

Pattern-mixture models for categorical outcomes with non-monotone missingness

Peer-reviewed author version

JANSEN, Ivy & MOLENBERGHS, Geert (2010) Pattern-mixture models for categorical outcomes with non-monotone missingness. In: JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 80 (11). p. 1279-1296.

DOI: 10.1080/00949650903062566

Handle: <http://hdl.handle.net/1942/11408>

# Pattern-Mixture Models for Categorical Outcomes with Non-Monotone Missingness

**Ivy Jansen, Geert Molenberghs**

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

*Email: ivy.jansen@uhasselt.be*

## SUMMARY

Whereas most models for incomplete longitudinal data are formulated within the selection model framework, pattern-mixture models have gained considerable interest in recent years (Little, 1993, 1994), since it is often argued that selection models, although many are identifiable, should be approached with caution, especially in the context of MNAR models (Glynn, Laird and Rubin, 1986). In this paper, focus is on several strategies to fit pattern-mixture models for non-monotone categorical outcomes. The issue of under-identification in pattern-mixture models is addressed through identifying restrictions. Attention will be given to the derivation of the marginal covariate effect in pattern-mixture models for non-monotone categorical data, which is less straightforward than in the case of linear models for continuous data. The techniques developed will be used to analyze data from a clinical study in psychiatry.

*Key words:* Categorical Data; Identifying Restrictions; Multivariate Dale Model; Non-monotone Missingness; Pattern Mixture Models.

## 1 Introduction

A vast number of studies collect data longitudinally. In such studies, measurement sequences are prone to incompleteness, an issue requiring attention. A model for incomplete data starts from the joint distribution of the outcomes,  $\mathbf{Y}$  say, and the non-response process,  $\mathbf{R}$  say. This joint distribution  $f(\mathbf{y}, \mathbf{r} | \boldsymbol{\theta}, \boldsymbol{\psi})$  can be factorized in several ways. A selection model is based on the factorization  $f(\mathbf{y} | \boldsymbol{\theta})f(\mathbf{r} | \mathbf{y}, \boldsymbol{\psi})$ , whereas the reverse factorization  $f(\mathbf{y} | \mathbf{r}, \boldsymbol{\theta})f(\mathbf{r} | \boldsymbol{\psi})$  is referred to as

a pattern-mixture model (Little, 1993, 1994). When a common set of random-effects is thought to influence both the  $\mathbf{Y}$  and  $\mathbf{R}$  processes, conditional upon which they are independent, then the so introduced model is referred to as a shared-parameter model. For reviews, see Kenward and Molenberghs (1998); Little (1995); Wu and Carroll (1988). The non-response process  $\mathbf{R}$  can either be monotone, also called dropout, or non-monotone when there are intermittent missing values. For each of these processes different modeling strategies are needed. For details see 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 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 analysis with those from 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,b).

An important issue is that pattern-mixture models are by construction under-identified. Little (1993, 1994) solved this problem through the use of identifying restrictions: inestimable parameters of the incomplete patterns are set equal to (functions of) the parameters describing the distribution of the completers. In this way, the conditional distribution of the unobserved measurements, given the observed ones in a specific pattern, is identified. Molenberghs *et al.* (1998) proposed a particular set of restrictions for the monotone case which corresponds to MAR and in Thijs *et al.* (2002) a formal way for how to handle this kind of restrictions is introduced. Alternatively, several types of simplified (identified) models can be considered. The advantage is that the number of parameters decreases, which is generally an issue with pattern-mixture models. Hogan and Laird (1997) noted that in order to estimate the large number of parameters in general pattern-mixture models, one has to make the awkward requirement that each dropout pattern is sufficiently “filled”, in other

words, one has to require large numbers of dropouts. This problem is less prominent in simplified models. Note, however, that simplified models, qualified as “assumption rich” by Sheiner, Beal and Dunne (1997), are also making untestable assumptions and therefore reiterate that no models are able to circumvent the need for making assumptions; pattern-mixture models are no exception. Notwithstanding this, pattern-mixture models render the need for making assumptions and the implications thereof more obvious, since it is clear that for some components there is no information.

On a parallel research track, categorical data modeling has received a lot of attention during the past decades (Agresti, 2002). More recently, quite a bit of attention has been devoted to repeated categorical data (Diggle *et al.*, 2002; Fahrmeir and Tutz, 1994; Molenberghs and Verbeke, 2005). Combining both strands of research, methods have been developed to analyze non-monotone missing categorical data (Jansen *et al.*, 2003; Jansen and Molenberghs, 2008). However, these models all belong to the selection model framework. Pattern-mixture models for monotone missing categorical outcomes have been treated by Michiels, Molenberghs and Lipsitz (1999a,b). In this paper, focus will be on pattern-mixture models to analyze non-monotone missing categorical data. One of the major differences with work for continuous data, such as in Thijs *et al.* (2002) is that, owing to the presence of a non-linear link function in most models commonly considered, unlike in the normal case, marginalization of pattern-specific effects is less than straightforward. In particular, such marginalizations do not merely follow as simple weighted averages of the pattern-specific parameters.

However, the most important difference is between the monotone and non-monotone cases. Indeed, Diggle and Kenward (1994) and Molenberghs, Kenward and Lesaffre (1997) were able to exploit monotonicity through the use of iterated logistic regressions for the dropout model. One important contribution of this paper is that we use a Dale model, a model for multivariate categorical data (Molenberghs and Lesaffre, 1999), not only for the outcomes of interest, but also for the missingness

model.

The rest of the paper is organized as follows. In Section 2, the general context of pattern-mixture models will be sketched, together with the strategy of identifying restrictions. Section 3 gives attention to the use of the multivariate Dale model to fit pattern-mixture models for categorical outcomes, while Section 4 will discuss the assumptions needed when intermittent missingness is present. Section 5 focuses on the derivation of marginal effects in pattern-mixture models for categorical outcomes. Finally, in Section 6, the techniques developed are applied to data from a psychiatric study.

## 2 Pattern-Mixture Modeling in the Monotone Case

The family of pattern-mixture models is based on the factorization

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

where dependence on covariates is suppressed from notation. Thus, the conditional density of the measurements given the missingness process is combined with the marginal density describing the missingness mechanism. Both factors can depend on covariates. It is, of course, possible to have different covariate dependencies in either components of the factorization.

Pattern-mixture have received quite a bit of attention (Rubin, 1987; Little, 1993, 1994). There are three *steps* in the analysis procedure: (a) the *initial* model will be fitted to the observed data within each of the patterns; (b) a process of *identification* (or data augmentation) takes place to identify the distribution of the unobserved measurements, given the observed ones; (c) a model fitted to the so-augmented data. Fitting such models can be conveniently done using multiple imputation. Then, the identification step corresponds to the so-called *imputation task*. The final model corresponds to the MI *analysis task*. Finally, combining the inferences from the, say,  $M$ , imputations into a

single one is the *inference task*. Details can be found in Rubin (1987); Schafer (1997); Verbeke and Molenberghs (2000).

The choice of final model is done based on carefully weighing the target of inference and the nature of the multiple imputation procedure. Regarding the latter, Rubin (1987) refers to proper imputation 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. Regarding the former point, if one is interested in an overall (treatment) effect, it would be natural to use an SEM. However, this may then clash with the desire for proper imputation, in which case considering a PMM as well for the analysis task might be considered wise, even when interest lies in a marginal effect. In this case, the target effect needs to be obtained by marginalizing the PMM. The latter is more challenging with nonlinear models, including generalized linear models; only the linear case allows for forthright marginalization. Additional reasons to choose for PMM in the analysis include (1) as a sensitivity analysis viz. the direct use of SEM; (2) as a response to questions that are inherently of a PMM type, such as the wish to study differences between, say, completers and early dropouts. In summary, one needs to, and has the ability to, balance the desire for proper imputation with the nature of the scientific question.

Pattern-mixture modeling for the general monotone case is summarized in Molenberghs and Verbeke (2005). In the next section, we will focus on binary data and monotone and non-monotone missingness, respectively.

### **3 Pattern-Mixture Models for Categorical Outcomes in the Monotone Case**

Let us focus on the case of three binary measurements. Extension to more than three 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 identifiable densities. For the completers (pattern 3), a trivariate Dale model will be used, for pattern 2 a bivariate Dale model, and a univariate Dale model (which is equal to conventional logistic regression) for pattern 1. We will term this the *minimal approach*. 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 measurements, i.e., for the group of completers, this results in the following logistic-regression and odds-ratio formulations (subject-specific indices  $i$  are removed for the ease of notation, as is an index to pattern 3 that should be present, strictly speaking, in the  $\eta$ ,  $p$ , and  $X$  functions):

$$\begin{aligned}
\eta_1 &= \ln \left( \frac{p_{1++}}{1 - p_{1++}} \right) = X_1 \boldsymbol{\theta}, \\
\eta_2 &= \ln \left( \frac{p_{+1+}}{1 - p_{+1+}} \right) = X_2 \boldsymbol{\theta}, \\
\eta_3 &= \ln \left( \frac{p_{++1}}{1 - p_{++1}} \right) = X_3 \boldsymbol{\theta}, \\
\eta_4 &= \ln \varphi_{12} = \ln \left( \frac{p_{11+}(1 - p_{1++} - p_{+1+} + p_{11+})}{(p_{1++} - p_{11+})(p_{+1+} - p_{11+})} \right) = X_4 \boldsymbol{\theta}, \\
\eta_5 &= \ln \varphi_{13} = \ln \left( \frac{p_{1+1}(1 - p_{1++} - p_{++1} + p_{1+1})}{(p_{1++} - p_{1+1})(p_{++1} - p_{1+1})} \right) = X_5 \boldsymbol{\theta}, \\
\eta_6 &= \ln \varphi_{23} = \ln \left( \frac{p_{+11}(1 - p_{++1} - p_{+1+} + p_{+11})}{(p_{++1} - p_{+11})(p_{+1+} - p_{+11})} \right) = X_6 \boldsymbol{\theta}, \\
\eta_7 &= \ln \varphi_{123} = \ln \left( \frac{p_{111}p_{122}p_{212}p_{221}}{p_{112}p_{121}p_{211}p_{222}} \right) = X_7 \boldsymbol{\theta},
\end{aligned}$$

with  $p_{ijk} = P(Y_1 = i, Y_2 = j, Y_3 = k)$ ,  $(i, j, k = 1, 2)$ , and a  $+$  in lieu of a subscript indicating that the marginal probability over this index needs to be used. Therefore, the incomplete patterns provide information neither about the unobserved outcomes nor about the associations involving those unobserved outcomes. Thus, for pattern 2, an analogous model only involving  $\eta_1$ ,  $\eta_2$ , and  $\eta_4$  can be obtained from the data, while for pattern 1 only  $\eta_1$  will be available. Of course, the  $\eta$  functions are specific to the 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 the 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 underlying probabilities  $p_{y_1 y_2 y_3|3} = P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3 | \text{pattern} = 3)$ ,  $p_{y_1 y_2|2} = P(Y_1 = y_1, Y_2 = y_2 | \text{pattern} = 2)$  and  $p_{y_1|1} = P(Y_1 = y_1 | \text{pattern} = 1)$  can be estimated. The corresponding counts are indicated by the symbol  $Z$ . 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  $p_{y_3|y_1 y_2,3} = P(Y_3 = y_3 | Y_1 = y_1, Y_2 = y_2, \text{pattern} = 3)$  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. 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 (??), the joint probability of  $y_1$ ,  $y_2$  and  $y_3$  in pattern 1 can be written as

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

where specific choices of  $\omega$  lead to three sets of identifying restrictions: (a) *complete case missing*



values, *CCMV*, where identification is always done from the completers' pattern; (b) *neighboring case missing values*, *NCMV*, where the pattern with one more observation is used for identification; and (c) *available case missing values*, *ACMV*, where a particular linear combination of the information coming from all patterns with strictly more measurements taken is used:

$$\text{CCMV} : p_{y_1|1} p_{y_2|y_1,3} p_{y_3|y_1y_2,3},$$

$$\text{NCMV} : p_{y_1|1} p_{y_2|y_1,2} p_{y_3|y_1y_2,3},$$

$$\text{ACMV} : \omega = \frac{\pi_2 p_{y_1|2}}{\pi_2 p_{y_1|2} + \pi_3 p_{y_1|3}},$$

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

$$\text{CCMV} : Z_{y_1y_2y_3|1}^* = Z_{y_1|1} p_{y_2|y_1,3} p_{y_3|y_1y_2,3}, \quad (3.1)$$

$$\text{NCMV} : Z_{y_1y_2y_3|1}^* = Z_{y_1|1} p_{y_2|y_1,2} p_{y_3|y_1y_2,3}, \quad (3.2)$$

$$\text{ACMV} : Z_{y_1y_2y_3|1}^* = Z_{y_1|1} \left[ \frac{\pi_2 p_{y_1y_2|2} + \pi_3 p_{y_1y_2|3}}{\pi_2 p_{y_1|2} + \pi_3 p_{y_1|3}} \right] p_{y_3|y_1y_2,3}. \quad (3.3)$$

As shown by Molenberghs *et al.* (1998), ACVM is the PMM counterpart to MAR.

Multiple imputations are then generated by drawing uniformly from Bernoulli variables with the probabilities embedded in (3.1)–(3.3). As stated earlier, once the imputations have been generated, the analysis task model (final model) can be fitted and multiple-imputation inference conducted. Using conventional multiple-imputation machinery, obtaining parameter and precision estimates is straightforward. In particular, the asymptotic covariance matrix of the form

$$\mathbf{V} = \mathbf{W} + \left( \frac{M+1}{M} \right) \mathbf{B}, \quad (3.4)$$

where  $\mathbf{W}$  is the average *within*-imputation variance,  $\mathbf{B}$  the *between*-imputation variance, and  $M$  the number of imputations (Rubin, 1987).

## 4 Assumptions Needed for the Intermittent Missingness Case

We will extend the above strategy to the situation of non-monotone missing data. It is therefore useful to label the patterns by a triple-index system, The first three patterns are the monotone ones, which have been discussed already in Section 4. Pattern 3 becomes 111, the fully observed pattern, which does not need any imputation. Patterns 1 and 2, relabeled as 100 and 110, respectively, will be considered again in this section, since many more possibilities will be available now to impute the unobserved data.

Let us first consider the patterns for which only one measurement is missing, namely patterns 110, 101, and 011, where the third, the second and the first outcome, respectively, are unobserved. A bivariate Dale model can be used to fit the observed data densities  $f_{110}(y_1, y_2)$ ,  $f_{101}(y_1, y_3)$  and  $f_{011}(y_2, y_3)$ . Since it is recommended to use as much of the available data as possible to impute the conditional distributions of the unobserved outcomes, given the observed ones, we can only use information from pattern 111 for imputation. This results in the following complete data densities:

$$\begin{aligned} f_{110}(y_1, y_2, y_3) &= f_{110}(y_1, y_2)f_{111}(y_3|y_1, y_2), \\ f_{101}(y_1, y_2, y_3) &= f_{101}(y_1, y_3)f_{111}(y_2|y_1, y_3), \\ f_{011}(y_1, y_2, y_3) &= f_{011}(y_2, y_3)f_{111}(y_1|y_2, y_3). \end{aligned}$$

Next, patterns 100, 010, and 001 will be discussed. Here, only one out of the three outcomes is measured, and a univariate Dale model can be used to obtain  $f_{100}(y_1)$ ,  $f_{010}(y_2)$ , and  $f_{001}(y_3)$ . First, we have to decide which of the two unobserved outcomes will be imputed first. In the case of monotone missingness, the obvious choice for pattern 100 was to first impute  $y_2$  and then  $y_3$ . In the case of non-monotone missingness, there is no such obvious choice. Therefore, we will consider both possibilities simultaneously. For pattern 001, for example, we can first consider the conditional

density of  $y_1$ , given  $y_3$ . Information on this density can be borrowed from either the completers (pattern 111) or the neighbors (pattern 101), or a combination of both densities. The conditional density of  $y_2$ , given  $y_1$  and  $y_3$ , can only be borrowed from pattern 111. Similarly, the conditional density of  $y_2$ , given  $y_3$ , can be obtained first, using one of the available identifying restrictions, and afterwards the conditional density of  $y_1$ , given  $y_2$  and  $y_3$ . In fact, these choices are merely two options out of a continuum, where all information available is being used. This produces the following complete-data densitites for patterns 100, 010, and 001:

$$\begin{aligned} f_{100}(y_1, y_2, y_3) = & f_{100}(y_1) \times \{\omega_{100,1}f_{110}(y_2|y_1)f_{111}(y_3|y_1, y_2) \\ & + \omega_{100,2}f_{101}(y_3|y_1)f_{111}(y_2|y_1, y_3) \\ & + \omega_{100,3}f_{111}(y_2, y_3|y_1)\}, \end{aligned}$$

$$\begin{aligned} f_{010}(y_1, y_2, y_3) = & f_{010}(y_2) \times \{\omega_{010,1}f_{110}(y_1|y_2)f_{111}(y_3|y_1, y_2) \\ & + \omega_{010,2}f_{011}(y_3|y_2)f_{111}(y_1|y_2, y_3) \\ & + \omega_{010,3}f_{111}(y_1, y_3|y_2)\}, \end{aligned}$$

$$\begin{aligned} f_{001}(y_1, y_2, y_3) = & f_{001}(y_3) \times \{\omega_{001,1}f_{101}(y_1|y_3)f_{111}(y_2|y_1, y_3) \\ & + \omega_{001,2}f_{011}(y_2|y_3)f_{111}(y_1|y_2, y_3) \\ & + \omega_{001,3}f_{111}(y_1, y_2|y_3)\}. \end{aligned}$$

Now, CCMV results from setting  $\omega_{100,3} = \omega_{010,3} = \omega_{001,3} = 1$ , with the others equal to zero. NCMV corresponds to all situations where  $\omega_{100,3} = \omega_{010,3} = \omega_{001,3} = 0$ . In the non-monotone case, an explicit expression for ACMV is not straightforward.

Note that there is an additional pattern 000, corresponding to no follow-up information. When no covariate information is available, it follows that this pattern provides no information and would

effectively cancel out. When covariates are observed for pattern 000, this corresponds to having more outcome variables, in which case the pattern is not really empty.

## 5 Marginal Effects Across Patterns

In Section 2 we asserted that in the analysis task there is room for choice. For example, one can have an interest in a marginal effect and therefore opt for an SEM. Alternatively, scientific interest can be pattern-specific, in which case a PMM is a more natural choice. In neither case is it necessary to marginalize a PMM. However, when a strong emphasis is put on proper imputation, or when one considers the PMM by way of sensitivity assessment for the SEM, then it is necessary to marginalize the PMM. In the case of continuous data, where linear models are used, the overall effect is simply a weighted average of the pattern-specific effects. We will show that this is not true for categorical data. We therefore assume that the logistic regression

$$P(Y_{ij} = 1 | \text{pattern } k) = \frac{e^{\alpha_k + \beta_k T_i}}{1 + e^{\alpha_k + \beta_k T_i}},$$

where  $i$  refers to subject and  $j$  to measurement occasion, is used to model the data from pattern  $k = 1, \dots, K$  (as in the multivariate Dale model).  $\alpha$  and  $\beta$  can depend on  $j$ , but we suppress this index from notation.

Assume interest is in one particular effect  $T$ , e.g., treatment effect at the last occasion, and assume  $\pi_k$  to be the pattern probability as defined before. The marginal success probability is then equal to

$$\sum_{k=1}^K \pi_k \frac{e^{\alpha_k + \beta_k T}}{1 + e^{\alpha_k + \beta_k T}}. \quad (5.1)$$

There are three ways to calculate the marginal treatment effect at the last occasion from this. First,

the direct linear approach (Park and Lee, 1999) can be used, where

$$\beta \simeq \sum_k \pi_k \beta_k, \quad (5.2)$$

but this is not always correct. It is a correct approach: (a) when data are continuous and linear, usually Gaussian, methods are used; (b) when models for non-Gaussian data are used but with an identity link; and (c) locally in a number of general situations, such as with logistic and probit links, in the vicinity of the probability of 0.5.

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 (GLMM), etc. Third, classical averaging can be performed. To this effect, keep function (5.1) as is and compute and graph, or sample. Note that averaging in this way will be similar to the marginalization of random-effects models (e.g., GLMM to GEE). Here, the marginalization is over pattern, rather than over random effects. When a GLMM is used in each pattern, then there is a double marginalization, one over the random effects and one over the patterns. It is our view that both the second and third approach are broadly useful. The second one has the advantage of providing parametric functions for the marginal effect, at the expense of approximation and the need for additional calculation. The third approach immediately follows, but then does not lead to specific parameters, such as, for example, a marginal treatment effect.

We will now discuss and illustrate the the second approach in more detail, using a marginal logistic model.

Let us approximate (5.1) by a logistic regression, with obvious notation:

$$f(T) = \sum_k \pi_k \frac{e^{\alpha_k + \beta_k T}}{1 + e^{\alpha_k + \beta_k T}} \cong \frac{e^{A+BT}}{1 + e^{A+BT}}. \quad (5.3)$$

Then, the logit of  $f(T)$  can be approximated by

$$F(T) = \text{logit}(f(T)) \cong A + BT.$$

Using a first order Taylor expansion results in

$$F(0) + \left. \frac{\partial F}{\partial T} \right|_{T=0} T \cong A + BT,$$

such that

$$A \simeq F(T = 0) = \text{logit} \left( \sum_k \pi_k \frac{e^{\alpha_k}}{1 + e^{\alpha_k}} \right).$$

It is easily shown that

$$\frac{\partial \text{logit}(x)}{\partial x} = \frac{1}{x(1-x)}$$

and

$$\frac{\partial f}{\partial T} = \sum_k \pi_k \frac{(e^{\alpha_k + \beta_k T}) \beta_k}{(1 + e^{\alpha_k + \beta_k T})^2},$$

such that

$$\left. \frac{\partial F}{\partial T} \right|_{T=0} = \frac{1}{\sum_k \pi_k \frac{e^{\alpha_k}}{1 + e^{\alpha_k}}} \frac{1}{\sum_k \pi_k \frac{1}{1 + e^{\alpha_k}}} \sum_k \beta_k \pi_k \frac{e^{\alpha_k}}{(1 + e^{\alpha_k})^2},$$

and equivalently

$$B \simeq \frac{\sum_k \beta_k \pi_k \frac{e^{\alpha_k}}{1 + e^{\alpha_k}} \frac{1}{1 + e^{\alpha_k}}}{\left( \sum_k \pi_k \frac{e^{\alpha_k}}{1 + e^{\alpha_k}} \right) \left( \sum_k \pi_k \frac{1}{1 + e^{\alpha_k}} \right)}.$$

Let  $P_k = \frac{e^{\alpha_k}}{1 + e^{\alpha_k}}$ , then the approximate marginalized treatment effect can be estimated using

$$B \simeq \frac{\sum_k \beta_k \pi_k P_k (1 - P_k)}{(\sum_k \pi_k P_k) [\sum_k \pi_k (1 - P_k)]}. \quad (5.4)$$

Note that direct expansion of (5.3), without taking the logit first, leads to exactly the same expression.

Let us now consider the special case where the treatment effect is the same in each pattern ( $\beta_k = \beta, \forall k$ ), then

$$B \simeq \beta \frac{\sum_k \pi_k P_k (1 - P_k)}{(\sum_k \pi_k P_k) (\sum_k \pi_k (1 - P_k))}, \quad (5.5)$$

such that

$$|B| \leq |\beta|, \quad (5.6)$$

which follows from Jensen's inequality, given that (5.5) is of the form  $B \simeq \beta E h(p) / h(E(p))$ .

This means that the marginal treatment effect at the last occasion, obtained through approximation (5.3), will not be larger in absolute value than the marginal treatment effect, obtained from the direct linear approach (5.2), when the treatment effects are equal across patterns.

Marginalization when the  $\beta_k$ 's are different, may both increase and decrease the effect, in absolute value. Let us consider the example of two patterns ( $K = 2$ ). Set  $\pi_1 = \pi$ ,  $\pi_2 = 1 - \pi$ ,  $\beta_1 = 1$  and  $\beta_2 = \rho$ . Expressions (5.2) and (5.4) then reduce to

$$\pi + (1 - \pi)\rho \quad \text{and} \quad \frac{\pi P_1(1 - P_1) + \rho(1 - \pi)P_2(1 - P_2)}{[\pi P_1 + (1 - \pi)P_2][\pi(1 - P_1) + (1 - \pi)(1 - P_2)]}. \quad (5.7)$$

Let  $N = [\pi P_1 + (1 - \pi)P_2][\pi(1 - P_1) + (1 - \pi)(1 - P_2)]$ . Choosing  $\rho$  such that the equality between both expressions in (5.7) holds, results in

$$\rho = \frac{\pi[N - P_1(1 - P_1)]}{(1 - \pi)[P_2(1 - P_2) - N]}. \quad (5.8)$$

Since  $\rho \in \mathbb{R}$ , setting  $\rho$  equal to this value is sufficient to have both equations equal.  $\rho + \varepsilon$  and  $\rho - \varepsilon$  will then make the relative positions of both quantities go either way.

If  $P_1 = P_2 = P$  then the right hand side expression in (5.7) reduces to

$$\frac{\pi P(1 - P) + \rho(1 - \pi)P(1 - P)}{P[\pi + (1 - \pi)](1 - P)[\pi + (1 - \pi)]},$$

which is equal to  $\pi + (1 - \pi)\rho$ , and hence, for all  $\rho$ , both expressions in (5.7) are the same. Note also that then in (5.8) the numerator and denominator are both equal to zero, confirming that the result applies to every  $\rho$ . Thus, the difference emerges from a difference in background success probability  $P_k$ . Deriving the sign of  $\rho$  is possible but tedious and deferred to the Appendix.

Given all of these considerations, it is clear that determining a marginal effect across patterns in the case of non-Gaussian data, is less straightforward than in the Gaussian case. One should bear in mind that the direct linear approach (Park and Lee, 1999) is invalid in the case of categorical data, and that this method can neither be considered to be conservative nor liberal.

## 6 The Fluvoxamine Data

These data come from a multicenter, postmarketing study involving 315 patients that were treated by fluvoxamine for psychiatric symptoms described as possibly resulting from a dysregulation of serotonin in the brain. The data are discussed in Molenberghs and Lesaffre (1994), Kenward, Lesaffre and Molenberghs (1994), Molenberghs, Kenward and Lesaffre (1997), Michiels and Molenberghs (1997), Molenberghs *et al.* (1999), and Jansen *et al.* (2003).

After enrollment into the study, a number of baseline characteristics were scored, and the patient was assessed at four follow-up visits. The therapeutic effect and the extent of worsening side effects were scored at each visit on an ordinal scale. A side effect occurs if new symptoms appear while there is therapeutic effect if old symptoms disappear. We will focus on a dichotomized version (present/absent) of side effects. The first, second, and last visit will be considered, because three time points is the smallest number that allows to usefully illustrate all similarities and differences between various choices.

Note that a number of prior analyzes have been done, many summarized in Molenberghs and Verbeke (2005) and Molenberghs and Kenward (2007). Among others, Markov models have been fitted to a dichotomized version of the outcomes, as well as a Dale model, combined with logistic regression to describe dropout, ignoring the non-monotone patterns. Interestingly, a pattern-mixture model had been fitted to a dichotomized version of the data, combined with multiple imputation, and assuming dropout. At the analysis stage, an SEM was compared with a PMM and noticeable differences were



found, underscoring the usefulness in disposing of choice between both. Also a number of sensitivity analysis tools were applied. We refer the reader to these references for more detail.

We will account for patients' gender (0 = males, 1 = females), a covariate recorded for all. There are 224 completers (pattern 111), 44 patients missed the last visit (pattern 110), 31 only appeared at the first visit (pattern 100), 1 person belongs to pattern 011, 1 to pattern 001, and the remaining 14 patients do not have any observations at all (pattern 000). For those 14 patients, there is no reasonable way to impute the missing outcomes, and therefore, they will not be considered for analyses. Pattern 011 and pattern 001 both only contain 1 patient (0.33% of the total number of subjects in the study), so their effect on the results can be ignored. This leaves 299 patients in the study. The data are summarized in Table 1.

As described in Section 3, we start by fitting a trivariate Dale model to the completers, a bivariate Dale model to pattern 110, and a logistic regression to pattern 100. Then, an identifying restriction is chosen to define the conditional distributions of the unobserved outcomes, given the observed ones. Thereafter, we draw multiple imputations ( $M = 10$ ). We thus obtain, for each choice of identifying restriction strategy, ten multiply-imputed sets of data, which can then be analyzed, using several possible models.

Let us first discuss the results reported in Table 2. A single trivariate Dale model is fitted, producing a SEM, with a log odds ratio independent of covariate effects, for each of the possible associations between outcomes, and a possible effect of gender on the marginal probabilities. We notice that the estimates for the association parameters are very close under the three possible identifying restrictions. The associations  $\varphi_{12}$ ,  $\varphi_{13}$  and  $\varphi_{23}$  are highly significant ( $p < 0.0001$ ), while  $\varphi_{123}$  is borderline significant ( $p \approx 0.045$ ). Also, the estimates for the first marginal probability are almost equal under CCMV, NCMV, and ACMV. This was to be expected, since the first outcome was observed for all subjects that were included for analysis. The parameter estimates for the logistic

regression of the third marginal probability are also quite similar. This is due to the fact that all identifying restrictions implied the same conditional density for the third outcome, given the first and second ones, namely to borrow it from the completers. The small difference that is observed nevertheless, results from a difference in imputation for the second outcome, since the imputation of the third outcome is conditional on the second one. And as we can see, the estimates for the second marginal probability differ much between the three identifying restrictions. The CCMV and NCMV estimates, for the intercept as well as for gender, are lying furthest apart. ACMV estimates are closer to CCMV estimates, since many more completers are available than neighbors, thus  $\omega$  will be smaller than 0.5. Finally, we will contemplate the effect of gender. We observe that the estimate is negative for the first marginal probability, approximately zero for the second one, and positive for the third one, meaning that the probability of no side effects is larger, equal or smaller for males than for females, for the first, second and last measurement occasion, respectively. However, the effect of gender on the marginal probabilities is not significant.

Next, a more extended trivariate Dale model is presented in Table 3. Parameter estimates, specific to each pattern, are supplemented with marginalized parameters added. Now, pattern-specific intercepts are allowed in the logistic regressions for the marginal probabilities. The gender effect is assumed to be the same for all patterns, and the associations between outcomes are still independent of covariate effects. The parameter `intercepti` is the intercept in the logistic regression for the  $i$ th marginal probability for pattern 111. `pattern1i` and `pattern2i` are dummy variables, such that they correspond to the difference in intercept between pattern 3 and pattern 1 or pattern 2, respectively. For the first marginal probability there is no significant difference between the pattern-specific intercepts. Only in the NCMV case, a borderline non-significant difference ( $p \approx 0.077$ ) is observed between pattern 100 and pattern 111. We notice that the intercept for pattern 111 is higher than for the other patterns, resulting in a higher probability of no side effects at the first measurement

occasion for the completers. Similar conclusions can be found for the second and last occasions. Taking a closer look at the results for the second marginal probability, we notice that the intercepts for pattern 110 and 111 are significantly different ( $p \approx 0.035$ ), while patterns 100 and 111 are only borderline significantly different ( $p \approx 0.05$ ) when NCMV is used, and not significant when CCMV or ACMV are used. This can be explained by the fact that for NCMV, pattern 1 borrows all information from pattern 110 and thus takes distance from pattern 111, while under CCMV and ACMV all, or most, of the information is borrowed from pattern 111, and therefore there is only little distance between pattern 100 and 111. For the third marginal probability, there is no significant difference between the three patterns for all identifying restrictions, since the missing information is always identified from pattern 111. CCMV, NCMV, and ACMV lead to almost the same estimates for all parameters concerning the third marginal probability. Finally, the effect of gender changes over the different measurement occasions as before, and again its effect on the marginal probabilities is not significant. The associations  $\varphi_{12}$ ,  $\varphi_{13}$ , and  $\varphi_{23}$  are highly significant ( $p < 0.0001$ ), while  $\varphi_{123}$  is now borderline significant ( $p \approx 0.045$ ) only for NCMV, and borderline non-significant ( $p \approx 0.064$ ) for CCMV and ACMV.

Third, Table 4 contains parameter estimates of a trivariate Dale model where now not only the intercept, but also the gender effect is allowed to be different in the three patterns. Also here `intercepti` corresponds to the effect of pattern 111, while the dummy variables `pattern1i` and `pattern2i` model the difference in success probability between pattern 111 and pattern 100 or 110, respectively. `genderi` represents the gender effect in pattern 111, while the interactions between the dummies and gender refer to the difference in gender effect between pattern 3 and pattern 100 or 110, respectively. The parameter estimates for the logistic regression of  $p_{1++}$  reveal the following results. The probability of no side effects is borderline significantly different ( $p \approx 0.05$ ) between pattern 110 and 111, but not significantly different between pattern 100 and 111. Gender is borderline

non-significant ( $p \approx 0.058$ ) in pattern 111, and a borderline non-significant different gender effect occurred between pattern 2 and 3. For  $p_{+1+}$ , similar conclusions are reached, but now the difference in probability of no side effects is highly significant ( $p \approx 0.006$ ) between pattern 110 and 111, and under NCMV borderline significant ( $p \approx 0.05$ ) between pattern 100 and 3. The gender effect in pattern 111 is not significant anymore. Finally, the success probability  $p_{++1}$  is not different in the three patterns, and the gender effect is not significant. The associations  $\varphi_{12}$ ,  $\varphi_{13}$  and  $\varphi_{23}$  are again highly significant ( $p < 0.0001$ ), while  $\varphi_{123}$  is borderline significant ( $p \approx 0.044$ ) only for NCMV, and borderline non-significant ( $p \approx 0.059$ ) for CCMV and ACMV.

Finally, a trivariate Dale model is fitted to each of the patterns separately, with marginal probabilities depending on gender, and associations between outcomes independent of covariate effects. These results are summarized in Table 5. If the previous model was further extended, with, for the three patterns, different associations between outcomes, the same estimates would have been obtained, as in Table 5. We will now discuss the estimates that were obtained by fitting a separate trivariate Dale model to each pattern. For pattern 111, of course, there is no difference between the initial estimates and the multiple imputation estimates, since no imputation was necessary in this pattern. For patterns 100 and 110, several estimates are tending to infinity, since a lot of sparse or empty cells were present in the multiply-imputed sets of data, because the 13 males and 18 females in pattern 100, and the 20 males and 24 females in pattern 110, had to be distributed over 8 cells, with one more likely to be filled than the other. Especially the association parameters suffer from those empty cells. Therefore, it is hard to draw conclusions for patterns 100 and 110. Also, it is no avail trying to find the marginal effects of the covariate gender, using the technique of Section 5. If, however, the proportion of subjects were equal in each pattern, then the marginal gender effects, obtained by using those techniques, would correspond to the gender effects that resulted from the first model that was fitted.

The above results indicate, repeatedly, that the conclusions derived under different identifying restrictions, such as CCMV, NCMV, and ACMV, under otherwise identical models, may differ considerably. This begs the question as to how to choose between such restrictions. Unfortunately, this is not formally possible purely on statistical grounds. Indeed, the difference between the models is entirely in terms of the unobserved data, given the observed ones. As Molenberghs *et al.* (2008) have shown, it is not possible to choose between such models, because they fit the observed data equally well. The only difference is in terms of their prediction of the unobserved data, given the observed ones. Arguably, this is where careful negotiation with the substantive scientists comes in. Since the empty cells occurring through multiple imputation can be seen as sampling zeroes instead of structural zeroes, a continuity correction (adding  $\frac{1}{2}$  to each cell count, which is possible given that the data are organized as contingency tables for each covariate level) is advisable to overcome the problem of estimates tending to  $\infty$  (Agresti, 2002). Table 6 shows the results of a pattern-specific analysis of these continuity corrected data. When comparing the parameter estimates of the imputed data with the initial estimates, we observe that there is some deviation, probably due to the continuity correction of 0.5. This, in itself, is interesting since it may point to some sensitivity in the results owing to small cell counts. A sensitivity analysis can be performed by repeating the analysis for continuity corrections of various sizes ( $10^{-8}$ ,  $10^{-4}$ ,  $10^{-2}$ ,  $10^{-1}$ ) to explore their effect on the parameter estimates.

From all the analyses performed here, we conclude that the first model is overly simple, since all patterns are treated equally, and from more complex models, we conclude that some difference in success probability exists between the patterns. Thus, this should at least be taken into account. The last model, however, is very complex, and a continuity correction was needed before convergence was reached. Also, marginal covariate effects cannot be obtained so easily. A sensible compromise has to be chosen between the simplest and most complex model, ensuring that non-significant

covariate effects be removed from the model.

## 7 Conclusions

In this paper, we reviewed the general concepts of pattern-mixture models and the technique of identifying restrictions to specify the conditional distribution of the unobserved measurements, given the observed ones. Then, these concepts were extended to categorical outcomes, subject to intermittent missingness. The same identification families as employed with monotone missingness are employed here as well.

Interest is often in an overall, marginal covariate effect. Even in such cases, the analysis model might be chosen to be of the PMM type, for reasons of proper imputation and/or sensitivity analysis. One then needs to marginalize the pattern-specific covariate effects. Since such an overall effect cannot be obtained as simply as in the case of Gaussian data by averaging the pattern-specific effects, attention was devoted to the derivation of a marginal effect of interest. Generally, the intuitive but naive method of averaging can lead to both deflated as well as overestimated effects; it should not be seen as a conservative method.

In addition to the arguments developed for selecting a PMM approach or rather, say, a SEM, it is important to realize there can be an impact on efficiency. Especially when there are many different patterns and/or patterns are sparsely filled, there may be a large number of parameters and limited information. On the other hand, just like it can be beneficial to correct for auxiliary variables, such as age and sex, in a regression context, ‘correcting’ for pattern may afford a more focused, less imprecise assessment of, say, the effect of treatment. These considerations counteract each other, and whether or not a PMM is a sensible choice will depend on the study at hand, as well as on the other considerations contemplated earlier.

The fluvoxamine data were analyzed, using the method of pattern-mixture models, including identifying restrictions. Several models were fitted to the multiply-imputed sets of data. Some were too simple, others too complex, leading to sparse or even empty cells for the originally incomplete patterns, and resulting in convergence problems. Nevertheless, the different ways in which the data were analyzed, can be seen as a sensitivity analysis. Especially the use of different identifying restrictions is a first step in assessing the sensitivity of the assumptions made.

Throughout, we have chosen to analyze three time points for a binary version of the outcome. While extension to more time points and/or to categorical (ordinal or nominal) outcomes is in principle straightforward, the algebra is tedious and would deflect from the key messages. Nevertheless, when there are more than three time points, one can proceed sequentially, by imputing the next missing measurement given the observed or already imputed earlier ones. When there are more time points and/or non-monotone missingness reaches non-negligible proportions, the number of patterns and hence parameters might proliferate, always in the initial model and in the final model when a PMM is chosen for the analysis task. The user can then opt for allowing some parameters to be common across patterns, or for them to vary in a parametric way. Arguably, there is then a tradeoff between transparency and clarity on the one hand, and parsimony on the other hand. While the proposed methodology is general and can encompass a wide variety of choices, models, as well as inferential goals, building further empirical evidence as to its performance in practice would seem desirable. Such an assessment is a topic of further research.

# Appendix

## A Sign of $\rho$

As indicated in Section 6, we will determine the sign of  $\rho$  for  $P_1 \neq P_2$ . Denote the coefficient of  $\pi$  in the numerator of  $\rho$  by  $f_1$ , and the coefficient of  $(1 - \pi)$  in the denominator by  $f_2$ . Then,

$$f_1 = [\pi P_1 + (1 - \pi)P_2] [\pi(1 - P_1) + (1 - \pi)(1 - P_2)] - P_1(1 - P_1)$$

and

$$f_2 = -[\pi P_1 + (1 - \pi)P_2] [\pi(1 - P_1) + (1 - \pi)(1 - P_2)] + P_2(1 - P_2).$$

Since for  $\pi = 0$ ,  $f_1 = P_2(1 - P_2) - P_1(1 - P_1) = Q$  and  $f_2 = 0$ , and for  $\pi = 1$ ,  $f_1 = 0$  and  $f_2 = P_2(1 - P_2) - P_1(1 - P_1) = Q$ , both functions evolve in the interval  $[0, Q]$  when they are monotone. To determine whether there are internal extrema in  $f_1$  and  $f_2$ , we calculate

$$\frac{\partial f_1}{\partial \pi} = (P_1 - P_2) [\pi(1 - 2P_1) + (1 - \pi)(1 - 2P_2)],$$

which, since  $P_1 \neq P_2$  by assumption, equals 0 for  $\pi$  equal to

$$\pi^* = \frac{2P_2 - 1}{2(P_2 - P_1)}.$$

$\frac{\partial f_2}{\partial \pi}$  equals zero at the same point  $\pi^*$ . By calculating the second order derivatives of  $f_1$  and  $f_2$  to  $\pi$ ,

$$\begin{aligned} \frac{\partial^2 f_1}{\partial \pi^2} &= -2(P_1 - P_2)^2 < 0, \\ \frac{\partial^2 f_2}{\partial \pi^2} &= 2(P_1 - P_2)^2 > 0, \end{aligned}$$

we see that  $f_1$  reaches a maximum in  $\pi^*$ , while  $f_2$  is minimal in  $\pi^*$ . At  $\pi^*$ ,  $f_1 = \frac{1}{2} \cdot \frac{1}{2} - P_1(1 - P_1) \geq 0$  and  $f_2 = -\frac{1}{2} \cdot \frac{1}{2} + P_2(1 - P_2) \leq 0$ . Note that  $\pi^*$  is a valid extremum in  $[0, 1]$  if for  $P_1 \leq \frac{1}{2} \leq P_2$  and  $P_1 < P_2$ , and for  $P_2 \leq \frac{1}{2} \leq P_1$  with  $P_2 < P_1$ . In those situations,  $f_1 > 0$  and  $f_2 < 0$ , and hence



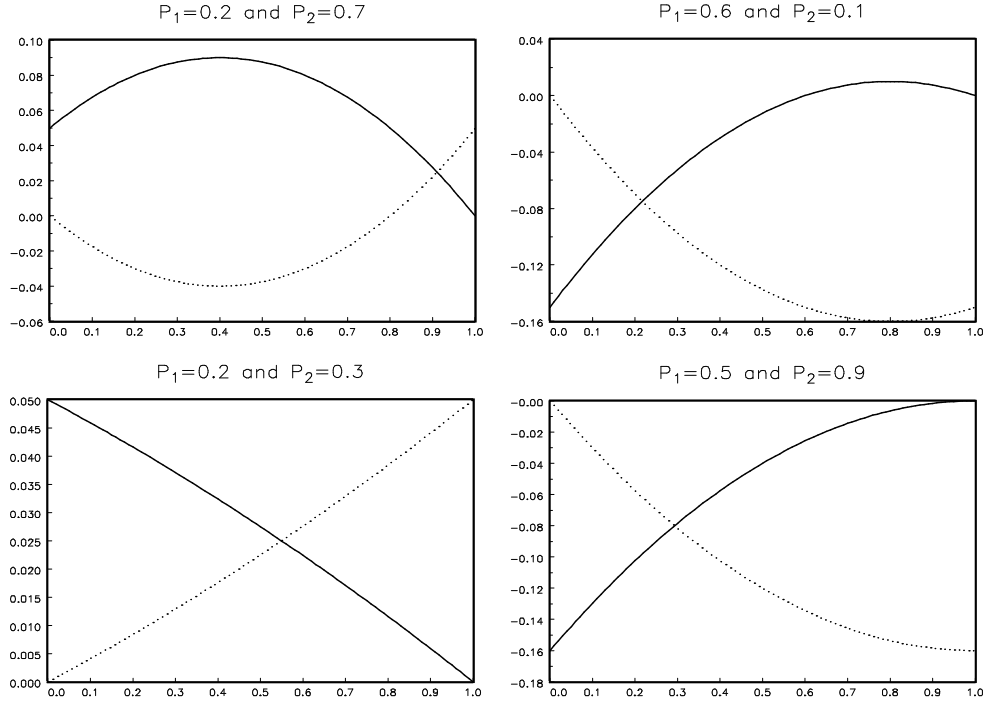


Figure 1: Graphical representation of the  $f_1$  (solid line) and  $f_2$  (dotted line) curves for several values of  $P_1$  and  $P_2$ . In the top panels,  $\pi^* \in [0, 1]$ , in the bottom panels,  $\pi^* \notin [0, 1]$ .

$\rho < 0$ , in a neighborhood of  $\pi^*$ . When  $\pi^* \notin [0, 1]$ , then  $f_1$  and  $f_2$  are monotonic and both of the same sign, such that  $\rho$  is nonnegative. Then there exist treatment effects  $(1, \rho)$  such that there is no dilution of effect, but equality or inflation. Figure 1 shows the curves of  $f_1$  and  $f_2$  for several values of  $P_1$  and  $P_2$ . Two of those examples are further studied in detail in B and C.

## B Example 1

Assume  $P_1 = 0.2$  and  $P_2 = 0.7$ . Set  $\pi$  equal to  $\pi^* = 0.4$ . Then  $f_1 = 0.09$  and  $f_2 = -0.04$  (see also Figure 1), such that

$$\rho = \frac{\pi}{1 - \pi} \cdot \frac{f_1}{f_2} = -\frac{0.4}{0.6} \cdot \frac{0.09}{0.04} = -1.5.$$

In this case, the treatment effects,  $(1; -1.5)$  are in the opposite direction. Since  $\pi P_1(1 - P_1) = 0.064$ ,  $(1 - \pi)P_2(1 - P_2) = 0.126$ ,  $\pi P_1 + (1 - \pi)P_2 = 0.5$  and  $\pi(1 - P_1) + (1 - \pi)(1 - P_2) = 0.5$ , the marginal treatment effect, calculated by (6.2) and (6.4) can be summarized as follows, for several values of  $\rho$ :

	$ B $	versus	$ \beta $
	$ 0.256 + 0.504\rho $		$ 0.4 + 0.6\rho $
$\rho = -2$	$ -0.752 $	$<$	$ -0.8 $
$\rho = -1.5$	$ -0.5 $	$=$	$ -0.5 $
$\rho = -1$	$ -0.248 $	$>$	$ -0.2 $

## C Example 2

Assume now  $P_1 = 0.2$  and  $P_2 = 0.3$ . Then  $\pi^* = -2 \notin [0, 1]$ . We choose  $\pi = 0.5$ . Now  $f_1 = 0.0275$  and  $f_2 = 0.0225$ , such that

$$\rho = \frac{\pi}{1 - \pi} \cdot \frac{f_1}{f_2} = \frac{0.5}{0.5} \cdot \frac{0.0275}{0.0225} = \frac{11}{9}.$$

So, in this case, both treatment effects,  $(1; 1.22)$  are quite close to each other. Since  $\pi P_1(1 - P_1) = 0.08$ ,  $(1 - \pi)P_2(1 - P_2) = 0.105$ ,  $\pi P_1 + (1 - \pi)P_2 = 0.25$  and  $\pi(1 - P_1) + (1 - \pi)(1 - P_2) = 0.75$ , the marginal treatment effect, calculated by (6.2) and (6.4) can be summarized as follows, for several values of  $\rho$ :

	$ B $	versus	$ \beta $
	$ 128/300 + 0.56\rho $		$ 0.5 + 0.5\rho $
$\rho = 10/9$	$ 18.88/18 $	$<$	$ 19/18 $
$\rho = 11/9$	$ 10/9 $	$=$	$ 10/9 $
$\rho = 12/9$	$ 21.12/18 $	$>$	$ 21/18 $

## Acknowledgment

Ivy Jansen and Geert Molenberghs gratefully acknowledge support from *Fonds Wetenschappelijk Onderzoek-Vlaanderen* Research Project G.0002.98 “Sensitivity Analysis for Incomplete and Coarse

Data” and from IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

## References

- Agresti, A. (2002) *Categorical Data Analysis*. Hoboken, NJ: Wiley.
- Diggle, P. J., Heagerty, P. J., Liang, K.-Y. and Zeger, S. L. (2002) *Analysis of Longitudinal Data (2nd ed.)*. Oxford Science Publications. Oxford: Clarendon Press.
- Diggle, P. J. and Kenward, M. G. (1994) Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.
- Fahrmeir, L. and Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Heidelberg: Springer-Verlag.
- Glynn, R. J., Laird, N. M. and Rubin, D. B. (1986) Selection modeling versus mixture modeling with nonignorable nonresponse. In: *Drawing Inferences from Self-Selected Samples* (Ed. H. Wainer), pp. 115–142. New York: Springer-Verlag.
- Hogan, J. W. and Laird, N. M. (1997) Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, **16**, 239–257.
- Jansen, I. and Molenberghs, G. (2008) A flexible marginal modeling strategy for non-monotone missing data. *Journal of the Royal Statistical Society, Series A*, **171**, 347–373.
- 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**, 409–418.
- Kenward, M. G., Lesaffre, E. and Molenberghs, G. (1994) An application of maximum likelihood and estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics*, **50**, 945–953.

- Kenward, M. G. and Molenberghs, G. (1998) Likelihood based frequentist inference when data are missing at random. *Statistical Science*, **12**, 236–247.
- 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.
- Michiels, B. and Molenberghs, G. (1997) Protective estimation of longitudinal categorical data with non-random dropout. *Communications in Statistics: Theory and Methods*, **26**, 65–94.
- Michiels, B., Molenberghs, G., Bijmens, 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., Beunckens, C., Sotto, C. and Kenward, M. M. G. (2008) Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, **70**, 371–388.

- Molenberghs, G., Goetghebeur, E., Lipsitz, S. R., Kenward, M. G., Lesaffre, E. and Michiels, B. (1999) Missing data perspectives of the fluvoxamine data set: A review. *Statistics in Medicine*, **18**, 2449–2464.
- Molenberghs, G. and Kenward, M. (2007) *Missing Data in Clinical Studies*. New York: John Wiley.
- Molenberghs, G., Kenward, M. G. and Lesaffre, E. (1997) The analysis of longitudinal ordinal data with nonrandom dropout. *Biometrika*, **84**, 33–44.
- Molenberghs, G. and Lesaffre, E. (1994) Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, **89**, 633–644.
- Molenberghs, G. and Lesaffre, E. (1999) Marginal modelling of multivariate categorical data. *Statistics in Medicine*, **18**, 2237–2255.
- Molenberghs, G., Michiels, B., Kenward, M. G. and Diggle, P. J. (1998) Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, **52**, 153–161.
- Molenberghs, G. and Verbeke, G. (2005) *Discrete Longitudinal Data*. New York: Springer-Verlag.
- Park, T. and Lee, S.-Y. (1999) Simple pattern-mixture models for longitudinal data with missing observations: Analysis of urinary incontinence data. *Statistics in Medicine*, **18**, 2933–2941.
- Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Schafer, J. L. (1997) *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.
- Sheiner, L. B., Beal, S. L. and Dunne, A. (1997) Analysis of non-randomly censored ordered categorical longitudinal data from analgesic trials. *Journal of the American Statistical Association*, **92**, 1235–1244.

- Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G. and Curran, D. (2002) Strategies to fit pattern-mixture models. *Biostatistics*, **3**, 245–265.
- Verbeke, G., Lesaffre, E. and Spiessens, B. (2001) The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, **35**, 419–439.
- Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Wu, M. C. and Carroll, R. J. (1988) Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, 175–188.

Table 1: *Fluvoxamine Data*. ‘Side effects’ (yes/no) at the first (horizontal), second (vertical) and last visit. Top table for males, bottom table for females.

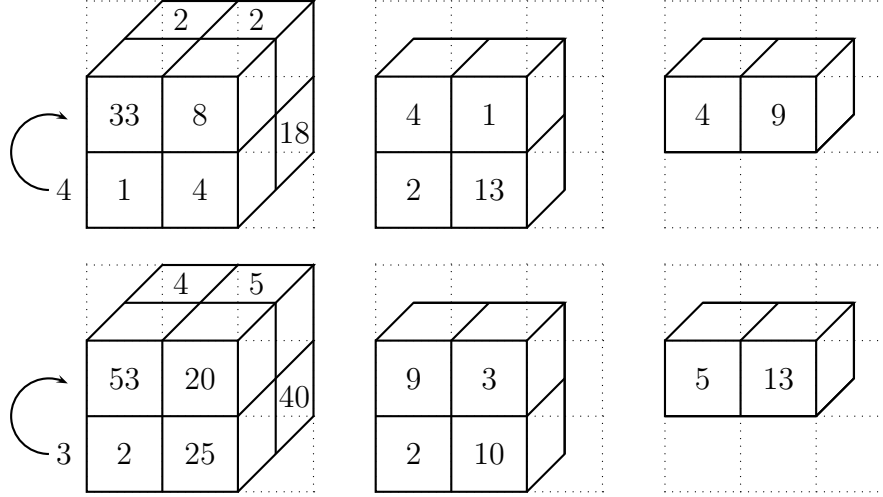


Table 2: *Fluvoxamine Data*. Multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on gender, and constant associations.

	CCMV	NCMV	ACMV
intercept <sub>1</sub>	-0.1259(0.1949)	-0.1266(0.1951)	-0.1230(0.1949)
gender <sub>1</sub>	-0.2528(0.2423)	-0.2516(0.2429)	-0.2574(0.2424)
intercept <sub>2</sub>	0.1180(0.1995)	0.0385(0.1984)	0.1060(0.2005)
gender <sub>2</sub>	-0.0022(0.2536)	0.0375(0.2435)	0.0020(0.2531)
intercept <sub>3</sub>	0.3245(0.2134)	0.2901(0.2139)	0.3120(0.2166)
gender <sub>3</sub>	0.2816(0.2675)	0.3159(0.2700)	0.2968(0.2703)
$\varphi_{12}$	3.1051(0.3433)	3.1218(0.3284)	3.1178(0.3386)
$\varphi_{13}$	2.0288(0.3072)	2.0047(0.3077)	2.0220(0.3121)
$\varphi_{23}$	2.8687(0.3583)	2.9588(0.3521)	2.8639(0.3548)
$\varphi_{123}$	1.8446(0.9272)	1.9283(0.9269)	1.8524(0.9386)

Table 3: *Fluvoxamine Data. Multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on a pattern-specific intercept and a fixed gender effect, and constant associations.*

	CCMV	NCMV	ACMV
intercept <sub>1</sub>	0.0215(0.2134)	0.0266(0.2131)	0.0238(0.2133)
pattern1 <sub>1</sub>	-0.6731(0.4209)	-0.7339(0.4151)	-0.6736(0.4205)
pattern2 <sub>1</sub>	-0.3418(0.3379)	-0.3429(0.3376)	-0.3426(0.3379)
gender <sub>1</sub>	-0.3027(0.2458)	-0.3013(0.2459)	-0.3060(0.2457)
intercept <sub>2</sub>	0.3164(0.2250)	0.2935(0.2187)	0.3172(0.2240)
pattern1 <sub>2</sub>	-0.4485(0.4777)	-0.9597(0.4906)	-0.5451(0.4927)
pattern2 <sub>2</sub>	-0.6989(0.3324)	-0.6914(0.3323)	-0.7004(0.3325)
gender <sub>2</sub>	-0.0709(0.2629)	-0.0424(0.2514)	-0.0725(0.2608)
intercept <sub>3</sub>	0.4713(0.2326)	0.4503(0.2346)	0.4607(0.2321)
pattern1 <sub>3</sub>	-0.2846(0.5761)	-0.4108(0.5997)	-0.3162(0.5311)
pattern2 <sub>3</sub>	-0.5498(0.4615)	-0.5469(0.4639)	-0.5457(0.4620)
gender <sub>3</sub>	0.2309(0.2778)	0.2654(0.2812)	0.2476(0.2779)
$\varphi_{12}$	3.1343(0.3469)	3.1410(0.3361)	3.1406(0.3444)
$\varphi_{13}$	2.0304(0.3084)	2.0168(0.3134)	2.0208(0.3112)
$\varphi_{23}$	2.8706(0.3589)	2.9654(0.3573)	2.8624(0.3561)
$\varphi_{123}$	1.7910(0.9649)	1.9351(0.9666)	1.8100(0.9778)



Table 4: *Fluvoxamine Data. Multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on a pattern-specific intercept and a pattern-specific gender effect, and constant associations.*

	CCMV	NCMV	ACMV
intercept <sub>1</sub>	0.1878(0.2364)	0.1933(0.2356)	0.1882(0.2361)
pattern1 <sub>1</sub>	-0.9785(0.6495)	-1.0467(0.6485)	-0.9689(0.6510)
pattern2 <sub>1</sub>	-1.0648(0.5433)	-1.0659(0.5424)	-1.0664(0.5431)
gender <sub>1</sub>	-0.5432(0.2866)	-0.5438(0.2862)	-0.5434(0.2865)
pattern1 × gender <sub>1</sub>	0.4447(0.8463)	0.4748(0.8506)	0.4224(0.8521)
pattern2 × gender <sub>1</sub>	1.2278(0.6989)	1.2271(0.6981)	1.2297(0.6991)
intercept <sub>2</sub>	0.5089(0.2448)	0.5067(0.2456)	0.5087(0.2448)
pattern1 <sub>2</sub>	-0.8513(0.7509)	-1.5285(0.7823)	-0.9168(0.7770)
pattern2 <sub>2</sub>	-1.4699(0.5386)	-1.4657(0.5382)	-1.4711(0.5390)
gender <sub>2</sub>	-0.3519(0.2937)	-0.3517(0.2943)	-0.3519(0.2938)
pattern1 × gender <sub>2</sub>	0.6298(1.1582)	0.8790(0.9177)	0.5870(1.1482)
pattern2 × gender <sub>2</sub>	1.3098(0.6927)	1.3032(0.6929)	1.3095(0.6934)
intercept <sub>3</sub>	0.5916(0.2445)	0.5942(0.2446)	0.5922(0.2446)
pattern1 <sub>3</sub>	-0.5736(0.7847)	-0.8706(0.8602)	-0.6826(0.8516)
pattern2 <sub>3</sub>	-0.9877(0.6146)	-0.9937(0.6158)	-0.9868(0.6134)
gender <sub>3</sub>	0.0561(0.2979)	0.0542(0.2978)	0.0559(0.2979)
pattern1 × gender <sub>3</sub>	0.4706(0.9388)	0.7796(1.0740)	0.5907(0.9538)
pattern2 × gender <sub>3</sub>	0.7610(0.8722)	0.7683(0.8693)	0.7612(0.8700)
$\varphi_{12}$	3.1328(0.3456)	3.1271(0.3359)	3.1412(0.3424)
$\varphi_{13}$	2.0235(0.3102)	2.0092(0.3140)	2.0143(0.3139)
$\varphi_{23}$	2.9035(0.3702)	2.9732(0.3564)	2.8943(0.3669)
$\varphi_{123}$	1.7912(0.9537)	1.9162(0.9506)	1.8026(0.9504)
Marginalized effects			
intercept <sub>1</sub>	-0.0591	-0.0600	-0.0580
gender <sub>1</sub>	-0.3288	-0.3263	-0.3308
intercept <sub>2</sub>	0.2074	0.1431	0.2004
gender <sub>2</sub>	-0.1076	-0.0930	-0.1123
intercept <sub>3</sub>	0.3795	0.3495	0.3686
gender <sub>3</sub>	0.2164	0.2457	0.2286

Table 5: *Fluvoxamine Data. Estimates from the initial Dale models for the incomplete data, together with multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on gender, and constant associations, fitted for each pattern separately.*

	initial	CCMV	NCMV	ACMV
<i>Pattern 100</i>				
intercept <sub>1</sub>	-0.8109 (0.6009)	-0.8329 (0.4665)	-1.1210 (14.110)	-0.7455 (0.5671)
gender <sub>1</sub>	-0.1446 (0.7988)	-0.0794 (0.6389)	6.0180 (21.981)	-0.3621 (0.9958)
intercept <sub>2</sub>		-0.4042 (0.5910)	-13.062 (40.395)	-0.5381 (0.7385)
gender <sub>2</sub>		0.2961 (0.9318)	0.5644 (0.8026)	0.3274 (0.9198)
intercept <sub>3</sub>		0.0394 (0.7034)	-4.3260 (13.075)	-0.0734 (0.7340)
gender <sub>3</sub>		0.5328 (0.8321)	12.117 (37.712)	0.6421 (0.8248)
$\varphi_{12}$		6.7429 (979.38)	-4010.0 (13044)	3.9641 (49.283)
$\varphi_{13}$		20.529 (2.86E6)	13.581 (2462i)	4.8504 (115.91)
$\varphi_{23}$		2.6890 (2.0828)	62.209 (205.04)	3.3699 (217.70)
$\varphi_{123}$		-4.5287 (7.69E43)	226.80 (45980)	-1.3710 (707.5i)
<i>Pattern 110</i>				
intercept <sub>1</sub>	-0.8473 (0.4880)	1.4118 (7.2243)	1.4118 (7.2202)	1.4118 (7.2269)
gender <sub>1</sub>	0.6802 (0.6371)	-1.9439 (8.2479)	-1.9439 (8.2396)	-1.9439 (8.2498)
intercept <sub>2</sub>	-1.0986 (0.5164)	-4.4028 (12.501)	-4.4028 (12.497)	-4.4028 (12.500)
gender <sub>2</sub>	1.0986 (0.6583)	4.1350 (12.071)	4.1350 (12.064)	4.1350 (12.070)
intercept <sub>3</sub>		13.651 (46.566)	13.651 (46.555)	13.651 (46.566)
gender <sub>3</sub>		-13.221 (46.112)	-13.221 (46.107)	-13.221 (46.111)
$\varphi_{12}$	2.9199 (0.8145)	3.9217 (28.152)	3.9217 (62.726)	3.9217 (48.873)
$\varphi_{13}$		2596.9 (2.86E6)	2596.9 (8249.1)	2596.9 (8609.7)
$\varphi_{23}$		-9.8258 (54.313)	-9.8258 (73.867)	-9.8258 (224.30)
$\varphi_{123}$		2581.9 (7.69E43)	2581.9 (2.62E22)	2581.9 (2.62E22)
<i>Pattern 111</i>				
intercept <sub>1</sub>	0.1956 (0.2376)	0.1956 (0.2376)	0.1956 (0.2376)	0.1956 (0.2376)
gender <sub>1</sub>	-0.5525 (0.2886)	-0.5525 (0.2886)	-0.5525 (0.2886)	-0.5525 (0.2886)
intercept <sub>2</sub>	0.5107 (0.2437)	0.5107 (0.2437)	0.5107 (0.2437)	0.5107 (0.2437)
gender <sub>2</sub>	-0.3522 (0.2929)	-0.3522 (0.2929)	-0.3522 (0.2929)	-0.3522 (0.2929)
intercept <sub>3</sub>	0.5824 (0.2447)	0.5824 (0.2447)	0.5824 (0.2447)	0.5824 (0.2447)
gender <sub>3</sub>	0.0679 (0.2987)	0.0679 (0.2987)	0.0679 (0.2987)	0.0679 (0.2987)
$\varphi_{12}$	3.1325 (0.3889)	3.1325 (0.3889)	3.1325 (0.3889)	3.1325 (0.3889)
$\varphi_{13}$	2.1026 (0.3533)	2.1026 (0.3533)	2.1026 (0.3533)	2.1026 (0.3533)
$\varphi_{23}$	2.9471 (0.3726)	2.9471 (0.3726)	2.9471 (0.3726)	2.9471 (0.3726)
$\varphi_{123}$	1.2110 (0.9510)	1.2110 (0.9510)	1.2110 (0.9510)	1.2110 (0.9510)
Marginalized effects				
intercept <sub>1</sub>		-0.0128	-0.0146	-0.0118
gender <sub>1</sub>		-0.3529	-0.3495	-0.3548
intercept <sub>2</sub>		0.2331	0.1813	0.2250
gender <sub>2</sub>		-0.1275	-0.1143	-0.1270
intercept <sub>3</sub>		0.3834	0.3598	0.3743
gender <sub>3</sub>		0.1960	0.2185	0.2071

Table 6: *Fluvoxamine Data. Estimates from the initial Dale models for the incomplete data, together with multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on gender, and constant associations, fitted for each pattern separately. A continuity correction of  $\frac{1}{2}$  is used to overcome the problem of sampling zeroes.*

	initial	CCMV	NCMV	ACMV
<i>Pattern 100</i>				
intercept <sub>1</sub>	-0.8109 (0.6009)	-0.6141 (0.5074)	-0.6327 (0.5085)	-0.6031 (0.5070)
gender <sub>1</sub>	-0.1446 (0.7988)	-0.1426 (0.6797)	-0.1074 (0.6778)	-0.1622 (0.6812)
intercept <sub>2</sub>		-0.2541 (0.5666)	-0.7844 (0.6024)	-0.3323 (0.5774)
gender <sub>2</sub>		0.2087 (0.8750)	0.3775 (0.7287)	0.2146 (0.8557)
intercept <sub>3</sub>		0.0221 (0.6105)	-0.1987 (0.6374)	-0.0625 (0.6462)
gender <sub>3</sub>		0.4174 (0.7459)	0.6426 (0.8118)	0.5247 (0.7506)
$\varphi_{12}$		2.2922 (1.0342)	2.4398 (0.8669)	2.3648 (0.9745)
$\varphi_{13}$		1.4683 (0.9096)	1.2864 (0.9326)	1.4161 (0.9922)
$\varphi_{23}$		1.5591 (0.9562)	1.8639 (0.9715)	1.4965 (1.0034)
$\varphi_{123}$		1.0355 (2.4118)	1.9070 (2.4134)	1.2348 (2.5027)
<i>Pattern 110</i>				
intercept <sub>1</sub>	-0.8473 (0.4880)	-0.6859 (0.4294)	-0.6859 (0.4294)	-0.6859 (0.4294)
gender <sub>1</sub>	0.6802 (0.6371)	0.5381 (0.5675)	0.5381 (0.5675)	0.5381 (0.5675)
intercept <sub>2</sub>	-1.0986 (0.5164)	-0.8357 (0.4337)	-0.8357 (0.4337)	-0.8357 (0.4337)
gender <sub>2</sub>	1.0986 (0.6583)	0.8075 (0.5716)	0.8075 (0.5716)	0.8075 (0.5716)
intercept <sub>3</sub>		-0.3284 (0.4858)	-0.3284 (0.4858)	-0.3284 (0.4858)
gender <sub>3</sub>		0.6901 (0.7085)	0.6901 (0.7085)	0.6901 (0.7085)
$\varphi_{12}$	2.9199 (0.8145)	2.3211 (0.6719)	2.3211 (0.6719)	2.3211 (0.6719)
$\varphi_{13}$		1.2766 (0.6825)	1.2766 (0.6825)	1.2766 (0.6825)
$\varphi_{23}$		2.3965 (0.8353)	2.3965 (0.8353)	2.3965 (0.8353)
$\varphi_{123}$		2.0700 (2.0344)	2.0700 (2.0344)	2.0700 (2.0344)