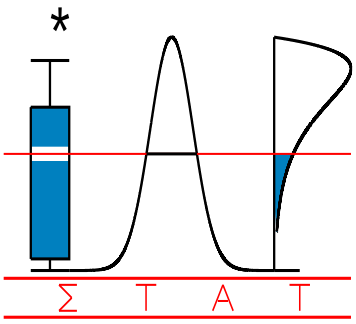
T E C H N I C A L

R E P O R T

0654

PATTERN-MIXTURE MODELS FOR CATEGORICAL OUTCOMES WITH NON-MONOTONE MISSINGNESS

JANSEN I., and G. MOLENBERGHS



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

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, $\boldsymbol{Y}$ say, and the non-response process, $\boldsymbol{R}$ say. This joint distribution $f(\boldsymbol{y}, \boldsymbol{r}|\boldsymbol{\theta}, \boldsymbol{\psi})$ can be factorized in several ways. A selection model is based

on the factorization $f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{r}|\boldsymbol{y},\boldsymbol{\psi})$, whereas the reverse factorization $f(\boldsymbol{y}|\boldsymbol{r},\boldsymbol{\theta})f(\boldsymbol{r}|\boldsymbol{\psi})$ is referred to as a pattern-mixture model (Little, 1993, 1994). When a common set of random-effects is thought to influence both the $\boldsymbol{Y}$ and $\boldsymbol{R}$ 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 $\boldsymbol{R}$ 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 esti-

mate 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 make the need for making assumptions and the implications thereof more obvious.

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, 2007). 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.

The rest of the paper is organized as follows. Section 2 introduces the data from a psychiatric study. In Section 3, the general context of pattern-mixture models will be sketched, together with the strategy of identifying restrictions. Section 4 gives attention to the use of the multivariate Dale model to fit pattern-mixture models for categorical outcomes, while Section 5 will discuss the assumptions needed when intermittent missingness is present. Section 6 focuses on the derivation of marginal effects in pattern-mixture models for categorical outcomes. Finally, in Section 7, the techniques developed will be used to analyze the fluvoxamine data.

2 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 serotonine 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.

3 Pattern-Mixture Modeling

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

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

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. Note that the second factor can depend on covariates, but not on the random outcomes. It is, of course, possible to have different covariate dependencies in either components of the factorization.

The measurement model has to reflect dependence on the missingness mechanism. Therefore, we will assume a different distribution for each missingness pattern. However, although

the parameters can be simply estimated from the data in each missingness pattern, not all of them can be identified without making further assumptions. It is also clear which information is provided by each missingness pattern and, consequently, where assumptions are needed to predict the behavior of the unobserved responses, and to obtain marginal models for the response. However, this model contains under-identified members since it describes the full set of measurements in each pattern, even though there are no measurements at certain occasions. At first sight, this leaves them open to the same criticism as selection models, but Little (1993) claims that the pattern-mixture approach is more forthright, because (conditional) distributions for which the data provide information are clearly distinguished from parameters for which there is no information at all. Next, we will introduce the identifying-restrictions strategy advocated by Little (1993, 1994).

For the time being, we restrict attention to monotone patterns, i.e., only dropout is present, no intermittent missingness. In general, let us assume that we have patterns t ($t = 1, \ldots, n$, but not necessary all of them are present) for which the complete data density is given by

$$f_t(y_1, \ldots, y_n) = f_t(y_1, \ldots, y_t) f_t(y_{t+1}, \ldots, y_n | y_1, \ldots, y_t). \tag{3.1}$$

The first factor is clearly identified from the observed data, while the second factor is not. It is assumed that the first factor is known or, more realistically, can be modeled using the observed data. Then, identifying restrictions are applied in order to identify the second component.

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

$$f_t(y_s | y_1, \ldots y_{s-1}) = \sum_{j=s}^{n} \omega_{sj} f_j(y_s | y_1, \ldots y_{s-1}), \quad s = t+1, \ldots, n. \tag{3.2}$$

Let $\boldsymbol{\omega}_s = (\omega_{ss}, \ldots, \omega_{sn})'$. Every $\boldsymbol{\omega}_s$ with components summing to one, provides a valid identification scheme.

Three special but important cases are considered. Little (1993) proposed *complete case missing value* (CCMV), which uses the following identification:

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

In other words, the conditional distribution beyond time t is always borrowed from the conditional distribution from the completers. This strategy can be defended in cases where the bulk of the subjects are complete and only small proportions are assigned to the various dropout patterns. Also, extension of this approach to non-monotone patterns is particularly easy. Alternatively, the nearest identified pattern can be used:

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

We will refer to these restrictions as *neighboring case missing value* (NCMV). The third special case of (3.2) will be *available case missing value* (ACMV). It has been shown in Molenberghs *et al.* (1998) that, for monotone missing data, ACMV in the pattern-mixture context is equivalent with MAR in the selection model framework. The derivation of the corresponding $\boldsymbol{\omega}_s$ vectors is straightforward and results in

$$\omega_{sj} = \frac{\pi_j f_j(y_1, \ldots, y_{s-1})}{\sum_{\ell=s}^{n} \pi_\ell f_\ell(y_1, \ldots, y_{s-1})}, \tag{3.3}$$

where π_j is the fraction of observations in pattern j (Molenberghs *et al.*, 1998). Clearly, $\boldsymbol{\omega}_s$ defined by (3.3) consists of components which are nonnegative and sum to one. In other words, a valid density function is defined.

Restrictions (3.2), with the CCMV, NCMV, and ACMV forms as special cases, can be incorporated in a comprehensive strategy to fit pattern-mixture models. These are the steps needed to be followed. First, fit a model to the pattern-specific identifiable densities:

$f_t(y_1, \ldots, y_t)$. Next, select an identification method of choice. Third, using this identification method, determine the conditional distributions of the unobserved outcomes, given the observed ones. Fourth, draw multiple imputations for the unobserved components, given the observed outcomes and the correct pattern-specific density. Fifth, analyze the multiply-imputed sets of data using the method of choice. This can be another pattern-mixture model, but also a selection model or any other desired model. Finally, inferences can be conducted in the standard multiple imputation way (Rubin, 1987; Schafer, 1997; Verbeke and Molenberghs, 2000).

4 Pattern-Mixture Models for Categorical Outcomes

From now on, we will specialize to 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, this results in the following logistic-regression and odds-

ratio formulations (subject-specific indices i are removed for the ease of notation):

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

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, only η_1, η_2 and η_4 can be obtained from the data, while for pattern 1 only η_1 will be available.

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 $p_{y_1y_2y_3|3} = P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3|\text{pattern} = 3)$, $p_{y_1y_2|2} = P(Y_1 = y_1, Y_2 = y_2|\text{pattern} = 2)$ and $p_{y_1|1} = P(Y_1 = y_1|\text{pattern} = 1)$ can be estimated. 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 $Z_{y_1y_2|2}$ and the conditional probabilities $p_{y_3|y_1y_2,3} = P(Y_3 = y_3|Y_1 = y_1, Y_2 = y_2, \text{pattern} = 3)$ have to be used to identify $Z^*_{y_1y_2y_3|2}$ as $Z_{y_1y_2|2} \times p_{y_3|y_1y_2,3}$. 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 (3.2), the joint probability of y_1, y_2 and y_3 in pattern 1 can be written as

$$p_{y_1y_2y_3|1} \;=\; p_{y_1|1}\left[\omega p_{y_2|y_1,2} + (1-\omega)p_{y_2|y_1,3}\right] p_{y_3|y_1y_2,3},$$

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

$$\begin{aligned}
\text{CCMV} &: p_{y_1|1}\, p_{y_2|y_1,3}\, p_{y_3|y_1y_2,3},\\
\text{NCMV} &: p_{y_1|1}\, p_{y_2|y_1,2}\, p_{y_3|y_1y_2,3},\\
\text{ACMV} &: \omega = \frac{\pi_2 p_{y_1|2}}{\pi_2 p_{y_1|2} + \pi_3 p_{y_1|3}},
\end{aligned}$$

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

$$\begin{aligned}
\text{CCMV} &: Z^*_{y_1y_2y_3|1} = Z_{y_1|1}\, p_{y_2|y_1,3}\, p_{y_3|y_1y_2,3},\\
\text{NCMV} &: Z^*_{y_1y_2y_3|1} = Z_{y_1|1}\, p_{y_2|y_1,2}\, p_{y_3|y_1y_2,3},\\
\text{ACMV} &: Z^*_{y_1y_2y_3|1} = Z_{y_1|1}\left[\frac{\pi_2 p_{y_1y_2|2} + \pi_3 p_{y_1y_2|3}}{\pi_2 p_{y_1|2} + \pi_3 p_{y_1|3}}\right] p_{y_3|y_1y_2,3}.
\end{aligned}$$

To perform the corresponding imputations, we use a uniform random number generator. Suppose the count Z is to be distributed over the cells Z_k, $k = 1, \ldots, c$. Then, the cumulative probabilities $\lambda_0, \ldots, \lambda_c$ are calculated and Z draws U_t from a uniform $U[0,1]$ distribution are made. Next, Z_k is set equal to $\sum_t(\lambda_{k-1} < U_t \leq \lambda_k)$.

From the completed counts $Z^*_{y_1y_2y_3|1}$ and $Z^*_{y_1y_2y_3|2}$, and from $Z_{y_1y_2y_3|3}$, one can estimate the parameters of interest, for example a trivariate Dale model for the three patterns separately, or a trivariate Dale model where pattern is included as a covariate. Also, other possible models can be fitted to the completed counts.

Although parameter estimation is very elegant and computationally simple with the above two-step procedure, precision estimation is less straightforward. Indeed, treating the filled-in table as if it represented observed data fails to reflect random variability in the unobserved counts. Therefore, multiple imputation will be used to construct an asymptotic covariance matrix of the form

$$\boldsymbol{V} = \boldsymbol{W} + \left(\frac{M+1}{M}\right)\boldsymbol{B}, \tag{4.1}$$

where $\boldsymbol{W}$ is the average *within*-imputation variance, $\boldsymbol{B}$ the *between*-imputation variance, and M the number of imputations.

5 Assumptions Needed for the Intermittent Missingness Case

We will extend the above strategy to the situation of non-monotone missing data. In Figure 1, a three-dimensional graphical representation is given of all possible patterns for three binary outcomes when intermittent missingness is allowed. Figure 2 gives an equivalent two-dimensional representation. The first three patterns are the monotone ones, which have been discussed already in Section 4. Pattern 3 is the fully observed pattern, and does not need any imputation. Patterns 1 and 2 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 2, 4, and 5, where the third, the second and the first outcome, respectively, are not

observed. A bivariate Dale model can be used to fit the observed data densities $f_1(y_1, y_2)$, $f_4(y_1, y_3)$ and $f_5(y_2, y_3)$. 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 3 to impute the unobserved data. This results in the following complete data densities:

$$\begin{aligned} f_2(y_1, y_2, y_3) &= f_2(y_1, y_2) f_3(y_3|y_1, y_2), \\ f_4(y_1, y_2, y_3) &= f_4(y_1, y_3) f_3(y_2|y_1, y_3), \\ f_5(y_1, y_2, y_3) &= f_5(y_2, y_3) f_3(y_1|y_2, y_3). \end{aligned}$$

Next, patterns 1, 6, and 7 will be discussed. Here, only one out of the three outcomes is measured, and a univariate Dale model can be used to obtain $f_1(y_1)$, $f_6(y_3)$, and $f_7(y_2)$. 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 1 was to impute first 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. For pattern 6, 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 3) or the neighbors (pattern 4), or a combination of both densities. Thus, we obtain the previously defined identifying restrictions CCMV, NCMV and ACMV. The conditional density of y_2, given y_1 and y_3, can only be borrowed from pattern 3. 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. Thus, the complete data densities for patterns 1, 6, and 7 can be written as:

$$\begin{aligned} f_1(y_1, y_2, y_3) &= \begin{cases} f_1(y_1)\left[\omega f_2(y_2|y_1) + (1-\omega) f_3(y_2|y_1)\right] f_3(y_3|y_1, y_2), \\ f_1(y_1)\left[\omega f_4(y_3|y_1) + (1-\omega) f_3(y_3|y_1)\right] f_3(y_2|y_1, y_3), \end{cases} \\ f_6(y_1, y_2, y_3) &= \begin{cases} f_6(y_3)\left[\omega f_4(y_1|y_3) + (1-\omega) f_3(y_1|y_3)\right] f_3(y_2|y_1, y_3), \\ f_6(y_3)\left[\omega f_5(y_2|y_3) + (1-\omega) f_3(y_2|y_3)\right] f_3(y_1|y_2, y_3), \end{cases} \\ f_7(y_1, y_2, y_3) &= \begin{cases} f_7(y_2)\left[\omega f_2(y_1|y_2) + (1-\omega) f_3(y_1|y_2)\right] f_3(y_3|y_1, y_2), \\ f_7(y_2)\left[\omega f_5(y_3|y_2) + (1-\omega) f_3(y_3|y_2)\right] f_3(y_1|y_2, y_3), \end{cases} \end{aligned}$$

where, in all cases, $\omega = 0$ corresponds to CCMV, $\omega = 1$ to NCMV, and ω as in (3.3), with the corresponding densities and pattern probabilities, corresponds to ACMV. So, we can either choose one of the two possibilities to determine the complete data density, or use a linear combination of both expressions. This combination is a topic of further research.

Finally, pattern 8 does not contain any observed data, such that it is not possible to impute the unobserved data conditional on the observed data. This pattern will therefore be ignored, generally the only sensible route.

6 Marginal Effects Across Patterns

We already mentioned that several strategies can be followed to analyze the imputed data sets. When a single model is used, an overall effect of the covariates of interest (e.g., treatment effect) is obtained immediately from the model. When, however, a separate model for each pattern is used to analyze the multiply imputed sets of data, the overall covariate effect cannot be obtained directly. 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

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

is used to model the data from pattern $k = 1, \ldots, K$ (as in the multivariate Dale model). α and β 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 π_k to be the pattern probability as defined before. The marginal success probability is then equal to

$$\sum_{k=1}^{K} \pi_k \frac{e^{\alpha_k + \beta_k T}}{1 + e^{\alpha_k + \beta_k T}}. \tag{6.1}$$

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

$$\beta \simeq \sum_k \pi_k \beta_k, \tag{6.2}$$

but this is clearly incorrect. 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), ... Third, classical averaging can be performed. To this effect, keep function (6.1) 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. We will focus on the second approach, using a marginal logistic model.

Let us approximate (6.1) by a logistic regression, with obvious notation:

$$f(T) = \sum_k \pi_k \frac{e^{\alpha_k + \beta_k T}}{1 + e^{\alpha_k + \beta_k T}} \cong \frac{e^{A+BT}}{1 + e^{A+BT}}. \tag{6.3}$$

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

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

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

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

such that

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

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

$$B \simeq \frac{\sum_k \beta_k \pi_k P_k(1-P_k)}{\left(\sum_k \pi_k P_k\right)\left[\sum_k \pi_k (1-P_k)\right]}. \tag{6.4}$$

Note that direct expansion of (6.3), 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

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

such that (proof is in A.1)

$$|B| \leq |\beta|. \tag{6.5}$$

This means that the marginal treatment effect at the last occasion, obtained through approximation (6.3), will not be larger in absolute value than the marginal treatment effect, obtained from the direct linear approach (6.2), when the treatment effects are equal across patterns.

Marginalization when the β_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 (6.2) and (6.4) then reduce to

$$\pi + (1-\pi)\rho \quad \text{and} \quad \frac{\pi P_1(1-P_1) + \rho(1-\pi)P_2(1-P_2)}{[\pi P_1 + (1-\pi)P_2]\,[\pi(1-P_1) + (1-\pi)(1-P_2)]}. \tag{6.6}$$

Let $N = [\pi P_1 + (1-\pi)P_2]\,[\pi(1-P_1) + (1-\pi)(1-P_2)]$. Choosing ρ such that the equality between both expressions in (6.6) holds, results in

$$\rho = \frac{\pi\,[N - P_1(1-P_1)]}{(1-\pi)\,[P_2(1-P_2) - N]}. \tag{6.7}$$

Since $\rho \in \mathbb{R}$, setting ρ 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 (6.6) 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 ρ, both expressions in (6.6) are the same. Note also that then in (6.7) the numerator and denominator are both equal to zero, confirming that the result applies to every ρ. Thus, the difference emerges from a difference in background success probability P_k.

Now we will determine the sign of ρ for $P_1 \neq P_2$. Denote the coefficient of π in the numerator of ρ 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 $[0, Q]$. To determine

whether there are internal extrema in f_1 and f_2, we calculate

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

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

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

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

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

we see that f_1 reaches a maximum in π^*, while f_2 is minimal in π^*. At π^*, $f_1 = \frac{1}{2}.\frac{1}{2} - P_1(1 - P_1) \geq 0$ and $f_2 = -\frac{1}{2}.\frac{1}{2} + P_2(1 - P_2) \leq 0$. Note that π^* is a valid extremum in $[0, 1]$ if for $P_1 < P_2$, $P_1 \leq \frac{1}{2} \leq P_2$, and for $P_1 > P_2$, $P_2 \leq \frac{1}{2} \leq P_1$. In those situations, $f_1 > 0$ and $f_2 < 0$, and hence $\rho < 0$, in a neighborhood of π^*. When $\pi^* \notin [0, 1]$, then f_1 and f_2 are monotonic and both of the same sign, such that ρ is nonnegative. Then there exist treatment effects $(1, \rho)$ such that there is no dilution of effect, but equality or inflation. Figure 3 (in Appendix) 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 A.2 and A.3.

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.

7 Models Fitted to the Fluvoxamine Data

In this section, we will analyze the fluvoxamine data, presented in Section 2. We will account for patients' gender (0 = males, 1 = females), a covariate recorded for all. There are 224 completers (pattern 3), 44 patients missed the last visit (pattern 2), 31 only appeared at the first visit (pattern 1), 1 person belongs to pattern 5, 1 to pattern 6, and the remaining 14 patients do not have any observations at all (pattern 8). For those 14 patients, there is no reasonable way to impute the missing outcomes, and therefore, they will not be considered for analyses. Pattern 5 and pattern 6 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 4, we start by fitting a trivariate Dale model to the completers, a bivariate Dale model to pattern 2, and a logistic regression to pattern 1. 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, with a constant log odds ratio 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 φ_{12}, φ_{13} and φ_{23} are highly significant ($p < 0.0001$), while φ_{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 ω 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. 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 constant. The parameter $\texttt{intercept}_i$ is the intercept in the logistic regression for the ith marginal probability for pattern 3. $\texttt{pattern1}_i$ and $\texttt{pattern2}_i$ 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 1 and pattern 3. We notice that the intercept for pattern 3 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 2 and 3 are significantly different ($p \approx 0.035$), while patterns 1 and 3 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 2 and thus takes distance from pattern 3, while under CCMV and ACMV all, or most, of the information is borrowed from pattern 3, and therefore there is only little distance between pattern 1 and 3. 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 3. 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 φ_{12}, φ_{13}, and φ_{23} are highly significant ($p < 0.0001$), while φ_{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 $\texttt{intercept}_i$ corresponds to the effect of pattern 3, while the dummy variables $\texttt{pattern1}_i$ and $\texttt{pattern2}_i$ model the difference in success probability between pattern 3 and pattern 1 or 2, respectively. $\texttt{gender}_i$ represents the gender effect in pattern 3, while the interactions between the dummies and gender refer to the difference in gender effect between pattern 3 and pattern 1 or 2, 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 2 and 3, but not significantly different between pattern 1 and 3. Gender is borderline non-significant ($p \approx 0.058$) in pattern 3, 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 2 and 3, and under NCMV borderline significant ($p \approx 0.05$) between pattern 1 and 3. The gender effect in pattern 3 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 φ_{12}, φ_{13} and φ_{23} are again highly significant ($p < 0.0001$), while φ_{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 constant associations between outcomes. 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 3, 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 1 and 2, 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 1, and the 20 males and 24 females in pattern 2, 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 1 and 2. Also, it is no avail trying to find the marginal effects of the covariate gender, using the technique of Section 6. 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.

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) is advisable to overcome the problem of estimates tending to ∞ (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 $\frac{1}{2}$. A sensitivity analysis can be performed by repeating the analysis for continuity corrections of various sizes (10^{-8}, 10^{-4}, 10^{-2}, 10^{-1}) in order 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.

8 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 suggested here as well.

Since interest is often in an overall covariate effect, and not in the pattern-specific effects only,

and since this overall effect cannot be obtained as simple 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. It was also shown that 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.

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.

Further research can be devoted to the analysis of other data sets with more intermittent missingness, such that the suggestions made in this chapter to identify those missing values, can be explored.

Acknowledgment

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

References

Agresti, A. (2002) *Categorical Data Analysis.* Hoboken, NJ: Wiley.

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.

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. (2007) A flexible marginal modeling strategy for non-monotone missing data. *Journal of the American Statistical Association*, Submitted.

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., 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., 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., 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.

A Appendix

A.1 Proof of Equation (6.5)

Let

$$g = \left(\sum_k \pi_k P_k\right)\left(\sum_k \pi_k (1-P_k)\right) - \sum_k \pi_k P_k (1-P_k)$$

and

$$H = \sum_k \pi_k P_k.$$

To find the extrema of g, we calculate

$$\begin{aligned}\frac{\partial g}{\partial P_\ell} &= \pi_\ell(1-H) - \pi_\ell H - \pi_\ell(1-2P_\ell) \\ &= 2\pi_\ell(P_\ell - H) \\ &= 2\pi_\ell\left(P_\ell - \sum_k \pi_k P_k\right).\end{aligned}$$

g then reaches an extremum if $\dfrac{\partial g}{\partial P_\ell} = 0$ for all ℓ. Thus,

$$\pi_\ell\left(P_\ell - \sum_k \pi_k P_k\right) = 0.$$

We can exclude $\pi_\ell = 0$, since such a pattern would vanish. Thus, g reaches an extremum if

$$P_\ell = \sum_k \pi_k P_k \;\; \forall \ell \quad \Leftrightarrow \quad P_1 = \ldots = P_K \equiv P$$

and hence

$$\sum_k \pi_k P_k = P \sum_k \pi_k = P.$$

At this extremum, $g = P(1-P) - P(1-P) = 0$. Now we still have to check whether this extremum is a minimum or a maximum. Therefore we calculate the second order derivatives of g.

$$\left.\begin{array}{rcl} \dfrac{\partial^2 g}{\partial P_\ell^2} & = & 2\pi_\ell(1-\pi_\ell) \\ \dfrac{\partial^2 g}{\partial P_\ell \partial P_m} & = & 2\pi_\ell\,(-\pi_m) = -2\pi_\ell\pi_m \end{array}\right\} \Rightarrow \frac{\partial^2 g}{\partial \boldsymbol{P}\partial \boldsymbol{P}'} = 2\,[\mathrm{diag}\boldsymbol{\pi} - \boldsymbol{\pi}\boldsymbol{\pi}']$$

which is a positive definite matrix. We can conclude that this extremum is a minimum, and thus $g \geq 0$, which means $0 \leq \dfrac{\sum_k \pi_k P_k(1-P_k)}{(\sum_k \pi_k P_k)\,(\sum_k \pi_k(1-P_k))} \leq 1$, and $|B| \leq |\beta|$. ■

A.2 Example 1

Assume $P_1 = 0.2$ and $P_2 = 0.7$. Set π equal to $\pi^* = 0.4$. Then $f_1 = 0.09$ and $f_2 = -0.04$ (see also Figure 3), such that

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

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 ρ:

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

A.3 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

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

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 ρ:

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

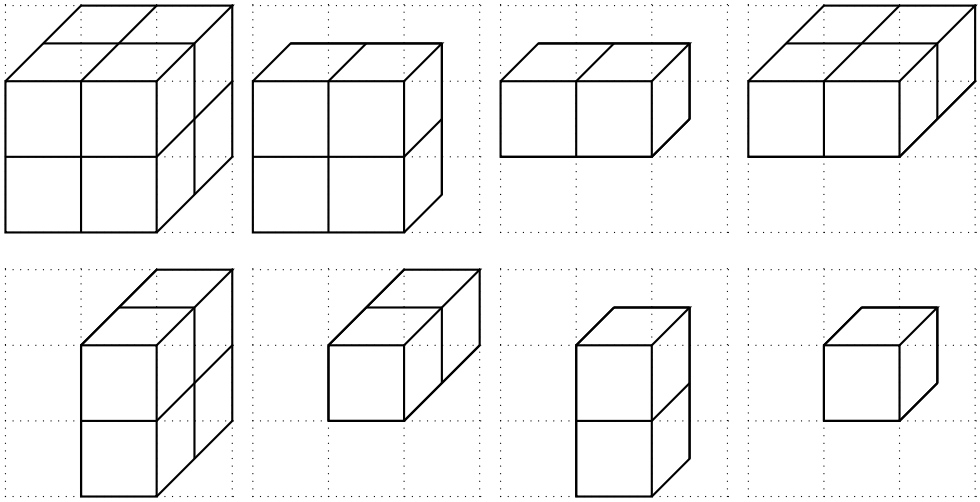


Figure 1: *Three-dimensional representation of all possible patterns for three binary outcomes with intermittent missingness. The horizontal axis displays the first measurement, the vertical axis corresponds to the second measurement, and the third axis to the last measurement.*

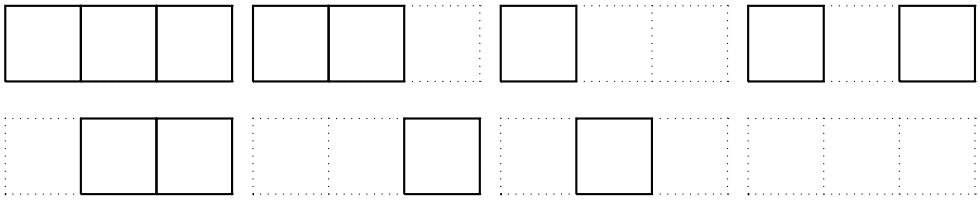


Figure 2: *Two-dimensional representation of all possible patterns for three outcomes with intermittent missingness, in the same order as in Figure 1. A solid square represents an observed measurement. From left to right, and from top to bottom, we have patterns 3, 2 and 1 as defined before, and further the non-monotone patterns 4, 5, 6, 7, and 8.*

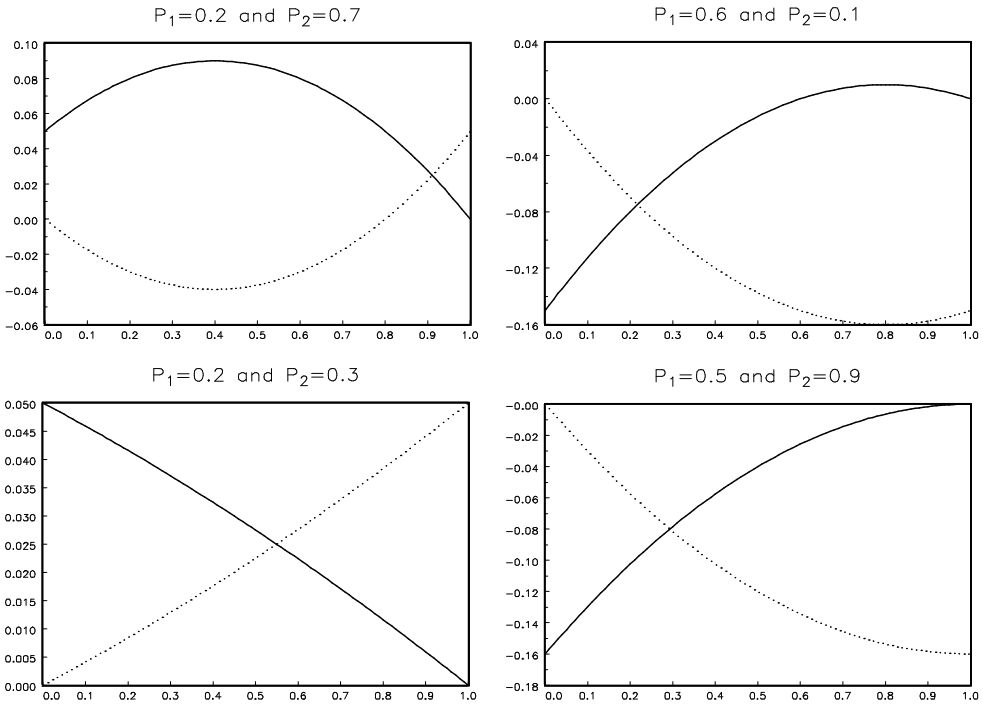


Figure 3: *Graphical representation of the f_1 (solid line) and f_2 (dotted line) curves for several values of P_1 and P_2. In the top panels, $\pi^* \in [0,1]$, in the bottom panels, $\pi^* \notin [0,1]$.*

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.*

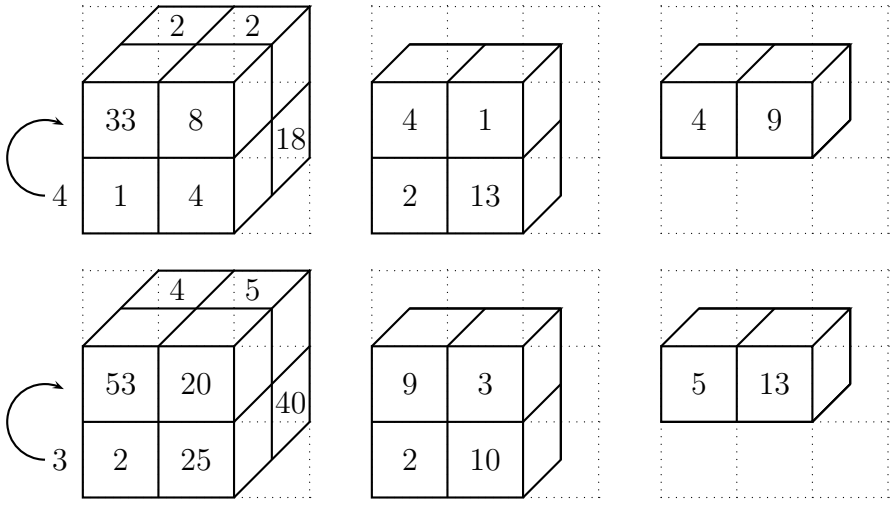


Table 2: *Fluvoxamine Data. Multiple imputation estimates and standard errors for CCMV, NCMV and ACMV. A trivariate Dale model, with marginal probabilities depending on gender, and constant associations.*

	CCMV	NCMV	ACMV
intercept_1	-0.1259(0.1949)	-0.1266(0.1951)	-0.1230(0.1949)
gender_1	-0.2528(0.2423)	-0.2516(0.2429)	-0.2574(0.2424)
intercept_2	0.1180(0.1995)	0.0385(0.1984)	0.1060(0.2005)
gender_2	-0.0022(0.2536)	0.0375(0.2435)	0.0020(0.2531)
intercept_3	0.3245(0.2134)	0.2901(0.2139)	0.3120(0.2166)
gender_3	0.2816(0.2675)	0.3159(0.2700)	0.2968(0.2703)
φ_{12}	3.1051(0.3433)	3.1218(0.3284)	3.1178(0.3386)
φ_{13}	2.0288(0.3072)	2.0047(0.3077)	2.0220(0.3121)
φ_{23}	2.8687(0.3583)	2.9588(0.3521)	2.8639(0.3548)
φ_{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_1	0.0215(0.2134)	0.0266(0.2131)	0.0238(0.2133)
pattern1_1	-0.6731(0.4209)	-0.7339(0.4151)	-0.6736(0.4205)
pattern2_1	-0.3418(0.3379)	-0.3429(0.3376)	-0.3426(0.3379)
gender_1	-0.3027(0.2458)	-0.3013(0.2459)	-0.3060(0.2457)
intercept_2	0.3164(0.2250)	0.2935(0.2187)	0.3172(0.2240)
pattern1_2	-0.4485(0.4777)	-0.9597(0.4906)	-0.5451(0.4927)
pattern2_2	-0.6989(0.3324)	-0.6914(0.3323)	-0.7004(0.3325)
gender_2	-0.0709(0.2629)	-0.0424(0.2514)	-0.0725(0.2608)
intercept_3	0.4713(0.2326)	0.4503(0.2346)	0.4607(0.2321)
pattern1_3	-0.2846(0.5761)	-0.4108(0.5997)	-0.3162(0.5311)
pattern2_3	-0.5498(0.4615)	-0.5469(0.4639)	-0.5457(0.4620)
gender_3	0.2309(0.2778)	0.2654(0.2812)	0.2476(0.2779)
φ_{12}	3.1343(0.3469)	3.1410(0.3361)	3.1406(0.3444)
φ_{13}	2.0304(0.3084)	2.0168(0.3134)	2.0208(0.3112)
φ_{23}	2.8706(0.3589)	2.9654(0.3573)	2.8624(0.3561)
φ_{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_1	0.1878(0.2364)	0.1933(0.2356)	0.1882(0.2361)
pattern1_1	-0.9785(0.6495)	-1.0467(0.6485)	-0.9689(0.6510)
pattern2_1	-1.0648(0.5433)	-1.0659(0.5424)	-1.0664(0.5431)
gender_1	-0.5432(0.2866)	-0.5438(0.2862)	-0.5434(0.2865)
$\text{pattern1}\times\text{gender}_1$	0.4447(0.8463)	0.4748(0.8506)	0.4224(0.8521)
$\text{pattern2}\times\text{gender}_1$	1.2278(0.6989)	1.2271(0.6981)	1.2297(0.6991)
intercept_2	0.5089(0.2448)	0.5067(0.2456)	0.5087(0.2448)
pattern1_2	-0.8513(0.7509)	-1.5285(0.7823)	-0.9168(0.7770)
pattern2_2	-1.4699(0.5386)	-1.4657(0.5382)	-1.4711(0.5390)
gender_2	-0.3519(0.2937)	-0.3517(0.2943)	-0.3519(0.2938)
$\text{pattern1}\times\text{gender}_2$	0.6298(1.1582)	0.8790(0.9177)	0.5870(1.1482)
$\text{pattern2}\times\text{gender}_2$	1.3098(0.6927)	1.3032(0.6929)	1.3095(0.6934)
intercept_3	0.5916(0.2445)	0.5942(0.2446)	0.5922(0.2446)
pattern1_3	-0.5736(0.7847)	-0.8706(0.8602)	-0.6826(0.8516)
pattern2_3	-0.9877(0.6146)	-0.9937(0.6158)	-0.9868(0.6134)
gender_3	0.0561(0.2979)	0.0542(0.2978)	0.0559(0.2979)
$\text{pattern1}\times\text{gender}_3$	0.4706(0.9388)	0.7796(1.0740)	0.5907(0.9538)
$\text{pattern2}\times\text{gender}_3$	0.7610(0.8722)	0.7683(0.8693)	0.7612(0.8700)
φ_{12}	3.1328(0.3456)	3.1271(0.3359)	3.1412(0.3424)
φ_{13}	2.0235(0.3102)	2.0092(0.3140)	2.0143(0.3139)
φ_{23}	2.9035(0.3702)	2.9732(0.3564)	2.8943(0.3669)
φ_{123}	1.7912(0.9537)	1.9162(0.9506)	1.8026(0.9504)

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
Pattern 1				
intercept_1	-0.8109 (0.6009)	-0.8329 (0.4665)	-1.1210 (14.110)	-0.7455 (0.5671)
gender_1	-0.1446 (0.7988)	-0.0794 (0.6389)	6.0180 (21.981)	-0.3621 (0.9958)
intercept_2		-0.4042 (0.5910)	-13.062 (40.395)	-0.5381 (0.7385)
gender_2		0.2961 (0.9318)	0.5644 (0.8026)	0.3274 (0.9198)
intercept_3		0.0394 (0.7034)	-4.3260 (13.075)	-0.0734 (0.7340)
gender_3		0.5328 (0.8321)	12.117 (37.712)	0.6421 (0.8248)
φ_{12}		6.7429 (979.38)	-4010.0 (13044)	3.9641 (49.283)
φ_{13}		20.529 (2.86E6)	13.581 (2462i)	4.8504 (115.91)
φ_{23}		2.6890 (2.0828)	62.209 (205.04)	3.3699 (217.70)
φ_{123}		-4.5287 (7.69E43)	226.80 (45980)	-1.3710 (707.5i)
Pattern 2				
intercept_1	-0.8473 (0.4880)	1.4118 (7.2243)	1.4118 (7.2202)	1.4118 (7.2269)
gender_1	0.6802 (0.6371)	-1.9439 (8.2479)	-1.9439 (8.2396)	-1.9439 (8.2498)
intercept_2	-1.0986 (0.5164)	-4.4028 (12.501)	-4.4028 (12.497)	-4.4028 (12.500)
gender_2	1.0986 (0.6583)	4.1350 (12.071)	4.1350 (12.064)	4.1350 (12.070)
intercept_3		13.651 (46.566)	13.651 (46.555)	13.651 (46.566)
gender_3		-13.221 (46.112)	-13.221 (46.107)	-13.221 (46.111)
φ_{12}	2.9199 (0.8145)	3.9217 (28.152)	3.9217 (62.726)	3.9217 (48.873)
φ_{13}		2596.9 (2.86E6)	2596.9 (8249.1)	2596.9 (8609.7)
φ_{23}		-9.8258 (54.313)	-9.8258 (73.867)	-9.8258 (224.30)
φ_{123}		2581.9 (7.69E43)	2581.9 (2.62E22)	2581.9 (2.62E22)
Pattern 3				
intercept_1	0.1956 (0.2376)	0.1956 (0.2376)	0.1956 (0.2376)	0.1956 (0.2376)
gender_1	-0.5525 (0.2886)	-0.5525 (0.2886)	-0.5525 (0.2886)	-0.5525 (0.2886)
intercept_2	0.5107 (0.2437)	0.5107 (0.2437)	0.5107 (0.2437)	0.5107 (0.2437)
gender_2	-0.3522 (0.2929)	-0.3522 (0.2929)	-0.3522 (0.2929)	-0.3522 (0.2929)
intercept_3	0.5824 (0.2447)	0.5824 (0.2447)	0.5824 (0.2447)	0.5824 (0.2447)
gender_3	0.0679 (0.2987)	0.0679 (0.2987)	0.0679 (0.2987)	0.0679 (0.2987)
φ_{12}	3.1325 (0.3889)	3.1325 (0.3889)	3.1325 (0.3889)	3.1325 (0.3889)
φ_{13}	2.1026 (0.3533)	2.1026 (0.3533)	2.1026 (0.3533)	2.1026 (0.3533)
φ_{23}	2.9471 (0.3726)	2.9471 (0.3726)	2.9471 (0.3726)	2.9471 (0.3726)
φ_{123}	1.2110 (0.9510)	1.2110 (0.9510)	1.2110 (0.9510)	1.2110 (0.9510)

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
Pattern 1				
intercept_1	-0.8109 (0.6009)	-0.6141 (0.5074)	-0.6327 (0.5085)	-0.6031 (0.5070)
gender_1	-0.1446 (0.7988)	-0.1426 (0.6797)	-0.1074 (0.6778)	-0.1622 (0.6812)
intercept_2		-0.2541 (0.5666)	-0.7844 (0.6024)	-0.3323 (0.5774)
gender_2		0.2087 (0.8750)	0.3775 (0.7287)	0.2146 (0.8557)
intercept_3		0.0221 (0.6105)	-0.1987 (0.6374)	-0.0625 (0.6462)
gender_3		0.4174 (0.7459)	0.6426 (0.8118)	0.5247 (0.7506)
φ_{12}		2.2922 (1.0342)	2.4398 (0.8669)	2.3648 (0.9745)
φ_{13}		1.4683 (0.9096)	1.2864 (0.9326)	1.4161 (0.9922)
φ_{23}		1.5591 (0.9562)	1.8639 (0.9715)	1.4965 (1.0034)
φ_{123}		1.0355 (2.4118)	1.9070 (2.4134)	1.2348 (2.5027)
Pattern 2				
intercept_1	-0.8473 (0.4880)	-0.6859 (0.4294)	-0.6859 (0.4294)	-0.6859 (0.4294)
gender_1	0.6802 (0.6371)	0.5381 (0.5675)	0.5381 (0.5675)	0.5381 (0.5675)
intercept_2	-1.0986 (0.5164)	-0.8357 (0.4337)	-0.8357 (0.4337)	-0.8357 (0.4337)
gender_2	1.0986 (0.6583)	0.8075 (0.5716)	0.8075 (0.5716)	0.8075 (0.5716)
intercept_3		-0.3284 (0.4858)	-0.3284 (0.4858)	-0.3284 (0.4858)
gender_3		0.6901 (0.7085)	0.6901 (0.7085)	0.6901 (0.7085)
φ_{12}	2.9199 (0.8145)	2.3211 (0.6719)	2.3211 (0.6719)	2.3211 (0.6719)
φ_{13}		1.2766 (0.6825)	1.2766 (0.6825)	1.2766 (0.6825)
φ_{23}		2.3965 (0.8353)	2.3965 (0.8353)	2.3965 (0.8353)
φ_{123}		2.0700 (2.0344)	2.0700 (2.0344)	2.0700 (2.0344)