2007). Several authors have contrasted selection models and pattern-mixture models, to either compare the answer to the same scientific question, such as a marginal treatment effect or time evolution, as a form of sensitivity analysis, or to gain additional insight by supplementing the results from a selection model with those of a pattern-mixture approach. Examples can be found in Verbeke, Lesaffre and Spiessens

(2001) and Michiels *et al.* (2002) for continuous outcomes, while categorical outcomes have been treated by Michiels, Molenberghs, and Lipsitz (1999a); Michiels, Molenberghs and Lipsitz (1999b). On a more directed slant, i.e., not necessarily *vis a vis* the selection modeling approach, Thijs *et al.* (2002) discussed various strategies in fitting pattern-mixture models for continuous outcomes via model simplification, while Jansen and Molenberghs (2007) explored strategies for the categorical case using identifying restrictions.

In this paper, we focus on pattern-mixture models (PMM) as applied to categorical data with monotone missingness using the approach described by Jansen and Molenberghs (2007). Our contribution here is to apply the procedure over simulated settings, so as to be able to assess the performance of the method. Moreover, while Jansen and Molenberghs (2007) restricted their formulae to marginalized point estimates, we additionally introduce asymptotic variance expressions. In Section 2, we start with a brief overview on pattern-mixture models, followed by a description of various identification schemes. The method for fitting a PMM to categorical data with monotone missingness is described in Section 2.2. Section 3 describes the design and the results of the simulation study, and finally, some points of discussion and concluding remarks are given in Section 4.

2 Pattern-Mixture Models

Pattern-mixture models (PMM) (Little, 1993, 1994, 1995; Molenberghs and Verbeke, 2005; Molenberghs and Kenward, 2007) employ a different response model for each pattern of the missing values, the observed data being a mixture of these, weighted by the probability of each missing value or dropout pattern. Thus, the family of pattern-mixture models is based on the factorization

$$f(\boldsymbol{y}, \boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}|\boldsymbol{r}, \boldsymbol{\theta}) f(\boldsymbol{r}|\boldsymbol{\psi}), \tag{1}$$

where dependence on covariates is suppressed from notation. The conditional density of the measurements given the missingness process is therefore combined with the marginal density describing the missingness mechanism. Note that the second factor can depend on covariates, but not on the random outcomes. In addition, it is possible to have different covariate dependencies in either component of the factorization.

A key issue in this modeling framework is that pattern-mixture models, by construction, are under-identified, that is, over-specified. Little (1993, 1994) addresses this problem with the use of so-called identifying restrictions, whereby inestimable parameters of the incomplete patterns are set equal to (functions of) the parameters describing the distribution of the completers. Under *complete case missing values* (CCMV), information that is unavailable is always borrowed from the completers. Alternatively, the nearest identified pattern can be used as a donor (*neighboring case missing values*, NCMV). Using such identification schemes, within a specific pattern, the conditional distribution of the unobserved measurements, given the observed ones becomes identified. Yet a third possibility is to borrow information for an unidentified distribution from all patterns for which it is identified (*available case missing values*, ACMV). A further set of restrictions is presented in Kenward, Molenberghs, and Thijs (2003). The concept of restrictions will be elaborated on in what follows.

On the other hand, model simplification can be considered as an alternative to the use of identifying restrictions in addressing the under-identification in pattern-mixture

models. This approach might, for instance, restrict trends to functional forms supported by the observed data within a pattern, e.g., linear or quadratic time trend. One can also take the route of allowing the parameters to vary in a controlled parametric way, and, for example, assume that the time evolution within each pattern is unstructured but parallel across patterns. Thijs *et al.* (2002) discussed these sub-strategies in detail, in the context of continuous longitudinal outcomes. Examples can be found in Molenberghs and Kenward (2007).

2.1 Identification Schemes

Restricting attention to the case of monotone missingness or dropout, let us assume that we have n patterns ($t = 1, 2, \ldots, n$), for which the complete data density is given by

$$f_t(y_1, y_2, \ldots, y_n) = f_t(y_1, y_2, \ldots, y_t) f_t(y_{t+1}, y_{t+2}, \ldots, y_n | y_1, y_2, \ldots, y_t). \quad (2)$$

Owing to the monotone nature of incompleteness, the number of repeated measurements is equal to the number of potential patterns, even though some of the latter may turn out to be empty. Density (2) describes all n outcomes, conditional on pattern t being observed. Although the first factor on the r.h.s. is clearly identified from the observed data, and can hence be modeled using the observed data, the second is not, necessitating the application of identifying restrictions.

While, in principle, completely arbitrary restrictions can be used by means of any valid density function over the appropriate support, strategies which relate back to the observed data deserve privileged interest. One can base identification on all patterns for which a given component y_s is identified. A general expression for this is

$$f_t(y_s | y_1, y_2, \ldots, y_{s-1}) = \sum_{j=s}^{n} \omega_{sj} f_j(y_s | y_1, y_2, \ldots, y_{s-1}), \quad s = t+1, t+2, \ldots, n. \quad (3)$$

Let $\boldsymbol{\omega}_s = (\omega_{ss}, \omega_{s,s+1}, \ldots, \omega_{sn})'$. Every $\boldsymbol{\omega}_s$ with components summing to one provides a valid identification scheme. Three special and important cases are considered. Little (1993) proposed *complete case missing values* (CCMV), which uses the following identification:

$$f_t(y_s | y_1, y_2, \ldots, y_{s-1}) = f_n(y_s | y_1, y_2, \ldots, y_{s-1}), \quad s = t+1, t+2, \ldots, n. \quad (4)$$

In other words, the conditional distribution beyond time t is always borrowed from the corresponding conditional distribution of the completers, i.e., $\omega_{sn} = 1$ and all other ω_{sj}'s set to zero. This identification scheme is perhaps most reasonable and applicable when a large bulk of the subjects have complete data and only small proportions fall within the various dropout patterns. Extension of this approach to non-monotone patterns is also particularly easy. Alternatively, the nearest identified pattern can be used to identify missing components:

$$f_t(y_s | y_1, y_2, \ldots, y_{s-1}) = f_s(y_s | y_1, y_2, \ldots, y_{s-1}), \quad s = t+1, t+2, \ldots, n. \quad (5)$$

Such restrictions are referred to as *neighboring case missing values* (NCMV), with $\omega_{ss} = 1$ and the other ω_{sj}'s set to zero. The third special case of (3) is *available case*

missing values (ACMV), under which derivation of the corresponding $\boldsymbol{\omega}_s$ vectors is straightforward and results in

$$\omega_{sj} = \frac{\pi_j f_j(y_1, y_2, \ldots, y_{s-1})}{\sum_{\ell=s}^{n} \pi_\ell f_\ell(y_1, y_2, \ldots, y_{s-1})}, \tag{6}$$

where π_j is the fraction of observations in pattern j (Molenberghs *et al.*, 1998). Clearly, $\boldsymbol{\omega}_s$ defined by (6) consists of components which are nonnegative and sum to one, i.e., a valid density function is obtained. Molenberghs *et al.* (1998) showed that, for monotone missing data, ACMV in the pattern-mixture framework is the natural counterpart of MAR in the selection model framework.

Note that identifying restrictions are unverifiable assumptions, which also follows from the unidentified nature of the ω parameters, affording sensitivity analysis rather than providing an unambiguous answer to the incompleteness issue. One might, for example, prefer CCMV in cases where the bulk of the data is complete, or perhaps ACMV since, in the monotone case, it is the counterpart of MAR (Molenberghs *et al.*, 1998). However, it is wise not to put too much emphasis on one particular set of restrictions and to consider several. Note that family (3) is termed "interior" by Kenward, Molenberghs, and Thijs (2003) since identification is done based on distributions following from the data itself. One could, of course, envisage restrictions coming from external sources, such as expert opinions, historical studies, etc. These authors also consider identifying restrictions that ensure dropout does not depend on future, unobserved values, termed *missing not future dependent* (MNFD).

2.2 PMM for Categorical Outcomes in the Monotone Case

In this section, we present a procedure for fitting PMM to categorical data with monotone missingness. Jansen and Molenberghs (2007) describe this procedure in detail, with an application to actual data. In general, the method can be broken down into three stages. In the first stage, initial models are fitted to the observed data: a univariate model for subjects with one measurement; a bivariate model for subjects with two measurements; and so on. Identification occurs at the second stage, in which a particular identifying restriction is chosen. Using the initial models obtained in stage 1, the required probabilities for the chosen identifying restriction are computed and subsequently used to perform multiple imputation (Molenberghs and Kenward, 2007) of the missing data to complete each of the patterns. At the last stage, analysis is conducted by fitting full-vector (n-variate) models to the completed data of each pattern and pooling these analyses over the multiple imputations. The mixture of these pattern-specific models comprises the final pattern-mixture model.

In line with Jansen and Molenberghs (2007), we zoom in on the case of three binary outcomes and outline the procedure in detail. Extension to more outcomes and/or to more than two outcome categories is straightforward. The multivariate Dale model (Molenberghs and Lesaffre, 1994) will be used to estimate the parameters of the identified densities. For completers (pattern 3), a trivariate Dale model will be used, for pattern 2, a bivariate Dale model, and for pattern 1, a univariate Dale model, which is equivalent to conventional logistic regression. This is referred to as the *minimal approach* (Jansen and Molenberghs, 2007). The multivariate Dale model combines logistic regression for each of the measurements with marginal global odds ratios to describe the association between outcomes. For three observed measurements, i.e., for the group

of completers in our case, this results in the following logistic-regression and odds-ratio formulations (subject-specific indices i are removed for ease of notation, as is an index to pattern 3 that should be present, strictly speaking, in the p and X functions):

$$\ln\left(\frac{p_{0++}}{1-p_{0++}}\right) = \boldsymbol{X}_1\boldsymbol{\theta}, \tag{7a}$$

$$\ln\left(\frac{p_{+0+}}{1-p_{+0+}}\right) = \boldsymbol{X}_2\boldsymbol{\theta}, \tag{7b}$$

$$\ln\left(\frac{p_{++0}}{1-p_{++0}}\right) = \boldsymbol{X}_3\boldsymbol{\theta}, \tag{7c}$$

$$\ln\psi_{12} = \ln\left(\frac{p_{00+}(1-p_{0++}-p_{+0+}+p_{00+})}{(p_{0++}-p_{00+})(p_{+0+}-p_{00+})}\right) = \boldsymbol{X}_4\boldsymbol{\theta}, \tag{7d}$$

$$\ln\psi_{13} = \ln\left(\frac{p_{0+0}(1-p_{0++}-p_{++0}+p_{0+0})}{(p_{0++}-p_{0+0})(p_{++0}-p_{0+0})}\right) = \boldsymbol{X}_5\boldsymbol{\theta}, \tag{7e}$$

$$\ln\psi_{23} = \ln\left(\frac{p_{+00}(1-p_{+0+}-p_{++0}+p_{+00})}{(p_{+0+}-p_{+00})(p_{++0}-p_{+00})}\right) = \boldsymbol{X}_6\boldsymbol{\theta}, \tag{7f}$$

$$\ln\psi_{123} = \ln\left(\frac{p_{000}p_{011}p_{101}p_{110}}{p_{001}p_{010}p_{100}p_{111}}\right) = \boldsymbol{X}_7\boldsymbol{\theta}, \tag{7g}$$

with $p_{ijk} = P(Y_1 = i, Y_2 = j, Y_3 = k), i, j, k = 0, 1$, and a + in lieu of a subscript indicating that the marginal probability over this index needs to be used. It is worthwhile to note that, using the above formulations, the incomplete patterns provide information neither about the unobserved outcomes nor about the associations involving those unobserved outcomes. Thus, for pattern 2, only an analogous model involving (7a), (7b) and (7d) can be obtained from the data, while for pattern 1 only (7a) will be available. Of course, the functions (7a)-(7g) defined above are specific to a particular pattern and therefore also the design matrices may change from pattern to pattern. This is necessary, among others, when different patterns correspond to different sets of parameters.

Also in this setting, one is interested in model parameters for the full set of repeated outcomes, and thus identifying restrictions are necessary to determine the unknown probabilities by equating them to functions of known probabilities. In the normal case, restrictions are very natural to apply, because marginal as well as conditional distributions can be expressed as simple functions of the mean vector and the covariance matrix components. For categorical data in general and for the Dale model in particular, there is no easy transition from marginal to conditional distributions in terms of the model parameters.

First, the minimal approach is followed in the sense that a trivariate Dale model for the complete pattern is combined with a bivariate and univariate Dale model for the incomplete patterns. This results in densities $f_3(y_1, y_2, y_3)$, $f_2(y_1, y_2)$, and $f_1(y_1)$, respectively. From this approach, the following underlying probabilities can be estimated:

$$\begin{aligned} p_{y_1,y_2,y_3|3} &= P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3|t = 3), \\ p_{y_1,y_2|2} &= P(Y_1 = y_1, Y_2 = y_2|t = 2), \\ p_{y_1|1} &= P(Y_1 = y_1|t = 1). \end{aligned}$$

For pattern 2, there is only one possibility to impute the missing cell counts, since information on the third measurement can only be borrowed from pattern 3. So, the partial counts $Z_{y_1,y_2|2}$ and the conditional probabilities $p_{y_3|y_1,y_2,3} = P(Y_3 = y_3|Y_1 = y_1, Y_2 = y_2, t = 3)$ have to be used to identify $Z^*_{y_1,y_2,y_3|2}$ as $Z_{y_1,y_2|2} \times p_{y_3|y_1,y_2,3}$. The asterisk refers to a completed count. The corresponding counts are indicated by the symbol Z. For pattern 1, we have several possibilities to impute the missing cell counts, since information on the second measurement can be borrowed from pattern 2 as well as from pattern 3. Using (3), the joint probability of y_1, y_2, and y_3 in pattern 1 can be written as:

$$p_{y_1,y_2,y_3|1} = p_{y_1|1}\left[\omega p_{y_2|y_1,2} + (1-\omega)p_{y_2|y_1,3}\right] p_{y_3|y_1,y_2,3},$$

where specific choices of ω lead to the previously defined identifying restrictions, i.e., CCMV, NCMV, and ACMV:

$$\begin{aligned}
\text{CCMV} &: p_{y_1|1} \cdot p_{y_2|y_1,3} \cdot p_{y_3|y_1,y_2,3},\\
\text{NCMV} &: p_{y_1|1} \cdot p_{y_2|y_1,2} \cdot p_{y_3|y_1,y_2,3},\\
\text{ACMV} &: \omega = \frac{\pi_2 p_{y_1|2}}{\pi_2 p_{y_1|2} + \pi_3 p_{y_1|3}},
\end{aligned}$$

such that the missing cell counts can be identified as follows:

$$\text{CCMV} : \widehat{Z}^*_{y_1,y_2,y_3|1} = Z_{y_1|1} \cdot \widehat{p}_{y_2|y_1,3} \cdot \widehat{p}_{y_3|y_1,y_2,3}, \tag{8}$$

$$\text{NCMV} : \widehat{Z}^*_{y_1,y_2,y_3|1} = Z_{y_1|1} \cdot \widehat{p}_{y_2|y_1,2} \cdot \widehat{p}_{y_3|y_1,y_2,3}, \tag{9}$$

$$\text{ACMV} : \widehat{Z}^*_{y_1,y_2,y_3|1} = Z_{y_1|1}\left[\frac{\widehat{\pi}_2\widehat{p}_{y_1,y_2|2} + \widehat{\pi}_3\widehat{p}_{y_1,y_2|3}}{\widehat{\pi}_2\widehat{p}_{y_1|2} + \widehat{\pi}_3\widehat{p}_{y_1|3}}\right]\widehat{p}_{y_3|y_1,y_2,3}. \tag{10}$$

Multiple imputations are then generated by drawing uniformly from Bernoulli variables with the probabilities embedded in (8)–(10). As stated earlier, once the imputations have been generated, the so-called analysis task model, i.e., the final model, can be fitted and multiple-imputation inference conducted. Using conventional multiple-imputation machinery, obtaining parameter and precision estimates is straightforward. In particular, suppose M imputations are performed to obtain M completed datasets, each of which are analyzed, yielding M sets of analyses on some parameter ϕ of interest. Letting $\widehat{\phi}^m$ and $\widehat{S}^m$, for $m = 1, \ldots, M$, denote, respectively, the parameter estimate and corresponding variance estimate for the m^{th} completed dataset, analyses for the M imputed datasets can be pooled into a single inference by

$$\overline{\widehat{\phi}} = \frac{1}{M}\sum_{m=1}^{M}\widehat{\phi}^m \quad \text{and} \quad \widehat{\boldsymbol{V}} = \widehat{\boldsymbol{W}} + \left(\frac{M+1}{M}\right)\widehat{\boldsymbol{B}},$$

where

$$\widehat{\boldsymbol{W}} = \frac{\sum_{m=1}^{M}\widehat{S}^m}{M} \quad \text{and} \quad \widehat{\boldsymbol{B}} = \frac{\sum_{m=1}^{M}(\widehat{\phi}^m - \overline{\widehat{\phi}})(\widehat{\phi}^m - \overline{\widehat{\phi}})'}{M-1},$$

with $\widehat{\boldsymbol{W}}$ denoting the estimate of the average *within*-imputation variance and $\widehat{\boldsymbol{B}}$ the estimated *between*-imputation variance (Rubin, 1987).

Although Jansen and Molenberghs (2007) described an extension of the above procedure to the case of non-monotone missingness, their analyses were confined to the monotone case, owing to an almost negligible proportion of intermittent missingness in

the data they considered. They further mention that generalization of the procedure to the non-monotone case would result in a proliferation of parameters, leading to a trade-off between clarity and parsimony.

2.3 Marginalized Effects Across Patterns

By way of its factorization of the joint distribution of $\boldsymbol{Y}$ and $\boldsymbol{R}$, as specified in (1), a PMM gives rise to pattern-specific estimates. In cases where the scientific interest is on such, a PMM would, of course, be a natural choice. If, however, interest lies on marginal effects, one might opt for a selection model instead. It is also possible though that one wants to obtain, from a fitted PMM, so-called "marginalized" effects, i.e., which are the pattern-specific effects marginalized across all patterns. This might be the case, for example, when one wants to compare PMM results with their selection model counterparts. In the case of continuous data, the overall (or marginalized) effect is simply a weighted average of the pattern-specific effects. For categorical data, however, the marginalization is less straightforward. Jansen and Molenberghs (2007) propose an approximation for the marginalized effects of a PMM for a logistic model. While they presented expressions for marginalized point estimates, we additionally establish asymptotic variance expressions.

Suppose that to model the data from pattern t, we consider a logistic regression of a binary response Y_{ij} on some treatment of interest, X_i, of the form

$$P(Y_{ij} = 1|t) = \frac{e^{\alpha_t + \beta_t X_i}}{1 + e^{\alpha_t + \beta_t X_i}}, \tag{11}$$

where the subscript i refers to the subject and j to the measurement occasion. In general, $t = 1, 2, \ldots, K$, where $K = n$ for the monotone case (as in Section 2.1), while $K = 2^n$ for non-monotone missingness, in which case, for instance, the multivariate Dale model might be used to model the non-response. The parameters α_t and β_t can depend on j, but we suppress this index from notation. Assuming that interest is on one particular treatment effect X, e.g., treatment effect at the last occasion, and that π_t denote the pattern probabilities as defined before, the marginal success probability over all patterns, is then equal to

$$P(Y_{ij} = 1) = \sum_{t=1}^{K} \pi_t \frac{e^{\alpha_t + \beta_t X}}{1 + e^{\alpha_t + \beta_t X}}. \tag{12}$$

From this formulation, there are three ways to calculate the marginal effects (e.g., the intercept A and treatment effect B) at the last occasion. First, using a direct linear approach (Park and Lee, 1999), where a weighted average over all patterns is taken, i.e.,

$$A_{PL} := \sum_t \pi_t \alpha_t \quad \text{and} \quad B_{PL} := \sum_t \pi_t \beta_t. \tag{13}$$

Though seemingly logical, this marginalization may not be entirely appropriate for binary responses, as we will show shortly. Second, the marginal probability can be approximated via a logistic regression, a probit model, or by fully using the longitudinal nature of the design, through a Dale model, a generalized linear mixed model, etc. Third, classical averaging can be performed. To this effect, function (12) is kept as is and then computed, graphed, or sampled from. Note that averaging in this last

way will be similar to the marginalization of generalized linear mixed-effects model (Molenberghs and Verbeke, 2005). Here, the marginalization is over pattern, rather than over random effects. When a GLMM is used in each pattern, then a double marginalization is necessary, one over the random effects and another over the patterns. In this paper, although we shall focus on the second approach, using a Dale model, comparison will also be made with the less appropriate but sometimes used direct linear approach.

Jansen and Molenberghs (2007) approximate (12) by a logistic regression, i.e., letting A_{JM} and B_{JM} denote the marginalized effects,

$$f(X) = \sum_t \pi_t \frac{e^{\alpha_t + \beta_t X}}{1 + e^{\alpha_t + \beta_t X}} \cong \frac{e^{A_{JM} + B_{JM} X}}{1 + e^{A_{JM} + B_{JM} X}}, \tag{14}$$

and derived the following expressions for A_{JM} and B_{JM}:

$$\begin{aligned} A_{JM} &:= \text{logit}\left(\sum_t \pi_t \frac{e^{\alpha_t}}{1 + e^{\alpha_t}}\right) \text{ and} \\ B_{JM} &:= \frac{\sum_t \pi_t \beta_t \frac{e^{\alpha_t}}{1 + e^{\alpha_t}} \frac{1}{1 + e^{\alpha_t}}}{\left(\sum_t \pi_t \frac{e^{\alpha_t}}{1 + e^{\alpha_t}}\right)\left(\sum_t \pi_t \frac{1}{1 + e^{\alpha_t}}\right)}. \end{aligned} \tag{15}$$

They further showed that when the treatment effects are equal across patterns, the marginalized treatment effect at the last occasion obtained through approximation (14), will not be larger in absolute value than that obtained using the direct linear approach (13). Moreover, marginalization may both increase or decrease the effect in absolute value when the treatment effects differ across patterns.

Considering now the direct linear approach, Park and Lee (1999) assume that the pattern-specific success probability (11) is approximately linear, i.e.,

$$P(Y_{ij} = 1|t) = \frac{e^{\alpha_t + \beta_t X_i}}{1 + e^{\alpha_t + \beta_t X_i}} \cong \alpha_t + \beta_t X_i, \tag{16}$$

whereby averaging over all patterns yields the following expression for the marginal success probability (12):

$$f(X) = \sum_t \pi_t \frac{e^{\alpha_t + \beta_t X}}{1 + e^{\alpha_t + \beta_t X}} \cong A_{PL} + B_{PL} X, \tag{17}$$

with A_{PL} and B_{PL} as defined in (13). Clearly, the assumption that the pattern-specific success probability is linear, though probable in certain cases, is generally not realistic for most scenarios. Moreover, approximation (16) essentially ignores the presence of the link function in modeling the binary response. Though this approach to obtain marginal effects is entirely appropriate for Gaussian outcomes, for which the response is modeled directly, i.e., with linear link function, such an approximation can easily fail when modeling is done via a link, especially when the link entails a highly non-linear transformation of the response.

Let us now denote the maximum likelihood estimates of the pattern-specific parameters, α_t and β_t, of the trivariate Dale model for pattern t, for $t = 1, 2, 3$, by $\widehat{\alpha}_t$

and $\widehat{\beta}_t$, respectively. Further, we let $\widehat{\pi}_t$ denote the sample proportion for pattern t, for $t = 1, 2, 3$. Note that the sample proportions are also the maximum likelihood estimates for the true pattern proportions π_t in a multinomial model. Substituting $\widehat{\alpha}_t, \widehat{\beta}_t$, and $\widehat{\pi}_t$ into expressions (13) and (15), the estimates for the marginalized effects are given by

$$\widehat{A}_{PL} = \sum_t \widehat{\pi}_t \widehat{\alpha}_t \qquad \text{and} \qquad \widehat{B}_{PL} = \sum_t \widehat{\pi}_t \widehat{\beta}_t \tag{18}$$

$$\widehat{A}_{JM} = \text{logit}\left(\sum_t \widehat{\pi}_t \frac{e^{\widehat{\alpha}_t}}{1+e^{\widehat{\alpha}_t}}\right) \text{ and } \widehat{B}_{JM} = \frac{\sum_t \widehat{\pi}_t \widehat{\beta}_t \frac{e^{\widehat{\alpha}_t}}{1+e^{\widehat{\alpha}_t}} \frac{1}{1+e^{\widehat{\alpha}_t}}}{\left(\sum_t \widehat{\pi}_t \frac{e^{\widehat{\alpha}_t}}{1+e^{\widehat{\alpha}_t}}\right)\left(\sum_t \widehat{\pi}_t \frac{1}{1+e^{\widehat{\alpha}_t}}\right)} \tag{19}$$

The asymptotic variances of these estimators can be obtained using the delta method, for which we assume independence of the pattern-specific estimates across patterns. Precisely, the variances are defined by

$$\begin{aligned} U \equiv f_1(\boldsymbol{\theta}) &= \sum_{k=1}^{3} \pi_k \frac{e^{\alpha_k}}{1+e^{\alpha_k}}, \\ V \equiv f_2(\boldsymbol{\theta}) &= \sum_{k=1}^{3} \pi_k \frac{1}{1+e^{\alpha_k}}, \quad \text{and} \\ W \equiv f_3(\boldsymbol{\theta}) &= \sum_{k=1}^{3} \pi_k \beta_k \frac{e^{\alpha_k}}{1+e^{\alpha_k}} \frac{1}{1+e^{\alpha_k}}, \end{aligned}$$

with corresponding estimators given by:

$$\begin{aligned} \widehat{U} \equiv f_1(\widehat{\boldsymbol{\theta}}) &= \sum_{k=1}^{3} \widehat{\pi}_k \frac{e^{\widehat{\alpha}_k}}{1+e^{\widehat{\alpha}_k}}, \\ \widehat{V} \equiv f_2(\widehat{\boldsymbol{\theta}}) &= \sum_{k=1}^{3} \widehat{\pi}_k \frac{1}{1+e^{\widehat{\alpha}_k}}, \qquad \text{and} \\ \widehat{W} \equiv f_3(\widehat{\boldsymbol{\theta}}) &= \sum_{k=1}^{3} \widehat{\pi}_k \widehat{\beta}_k \frac{e^{\widehat{\alpha}_k}}{1+e^{\widehat{\alpha}_k}} \frac{1}{1+e^{\widehat{\alpha}_k}}. \end{aligned}$$

Such an assumption is entirely justified in a PMM since, by (1), the outcome vector is modeled conditionally on a given dropout pattern and thus, the pattern-specific parameters for one particular pattern can be viewed as being estimated independently of those in other patterns. In fact, the model for the outcomes given a pattern can differ across patterns. Note that we need to be concerned, a priori, with potential dependencies between observations after the imputation process has taken place, because, for example, information on the completers is used to impute sequences that are incomplete. However, our application of multiple imputation follows the general theory (Little and Rubin, 2002), where the observed data are used to estimate the parameters from the imputation distribution. Observations would definitely become dependent if this parameter were used as if it were known. To alleviate this problem, first, draws are made from the parameter's posterior distribution, separately for each of the multiple imputations, and only thereafter are imputations generated. While not

straightforward to establish independence in a particular situation like ours, general theory (Rubin, 1987) states that this procedure removes or at least alleviates this problem. Moreover, it is sensible to assume that the correlations are mild to begin with. In our case, this rests on the following assumptions:

– Independence of pattern-specific estimates across patterns implies, $\forall k \neq \ell$,

$$\mathrm{Cov}(\widehat{\alpha}_k, \widehat{\alpha}_\ell) = 0, \quad \mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_\ell) = 0 \quad \text{and} \quad \mathrm{Cov}(\widehat{\alpha}_k, \widehat{\beta}_\ell) = 0.$$

– Independence of estimates for pattern probabilities and (measurement model) pattern-specific estimates further implies:

$$\begin{aligned} \mathrm{Cov}(\widehat{\alpha}_k, \widehat{\pi}_k) = 0 \quad \text{and} \quad \mathrm{Cov}(\widehat{\beta}_k, \widehat{\pi}_k) = 0, \quad \forall k = 1, 2, 3, \\ \mathrm{Cov}(\widehat{\alpha}_k, \widehat{\pi}_\ell) = 0 \quad \text{and} \quad \mathrm{Cov}(\widehat{\beta}_k, \widehat{\pi}_\ell) = 0, \quad \forall k \neq \ell. \end{aligned}$$

We further assume that the pattern probabilities and the pattern-specific parameters are estimated independently of each other. Again, this is plausible in a PMM since, by way of factorization (1), the dropout process, from which the pattern probabilities arise, is modeled independently of the outcomes, thereby rendering the estimates of the former independent of the pattern-specific estimates that characterize the latter. Finally, because we substitute the maximum likelihood estimates of α_t, β_t, and π_t, the above estimates (18) and (19) are consistent for (13) and (15), respectively, and this follows directly from standard maximum likelihood principles, under the usual regularity conditions (Welsh, 1996).

For our specific case of 3 outcomes, these aysmptotic variances are given by:

$$\begin{aligned} \mathrm{Var}(\widehat{A}_{PL}) &= \sum_{t=1}^{3} {\pi_t}^2 \mathrm{Var}(\widehat{\alpha}_t) + \sum_{t=1}^{3} {\alpha_t}^2 \mathrm{Var}(\widehat{\pi}_t) + 2 \sum_{t<\ell} \alpha_t \alpha_\ell \mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell) \\ \mathrm{Var}(\widehat{B}_{PL}) &= \sum_{t=1}^{3} {\pi_t}^2 \mathrm{Var}(\widehat{\beta}_t) + \sum_{t=1}^{3} {\beta_t}^2 \mathrm{Var}(\widehat{\pi}_t) + 2 \sum_{t<\ell} \beta_t \beta_\ell \mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell) \end{aligned} \tag{20}$$

$$\begin{aligned} \mathrm{Var}(\widehat{A}_{JM}) &= \frac{1}{U^2(1-U)^2} \mathrm{Var}(\widehat{U}) \\ \mathrm{Var}(\widehat{B}_{JM}) &= \frac{W^2}{U^4 V^2} \mathrm{Var}(\widehat{U}) + \frac{W^2}{U^2 V^4} \mathrm{Var}(\widehat{V}) + \frac{1}{U^2 V^2} \mathrm{Var}(\widehat{W}) + \\ &\quad \frac{2W^2}{U^3 V^3} \mathrm{Cov}(\widehat{U}, \widehat{V}) - \frac{2W}{U^3 V^2} \mathrm{Cov}(\widehat{U}, \widehat{W}) - \frac{2W}{U^2 V^3} \mathrm{Cov}(\widehat{V}, \widehat{W}) \end{aligned} \tag{21}$$

where:

$$\mathrm{Var}(\widehat{U}) = \sum_{t=1}^{3} {\pi_t}^2 \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^4}\mathrm{Var}(\widehat{\alpha}_t) + \sum_{t=1}^{3} \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^2}\mathrm{Var}(\widehat{\pi}_t) + 2\sum_{t<\ell} \frac{e^{\alpha_t}e^{\alpha_\ell}}{(1+e^{\alpha_t})(1+e^{\alpha_\ell})}\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell)$$

$$\mathrm{Var}(\widehat{V}) = \sum_{t=1}^{3} {\pi_t}^2 \frac{(-e^{\alpha_t})^2}{(1+e^{\alpha_t})^4}\mathrm{Var}(\widehat{\alpha}_t) + \sum_{t=1}^{3} \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^2}\mathrm{Var}(\widehat{\pi}_t) + 2\sum_{t<\ell} \frac{1}{(1+e^{\alpha_t})(1+e^{\alpha_\ell})}\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell)$$

$$\mathrm{Var}(\widehat{W}) = \sum_{t=1}^{3} {\pi_t}^2{\beta_t}^2 \frac{(e^{\alpha_t})^2(1-e^{\alpha_t})^2}{(1+e^{\alpha_t})^6}\mathrm{Var}(\widehat{\alpha}_t) + \sum_{t=1}^{3} {\pi_t}^2 \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^4}\mathrm{Var}(\widehat{\beta}_t) + \sum_{t=1}^{3} {\beta_t}^2 \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^4}\mathrm{Var}(\widehat{\pi}_t) + 2\sum_{t=1}^{3} {\pi_t}^2\beta_t \frac{(e^{\alpha_t})^2(1-e^{\alpha_t})}{(1+e^{\alpha_t})^5}\mathrm{Cov}(\widehat{\alpha}_t, \widehat{\beta}_t) + 2\sum_{t<\ell} \beta_t\beta_\ell \frac{e^{\alpha_t}e^{\alpha_\ell}}{(1+e^{\alpha_t})^2(1+e^{\alpha_\ell})^2}\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell)$$

$$\mathrm{Cov}(\widehat{U}, \widehat{V}) = \sum_{t=1}^{3} {\pi_t}^2(-1) \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^4}\mathrm{Var}(\widehat{\alpha}_t) + \sum_{t=1}^{3} \frac{e^{\alpha_t}}{(1+e^{\alpha_t})^2}\mathrm{Var}(\widehat{\pi}_t) + \sum_{t<\ell} \frac{e^{\alpha_t}+e^{\alpha_\ell}}{(1+e^{\alpha_t})(1+e^{\alpha_\ell})}\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell)$$

$$\mathrm{Cov}(\widehat{U}, \widehat{W}) = \sum_{t=1}^{3} {\pi_t}^2\beta_t \frac{(e^{\alpha_t})^2(1-e^{\alpha_t})}{(1+e^{\alpha_t})^5}\mathrm{Var}(\widehat{\alpha}_t) + \sum_{t=1}^{3} \beta_t \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^3}\mathrm{Var}(\widehat{\pi}_t) + \sum_{t=1}^{3} {\pi_t}^2 \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^4}\mathrm{Cov}(\widehat{\alpha}_t, \widehat{\beta}_t) + \sum_{t<\ell} e^{\alpha_t}e^{\alpha_\ell} \frac{\beta_t(1+e^{\alpha_\ell}) + \beta_\ell(1+e^{\alpha_t})}{(1+e^{\alpha_t})^2(1+e^{\alpha_\ell})^2}\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell)$$

$$\mathrm{Cov}(\widehat{V}, \widehat{W}) = \sum_{t=1}^{3} {\pi_t}^2(-1)\beta_t \frac{(e^{\alpha_t})^2(1-e^{\alpha_t})}{(1+e^{\alpha_t})^5}\mathrm{Var}(\widehat{\alpha}_t) + \sum_{t=1}^{3} \beta_t \frac{e^{\alpha_t}}{(1+e^{\alpha_t})^3}\mathrm{Var}(\widehat{\pi}_t) + \sum_{t=1}^{3} {\pi_t}^2(-1) \frac{(e^{\alpha_t})^2}{(1+e^{\alpha_t})^4}\mathrm{Cov}(\widehat{\alpha}_t, \widehat{\beta}_t) + \sum_{t<\ell} \frac{(1+e^{\alpha_t})\beta_\ell e^{\alpha_\ell} + (1+e^{\alpha_\ell})\beta_t e^{\alpha_t}}{(1+e^{\alpha_t})^2(1+e^{\alpha_\ell})^2}\mathrm{Cov}(\widehat{\pi}_t, \widehat{\pi}_\ell)$$

From (14) and (17) it is easy to see that (A_{JM}, B_{JM}) and (A_{PL}, B_{PL}) are two approximations for some true underlying marginal effects, say (A, B), which are unknown but might be estimated by fitting a selection model. The estimators $(\widehat{A}_{JM}, \widehat{B}_{JM})$ and $(\widehat{A}_{PL}, \widehat{B}_{PL})$ can thus be used to estimate these underlying marginal effects. For our

simulation study, the marginal effects were obtained by factoring the underlying data generating mechanism in a selection model (SEM) formulation. The procedure is described in what follows. In Section 3.2.3, we look at how these two approximations fare in estimating the true SEM-type marginal parameters.

We now describe how one might obtain the underlying marginal parameters from a PMM, with which the above-defined marginalized effects estimates, $(\widehat{A}_{PL}, \widehat{B}_{PL})$ and $(\widehat{A}_{JM}, \widehat{B}_{JM})$, can be subsequently compared. Though we illustrate the procedure for the specific case of 3 time points, generalization to more time points is straightforward. Recall that the general formulation of a PMM given in (1) allows us to compute the joint distribution of the binary responses, $\boldsymbol{Y}$, and the dropout pattern, t, conditional on the treatment indicator, X, as

$$\begin{aligned} P(\boldsymbol{Y} = \boldsymbol{y}, t|X) &= P(\boldsymbol{Y} = \boldsymbol{y}|t, X)P(t|X) \\ &= P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3|t, X)P(t|X) \\ &= P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3|t, X)\pi_{t|X}, \end{aligned}$$

where $\pi_{t|X}$ denote the pattern proportions conditional on the treatment indicator. The marginal success probability, say, at the last time point, given treatment X, can then be obtained by summing over all response values at all other time points and over all patterns, i.e.,

$$\begin{aligned} P(Y_3 = 1|X) &= \sum_{\forall y_1, y_2, t} P(\boldsymbol{Y} = \boldsymbol{y}, t|X) \\ &= \sum_{y_1=0}^{1} \sum_{y_2=0}^{1} \sum_{t=1}^{3} P(Y_1 = y_1, Y_2 = y_2, Y_3 = 1|t, X)\pi_{t|X} \\ &= \sum_{t=1}^{3} \pi_{t|X} P(Y_3 = 1|t, X). \end{aligned} \tag{22}$$

Assuming now a logistic regression of Y_3 on the treatment indicator, X, we have

$$\text{logit } P(Y_3 = 1|X) = A_3 + B_3 X,$$

which implies

$$\text{logit } P(Y_3 = 1|X = 0) = A_3 \quad \text{and} \quad \text{logit } P(Y_3 = 1|X = 1) = A_3 + B_3,$$

which can then be solved for the true underlying parameters, A_3 and B_3, as

$$\begin{aligned} A_3 &= \text{logit } P(Y_3 = 1|X = 0) \quad \text{and} \\ B_3 &= \text{logit } P(Y_3 = 1|X = 1) - \text{logit } P(Y_3 = 1|X = 0). \end{aligned}$$

Though interest is usually on the effects at the last time point, effects at intermediate time points can also be obtained analogously.

Note that $P(Y_3 = 1|t, X)$ in (22) is modeled logistically as specified by (7c) in the Dale model, i.e.,

$$\text{logit } P(Y_3 = 1|t, X) = \alpha_{t,3} + \beta_{t,3} X,$$

which is equivalent to

$$P(Y_3 = 1|t, X) = \frac{e^{\alpha_{t,3} + \beta_{t,3} X}}{1 + e^{\alpha_{t,3} + \beta_{t,3} X}},$$

which, upon substitution into (22) leads to

$$P(Y_3 = 1|X) = \sum_{t=1}^{3} \pi_{t|X} \frac{e^{\alpha_{t,3}+\beta_{t,3}X}}{1 + e^{\alpha_{t,3}+\beta_{t,3}X}}. \tag{23}$$

Clearly, this is not equivalent to (12), which uses the unconditional pattern proportions, π_t, as weights. Since $\widehat{A}_{JM}$ and $\widehat{B}_{JM}$, as well as $\widehat{A}_{PL}$ and $\widehat{B}_{PL}$, are both based on (12), it would be natural to expect bias for these when estimating the underlying SEM-type marginal effects. Thus, in our presentation of results, we look at MSE, rather than variances to assess the precision of the said estimates. Of course in the case that the conditional pattern proportions, $\pi_{t|X}$, are equal to the unconditional ones, π_t, e.g., under an MCAR mechanism, then the estimates of Jansen and Molenberghs (2007) are unbiased for the SEM-type marginal parameters.

3 Simulation Study

We first present the design and then the results of our primary simulation study. The last subsection presents results of additional simulations conducted to investigate convergence properties, as well as to gain more insight into the performance of the different identifying restrictions.

3.1 Design

For the simulation study, we considered 3 binary outcomes, (Y_1, Y_2, Y_3), and a single two-level covariate, X, possibly denoting a treatment indicator. To formulate an underlying PMM for the outcomes, we define, for each pattern, a distinct trivariate Dale model of the general form (7). For our case, we specify a simple logistic model for each outcome, consisting of an intercept and treatment effect, respectively denoted as α_t and β_t, $t = 1, 2, 3$. Further, we set the association parameters ψ's to be constant. This implies a 10-dimensional full parameter vector, i.e.,

$$\boldsymbol{\theta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3, \ln\psi_{12}, \ln\psi_{13}, \ln\psi_{23}, \ln\psi_{123})'.$$

Letting $\boldsymbol{C} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, the design matrices in (7) can be defined as:

$$\boldsymbol{X}_1 = \begin{pmatrix} \boldsymbol{C} & {}_2\mathbf{0}_8 \\ {}_8\mathbf{0}_2 & {}_8\mathbf{0}_8 \end{pmatrix}, \quad \boldsymbol{X}_2 = \begin{pmatrix} {}_2\mathbf{0}_2 & {}_2\mathbf{0}_2 & {}_2\mathbf{0}_6 \\ {}_2\mathbf{0}_2 & \boldsymbol{C} & {}_2\mathbf{0}_6 \\ {}_6\mathbf{0}_2 & {}_6\mathbf{0}_2 & {}_6\mathbf{0}_6 \end{pmatrix}, \quad \boldsymbol{X}_3 = \begin{pmatrix} {}_4\mathbf{0}_4 & {}_4\mathbf{0}_2 & {}_4\mathbf{0}_4 \\ {}_2\mathbf{0}_4 & \boldsymbol{C} & {}_2\mathbf{0}_4 \\ {}_4\mathbf{0}_4 & {}_4\mathbf{0}_2 & {}_4\mathbf{0}_4 \end{pmatrix},$$

while for $i = 4, 5, 6, 7$, we have

$$\boldsymbol{X}_4 = [x_4]_{ij}, \quad \text{where } x_4 = \begin{cases} 1, & \text{for } i = 7 \text{ and } j = 7 \\ 0, & \text{otherwise} \end{cases},$$

$$\boldsymbol{X}_5 = [x_5]_{ij}, \quad \text{where } x_5 = \begin{cases} 1, & \text{for } i = 8 \text{ and } j = 8 \\ 0, & \text{otherwise} \end{cases},$$

$$\boldsymbol{X}_6 = [x_6]_{ij}, \quad \text{where } x_6 = \begin{cases} 1, & \text{for } i = 9 \text{ and } j = 9 \\ 0, & \text{otherwise} \end{cases}, \quad \text{and}$$

$$\boldsymbol{X}_7 = [x_7]_{ij}, \quad \text{where } x_7 = \begin{cases} 1, & \text{for } i = 10 \text{ and } j = 10 \\ 0, & \text{otherwise} \end{cases},$$

Table 1 *Trivariate Dale model parameter values specified for the three dropout patterns.*

Pattern	α_1	β_1	α_2	β_2	α_3	β_3	ψ_{12}	ψ_{13}	ψ_{23}	ψ_{123}
1	0.214	0.096	0.182	0.067	0.142	0.090	1.4	1.1	1.6	1.2
2	0.155	0.135	0.130	0.084	0.100	0.083	1.6	1.2	1.5	1.7
3	0.109	0.033	0.155	0.028	0.142	0.056	2.2	1.8	2.3	1.5

which further implies

$$\begin{aligned}
\boldsymbol{X}_1\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\theta}_1 = \boldsymbol{C}(\alpha_1, \beta_1)' = (\alpha_1, \alpha_1 + \beta_1)', \qquad & \boldsymbol{X}_4\boldsymbol{\theta} = 1\theta_4 = \ln \psi_{12}, \\
\boldsymbol{X}_2\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\theta}_2 = \boldsymbol{C}(\alpha_2, \beta_2)' = (\alpha_2, \alpha_2 + \beta_2)', \qquad & \boldsymbol{X}_5\boldsymbol{\theta} = 1\theta_5 = \ln \psi_{13}, \\
\boldsymbol{X}_3\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\theta}_3 = \boldsymbol{C}(\alpha_3, \beta_3)' = (\alpha_3, \alpha_3 + \beta_3)', \qquad & \boldsymbol{X}_6\boldsymbol{\theta} = 1\theta_6 = \ln \psi_{23} \quad \text{and} \\
& \boldsymbol{X}_7\boldsymbol{\theta} = 1\theta_7 = \ln \psi_{123}.
\end{aligned}$$

The values chosen for $\boldsymbol{\theta}$, for each pattern, are given in Table 1. The mixture of the 3 trivariate Dale models defined by these sets of parameters gives rise to our underlying PMM.

To produce monotone missingness, we first need to define a dropout indicator, D, denoting the occasion at which dropout occurs, and make the convention that $D = n+1$ for a complete sequence. Letting X denote the covariate, in our case, the treatment indicator, missingness is then generated by formulating a dropout type mechanism using a logistic model of the form:

$$\text{logit } P(D = j | D \geq j, X) = \nu_0 + \nu_1 X,$$

where:

$$P(D = j|X) = \begin{cases} P(D = 2|D \geq 2, X), & j = 2, \\ P(D = 3|D \geq 3, X)[1 - P(D = 2|D \geq 2, X)], & j = 3, \\ [1 - P(D = 3|D \geq 3, X)]\,[1 - P(D = 2|D \geq 2, X)]\,, & j = 4. \end{cases}$$

Two dropout settings were considered for the above model, using the following sets of parameters: $(\nu_0, \nu_1) = (-2.2, 0.8)$ and $(\nu_0, \nu_1) = (-1.5, 0.8)$. The resulting percentages of subjects per pattern, as well as per covariate (treatment) level, are given in Table 2. Combining these mixing weights from each of these dropout models with the three trivariate Dale models defines two underlying PMMs.

Table 2 *Percentages of subjects per pattern for the two dropout settings.*

Pattern	Setting 1			Setting 2		
	$x = 0$	$x = 1$	Total	$x = 0$	$x = 1$	Total
1	4.99	9.89	14.88	9.12	16.59	25.71
2	4.49	7.93	12.42	7.46	11.09	18.54
3	40.52	32.17	72.70	33.42	22.32	55.75

The underlying SEM-type marginal parameters corresponding to the two PMMs were computed using the approach described at the end of Section 2.3 and are shown in Table 3.

Table 3 *True marginal parameter values for the two underlying pattern-mixture models. (The marginal intercept and marginal treatment effect for the j^{th} outcome, $j = 1, 2, 3$, are denoted A_j and B_j, respectively.)*

PMM	Y_1		Y_2		Y_3	
	A_1	B_1	A_2	B_2	A_3	B_3
1	0.1236	0.0746	0.1551	0.0457	0.1381	0.0640
2	0.1350	0.0950	0.1559	0.0558	0.1356	0.0700

On each of the two defined underlying PMMs, we generated $S = 500$ samples, each of size $N = 1000$ and applied the procedure described in Section 2.2. Marginalized parameter estimates using both (15) and the direct linear method (13) were also computed. Our choice of parameter values (Table 1), as well as the choice of sample size $N = 1000$, was dictated by limitations in fitting both the initial, as well as the final, analysis models. Fitting a trivariate or a bivariate Dale model, or even an ordinary logistic model for that matter, requires that all combinations of the outcome vector are present per level of the covariate. In generating each of the $S = 500$ samples, it was therefore necessary to ensure that all combinations were "observed" within each pattern. The values for $\boldsymbol{\theta}$ in Table 1 were thus chosen so that the probabilities for the trivariate Dale models (per pattern), when combined with the dropout probabilities, would result in joint probabilities that are still large enough to generate all combinations within one pattern for a sample of $N = 1000$.

Given the amount of replication, we believe the simulation error does not adversely interfere in important ways with our findings.

3.2 Primary Simulation Results

We organize the discussion of the results into those pertaining to the initial estimates, the pattern-specific estimates, and the marginalized effects estimates, respectively.

3.2.1 Initial Estimates

For the first stage, on the observed data of each sample, we fitted a trivariate Dale model for cases in pattern 3, a bivariate Dale model for the second pattern, and a logistic model for pattern 1. At this stage, no attempt is made to address the missingness yet; appropriate models are simply fit to the observed data. The resulting estimates, averaged over the $S = 500$ samples, were then compared with the true parameter values from the underlying trivariate Dale models, so as to obtain the bias. MSEs of the estimates were also computed. Both bias and MSE are presented in Table 4.

In both settings, parameter estimates exhibit very small bias and reasonably small MSEs. The results for pattern 3 under setting 1 are the most precise, as might be expected, since it is under this particular setting and within this particular pattern that a large proportion of the subjects fall. For the same pattern, but under setting 2, bias and MSEs are slightly larger, since setting 2 consists of more missingness and thus fewer subjects within pattern 3 (see Table 2). For patterns 1 and 2, under setting 1, although biases are quite acceptable, the treatment effects, as well as the association, seem to be less precisely estimated than the intercepts, but this improves in setting 2.

Table 4 *Bias and MSE of the parameter estimates for the initial models fitted, under each dropout setting, to the observed data: trivariate Dale model for pattern 3, bivariate Dale model for pattern 2 and logistic model for pattern 1.*

Setting	Parameter	Pattern 1		Pattern 2		Pattern 3	
		Bias	MSE	Bias	MSE	Bias	MSE
1	α_1	0.0044	0.0913	0.0099	0.0908	0.0045	0.0112
	β_1	0.0151	0.1284	-0.0302	0.1413	-0.0014	0.0236
	α_2	——	——	0.0030	0.0917	0.0049	0.0101
	β_2	——	——	-0.0123	0.1386	0.0001	0.0205
	α_3	——	——	——	——	0.0012	0.0104
	β_3	——	——	——	——	-0.0025	0.0223
	$\ln \psi_{12}$	——	——	0.0121	0.1354	-0.0042	0.0252
	$\ln \psi_{13}$	——	——	——	——	-0.0057	0.0222
	$\ln \psi_{23}$	——	——	——	——	0.0039	0.0217
	$\ln \psi_{123}$	——	——	——	——	0.0008	0.0994
2	α_1	0.0021	0.0476	0.0054	0.0543	0.0122	0.0139
	β_1	0.0005	0.0745	0.0006	0.0840	-0.0115	0.0345
	α_2	——	——	-0.0014	0.0570	-0.0005	0.0125
	β_2	——	——	0.0039	0.0858	-0.0027	0.0298
	α_3	——	——	——	——	0.0070	0.0128
	β_3	——	——	——	——	-0.0110	0.0311
	$\ln \psi_{12}$	——	——	0.0286	0.0976	0.0006	0.0310
	$\ln \psi_{13}$	——	——	——	——	-0.0034	0.0298
	$\ln \psi_{23}$	——	——	——	——	0.0047	0.0326
	$\ln \psi_{123}$	——	——	——	——	-0.0126	0.1166

Finally, in contrast with the results for pattern 3, the incomplete patterns 1 and 2 yield better results under setting 2, since there are more subjects within these patterns than there are under setting 1.

3.2.2 Pattern-Specific Estimates

The results from the initial models fitted to the observed data were then used to apply various identifying restrictions, under which multiple imputations ($M = 5$) were obtained to fill in the missing data. Under each setting and for each identifying restriction considered, three trivariate Dale models were then fitted to the completed data – one for each pattern. The bias and MSE of the pattern-specific parameter estimates under dropout settings 1 and 2 are given in Tables 5 and 6, respectively. Only the results for patterns 1 and 2 are presented, since the results for pattern 3 remain the same as in Table 4.

Let us set out by considering first the results for pattern 2. In general, under both dropout settings, substantial bias can be observed for the intercept of the third outcome, α_3, as well as the two-way association parameters involving this outcome, i.e., $\ln \psi_{13}$ and $\ln \psi_{23}$. This seems reasonable to expect since, for pattern 2, it is this outcome that is imputed, and thus more susceptible to bias. All other parameters for this pattern exhibit small bias. The MSEs, on the other hand, indicate that the intercepts α_1 and α_2 for outcomes 1 and 2, respectively, are slightly more precisely estimated than the treatment effects β_1 and β_2. For the association parameters, larger MSEs are obtained for those involving the third outcome, especially so for the three-way association $\ln \psi_{123}$. Finally, it seems that the treatment effect for the third outcome,

Table 5 *Bias and MSE of the pattern-specific parameter estimates for the trivariate Dale models fitted to the completed data using various identifying restrictions (dropout setting 1).*

Pattern	Parameter	CCMV		ACMV		NCMV	
		Bias	MSE	Bias	MSE	Bias	MSE
1	α_1	0.0037	0.0911	0.0049	0.0906	0.0047	0.0909
	β_1	0.0161	0.1279	0.0144	0.1270	0.0147	0.1281
	α_2	0.3139	0.1190	0.2645	0.0946	-0.0464	0.1190
	β_2	-0.0423	0.0311	-0.0454	0.0392	0.0162	0.1713
	α_3	0.4004	0.1830	0.3514	0.1421	0.2940	0.1087
	β_3	-0.0676	0.0387	-0.0739	0.0334	-0.0583	0.0354
	$\ln \psi_{12}$	0.4387	0.2346	0.4116	0.2115	0.1578	0.1899
	$\ln \psi_{13}$	0.4085	0.2268	0.4752	0.2733	0.4193	0.2286
	$\ln \psi_{23}$	0.2775	0.1349	0.3634	0.1818	0.3350	0.1642
	$\ln \psi_{123}$	0.3877	0.4634	0.2283	0.3062	0.2464	0.3099
2	α_1	0.0105	0.0901	0.0099	0.0907	0.0098	0.0903
	β_1	-0.0310	0.1397	-0.0302	0.1412	-0.0299	0.1406
	α_2	0.0030	0.0921	0.0037	0.0915	0.0030	0.0914
	β_2	-0.0123	0.1392	-0.0133	0.1375	-0.0124	0.1375
	α_3	0.3241	0.1288	0.3312	0.1350	0.3269	0.1318
	β_3	-0.0436	0.0416	-0.0578	0.0430	-0.0513	0.0394
	$\ln \psi_{12}$	0.0132	0.1355	0.0132	0.1355	0.0131	0.1352
	$\ln \psi_{13}$	0.3225	0.1626	0.3308	0.1716	0.3236	0.1631
	$\ln \psi_{23}$	0.4009	0.2148	0.3986	0.2167	0.3963	0.2110
	$\ln \psi_{123}$	-0.1062	0.2881	-0.0678	0.2934	-0.0876	0.2828

β_3, is the most precisely estimated of all parameters, as indicated by the smaller MSE for this parameter under both settings.

For pattern 1, in general, under both settings, we observe that the intercept parameters for the second and third outcomes, α_2 and α_3, respectively, as well as all association parameters, have fairly large bias. A particular exception can be seen for the NCMV case, under which α_2, exhibits much smaller bias compared to the CCMV and ACMV cases. On the other hand, the corresponding treatment effects for the last two outcomes, β_2 and β_3, have small bias. With respect to precision, all association parameters have relatively high MSE. In addition, the MSEs for the intercepts α_2 and α_3 are similar in magnitude, whereas those for the treatment effects β_2 and β_3 are considerably smaller. Finally, as was observed for pattern 2, the treatment effect for the third outcome, β_3, is also the most precisely estimated parameter in pattern 1.

Comparing now the results across the various identifying restrictions, for a particular parameter, all three identification schemes yield very similar values for bias and MSE for pattern 2, under both settings, as might be expected, since the three restrictions are in fact equivalent for this pattern. That is, information about the last outcome can only be borrowed from the completers' pattern 3, which is the neighboring pattern, as well as the only pattern for which such information is available. For pattern 1, however, the situation is quite different. Discrepancies in bias can be seen across the various identification schemes for a particular parameter, and these are larger in value than those observed under pattern 2. Looking further into the parameters related to the missing outcomes, under both settings, the smallest bias is observed under the NCMV case for $\alpha_2, \beta_2, \alpha_3, \beta_3$ and $\ln \psi_{12}$, while for $\ln \psi_{13}$ and $\ln \psi_{23}$, the CCMV restriction yields the smallest bias. For $\ln \psi_{123}$, it is the ACMV (NCMV) case that results in the least bias for setting 1 (2). In terms of precision, on the other hand, the

Table 6 *Bias and MSE of the pattern-specific parameter estimates for the trivariate Dale models fitted to the completed data using various identifying restrictions (dropout setting 2).*

Pattern	Parameter	CCMV		ACMV		NCMV	
		Bias	MSE	Bias	MSE	Bias	MSE
1	α_1	0.0023	0.0472	0.0025	0.0474	0.0025	0.0475
	β_1	0.0002	0.0740	-0.0001	0.0741	-0.0001	0.0744
	α_2	0.3107	0.1081	0.2072	0.0574	-0.0436	0.0731
	β_2	-0.0360	0.0195	-0.0225	0.0228	0.0103	0.1073
	α_3	0.3978	0.1719	0.3409	0.1272	0.2918	0.0983
	β_3	-0.0628	0.0228	-0.0619	0.0210	-0.0593	0.0259
	$\ln \psi_{12}$	0.4298	0.2166	0.4070	0.2025	0.1617	0.1381
	$\ln \psi_{13}$	0.3973	0.2196	0.5010	0.2942	0.4192	0.2242
	$\ln \psi_{23}$	0.2726	0.1412	0.3469	0.1681	0.3282	0.1560
	$\ln \psi_{123}$	0.3557	0.3809	0.2063	0.2328	0.2059	0.2291
2	α_1	0.0054	0.0542	0.0051	0.0545	0.0054	0.0542
	β_1	0.0006	0.0839	0.0010	0.0843	0.0006	0.0837
	α_2	-0.0019	0.0566	-0.0011	0.0569	-0.0017	0.0569
	β_2	0.0048	0.0847	0.0034	0.0857	0.0044	0.0857
	α_3	0.3239	0.1182	0.3228	0.1174	0.3223	0.1181
	β_3	-0.0405	0.0262	-0.0391	0.0250	-0.0459	0.0256
	$\ln \psi_{12}$	0.0284	0.0977	0.0286	0.0976	0.0287	0.0976
	$\ln \psi_{13}$	0.3360	0.1665	0.3354	0.1694	0.3390	0.1657
	$\ln \psi_{23}$	0.4087	0.2208	0.3986	0.2159	0.3969	0.2095
	$\ln \psi_{123}$	-0.1231	0.2246	-0.1334	0.2282	-0.1467	0.2555

parameters $\alpha_2, \beta_3, \ln \psi_{13}$ and $\ln \psi_{23}$ seem to be estimated fairly comparably across the identifying restrictions in both settings. The parameter β_2 is most precisely estimated under CCMV in setting 1, and under NCMV in setting 2, while α_3 is most precise under NCMV (both settings). Finally, $\ln \psi_{123}$ is most precise under ACMV (setting 1) and NCMV (setting 2).

To evaluate the effect of the amount of missingness on the fitting procedure, we now compare the results for the two dropout settings within a particular pattern. For pattern 1, the bias and MSEs of the estimates are generally smaller under setting 2, with more missingness, i.e., more subjects within this pattern. For pattern 2, the setting with more missingness shows smaller bias for the α and β parameters, but higher values for the association parameters, and generally smaller MSEs are obtained under this setting.

These results require careful qualification. Let us first weigh in on the results and how they relate to the identifying restrictions. It is clear from the previous observations that no particular identifying restriction consistently led to the most precise estimates. This is understandable as the underlying parameters (Table 1) do not reflect any particular identification scheme. In Section 3.3, we investigate further the performance of the various identifying restrictions by specifically choosing underlying parameters that represent an NCMV setting. Whereas it is entirely possible to explore this situation within the context of a simulation study, the choice of which restriction leads to superior results is more difficult when dealing with actual data, since the true underlying identification pattern is usually unverifiable.

Secondly, let us pay particular attention to the results with respect to the degree of incompleteness in the data. Our observations seem to indicate that, within reason, the amount of missingness does not really pose additional difficulty in applying the pro-

Table 7 *Bias and MSE of the marginalized parameter estimates using approximation (15) of Jansen and Molenberghs (2007) for the pattern-mixture model fitted to the completed data using various identifying restrictions for both dropout settings. (The marginalized intercept and treatment effect for the j^{th} outcome, $j = 1, 2, 3$, are denoted A_{JM_j} and B_{JM_j}, respectively.)*

Setting	Parameter	CCMV		ACMV		NCMV	
		Bias	MSE	Bias	MSE	Bias	MSE
1	A_{JM_1}	0.0109	0.0088	0.0110	0.0088	0.0110	0.0088
	B_{JM_1}	-0.0209	0.0163	-0.0211	0.0164	-0.0210	0.0164
	A_{JM_2}	0.0486	0.0099	0.0414	0.0096	-0.0039	0.0122
	B_{JM_2}	-0.0096	0.0143	-0.0101	0.0154	-0.0023	0.0209
	A_{JM_3}	0.0950	0.0164	0.0891	0.0149	0.0806	0.0141
	B_{JM_3}	-0.0119	0.0156	-0.0150	0.0150	-0.0123	0.0153
2	A_{JM_1}	0.0170	0.0091	0.0170	0.0091	0.0170	0.0091
	B_{JM_1}	-0.0322	0.0182	-0.0322	0.0181	-0.0323	0.0182
	A_{JM_2}	0.0773	0.0128	0.0519	0.0113	-0.0117	0.0154
	B_{JM_2}	-0.0150	0.0147	-0.0121	0.0177	-0.0045	0.0269
	A_{JM_3}	0.1603	0.0325	0.1462	0.0281	0.1338	0.0247
	B_{JM_3}	-0.0261	0.0164	-0.0259	0.0158	-0.0265	0.0171

posed method, at least up to cases with moderate incompleteness (e.g., around 50%). Results showed that better precision is obtained when the data contains more missingness. Initially, this may seem contrary to what might be expected, since dropout setting 2 consists of more missing data, and we would therefore expect worse results under this case. However, it is also necessary to consider the proportions of subjects within each pattern. It is useful to recall that the conditional probabilities used for the imputations are initially estimated from the observed data. When a pattern has very few subjects, as would be the case when there are few dropouts and most subjects are completers, the conditional probabilities for the incomplete patterns may be poorly determined, thus yielding imputations that are probably less reliable. Hogan and Laird (1997) suggested that each pattern needs to be sufficiently "filled," requiring large numbers of dropouts. This is further validated by inspection of the results for the completers (Table 4), which indicate better precision for setting 1, under which pattern 3 has more subjects. Hence, in assessing the effects of the amount of missingness on the pattern-specific estimates, it is essential that the particular pattern is considered, since more missingness in the data may imply less subjects in one pattern (e.g., completers), but more subjects in the other (incomplete) patterns.

3.2.3 Marginalized Effects Estimates

The estimates for the marginalized effects (15) proposed by Jansen and Molenberghs (2007) were also computed from the pattern-specific parameter estimates and these were compared with the true marginalized parameter values of the underlying PMM (Table 3). The results for each identifying restriction and for the two dropout settings are summarized in Table 7. Although scientifically speaking, interest might really be placed on the marginalized effects estimates for the last outcome, i.e., at the end of the series, at which point the treatment presumably does or does not take its desired effect, values for the other outcomes are nevertheless presented here for a more concise evaluation of the marginalization procedure. In general, the magnitudes of the bias are

Table 8 *Bias and MSE of the marginalized parameter estimates using the direct linear approach (13) of Park and Lee (1999) for the pattern-mixture model fitted to the completed data using various identifying restrictions for both dropout settings.(The marginalized intercept and treatment effect for the j^{th} outcome, $j = 1, 2, 3$, are denoted A_{PL_j} and B_{PL_j}, respectively.)*

Setting	Parameter	CCMV		ACMV		NCMV	
		Bias	MSE	Bias	MSE	Bias	MSE
1	A_{PL_1}	0.0122	0.0090	0.0123	0.0090	0.0122	0.0090
	B_{PL_1}	-0.0225	0.0168	-0.0227	0.0169	-0.0226	0.0169
	A_{PL_2}	0.0513	0.0102	0.0438	0.0100	-0.0029	0.0126
	B_{PL_2}	-0.0127	0.0147	-0.0131	0.0159	-0.0035	0.0222
	A_{PL_3}	0.0995	0.0174	0.0930	0.0157	0.0840	0.0148
	B_{PL_3}	-0.0169	0.0160	-0.0197	0.0153	-0.0167	0.0156
2	A_{PL_1}	0.0180	0.0092	0.0180	0.0092	0.0180	0.0092
	B_{PL_1}	-0.0333	0.0186	-0.0334	0.0185	-0.0334	0.0185
	A_{PL_2}	0.0803	0.0134	0.0538	0.0116	-0.0110	0.0158
	B_{PL_2}	-0.0173	0.0151	-0.0139	0.0182	-0.0052	0.0285
	A_{PL_3}	0.1650	0.0341	0.1499	0.0293	0.1372	0.0257
	B_{PL_3}	-0.0302	0.0168	-0.0296	0.0161	-0.0302	0.0175

small, and MSE values seem to indicate that the procedure for marginalization works pleasingly stably.

Comparing across identifying restrictions, under both settings, the bias and MSE are very similar for the marginalized parameters A_{JM_1} and B_{JM_1}, which is reasonable since this parameter relates to the first outcome, which is always observed and never imputed. For A_{JM_2} and B_{JM_2}, within both settings, although NCMV exhibits the least bias for these two parameters, their MSEs are largest under this restriction. For A_{JM_3} and B_{JM_3}, bias and MSE values are comparable across the identifying restrictions and no particular identifying restriction seems to show superiority. In addition, for setting 2, it can be observed that MSE values for A_{JM_3} are somewhat larger than those of the other parameters, implying that this parameter seems to be the least precisely estimated one. With respect to the direction of bias of a particular parameter, although generally consistent across the two settings within a given identifying restriction, the NCMV case differs from CCMV and ACMV for the parameter A_{JM_2}. Finally, it generally seems that the intercept parameters are overestimated, while the treatment effect parameters are underestimated.

Looking now across the two dropout settings, within a particular identifying restriction, bias and MSE values are smaller under setting 1, which has less missingness. Whereas the amount of missingness affects the pattern-specific estimates differently across patterns, its effects on the marginalized estimates are in line with what we would normally expect – that situations with less missing data will tend to show more accurate results.

Table 8 shows the bias and MSE for the marginalized effects estimates using the direct linear approach (13). Magnitudes of bias for almost all parameters are larger for this approach compared to those using the approximation proposed by Jansen and Molenberghs (2007) (in Table 7), as might be predicted, since the former is probably a less appropriate way of marginalizing the pattern-specific estimates. However, for the parameters A_{PL_2} and B_{PL_2}, under the NCMV case in both settings, the direct linear approach actually gives slightly smaller magnitudes of bias. For MSE, on the other

hand, in all cases, Jansen and Molenberghs (2007)'s approximation (15) gives more precise estimates.

3.3 Additional Simulation Results

In order to get deeper insight into the performance of the proposed method, we considered additional simulations. We defined a new underlying PMM which is clearly of an NCMV type. It can be recalled that for the case of 3 outcomes, the identifying restrictions are equivalent for pattern 2, and therefore, the choice of parameters is dictated primarily with reference to pattern 1. Hence, in order to reflect an NCMV setting, we choose for pattern 1 parameters that are more similar to pattern 2 than to pattern 3. The values of these chosen parameters are shown in Table 9. We further used an increased sample size of $N = 4000$ to be able to assess consistency of the previously defined estimates, as well as to be able to freely choose parameters in an NCMV way, without having to encounter samples with incomplete combination levels. With respect to missingness, we considered the same dropout settings as in our primary simulation (Table 2). The underlying SEM-type marginal effects from these new PMMs are given in Table 10. Finally, under these settings, we again generated $S = 500$ samples.

Table 9 *Trivariate Dale model parameter values specified for the three dropout patterns for the underlying pattern-mixture model having NCMV structure.*

Pattern	α_1	β_1	α_2	β_2	α_3	β_3	ψ_{12}	ψ_{13}	ψ_{23}	ψ_{123}
1	0.190	0.096	0.155	0.067	0.142	0.090	1.4	1.1	1.6	1.2
2	0.214	0.115	0.130	0.084	0.142	0.083	1.6	1.2	1.5	1.5
3	0.220	0.150	0.110	0.125	0.170	0.065	2.2	1.8	2.3	1.7

For conciseness' sake, we present only partial results from these new simulations. Inasmuch as the additional simulations exhibited the same results as the primary simulations did, at least as far as the effects of the different dropout settings are concerned, we restrict our presentation of results for these new simulations to those under setting 2, having more missingness. In addition, with respect to comparison of the identifying restrictions, we discuss primarily the results for pattern 1, since the choice of scheme is immaterial in pattern 2, under which they are equivalent. Full results are available from the authors upon request. In what follows, we present the results for the initial estimates, those for the pattern-specific estimates, and then the marginalized effects

Table 10 *True marginal parameter values for the two underlying pattern-mixture models having NCMV structure. (The marginal intercept and marginal treatment effect for the j^{th} outcome, $j = 1, 2, 3$, are denoted A_j and B_j, respectively.)*

PMM	Y_1		Y_2		Y_3	
	A_1	B_1	A_2	B_2	A_3	B_3
1	0.2165	0.1303	0.1163	0.1128	0.1647	0.0681
2	0.2136	0.1193	0.1212	0.1048	0.1607	0.0711

Table 11 *For the underlying pattern-mixture model having NCMV structure and under dropout setting 2, bias and MSE of the parameter estimates for the initial models fitted to the observed data: trivariate Dale model for pattern 3, bivariate Dale model for pattern 2 and logistic model for pattern 1.*

Parameter	Pattern 1		Pattern 2		Pattern 3	
	Bias	MSE	Bias	MSE	Bias	MSE
α_1	0.0005	0.0114	-0.0060	0.0138	0.0027	0.0031
β_1	-0.0056	0.0169	0.0150	0.0245	-0.0063	0.0077
α_2	——	——	0.0039	0.0136	-0.0008	0.0031
β_2	——	——	-0.0082	0.0227	-0.0017	0.0083
α_3	——	——	——	——	-0.0024	0.0028
β_3	——	——	——	——	0.0003	0.0068
$\ln \psi_{12}$	——	——	0.0068	0.0211	-0.0028	0.0073
$\ln \psi_{13}$	——	——	——	——	-0.0067	0.0083
$\ln \psi_{23}$	——	——	——	——	-0.0011	0.0075
$\ln \psi_{123}$	——	——	——	——	-0.0018	0.0333

estimates. Finally, at the end of this section, we also compare the asymptotic variances of the marginalized effects estimates with the variances obtained from the simulation study.

Table 11 summarizes the results for the initial estimates for the NCMV-type PMM under setting 2. Bias for all parameters are quite small and the MSEs are generally smaller than those for the previous simulation (Table 4), demonstrating the consistency of these estimates for the pattern-specific parameters of the initial models. Somewhat larger MSEs are observed for the incomplete patterns, as these consist of fewer subjects than the group of completers (pattern 3). Finally, though the MSEs indicate fairly precise estimation of the initial parameters, it seems that the treatment effects and the associations are less accurately estimated than are the intercepts.

The results for the pattern-specific estimates after imputation of the missing values are given in Table 12. As expected, for any given parameter, bias and MSE are more or less equal across the identification schemes in pattern 2. Moreover, as observed under the previous simulations, parameters involving the missing outcome, namely $\alpha_3, \psi_{13}, \psi_{23}$ and ψ_{123}, exhibit higher bias and MSEs. However, the treatment effect at the last time point, β_3, is quite precisely estimated in comparison with all the other estimates. Moving on the the results for pattern 1, it can be observed that both bias and MSE are smallest under the NCMV restriction for most parameters, except β_3 and ψ_{23}, both of which were most precisely estimated under the CCMV case. For pattern 1, NCMV borrows from pattern 2, but the latter pattern contains no information about these two parameters and hence the actual information propagates from pattern 3 instead. As a result, uncertainly increases and performance worsens. The differences, however, in the bias and MSE between the NCMV and CCMV case for the latter two parameters are much less pronounced than those for the other parameters. We also notice that the bias and MSEs are higher for the association parameters than the α and β parameters, but the smallest values for these are generally still obtained under the NCMV case.

Let us now consider the results for the marginalized effects estimates, which are tabulated in Table 13. We start by a comparison of these results with those for setting 2 in Tables 7 and 8. For both sets of estimates, substantially smaller MSEs are

Table 12 *For the underlying pattern-mixture model having NCMV structure and under dropout setting 2, bias and MSE of the pattern-specific parameter estimates for the trivariate Dale models fitted to the completed data using various identifying restrictions.*

Pattern	Parameter	CCMV		ACMV		NCMV	
		Bias	MSE	Bias	MSE	Bias	MSE
1	α_1	-0.0000	0.0113	0.0007	0.0113	0.0007	0.0113
	β_1	-0.0048	0.0169	-0.0059	0.0168	-0.0060	0.0168
	α_2	0.3086	0.0984	0.2222	0.0528	-0.0239	0.0159
	β_2	-0.0255	0.0055	-0.0178	0.0058	-0.0006	0.0250
	α_3	0.3734	0.1424	0.3330	0.1136	0.2831	0.0832
	β_3	-0.0550	0.0073	-0.0634	0.0085	-0.0669	0.0093
	$\ln \psi_{12}$	0.4351	0.1979	0.3792	0.1523	0.1455	0.0450
	$\ln \psi_{13}$	0.4223	0.1929	0.4653	0.2277	0.4056	0.1761
	$\ln \psi_{23}$	0.2512	0.0782	0.3187	0.1136	0.3039	0.1034
	$\ln \psi_{123}$	0.4761	0.2872	0.3549	0.1728	0.3380	0.1637
2	α_1	-0.0058	0.0137	-0.0057	0.0138	-0.0058	0.0137
	β_1	0.0146	0.0243	0.0145	0.0243	0.0147	0.0243
	α_2	0.0041	0.0135	0.0039	0.0135	0.0040	0.0136
	β_2	-0.0085	0.0225	-0.0082	0.0224	-0.0084	0.0227
	α_3	0.2857	0.0854	0.2853	0.0846	0.2882	0.0868
	β_3	-0.0485	0.0084	-0.0473	0.0081	-0.0507	0.0086
	$\ln \psi_{12}$	0.0068	0.0211	0.0068	0.0211	0.0068	0.0211
	$\ln \psi_{13}$	0.3144	0.1122	0.3113	0.1109	0.3129	0.1108
	$\ln \psi_{23}$	0.3733	0.1523	0.3726	0.1513	0.3713	0.1494
	$\ln \psi_{123}$	0.1187	0.0619	0.1206	0.0642	0.1278	0.0689

consistently observed for all parameters and for all identification schemes under this simulation using $N = 4000$. With respect to bias, under the NCMV case, all parameters exhibit smaller bias compared to the previous simulations with smaller sample size. Thus, at least under the NCMV strategy, both bias and MSE are considerably reduced for all parameters, which is compatible with these estimates' consistency. Under the CCMV and ACMV schemes, almost all parameter estimates yield smaller bias for this setting of increased sample size, with the exception of the estimates for the marginalized intercept for outcome 2, A_{JM_2} or A_{PL_2}. For this parameter, bias actually increased slightly for $N = 4000$ under CCMV and ACMV. Though this might seem contrary to intuition, as an increased sample size usually leads to a less biased estimate, the slight inflation of the bias might be attributed to the use of the incorrect identifying strategy.

In terms of the precision of the marginalized effects estimates relative to each other, it can be observed that the treatment effects generally seem to be more precisely estimated than the intercepts, for outcomes 2 and 3, under most schemes and for both sets of estimates – a result that is quite practical since interest is usually on the treatment effect rather than on the intercept, particulary so on that of the last outcome. The intercept for the last outcome, however, seems to be beset with a substantive amount of bias, and hence, a larger MSE, than the other parameters, for both the Jansen and Molenberghs (2007) and Park and Lee (1999) estimates. In comparison with the corresponding estimates under the simulations with smaller sample size, we see that this bias, though still substantial, has already improved in this setting with the use of a larger sample. However, since this improvement is not much, there would be no reason to believe that it will decrease much further under an even larger sample

Table 13 *For the underlying pattern-mixture model having NCMV structure and under dropout setting 2, bias and MSE of the marginalized parameter estimates using approximation (15) of Jansen and Molenberghs (2007) and using the direct linear approach (13) of Park and Lee (1999) for the pattern-mixture model fitted to the completed data using various identifying restrictions. (The additional subscript $j, j = 1, 2, 3$, is used to identify the outcome number.)*

Parameter	CCMV		ACMV		NCMV	
	Bias	MSE	Bias	MSE	Bias	MSE
A_{JM_1}	-0.0023	0.0021	-0.0021	0.0021	-0.0021	0.0021
B_{JM_1}	0.0086	0.0044	0.0083	0.0044	0.0083	0.0044
A_{JM_2}	0.0820	0.0085	0.0606	0.0057	-0.0019	0.0036
B_{JM_2}	-0.0108	0.0042	-0.0088	0.0045	-0.0047	0.0067
A_{JM_3}	0.1423	0.0217	0.1322	0.0190	0.1203	0.0160
B_{JM_3}	-0.0182	0.0038	-0.0201	0.0040	-0.0217	0.0039
A_{PL_1}	-0.0020	0.0021	-0.0018	0.0021	-0.0018	0.0021
B_{PL_1}	0.0082	0.0044	0.0079	0.0044	0.0080	0.0044
A_{PL_2}	0.0839	0.0088	0.0616	0.0059	-0.0017	0.0036
B_{PL_2}	-0.0116	0.0042	-0.0095	0.0046	-0.0051	0.0068
A_{PL_3}	0.1448	0.0225	0.1343	0.0195	0.1221	0.0164
B_{PL_3}	-0.0193	0.0039	-0.0212	0.0041	-0.0228	0.0040

size. This might be a reflection of our earlier qualification that the marginalized effects estimates may be biased for the SEM-type marginal parameters, particularly when the unconditional pattern proportions are not equal to the conditional ones. On the one hand, one might try to improve the bias by using the conditional pattern probabilities as weights in the proposed estimates (19) of Jansen and Molenberghs (2007). In contrast, if focus does not lie on the marginalized intercept but on the marginalized treatment effect, one may proceed to use the estimates as they are, since the above results seem to indicate that the treatment effects are fairly well estimated by the proposed procedure. Finally, with respect to comparisons across the two sets of estimates, both bias and MSE are slightly higher for the approximation by Park and Lee (1999).

To evaluate the performance of the marginalized effects estimates, we computed the asymptotic variances, as given in (20) and (21), using the underlying parameter values. Variances and/or covariances of the pattern-specific estimates are obtained from the simulation runs, while variances and/or covariances of the pattern proportions are obtained using standard results for the multinomial and/or binomial distributions. For instance, $Var(\widehat{\pi}_t) = \pi_t(1 - \pi_t)/n$. This was then compared with the simulation variance for the corresponding marginalized effects estimate. The results under each identification scheme are shown in Table 14. The very small differences, generally of order $(1 \times e^{-04})$, between the asymptotic and simulation variance demonstrate efficiency of the marginalized effects estimates.

4 Concluding Remarks

In this paper, we have applied, under simulated settings, a procedure for fitting pattern-mixture models to categorical data with monotone missingness via the use of identifying restrictions. Asymptotic variances for marginalized effects estimates, as proposed by Jansen and Molenberghs (2007) and Park and Lee (1999), were derived and the performance of these marginalized effects was subsequently investigated.

Table 14 *Asymptotic variance (Asy) and simulation variance (Sim) of the marginalized parameter estimates for the underlying pattern-mixture model having NCMV structure and under dropout setting 2*

Parameter	CCMV		ACMV		NCMV	
	Asy	Sim	Asy	Sim	Asy	Sim
A_{JM_1}	0.0022	0.0021	0.0021	0.0021	0.0022	0.0021
B_{JM_1}	0.0043	0.0043	0.0043	0.0043	0.0043	0.0043
A_{JM_2}	0.0016	0.0018	0.0016	0.0021	0.0024	0.0036
B_{JM_2}	0.0037	0.0041	0.0037	0.0044	0.0050	0.0067
A_{JM_3}	0.0012	0.0015	0.0012	0.0015	0.0012	0.0015
B_{JM_3}	0.0026	0.0035	0.0026	0.0036	0.0026	0.0035
A_{PL_1}	0.0022	0.0021	0.0022	0.0021	0.0022	0.0021
B_{PL_1}	0.0043	0.0044	0.0043	0.0043	0.0043	0.0043
A_{PL_2}	0.0016	0.0018	0.0016	0.0021	0.0024	0.0036
B_{PL_2}	0.0037	0.0041	0.0037	0.0045	0.0050	0.0068
A_{PL_3}	0.0012	0.0015	0.0012	0.0015	0.0012	0.0014
B_{PL_3}	0.0026	0.0035	0.0026	0.0036	0.0026	0.0034

It was observed that although the (maximum likelihood) estimates for the parameters of the pattern-specific initial models were consistent, precision was contingent on the amount of missingness within the pattern. This sparseness in some patterns may have further consequences on the resulting imputations required for the method, since within these patterns, conditional probabilities on which the imputations are based will probably be less reliable. In such situations, identifying strategies that are based on the more amply filled patterns may prove to be more effective. Similarly, the pattern-specific estimates of the trivariate Dale model parameters were also influenced by the amount of missing data in the particular pattern. However, even for the least filled pattern, the precision for the main parameters (e.g., intercepts and treatment effects) seemed to be quite reasonable. The association parameters, on the other hand, were generally seen to be poorly estimated, but as these are usually regarded as nuisance parameters, such a result is not too alarming. Moreover, these were specified in a simplistic manner – as constants – in our underlying Dale models, but one could always reformulate the Dale model in such a way that these associations are more meaningfully modeled in terms of some known covariates, thereby possibly improving their estimation. Interestingly, the results of the simulations showed that the treatment effect at the last time point, on which scientific interest usually lies, was the most precisely estimated parameter, under any dropout setting or any identification scheme.

With respect to the marginalized effects estimates, missingness again played a fairly important role, with decreased precision in the case of more incomplete data. In addition, the direct linear approach by Park and Lee (1999) yielded slightly more bias and less precision than the proposed estimates of Jansen and Molenberghs (2007). Yet again the primary parameter of interest, i.e., the marginalized treatment effect at the last outcome, was seen to be remarkably stable. The additional simulations allowed us to more objectively assess the behavior of these marginalized effects estimates in terms of the various identifying restrictions. The most favorable results were observed under the restriction that was consistent with the underlying identification scheme, demonstrating that the proposed method and its accompanying marginalization actually can be considered successful, provided the appropriate identification is used. However, the use of an identification scheme that is inconsistent with the underlying

one could induce bias in some of the less important parameters (e.g., intercept at the last time point). This, of course, poses a more difficult issue to deal with, in the sense that the underlying identification pattern is almost always unverifiable. At best, the bias might be reduced by using the conditional pattern proportions in the estimation of the marginalized effects, as opposed to the unconditional ones, whenever obtaining SEM-type marginal effects from a fitted PMM is the goal.

Regarding the choice of our simulation settings, two points deserve mention here. The first is with regard to the sample size, $N = 1000$, which is undoubtedly large for a typical clinical trial. Although not quite realistic except for very large clinical trials, such a sample size is not unusual for survey-type data, in which case the outcome vector would be a truly multivariate one, rather than a single outcome measured longitudinally. The applicability therefore extends beyond longitudinal data. As was already pointed out earlier, the large sample size was necessary to ensure that all combinations were present within each pattern, to allow fitting of the initial and final models; this is a feature, and sometimes a drawback, of the PMM framework, which is indeed a pragmatic concern. Though we were able to get around this issue in our simulation study with the use of a large sample size, one ought not forget that for real data analysis settings, sparsely filled or even empty levels are bound to occur. It is thus relevant to address rather than circumvent this issue. A second point that arises has to do with the type of missingness: monotone in our case. One might argue that such type of missingness is more difficult to define for a multivariate outcome vector (e.g., survey data) than in the longitudinal case, where one can rely on the natural time-ordering of the measurement sequence. However, when the multivariate outcomes within the vector can somehow be ranked, e.g., by order of importance of the variables in the survey, then defining monotone missingness is sometimes an option.

Jansen and Molenberghs (2007) pointed out that the final analysis model can, but does not have to be, equal to the initial model fitted to the observed data. That is, although one fits a pattern-mixture model to the observed data at the first stage, it is entirely possible to fit a selection model after the identification/imputation stage, i.e., after the data have been completed. This, however, raises the issue of "proper" imputation, which Rubin (1987) defines as one where the analysis model is in agreement with the imputation model. This means, broadly, that the imputation model ideally is a super-model of the analysis model. Fitting a selection model at the final stage would therefore pose a conflict in this regard. Thus, we have restricted our final models to pattern-mixture types to avoid so-called improper imputation. The nature of the scientific question, however, would, of course, be an important consideration as well. For instance, if marginal effects are of primary interest, then a selection model would be the natural choice for final analysis model. Hence, the final choice would have to be a balance between the desire for proper imputation and the nature of the scientific question.

Regarding application of the procedure to the non-monotone case, several concerns arise. The first issue has to do with parsimony. In the case of non-monotone missingness, a lot more patterns arise, leading to a proliferation of parameters. In this study, where we considered longitudinal sequences of three time points, only three patterns arise in the monotone case, each employing 10 parameters, and this is only assuming a very simple structure such as, for example, a single treatment indicator and/or association parameters that are not allowed to vary over covariate levels. For the non-monotone case, there will be 8 patterns, and, under the same simple structure we considered, the PMM will consist of a full set of 80 parameters! It is easy to see that increasing

the number of time points, even just slightly, will necessitate estimation of even much larger numbers of parameters. In addition, these patterns should be more or less sufficiently filled in order to get any useful information from them. And, unless the study employs a very large sample size, it is quite difficult to foresee such a situation. In line with this, complete combinations would also be necessary within each pattern, to fit the proposed models without any computational challenges, again requiring large numbers of subjects. On a somewhat different note, difficulties also arise in considering which outcome to impute first. Jansen and Molenberghs (2007) discussed that the formulations they present can be seen as merely a few points in a vast, continuous, design space. To compound the issue, in the non-monotone setting, there are no explicit expressions for ACMV, which might be the more popular choice of identifying restriction as it is the equivalent of MAR a pattern-mixture framework (Molenberghs *et al.*, 1998). For these reasons, we deemed it best to defer focus and discussion of the non-monotone case, perhaps to a separate study on its own, so that such issues can be adequately dealt with and properly discussed therein.

References

Cook, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B* **2**, 133–169.

Glynn, R.J., Laird, N.M., and Rubin, D.B. (1986). Selection modelling versus mixture modelling with non-ignorable nonresponse. In: *Drawing Inferences from Self Selected Samples*, H. Wainer (Ed.). New York: Springer-Verlag, 115–142.

Hogan, J.W. and Laird, N.M. (1997). Mixture models for the join distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239–258.

Jansen, I. and Molenberghs, G. (2007). Pattern-mixture models for categorical outcomes with non-monotone missingness. *Submitted for publication*.

Jansen, I., Molenberghs, G., Aerts, M., Thijs, H., and Van Steen, K. (2003). A Local influence approach applied to binary data from a psychiatric study. *Biometrics* **59**, 410–419.

Kenward, M.G., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53–71.

Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.

Little, R.J.A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483.

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

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data (2nd Ed.)*. New York: John Wiley and Sons.

Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T., and Thijs, H. (2002). Selection models and pattern-mixture models to analyze longitudinal quality of life data subject to dropout. *Statistics in Medicine*, **21**, 1023–1042.

Michiels, B., Molenberghs, G., and Lipsitz, S.R. (1999a). A pattern-mixture odds ratio model for incomplete categorical data. *Communications in Statistics: Theory and Methods* **28**, 2843–2869.

Michiels, B., Molenberghs, G., and Lipsitz, S.R. (1999b). Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics* **55**, 978–983.

Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. New York: John Wiley and Sons.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

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

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

Molenberghs, G. and Lesaffre, E. (1997). Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* **89**, 633-644.

Molenberghs, G., Michiels, B., Kenward, M.G., and Diggle, P.J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica* **52**, 153–161.

Park, T. and Lee, S.Y. (1999). Simple pattern-mixture models for longitudinal analysis with missing observations: analysis of urinary incontinence data. *Statistic in*

Medicine **18**, 2933-2941.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

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

Rubin, D.B. (1978). Multiple imputations in sample surveys – a phenomenological Bayesian approach to nonresponse. In: *Imputation and Editing of Faulty or Missing Survey Data.* Washington, DC: U.S. Department of Commerce, pp. 1–23.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley and Sons.

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3**, 245–265.

Welsh, A.H. (1996). *Aspects of Statistical Inference.* New York: Wiley.

Wu, M.C. and Bailey, K.R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939–955.

Wu, M.C. and Carrol, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics* **44**, 175–188.

Verbeke, G., Lesaffre, E. and Spiessens, B. (2001). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal* **35**, 419–434.