

Missing Not at Random Models and Their Missing at Random Counterparts, in Various Modelling Frameworks

Geert Molenberghs^{1,2}

Marc Aerts¹

Caroline Beunckens¹

An Creemers¹

Niel Hens¹

Michael G. Kenward³

Cristina Sotro¹

Geert Verbeke^{2,1}

¹ *I-BioStat, Universiteit Hasselt, Agoralaan 1, B-3590 Diepenbeek, Belgium;* ² *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium;* ³ *Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E7HT, UK* - geert.molenberghs@uhasselt.be

Resumo: Escreva aqui o resumo do artigo em Português.

Palavras-chave: palavra-chave1, palavra-chave2, palavra-chave3,...

Abstract: Over the last decade a variety of models to analyze incomplete multivariate and longitudinal data have been proposed, many of which allowing for the missingness to be not at random (MNAR), in the sense that the unobserved measurements influence the process governing missingness, in addition to influences coming from observed measurements and/or covariates. The fundamental problems implied by such models, to which we refer as sensitivity to unverifiable modelling assumptions, has, in turn, sparked off various strands of research in what is now termed *sensitivity analysis*. The nature of sensitivity originates from the fact that an MNAR model is not fully verifiable from the data, rendering the empirical distinction between MNAR and random missingness (MAR), where only covariates and observed outcomes influence missingness, hard or even impossible, unless one is prepared to accept the posited MNAR model in an unquestioning way. Based on Molenberghs *et al* (2007), we show that the empirical distinction between MAR and MNAR is not possible, in the sense that each MNAR model fit to a set of observed data can be reproduced exactly by an MAR counterpart. Of course, such a pair of models will produce different predictions of the unobserved outcomes, given the observed ones. While MAR are easy to characterize in the selection model and pattern-mixture modeling frameworks, this has not been the case for shared-parameter models (SPM). To mend this, based on Creemers *et al* (2008), we characterize MAR for the SPM framework, and study the form of the MAR counterpart for an MNAR model in this context. Also, for all settings combined, we examine the conditions that need to be imposed on models for longitudinal data to ensure that missingness does not depend on future occasions. Two illustrations are given, one based on the Slovenian Public Opinion survey, and one based on a clinical trial in onychomycosis.

Keywords: Contingency table; Ignorability; Missing completely at random; Pattern-mixture model; Selection model; Shared parameter model.

1 Introduction

Incomplete sets of data are common throughout all branches of empirical research. Incomplete data have always posed problems of imbalance in the data matrix, but more importantly incompleteness often destroys a trial's randomization justification or a survey's representativeness. The extent to which this happens depends on the nature of the missing data mechanism. Rubin (1976) distinguished between *missing complete at random* (MCAR), where the outcomes are independent of the mechanism governing missingness, *missing at random* (MAR), where there is dependence between both, but only in the sense that missingness may depend on the observed, but not further on the unobserved measurements. Finally, when a *missing not at random* (MNAR) mechanism operates, missingness depends on the unobserved outcomes, perhaps in addition to the observed ones.

During the same era, the *selection model* (SeM), *pattern-mixture model* (PMM), and *shared-parameter model* (SPM) frameworks have been established. In a selection model, the joint distribution of the i th subject's outcomes, denoted \mathbf{Y}_i , and vector of missingness indicators, written \mathbf{R}_i , is factored as the marginal outcome distribution and the conditional distribution of \mathbf{R}_i given \mathbf{Y}_i . A pattern-mixture approach starts from the reverse factorization. In a shared-parameter model, a set of latent variables, latent classes, and/or random effects is assumed to drive both the \mathbf{Y}_i and \mathbf{R}_i processes. An important version of such a model further asserts that, conditional on the latent variables, \mathbf{Y}_i and \mathbf{R}_i exhibit no further dependence. Rubin (1976) contributed the concept of *ignorability*, stating that under precise conditions, the missing data mechanism can be ignored when interest lies in inferences about the measurement process. Combined with regularity conditions, ignorability applies to MCAR and MAR combined, when likelihood or Bayesian inference routes are chosen, but the stricter MCAR condition is required for frequentist inferences to be generally valid.

Traditionally, such simple methods as a complete case analysis or simple forms of imputation (e.g., last observation carried forward) have been in use. While they have the advantage of restoring balance and/or a rectangular data matrix, it is sufficiently documented that such analyses are prone to severe bias and/or losses of efficiency (Molenberghs *et al*, 2004; Jansen *et al*, 2006) and should be avoided. Since a likelihood-based or Bayesian analysis is valid when the missing data mechanism is MAR, as long as all observed data are included into the analysis, the so-called *ignorability* property, so-called direct likelihood analyses, their Bayesian counterparts, or multiple imputation (Rubin, 1987), are regarded by many as candidates for the primary analyses of a study. When

semi-parametric inferences are desired, the methods proposed by Robins *et al* (1995) can be applied.

However, one can never exclude the possibility that MNAR models may be operating. Even though a variety of statistical models have been proposed for the MNAR situation (Diggle, and Kenward, 1994; Baker, 1995; Molenberghs *et al*, 1997; Troxel *et al*, 1998), and in spite of the dramatically increased computational power, such models are prone to considerable sensitivity. This was made clear by a variety of discussants to Diggle, and Kenward (1994), such as Laird (1994), Little (1994b), and Rubin (1994). Several authors have laid bare such sensitivities and proposed methods for informal and formal sensitivity analysis (Kenward, 1998; Robins *et al*, 1998; Molenberghs *et al*, 2001; Van Steen *et al*, 2001; Verbeke *et al*, 2001; Thijs *et al*, 2002; Jansen *et al*, 2003). Overviews are provided in Verbeke, and Molenberghs (2000) and Molenberghs, and Verbeke (2005).

One view is that testing the MAR null hypothesis against an MNAR alternative is of a conventional nature. While indeed Diggle, and Kenward (1994) have conducted such tests, it is very important to realize that they are conditional upon the alternative model holding.

One contribution of this paper, based on Molenberghs *et al* (2007), is to show that, strictly speaking, the correctness of the alternative model can only be verified in as far as it fits the *observed* data. Thus, evidence for or against MNAR can only be provided within a particular, predefined parametric family, the plausibility of which cannot be verified in empirical terms alone. We show that an overall (omnibus) assessment of MAR *versus* MNAR is not possible, since every MNAR model can be doubled up with a uniquely defined MAR counterpart, producing exactly the same fit as the original MNAR model, in the sense that it produces exactly the same predictions to the observed data (e.g., fitted counts in an incomplete contingency table) as the original MNAR model, and depending on exactly the same parameter vector. We show that, while this so-called MAR counterpart generally does not belong to a conventional parametric family, its existence has important ramifications. While this broad issue is still open to debate and even confusion, it has been pointed out in the literature. For example, the issue has been referred to, in general terms, by Little, and Rubin (2002) and, in a non- and semi-parametric context, by Gill, van der Laan, and Robins (1997). An excellent exposition, together with related references, can be found in Schafer and Graham (2002). Here, we focus on a general construction method for this counterpart, which we make explicit for the case of categorical data.

Now, the concept of MAR has typically been framed within the SeM framework, while Molenberghs *et al* (1998) provided a formulation in the PMM setting as well. For the particular case of longitudinal data with dropout, these authors derived a set of so-called identifying restrictions, to identify the model for the missing measurements given the observed ones within a missing-data pattern, consistent with MAR. Molenberghs *et al* (2007) showed that for every MNAR model, there is an MAR counterpart that produces exactly the same fit to the

observed data. Hence the original model and its MAR counterpart cannot be distinguished from one another. This can be viewed as a formalization of the ideas put forward in Jansen *et al* (2006). These authors focused on the SeM and PMM frameworks. Another contribution of this paper, based on Creemers *et al* (2008), we will characterize MAR in the SPM framework as well and a connection will be made with the MAR counterpart in the sense of Molenberghs *et al* (2007). To this end, a broad class of SPM will be defined. Implications for both non-monotone missing data as well as longitudinal data with dropout will be considered. In particular, in analogy with the PMM work by Kenward *et al* (2003), conditions will be derived to ensure future, unobserved measurements provide no information about dropout in addition to what is available from current and past measurements.

The rest of the paper is organized as follows. Section 2 introduces the two motivating case studies. In Section 3 we outline the necessary concepts, terminology, and notation. Section 4 presents our results regarding the MAR counterpart to MNAR models. In Section 5 the specific case of incomplete contingency tables is studied. Section 6 focuses on the specific case of shared-parameter models, while Section 7 examines what conditions need to be imposed on models for incomplete longitudinal data, to ensure that the missingness mechanism does not depend on future occasions. In Section 8 we apply the ideas developed to data from the Slovenian Public Opinion Survey, analyzed before by Rubin *et al* (1995) and Molenberghs *et al* (2001). Section 9 reports on the analysis of the onychomycosis data.

2 Motivating Case Studies

2.1 The Slovenian Public Opinion Survey

In 1991 Slovenians voted for independence from former Yugoslavia in a plebiscite. To prepare for this result, the Slovenian government collected data in the Slovenian Public Opinion Survey (SPO), a month prior to the plebiscite. Rubin *et al* (1995) studied the three fundamental questions added to the SPO and, in comparing it to the plebiscite's outcome, drew conclusions about the missing data process.

The three questions added were: (1) Are you in favour of Slovenian independence? (2) Are you in favour of Slovenia's secession from Yugoslavia? (3) Will you attend the plebiscite? In spite of their apparent equivalence, questions (1) and (2) are different since independence would have been possible in confederal form as well and therefore the secession question is added. Question (3) is highly relevant since the political decision was taken that not attending was treated as an effective NO to question (1). Thus, the primary estimand is the proportion θ of people that will be considered as voting YES, which is the fraction of people answering yes to both the attendance and independence question. The raw data are presented in Table 1. We will return to this question in Section 8.2.

Tabela 1: *Data from the Slovenian Public Opinion Survey. The Don't Know category is indicated by *.*

Secession	Attendance	Independence		
		Yes	No	*
Yes	Yes	1191	8	21
	No	8	0	4
	*	107	3	9
No	Yes	158	68	29
	No	7	14	3
	*	18	43	31
*	Yes	90	2	109
	No	1	2	25
	*	19	8	96

Molenberghs *et al* (2001) reanalyzed these data and used them as motivation to introduce their so-called *intervals of ignorance*, a formal way of incorporating uncertainty stemming from incompleteness into the analysis of incomplete contingency tables. To this end, they used the convenient model family proposed by Baker *et al* (1992). We will now introduce the model family.

2.2 An Onychomycosis Trial

The data introduced in this section were obtained from a randomized, double-blind, parallel group, multicenter study for the comparison of two oral treatments (in the sequel coded as *A* and *B*) for toenail dermatophyte onychomycosis (TDO), described in full detail by De Backer *et al* (1996). TDO is a common toenail infection, difficult to treat, affecting more than 2 out of 100 persons (Roberts, 1992). Anti-fungal compounds, classically used for treatment of TDO, need to be taken until the whole nail has grown out healthy. The development of new such compounds, however, has reduced the treatment duration to 3 months. The aim of the present study was to compare the efficacy and safety of 12 weeks of continuous therapy with treatment *A* or with treatment *B*.

In total, 2×189 patients, distributed over 36 centers, were randomized. Subjects were followed during 12 weeks (3 months) of treatment and followed further, up to a total of 48 weeks (12 months). Measurements were taken at baseline, every month during treatment, and every 3 months afterwards, resulting in a maximum of 7 measurements per subject. At the first occasion, the treating physician indicates one of the affected toenails as the target nail, the nail which will be followed over time. We will restrict our analyses to only those patients for which the target nail was one of the two big toenails. This

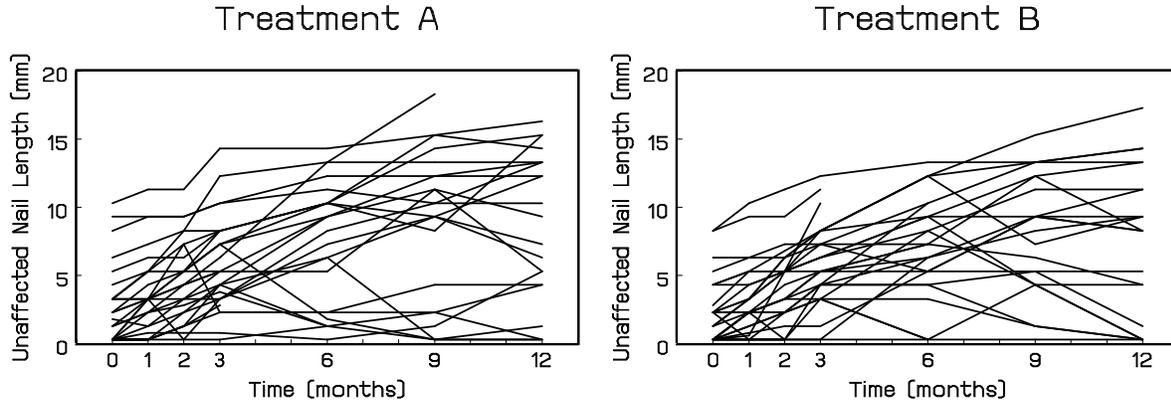


Figura 1: *Toenail Data. Individual profiles of 30 randomly selected subjects in each of the treatment groups in the toenail experiment.*

Tabela 2: *Toenail Data. Number and percentage of patients (N) with severe toenail infection, for each treatment arm separately.*

	Group A			Group B		
	# Severe	N	%	# Severe	N	%
Baseline	54	146	37.0%	55	148	37.2%
1 month	49	141	34.7%	48	147	32.6%
2 months	44	138	31.9%	40	145	27.6%
3 months	29	132	22.0%	29	140	20.7%
6 months	14	130	10.8%	8	133	6.0%
9 months	10	117	8.5%	8	127	6.3%
12 months	14	133	10.5%	6	131	4.6%

reduces our sample under consideration to 146 and 148 subjects, in group A and group B, respectively.

Figure 1 shows the observed profiles of 30 randomly selected subjects from treatment group A and treatment group B, respectively.

One of the responses of interest was the unaffected nail length, measured from the nail bed to the infected part of the nail, which is always at the free end of the nail, expressed in millimeters. This outcome has been studied extensively in Verbeke, and Molenberghs (2000). Another important outcome in this study was the severity of the infection, coded as 0 (not severe) or 1 (severe). The question of interest was whether the downward evolution of severe infection differs among treatment groups. A summary of the number of patients in the study at each time-point, and the number of patients with severe infections is given in Table 2. A graphical representation is given in Figure 2. Due to

Tabela 3: *Toenail Data. Number of available repeated measurements per subject, for each treatment arm separately.*

# Obs.	Group A		Group B	
	N	%	N	%
7	107	73.29%	117	79.05%
6	25	17.12%	14	9.46%
5	2	1.37%	8	5.41%
4	2	1.37%	4	2.70%
3	4	2.74%	3	2.03%
2	2	1.37%	1	0.68%
1	4	2.74%	1	0.68%
Total:	146	100%	148	100%

a variety of reasons, the outcome has been measured at all 7 scheduled time points, for only 224 (76%) out of the 298 participants. Table 3 summarizes the number of available repeated measurements per subject, for both treatment groups separately. We see that the occurrence of missingness is similar in both treatment groups.

3 Notation and Concepts

Let the random variable Y_{ij} denote the response of interest, for the i th study subject, designed to be measured at occasions t_{ij} , $i = 1, \dots, N$, $j = 1, \dots, n_i$. Independence across subjects is assumed. This setting covers both the longitudinal as well as the multivariate settings. In the latter case, $t_{ij} = t_j$ would merely be indicators for the various variables studied, and typically $n_i \equiv n$. The outcomes can conveniently be grouped into a vector $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})'$. In addition, define a vector of missingness indicators $\mathbf{R}_i = (R_{i1}, \dots, R_{in_i})'$ with $R_{ij} = 1$ if Y_{ij} is observed and 0 otherwise. In the specific case of dropout, \mathbf{R}_i can usefully be replaced by the dropout indicator

$$D_i = \sum_{j=1}^{n_i} R_{ij}.$$

Note that the concept of dropout refers to time-ordered variables, such as in longitudinal studies. For a complete sequence, $\mathbf{R}_i = \mathbf{1}$ and/or $D_i = n_i$. It is customary to split the vector \mathbf{Y}_i into observed (\mathbf{Y}_i^o) and missing (\mathbf{Y}_i^m) components, respectively. When \mathbf{R}_i is conditioned up, \mathbf{Y}_i^o and \mathbf{Y}_i^m explicitly refer to the observed and missing components. In the reverse case, they refer to an arbitrary partition of the outcome vector.

In principle, one would like to consider the density of the full data $f(\mathbf{y}_i, \mathbf{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi})$, where the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describe the measurement and missingness

Toenail data

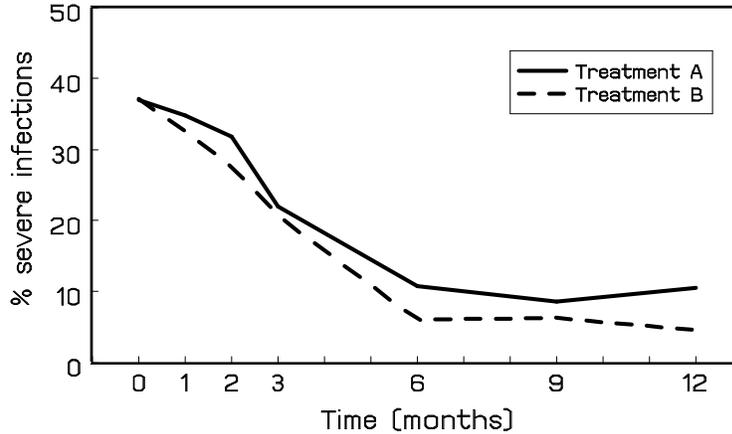


Figure 2: *Toenail Data*. Evolution of the observed percentage of severe toenail infections in the two treatment groups separately.

processes, respectively. Covariates are assumed to be measured and grouped in a vector \mathbf{x}_i but, throughout, are suppressed from notation. Although unusual, it is in principle possible for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ to have components in common.

This full density function can be factored in different ways, each leading to a different framework. They were mentioned briefly in the introduction. Here, we will present them more formally but in their standard form of appearance. In subsequent sections, they will be tailored to our needs, in particular the shared-parameter model.

The *selection model* (SeM) framework is based on the following factorization (Rubin, 1976; Little, and Rubin, 2002):

$$f(\mathbf{y}_i, \mathbf{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\psi}). \quad (1)$$

The first factor is the marginal density of the measurement process and the second one is the density of the missingness process, conditional on the outcomes. As an alternative, one can consider so-called *pattern-mixture models* (PMM; Little (1993, 1994a)) using the reversed factorization

$$f(\mathbf{y}_i, \mathbf{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \mathbf{r}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \boldsymbol{\psi}). \quad (2)$$

This can be seen as a mixture density over different populations, each of which is defined by the observed pattern of missingness.

Instead of using the selection modeling or pattern-mixture modeling frameworks, the measurement and the dropout process can be jointly modeled using a *shared-parameter model* (Wu, and Carroll, 1988; Wu, and Bailey, 1988, 1989;

TenHave *et al*, 1998; Follmann, and Wu, 1995; Little, 1995). One then might assume there exists a vector of random effects \mathbf{b}_i , conditional upon which the measurement and dropout processes are independent. This *shared-parameter model* (SPM) is formulated by way of the following factorization

$$f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{b}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{b}_i, \boldsymbol{\psi}), \quad (3)$$

and hence

$$f(\mathbf{y}_i, \mathbf{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = \int f(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{b}_i, \boldsymbol{\psi}) f(\mathbf{b}_i) d\mathbf{b}_i. \quad (4)$$

Here, \mathbf{b}_i are shared parameters, often considered to be random effects and following a specific parametric distribution. There are various other forms an SPM can take, and a more thorough discussion can be found in Section 6.

The taxonomy of missing data mechanisms, introduced by Rubin (1976) and informally described in the introduction, is customarily formalized using the second factor on the right hand side of selection-model factorization (1). A mechanism is MCAR if

$$f(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i | \boldsymbol{\psi}), \quad (5)$$

i.e., when the measurement and missingness processes are independent, perhaps conditional on covariates. For a given set of data, MAR holds when

$$f(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i | \mathbf{y}_i^o, \boldsymbol{\psi}), \quad (6)$$

strictly weaker than the MCAR condition, but still a simplification of the MNAR case, where missingness depends on the unobserved outcomes \mathbf{y}_i^m , regardless of the observed outcomes and the covariates.

Note that MCAR is equally trivial in the pattern-mixture model framework, where \mathbf{r}_i does not influence the mixture components, and in the shared-parameter model framework, where no random-effects are shared among the two factors in (3). The concept of MAR in the other framework is a different matter. As reviewed in the next section, a PMM characterization has been proposed by Molenberghs *et al* (1998). In Section 6, an SPM-based characterization will be provided, one of the contributions of this manuscript.

A final useful concept we need is *ignorability*. Note that the contribution to the likelihood of subject i , based on (1), equals

$$L_i = \int f(\mathbf{y}_i | \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi}) d\mathbf{y}_i^m. \quad (7)$$

In general, (7) does not simplify, but under MAR, we obtain:

$$L_i = f(\mathbf{y}_i^o | \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{y}_i^o, \boldsymbol{\psi}). \quad (8)$$

Hence, likelihood and Bayesian inferences for the measurement model parameters $\boldsymbol{\theta}$ can be made without explicitly formulating the missing data mechanism,

provided the parameters θ and ψ are distinct, meaning that their joint parameter space is the Cartesian product of the two component parameter spaces (Rubin, 1976). For Bayesian inferences, additionally the priors need to be independent (Little, and Rubin, 2002). It is precisely this result which makes so-called direct likelihood analyses, valid under MAR, viable candidates for the status of primary analysis in clinical trials and a variety of other settings (Molenberghs *et al*, 2004).

4 Every MNAR Model Has Got a MAR Counterpart

In this section, we will show that for every MNAR model fitted to a set of data, there is an MAR counterpart providing exactly the same fit to the data. Here, the concept of model fit should be understood as measured using such conventional methods as deviance measures and, of course, in as far as the observed data are concerned. The following steps are involved: (1) fitting an MNAR model to the data; (2) reformulating the fitted model in PMM form; (3) replacing the density or distribution of the unobserved measurements given the observed ones and given a particular response pattern by its MAR counterpart; (4) establishing that such an MAR counterpart uniquely exists. Throughout this section, we will suppress covariates \mathbf{x}_i from notation, but assume them to be present.

In the first step, we fit an MNAR model to the observed set of data. The observed data likelihood is:

$$L = \prod_i \int f(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i | \theta, \psi) d\mathbf{y}_i^m. \quad (9)$$

Upon denoting the obtained parameter estimates, e.g., obtained by likelihood-based or Bayesian methods, by $\hat{\theta}$ and $\hat{\psi}$ respectively, the fit to the hypothetical full data is

$$f(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i | \hat{\theta}, \hat{\psi}) = f(\mathbf{y}_i^o, \mathbf{y}_i^m | \hat{\theta}) f(\mathbf{r}_i | \mathbf{y}_i^o, \mathbf{y}_i^m, \hat{\psi}). \quad (10)$$

To undertake the second step, full density (10) can be re-expressed in PMM form as:

$$\begin{aligned} & f(\mathbf{y}_i^o, \mathbf{y}_i^m | \mathbf{r}_i, \hat{\theta}, \hat{\psi}) f(\mathbf{r}_i | \hat{\theta}, \hat{\psi}) \\ &= f(\mathbf{y}_i^o | \mathbf{r}_i, \hat{\theta}, \hat{\psi}) f(\mathbf{r}_i | \hat{\theta}, \hat{\psi}) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i, \hat{\theta}, \hat{\psi}). \end{aligned} \quad (11)$$

A similar reformulation can be considered for an SPM. In a PMM, the model will have been expressed in this form to begin with.

Note that, in line with PMM theory, the final term on the right hand side of (11), $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i, \hat{\theta}, \hat{\psi})$, is not identified from the observed data. In this case, it is determined solely from modelling assumptions. Within the PMM framework,

identifying restrictions have to be considered (Little, 1994a; Molenberghs *et al*, 1998; Kenward *et al*, 2003).

The third step requires replacing this factor by the appropriate MAR counterpart. To this end, we need the following lemma, formulating MAR equivalently within the PMM framework.

Lemma 1 *In the PMM framework, the missing data mechanism is MAR if and only if*

$$f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i^m | \mathbf{y}_i^o, \boldsymbol{\theta}). \quad (12)$$

This means that, in a given pattern, the conditional distribution of the unobserved components given the observed ones equals the corresponding distribution marginalized over the patterns. The proof, which is rather straightforward and similar to what can be found in Molenberghs *et al* (1998), is reported in Molenberghs *et al* (2007). Note that, owing to this result, MAR can be formulated in terms of R given Y , but also in terms of Y given R .

Using Lemma 1, it is clear that $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$ needs to be replaced with

$$h(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i) = h(\mathbf{y}_i^m | \mathbf{y}_i^o) = f(\mathbf{y}_i^m | \mathbf{y}_i^o, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}), \quad (13)$$

where the $h(\cdot)$ notation is used for shorthand purposes. Note that the density in (13) follows from the SeM-type marginal density of the complete data vector. Sometimes, therefore, it may be more convenient to replace the notation \mathbf{y}_i^o and \mathbf{y}_i^m by one that explicitly indicates which components are observed and missing in pattern \mathbf{r}_i under consideration:

$$h(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i) = h(\mathbf{y}_i^m | \mathbf{y}_i^o) = f[(y_{ij})_{r_j=0} | (y_{ij})_{r_j=1}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}]. \quad (14)$$

Thus, (14) provides a unique way of extending the model fit to the observed data, belonging to the MAR family. As stated before, the above construction does not lead to a member of a conventional parametric family. While this obviously implies limitations on its use, such is not dissimilar to the construction of some semi- and non-parametric estimators. Also, it helps to understand that an overall, definitive conclusion about the nature of the missing data mechanism is not possible, even though one can make progress if attention is confined to a given parametric family, in which one puts sufficiently strong prior belief. To show formally that the fit remains the same, we consider the observed-data likelihood based on (9) and (11):

$$\begin{aligned} \hat{L} &= \prod_i \int f(\mathbf{y}_i^o, \mathbf{y}_i^m | \hat{\boldsymbol{\theta}}) f(\mathbf{r}_i | \mathbf{y}_i^o, \mathbf{y}_i^m, \hat{\boldsymbol{\psi}}) d\mathbf{y}_i^m \\ &= \prod_i \int f(\mathbf{y}_i^o | \mathbf{r}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) f(\mathbf{r}_i | \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) d\mathbf{y}_i^m \\ &= \prod_i f(\mathbf{y}_i^o | \mathbf{r}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) f(\mathbf{r}_i | \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) \end{aligned}$$

$$= \prod_i \int f(\mathbf{y}_i^o | \mathbf{r}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) f(\mathbf{r}_i | \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}}) h(\mathbf{y}_i^m | \mathbf{y}_i^o) d\mathbf{y}_i^m.$$

The above results justify the following theorem:

Theorem 1 *Every fit to the observed data, obtained from fitting an MNAR model to a set of incomplete data, is exactly reproducible from an MAR decomposition.*

The key computational consequence is the need to compute $h(\mathbf{y}_i^m | \mathbf{y}_i^o)$ in (13) or (14). This means, for each pattern, the conditional density of the unobserved measurements given the observed ones needs to be extracted from the marginal distribution of the complete set of measurements. Molenberghs *et al* (1998) have shown that, for the case of dropout, the so-called *available case missing value restrictions* (ACMV) provide a practical computational scheme. Precisely, ACMV states that

$$\forall t \geq 2, \forall s < t : f(y_{it} | y_{i1}, \dots, y_{i,t-1}, d_i = s) = f(y_{it} | y_{i1}, \dots, y_{i,t-1}, d_i \geq t). \quad (15)$$

In other words, the density of a missing measurement, conditional on the measurement history, is determined from the corresponding density over all patterns for which all of these measurements are observed. For example, the density of the third measurement in a sequence, given the first and second ones, in patterns with only 1 or 2 measurements taken, is determined from the corresponding density over all patterns with 3 or more measurements. Thijs *et al* (2002) and Verbeke, and Molenberghs (2000) (p. 347) derived a practical computational method for the factors in (15):

$$\begin{aligned} & f(y_{it} | y_{i1}, \dots, y_{i,t-1}, d_i = s) \\ &= \frac{\sum_{d=s}^n \alpha_d f_d(y_{i1}, \dots, y_{is})}{\sum_{d=s}^n \alpha_d f_d(y_{i1}, \dots, y_{i,s-1})} \end{aligned} \quad (16)$$

$$= \sum_{d=s}^n \left(\frac{\alpha_d f_d(y_{i1}, \dots, y_{i,s-1})}{\sum_{d=s}^{n_i} \alpha_d f_d(y_{i1}, \dots, y_{i,s-1})} \right) f_d(y_s | y_{i1}, \dots, y_{i,s-1}). \quad (17)$$

Here, α_d is the probability to belong to pattern d .

The above identifications for the monotone case are useful in case an MNAR pattern-mixture model has been fitted to begin with, since then the identifications under MAR can be calculated from the pattern-specific marginal distributions. When a selection model has been fitted in the initial step, $f(y_{i1}, \dots, y_{i n_i} | \hat{\boldsymbol{\theta}})$ has been estimated, from which all conditional distributions, needed in (14), can be derived. When the initial model is an MNAR PMM model and the missing data patterns are non-monotone, then it is necessary to first rewrite the PMM in SeM form, and derive the required conditional distributions from the so-obtained SeM measurement model. This essentially comes down to calculating a weighted average of the pattern-specific measurement models. In some

cases, such as for contingency tables, this step can be done in an alternative way by fitting a saturated MAR selection model to the fit obtained from the PMM model.

We will illustrate and contrast the monotone and non-monotone cases using a bivariate and trivariate outcome with dropout on the one hand and a bivariate non-monotone outcome on the other hand. While the theorem applies to both the monotone and non-monotone settings, it is insightful to see that only for the former relatively simple and intuitively appealing expressions arise, while the latter setting involves the need for iterative computation. In the next section, the aforementioned general contingency table setting to which a PMM has been fitted, will be studied.

4.1 A Bivariate Outcome With Dropout

Here and in the following examples, we will present and equate the SeM and PMM decompositions, enabling us to derive expressions for the MAR counterparts. It is interesting and straightforward to derive results for the MCAR case, and hence these will be presented, too.

Dropping covariates, parameters, and the subject index i from notation, the SeM-PMM equivalence for the case of two outcomes, the first of which is always observed but the second one partially missing, is given by:

$$\begin{aligned} f(y_1, y_2)\tilde{g}(d=2|y_1, y_2) &= f_2(y_1, y_2)\tilde{\alpha}(d=2), \\ f(y_1, y_2)\tilde{g}(d=1|y_1, y_2) &= f_1(y_1, y_2)\tilde{\alpha}(d=1). \end{aligned}$$

Note that this is the setting considering by Glynn *et al* (1986). Here, $\tilde{g}(\cdot)$ is used for the SeM dropout model, with $\tilde{\alpha}(\cdot)$ denoting the PMM probabilities to belong to one of the patterns. Since $\tilde{\alpha}(d=1) + \tilde{\alpha}(d=2) = 1$ and a similar result holds for the $\tilde{g}(\cdot)$ functions, it is convenient to write:

$$f(y_1, y_2)g(y_1, y_2) = f_2(y_1, y_2)\alpha \quad (18)$$

$$f(y_1, y_2)[1 - g(y_1, y_2)] = f_1(y_1, y_2)[1 - \alpha]. \quad (19)$$

Assuming MCAR, it is clear that $\alpha = g(y_1, y_2)$, producing, without any difficulty:

$$f(y_1, y_2) = f_2(y_1, y_2) = f_1(y_1, y_2). \quad (20)$$

Under MAR, y_2 has to be removed from $g(\cdot)$ for incomplete observations, but since we assume a single parametric function for the missingness model, it follows that $g(y_1, y_2) = g(y_1)$ and hence (18) produces

$$f(y_1)f(y_2|y_1)g(y_1) = f_2(y_1)f_2(y_2|y_1)\alpha.$$

Upon reordering, we find:

$$\frac{f(y_1)g(y_1)}{f_2(y_1)\alpha} = \frac{f_2(y_2|y_1)}{f(y_2|y_1)}. \quad (21)$$

The same arguments can be applied to (19), from which we derive:

$$f(y_2|y_1) = f_2(y_2|y_1) = f_1(y_2|y_1). \quad (22)$$

Note that (22) is strictly weaker than (20). The last term in (22) is not identified by itself, and hence, we see it needs to be set equal to its counterpart from the completers which, in turn, is equal to the marginal distribution. This is in agreement with (14) as well as with the specific identifications applicable in the monotone and hence ACMV setting.

4.2 A Trivariate Outcome With Dropout

Note that identification (22) does not involve mixtures. This changes as soon as there are three or more outcomes. The equations corresponding to (18)–(19), specialized to the MAR case, are:

$$f(y_1, y_2, y_3)g_0 = f_0(y_1, y_2, y_3)\alpha_0, \quad (23)$$

$$f(y_1, y_2, y_3)g_1(y_1) = f_1(y_1, y_2, y_3)\alpha_1, \quad (24)$$

$$f(y_1, y_2, y_3)g_2(y_1, y_2) = f_2(y_1, y_2, y_3)\alpha_2, \quad (25)$$

$$f(y_1, y_2, y_3)g_3(y_1, y_2) = f_3(y_1, y_2, y_3)\alpha_3. \quad (26)$$

We have chosen to include pattern 0, the one without follow-up measurements, as well, and will return to this one. We could write $g_3(\cdot)$ as a function of y_3 as well, but because the sum of the $g_d(\cdot)$ equals one, it is clear that $g_3(\cdot)$ ought to be independent of y_3 . With arguments similar to the ones developed in the case of two measurements, we can rewrite (26) as:

$$\frac{f(y_1, y_2)}{f_3(y_1, y_2)} \cdot \frac{g_3(y_1, y_2)}{\alpha_3} = \frac{f_3(y_3|y_1, y_2)}{f(y_3|y_1, y_2)}.$$

Exactly the same consideration can be made based on (25), and hence

$$f_3(y_3|y_1, y_2) = f(y_3|y_1, y_2) = f_2(y_3|y_1, y_2). \quad (27)$$

The first factor identifies the second one, and hence also the third one. Starting from (24), we obtain:

$$f_1(y_2, y_3|y_1) = f(y_2, y_3|y_1),$$

which produces, in fact, two separate identities:

$$f_1(y_2|y_1) = f(y_2|y_1), \quad (28)$$

$$f_1(y_3|y_1, y_2) = f(y_3|y_1, y_2) = f_3(y_3|y_1, y_2) = f_2(y_3|y_1, y_2). \quad (29)$$

For the latter one, identity (27) has been used as well. The density $f(y_2|y_1)$, needed in (28), is determined from the general ACMV result (17):

$$f(y_2|y_1) = \frac{\alpha_2 f_2(y_2|y_1) + \alpha_3 f_3(y_2|y_1)}{\alpha_2 + \alpha_3}.$$

Finally, turning attention to (23), it is clear that $g_0 = \alpha_0$ and hence also $f_0(y_1, y_2, y_3) = f(y_1, y_2, y_3)$. From the latter density, only $f(y_1)$ has not been determined yet, but this one follows again very easily from the general ACMV result:

$$f(y_1) = \frac{\alpha_1 f_1(y_1) + \alpha_2 f_2(y_1) + \alpha_3 f_3(y_1)}{\alpha_1 + \alpha_2 + \alpha_3}.$$

In summary, the necessary MAR identifications easily follow from both the PMM and the SeM formulations of the model.

4.3 A Bivariate Outcome With Non-Monotone Missingness

The counterparts to (18)–(19) and (23)–(26) for a bivariate outcome with non-monotone missingness are

$$f(y_1, y_2)g_{00}(y_1, y_2) = f_{00}(y_1, y_2)\alpha_{00}, \quad (30)$$

$$f(y_1, y_2)g_{10}(y_1, y_2) = f_{10}(y_1, y_2)\alpha_{10}, \quad (31)$$

$$f(y_1, y_2)g_{01}(y_1, y_2) = f_{01}(y_1, y_2)\alpha_{01}, \quad (32)$$

$$f(y_1, y_2)g_{11}(y_1, y_2) = f_{11}(y_1, y_2)\alpha_{11}. \quad (33)$$

Clearly, under MCAR, the $g_{r_1 r_2}(\cdot)$ functions do not depend on the outcomes and hence $f_{r_1 r_2}(y_1, y_2) = f(y_1, y_2)$ for all four patterns. For the MAR case, (30)–(33) simplify to

$$f(y_1, y_2)g_{00} = f_{00}(y_1, y_2)\alpha_{00}, \quad (34)$$

$$f(y_1, y_2)g_{10}(y_1) = f_{10}(y_1, y_2)\alpha_{10}, \quad (35)$$

$$f(y_1, y_2)g_{01}(y_2) = f_{01}(y_1, y_2)\alpha_{01}, \quad (36)$$

$$f(y_1, y_2)g_{11}(y_1, y_2) = f_{11}(y_1, y_2)\alpha_{11}. \quad (37)$$

Observe there are four identifications across the $g_{r_1 r_2}(y_1, y_2)$ functions:

$$g_{00} + g_{10}(y_1) + g_{01}(y_2) + g_{11}(y_1, y_2) = 1,$$

for each (y_1, y_2) . Also $\sum_{r_1, r_2} \alpha_{r_1, r_2} = 1$. Applying the usual algebra to (34)–(37), we obtain three identifications for the unobservable densities:

$$f_{00}(y_1, y_2) = f(y_1, y_2), \quad (38)$$

$$f_{10}(y_1|y_2) = f(y_1|y_2), \quad (39)$$

$$f_{01}(y_2|y_1) = f(y_2|y_1). \quad (40)$$

Using these in conjunction with the identifiable parts of the distributions yields the MAR counterpart.

5 Incomplete Contingency Tables

In Sections 4.1–4.3 we have derived general identification schemes for an MAR extension of a fitted model to a binary or trivariate outcome with dropout, as well as to a bivariate outcome with non-monotone missingness. Whereas the monotone cases provide explicit expressions in terms of the pattern-specific densities, (38)–(40) provide an identification only in terms of the marginal probability. This in itself is not a problem, since the marginal density is always available, either directly when a SeM is fitted, or through marginalization when a PMM or an SPM is fitted.

In the specific case of contingency tables, further progress can be made. Indeed, we can show a saturated MAR model is always available, for any incomplete contingency table setting. This implies one can start from the fit of an MNAR model to the observed data, and then extend it, using this result, towards MAR. We will present the general result and then discuss its precise implications for practice.

Assume we have a $\prod_{k=1}^n c_k$ contingency table with supplemental margins, where k indexes the n dimensions in the table and c_k is the number of alternatives the k th categorical variable can take. The table of completers is indexed by $\mathbf{r} = \mathbf{1} = (1, \dots, 1)$. A particular incomplete table is indexed by a $\mathbf{r} \neq \mathbf{1}$. The full set of tables can but does not have to be present. The number of cells is:

$$\#\text{cells} = \sum_{\mathbf{r}} \prod_{k=1}^n c_k^{r_k}. \quad (41)$$

Denote the measurement model probabilities by $p_{\mathbf{j}} = p_{j_1 \dots j_n}$ for $j_k = 1, \dots, c_k$ and $k = 1, \dots, n$. Clearly, these probabilities sum to one. The missingness probabilities, assuming MAR, are:

$$p(\mathbf{r}|\mathbf{j}) = \begin{cases} p(\mathbf{r}|j_k \text{ with } r_k = 1) & \text{if } \mathbf{r} \neq \mathbf{1}, \\ 1 - \sum_{\mathbf{r} \neq \mathbf{1}} p(\mathbf{r}|\mathbf{j}) & \text{if } \mathbf{r} = \mathbf{1}. \end{cases} \quad (42)$$

Summing over \mathbf{r} implies summing over those patterns for which actual observations are available. The number of parameters in the saturated model is

$$\#\text{parameters} = \left(\prod_{k=1}^n c_k - 1 \right) + \sum_{\mathbf{r} \neq \mathbf{1}} \prod_{k=1}^n c_k^{r_k}. \quad (43)$$

The first term in (43) is for the measurement model, the second one is for the missingness model. Clearly, the number of parameters equals one less than the number of cells, establishing the claim. The situation where covariates are present is covered automatically, merely by considering one extra dimension in the contingency table, $j = 0$ say, with c_0 referring to the total number of covariate levels in the set of data.

We will now study the implications for the simple but important settings studied in Sections 4.1 and 4.3.

5.1 A Bivariate Contingency Table With Dropout

In Section 4.1 identifications have been derived for the bivariate case with monotone missingness. For contingency tables, these can be derived as well by further fitting the saturated MAR model, described in the previous section, to the fit obtained from the original MNAR model. Denote the counts obtained from the fit of the original model by $z_{2,jk}$ and $z_{1,j}$, for the completers and dropouts, respectively. Denote the measurement model probabilities by p_{jk} and the dropout probabilities by q_j . Then, due to ignorability, the likelihood factors into two components:

$$\ell_1 = \sum_{j,k} z_{2,jk} \ln p_{jk} + \sum_j z_{1,j} \ln p_{j+} - \lambda \left(\sum_{j,k} p_{jk} - 1 \right), \quad (44)$$

$$\ell_2 = \sum_{j,k} z_{2,jk} \ln q_j + \sum_j z_{1,j} \ln(1 - q_j). \quad (45)$$

We have used an undetermined Lagrange multiplier λ to incorporate the sum constraint on the marginal probabilities. Solving the score equations for (44) and (45) produces, with simple and well-known algebra:

$$\widehat{p}_{jk} = \frac{1}{n} z_{2,jk} \left(\frac{z_{2,j+} + z_{1,j}}{z_{2,j+}} \right), \quad (46)$$

$$\widehat{q}_j = \frac{z_{2,j+}}{z_{2,j+} + z_{1,j}}, \quad (47)$$

where n is the total sample size. Combining parameter estimates leads to the new, MAR-based, fitted counts:

$$\widehat{z}_{2,jk} = n \widehat{p}_{jk} \widehat{q}_j = z_{2,jk}, \quad (48)$$

$$\widehat{z}_{1,jk} = n \widehat{p}_{jk} (1 - \widehat{q}_j) = z_{1,j} \frac{z_{2,jk}}{z_{2,j+}}, \quad (49)$$

$$\widehat{z}_{1,j+} = z_{1,j+}. \quad (50)$$

From (48) and (50) it is clear that the fit in terms of the observed data has not changed. The expansion of the incomplete data into a complete one is described by (49). Equations (48) and (49) can be used to produce the MAR counterpart to the original model, without any additional calculations. This is not so simple for the non-monotone case, as we will show next.

5.2 A Bivariate Contingency Table With Non-Monotone Missingness

The counterparts to (44)–(45) for this case are:

$$\begin{aligned} \ell_1 = & \sum_{j,k} z_{11,jk} \ln p_{jk} + \sum_j z_{10,j} \ln p_{j+} + \sum_k z_{01,k} \ln p_{+k} \\ & + z_{00} \ln p_{++} - \lambda \left(\sum_{j,k} p_{jk} - 1 \right), \end{aligned} \quad (51)$$

$$\begin{aligned} \ell_2 = & \sum_{j,k} z_{11,jk} \ln(1 - q_{10,j} - q_{01,k} - q_{00}) + \sum_j z_{10,j} \ln q_{10,j} \\ & + \sum_k z_{01,k} \ln q_{01,k} + z_{00} \ln q_{00}. \end{aligned} \quad (52)$$

Notation has been modified in accordance with the design. The q quantities correspond to the $g(\cdot)$ model in Section 4.3.

While $p_{++} = 1$ and hence z_{00} does not contribute information to the measurement probabilities, it does add to the estimation of the missingness model.

Deriving the score equations from (51) and (52) is straightforward but, unlike in the previous section, no closed form exists. Chen, and Fienberg (1974) derived an iterative scheme for the probabilities p_{jk} , based on setting the expected sufficient statistics equal to their *complete-data* counterparts:

$$np_{jk} = z_{11,jk} + z_{10,j} \frac{p_{jk}}{p_{j+}} + z_{01,k} \frac{p_{jk}}{p_{+k}} + z_{00} \frac{p_{jk}}{p_{++}},$$

(with $p_{++} = 1$) and hence

$$(n - z_{00})p_{jk} = z_{11,jk} + z_{10,j} \frac{p_{jk}}{p_{j+}} + z_{01,k} \frac{p_{jk}}{p_{+k}}. \quad (53)$$

The same equation is obtained from the first derivative of (51). Chen and Fienberg's iterative scheme results from initiating the process with a set of starting values for the p_{jk} , e.g., from the completers, and then evaluating the right hand side of (53). Equating it to the left hand side provides an update for the parameters. The process is repeated until convergence.

While there are no closed-form counterparts to (46) and (47), the expressions equivalent to (48)–(50) are

$$\widehat{z_{11,jk}} = z_{11,jk}, \quad (54)$$

$$\widehat{z_{10,jk}} = z_{10,j} \frac{p_{jk}}{p_{j+}}, \quad (55)$$

$$\widehat{z_{01,jk}} = z_{01,k} \frac{p_{jk}}{p_{+k}}, \quad (56)$$

$$\widehat{z_{00,jk}} = z_{00} p_{jk}. \quad (57)$$

However, there is an important difference between (48)–(50) on the one hand and (54)–(57) on the other hand. In the monotone case, the expressions on the right hand side are in terms of the counts z only, whereas here the marginal probabilities p_{jk} intervene, which have to be determined from a numerical fit.

The practical use of the results in this section are illustrated next on data from the Slovenian Public Opinion Survey.

6 Shared-parameter Models and Missingness at Random

SPM's are closely linked to the joint modeling of longitudinal and time-to-event data, a class of models considered for at least three reasons. First, a time-to-event outcome may be measured in terms of a longitudinal covariate. Such a joint model then allows, in a natural way, for incorporation of measurement error present in the longitudinal covariate into the model. Second, a number of researchers have used joint modeling methods to exploit longitudinal markers as surrogates for survival (Tsiatis, DeGruttola, and Wulfsohn, 1995; Xu and Zeger, 2001a; Henderson, Diggle, and Dobson, 2000; Renard *et al*, 2002).

Third, and of most relevance here, such joint models can be used when incomplete longitudinal data are collected. Important early references to such models are Wu, and Carroll (1988), Wu, and Bailey (1988), and Wu, and Bailey (1989). Wu, and Bailey (1988) proposed such a model for what they termed informative right censoring. For a continuous response, Wu, and Carroll (1988) suggested using a conventional Gaussian random-coefficient model combined with an appropriate model for to time to dropout, such as proportional hazards, logistic or probit regression. The combination of probit and Gaussian responses allows explicit solution of the integral and was used in their application.

In a slightly different approach to modeling dropout time as a continuous variable in the latent variable setting, Schluchter (1992) and DeGruttola and Tu (1994) proposed joint multivariate Gaussian distributions for the latent variable(s) of the response process and a variable representing time to dropout. The correlation between these variables induces dependence between dropout and response. Rizopoulos, Verbeke, and Molenberghs (2007) study the impact of random-effects misspecification in a shared parameter model. Beunckens *et al* (2007a) combine continuous random effects with latent classes, leading to the simultaneous use of mixture and mixed-effects models ideas. It is very natural to handle random-coefficient models, and in particular shared-parameter models, in a Bayesian framework. Examples in the missing value setting are provided by Best *et al* (1996) and Carpenter, Pocock, and Lamm (2002). Further references include Pawitan and Self (1993); Taylor *et al* (1994); Faucett and Thomas (1996); Lavalley and DeGruttola (1996); Hogan and Laird (1997, 1998); Wulfsohn and Tsiatis (1997) and Xu and Zeger (2001b).

Models of this type handle non-monotone missingness quite conveniently through random effects. There are many ways in which such models can be extended and generalized. Nevertheless, these models seem to defy an easy,

elegant characterization of MAR, which is the topic of what follows.

In Section 3, the commonly used definition (3) of an SPM is presented. However, the preceding review makes clear that not all authors employ the same definition. Before passing on to the definition we will employ here, it is therefore instructive to take a more general position, also considered by Little (1995), based on augmenting the joint density of $(\mathbf{y}_i, \mathbf{r}_i)$ with a vector of random effects \mathbf{b}_i :

$$f(\mathbf{y}_i, \mathbf{r}_i, \mathbf{b}_i | \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}_i), \quad (58)$$

where \mathbf{x}_i is now explicitly included to parametrize the random-effects distribution. As before, covariates are allowed to be present, perhaps taking the form of different sets that each describe one of the three components. Again, they are suppressed from notation. Based on (58), one can still consider the selection-model factorization:

$$f(\mathbf{y}_i, \mathbf{r}_i, \mathbf{b}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\psi}) f(\mathbf{b}_i | \mathbf{x}_i) \quad (59)$$

and, likewise, the pattern-mixture model factorization:

$$f(\mathbf{y}_i, \mathbf{r}_i, \mathbf{b}_i | \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}_i) = f(\mathbf{y}_i | \mathbf{r}_i, \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{b}_i, \boldsymbol{\psi}) f(\mathbf{b}_i | \mathbf{x}_i). \quad (60)$$

The notation is the same as in Section 3, with in addition \mathbf{x}_i parameters describing the random-effects distribution. Little (1995) refers to such decompositions as random-coefficient selection and pattern-mixture models, respectively. Obviously, SeM (1) and PMM (2) follow by removing the random effects from (59) and (60), respectively or, at least, not having them in common between the models for \mathbf{Y}_i and \mathbf{R}_i .

An important simplification, leading to the already-defined SPM (3), arises when \mathbf{Y}_i and \mathbf{R}_i are assumed independent, given the random effects, *i.e.*, when conditional independence assumptions are made. Spelling out the model in full produces:

$$f(\mathbf{y}_i, \mathbf{r}_i, \mathbf{b}_i | \boldsymbol{\theta}, \boldsymbol{\psi}, \mathbf{x}_i) = f(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{b}_i, \boldsymbol{\psi}) f(\mathbf{b}_i | \mathbf{x}_i). \quad (61)$$

Model (61) corresponds to (3), but now also the distribution of the random effects has been spelled out explicitly. This model was entertained by Follmann, and Wu (1995). Note that, when \mathbf{b}_i is assumed to be discrete, a latent-class or mixture model follows.

We are now in a position to introduce the SPM framework needed for our purposes. Note that most formulations assume that a single, common set \mathbf{b}_i drives the entire process. Whilst holding on to the conditional-independence assumption, we will expand \mathbf{b}_i to a set of latent structures, as in the following definition.

Definition 1 (A General Shared-parameter Model Family.) *We define a general shared-parameter model as one of the form*

$$f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{q}_i), \quad (62)$$

where \mathbf{g}_i , \mathbf{h}_i , \mathbf{j}_i , \mathbf{k}_i , $\boldsymbol{\ell}_i$, \mathbf{m}_i , and \mathbf{q}_i are independent random-effects vectors (vectors of latent variables).

For convenience, write

$$\mathbf{b}_i = (\mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{k}_i, \boldsymbol{\ell}_i, \mathbf{m}_i, \mathbf{q}_i). \quad (63)$$

Several remarks are in place. First, this is the most general random-effects model that can be considered in the sense that \mathbf{g}_i is common to all three factors in (62), \mathbf{h}_i , \mathbf{j}_i , and \mathbf{k}_i are shared between a pair of factors, and $\boldsymbol{\ell}_i$, \mathbf{m}_i , and \mathbf{q}_i are restricted to a single factor. Depending on the application, one may choose to either retain all random effects or to omit some. It will then be useful to have a perspective on the implications of such simplifications, preferably also in terms of the missing data mechanism operating. This is why we will establish conditions under which MAR operates on the one hand, and missingness does not depend on future, unobserved measurements in a longitudinal context on the other hand. Second, in full generality, model (62) may come across as somewhat contrived. Our objective is not to postulate (62) as a model of use in every possible application of SPM, but rather as the most general SPM from which substantively appropriate models follow as sub-classes. Related to this, it appears (62) assumes two different distributions for the outcome vector, *i.e.*, divorcing the observed from the missing components. This is not entirely the case because \mathbf{g}_i and \mathbf{h}_i still tie both factors together. The impact of \mathbf{j}_i , \mathbf{k}_i , $\boldsymbol{\ell}_i$, and \mathbf{m}_i is to modify one's latent process in terms of missingness. In other words, the most general model assumes that observed and missing components are governed in part by common processes and partly by separate processes. Third, in principle, we could expand (62) with the densities of the random effects. This is generally not necessary for our purposes, though. Fourth, the assumption of independent random-effects vectors is not restrictive, because association is captured through the sets common to at least two factors. Fifth, conventional SPM formulation (61) follows by removing all random effects but \mathbf{g}_i .

Definition (62) will allow us to derive a general characterization of MAR in the SPM framework. It is instructive to set out by deriving an elegant set of sufficient conditions. Thereafter, necessity will be addressed. To this end, we can start from either the SeM-based definition (6) or the PMM characterization as laid out in Lemma 1.

Starting from the SeM definition, and assuming \mathbf{g}_i , \mathbf{h}_i , and \mathbf{k}_i are zero, we can show that MAR follows:

$$\begin{aligned} f(\mathbf{r}_i | \mathbf{y}_i^o, \mathbf{y}_i^m) &= \frac{f(\mathbf{r}_i, \mathbf{y}_i^o, \mathbf{y}_i^m)}{f(\mathbf{y}_i^o, \mathbf{y}_i^m)} \\ &= \frac{\int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) f(\mathbf{r}_i | \mathbf{j}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \end{aligned}$$

$$\begin{aligned}
&= \frac{\int f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) d\mathbf{m}_i \cdot \int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{r}_i | \mathbf{j}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) d\mathbf{m}_i \cdot \int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \\
&= \frac{f(\mathbf{y}_i^o, \mathbf{r}_i)}{f(\mathbf{y}_i^o)} = f(\mathbf{r}_i | \mathbf{y}_i^o),
\end{aligned}$$

where integration over \mathbf{b}_i is shorthand for integration over all component vectors making up \mathbf{b}_i , listed in (63), or an appropriate subset thereof. Hence, a sufficient condition for the SPM to be MAR is that the random effects driving the observed measurements and/or the missing-data process do not influence the missing measurements, given the observed ones. In other words, all information about the missing measurements, apart from covariates, stems from the observed measurements only. Clearly, the random effects \mathbf{m}_i are not identifiable; they are included for completeness only.

It is instructive to study the same set of sufficient conditions from the PMM perspective (Lemma 1), since it will lead us, at the end of the section, to the construction of an MAR counterpart:

$$\begin{aligned}
f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i) &= \frac{f(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i)}{f(\mathbf{y}_i^o, \mathbf{r}_i)} \\
&= \frac{\int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) f(\mathbf{r}_i | \mathbf{j}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int \int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) f(\mathbf{r}_i | \mathbf{j}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i d\mathbf{y}_i^m} \\
&= f(\mathbf{y}_i^m | \mathbf{y}_i^o) \cdot \frac{\int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{r}_i | \mathbf{j}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{r}_i | \mathbf{j}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \\
&= f(\mathbf{y}_i^m | \mathbf{y}_i^o),
\end{aligned}$$

not surprisingly leading to the same result.

These considerations at the same time define an important sub-class, establishing the ensuing result:

Definition 2 (A Sub-class of SPM Models.) *Define a sub-class of shared-parameter model (62):*

$$f(\mathbf{y}_i^o | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) f(\mathbf{r}_i | \mathbf{j}_i, \mathbf{q}_i), \quad (64)$$

where \mathbf{j}_i , $\boldsymbol{\ell}_i$, \mathbf{m}_i , and \mathbf{q}_i are independent random-effects vectors.

In other words, Definition 2 follows as a special case from Definition 1 by omitting the random effects \mathbf{g}_i , \mathbf{h}_i , and \mathbf{k}_i . The key rationale for this definition is, of course, the following result:

Theorem 2 (A Class of MAR-based SPM Models.) *The shared-parameter model (64) is missing at random.*

We have not addressed necessity thus far. To this effect, we need to derive general expressions for the left hand side and right hand side of (12), respectively. First, for the left hand side:

$$\begin{aligned} & f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i) \\ &= \frac{\int f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{q}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \end{aligned} \quad (65)$$

Second, for the right hand side, consider:

$$f(\mathbf{y}_i^m | \mathbf{y}_i^o) = \frac{\int f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{b}_i) d\mathbf{b}_i}. \quad (66)$$

Equating (65) and (66) and, for brevity, integrating over random effects that occur in one component only, produces the general conditions, laid out in the next theorem.

Theorem 3 (Characterization of MAR in SPM Family.) *A member of the general SPM family (62) is MAR if and only if*

$$\begin{aligned} & \frac{\int f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{j}_i) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \\ &= \frac{\int f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{f(\mathbf{y}_i^o)}. \end{aligned} \quad (67)$$

Evidently, again assuming that \mathbf{g}_i , \mathbf{h}_i , and \mathbf{k}_i cancel, reduces (67) to a tautological statement, showing that (64) satisfies Theorem 3.

There are situations where (67) is satisfied, without the triplet $(\mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i)$ vanishing, but these will necessarily be more *ad hoc* and less intuitively appealing than these laid out in Theorem 2. The existence of such singular solutions is not straightforward to establish, as is clear from the following pair of examples.

Example 1 (MAR Example in Line With Definition 1.) *For the purpose of the examples, drop the index i from notation. Consider a bivariate outcome (Y_1, Y_2) , where the first one is always observed, and the second component sometimes missing. This necessitates a scalar missing-data variable R only, leading to full-data vector (Y_1, Y_2, R) . Let $R = 0$ if the second component is missing and 1 otherwise. For $R = 1$, condition (67) is always fulfilled, since the key component, describing the distribution of the missing observations given the observed ones, is then empty. Therefore, we can concentrate on $R = 0$.*

For simplicity, assume that all random effects, describing one factor only, are absent, i.e., remove $\boldsymbol{\ell}_i$, \mathbf{m}_i , and \mathbf{q}_i . From the four remaining random-effects, retain only \mathbf{j}_i and \mathbf{k}_i , implying that the missing-data process is connected to both

response-related factors which, in turn, are unrelated to each other. Assume furthermore that both outcomes, Y_1 and Y_2 , are dichotomous, and that also both random effects are binary. This means that (67) can be simplified to:

$$\begin{aligned} & \left(\sum_j \pi_{y_1|j}^1 \pi_j \right) \cdot \left(\sum_{j,k} \pi_{y_1|j}^1 \pi_{y_2|y_1 k}^2 \pi_{r=0|jk} \pi_j \pi_k \right) \\ &= \left(\sum_{j,k} \pi_{y_1|j}^1 \pi_{r=0|jk} \pi_j \pi_k \right) \cdot \left(\sum_{j,k} \pi_{y_1|j}^1 \pi_{y_2|y_1 k}^2 \pi_j \pi_k \right), \quad (68) \end{aligned}$$

where the π 's are probabilities pertaining to the variables indicated by their corresponding indices. It is convenient to introduce some simplifying notation, making use of the fact that all key variables are dichotomous: set $\gamma = \pi_{j=0}$, $\varphi = \pi_{k=0}$, and $\rho_{jk} = \pi_{r=0|jk}$.

Expression (68) needs to be considered only for $(Y_1, Y_2) = (0, 0)$ and $(1, 0)$, since spelling out the ones for $(1, 0)$ and $(1, 1)$ and summing them with their counterparts lead to tautological statements. This implies that (68) produces two equations, i.e., there are two constraints to be satisfied. For the first equation, in $(Y_1, Y_2) = (0, 0)$, choose $x = \pi_{0|01}^2$ as the parameter to be determined. This means that (68) is a linear equation in x . Clearly, setting $\pi_{0|00}^2 = \pi_{0|01}^2$ solves the equation, based on two observations. First, a constant factor $\pi_{y_2|y_1}^2$ is common to both sides of the equation and cancels. Second, the remaining factors are pairwise equal: the first factor on the LHS then equals the second factor on the RHS; the second factor on the LHS equals the first factor on the RHS. The argument for $(Y_1, Y_2) = (1, 0)$ is entirely symmetric, and hence the unique solution implies that k vanishes from the distribution of Y_2 given Y_1 , in agreement with Definition 2.

Similar manipulations can be done for the cases: (1) where only \mathbf{g}_i is present; and (2) where only \mathbf{h}_i and \mathbf{j}_i are present. In these two cases, as well as in Example 1, a single random effect describes $\pi_{y_2|y_1}^2$. This is crucial to ensure accordance with Definition 1. The next example is different in that two independent random effects will influence the probability of the second component given the first one.

as is clear from the next example.

Example 2 (MAR Example Violating Definition 1.) *Retain the setting of Example 1, but now with the pair of random effects \mathbf{h}_i and \mathbf{k}_i present. This particular choice leads to a different simplification of (67):*

$$\left(\sum_h \pi_{y_1|h}^1 \pi_h \right) \cdot \left(\sum_{h,k} \pi_{y_1|h}^1 \pi_{y_2|y_1 hk}^2 \pi_{r=0|k} \pi_h \pi_k \right)$$

$$= \left(\sum_{h,k} \pi_{y_1|h}^1 \pi_{r=0|k} \pi_h \pi_k \right) \cdot \left(\sum_{h,k} \pi_{y_1|h}^1 \pi_{y_2|y_1hk}^2 \pi_h \pi_k \right). \quad (69)$$

We will conveniently use the following notation: $\eta = \pi_{h=0}$, $\varphi = \pi_{k=0}$, and $\rho_k = \pi_{r=0|k}$.

With similar logic as in Example 1, it easily follows that we only need to consider (69) for $(Y_1, Y_2) = (0, 0)$ and $(1, 0)$. Concentrating on the first of these, and singling out $\pi_{0|011}^2$ as the parameter to identify from the others, it follows that

$$\pi_{0|011}^2 = \frac{ab - de}{df - ac}, \quad (70)$$

with

$$\begin{aligned} a &= \pi_{0|0}^1 \eta + \pi_{0|1}^1 (1 - \eta), \\ b &= \pi_{0|0}^1 \pi_{000}^2 \rho_0 \eta \varphi + \pi_{0|0}^1 \pi_{001}^2 \rho_1 \eta (1 - \varphi) + \pi_{0|1}^1 \pi_{010}^2 \rho_0 (1 - \eta) \varphi, \\ c &= \pi_{0|1}^1 \rho_1 (1 - \eta) (1 - \varphi), \\ d &= \pi_{0|0}^1 \rho_0 \eta \varphi + \pi_{0|0}^1 \rho_1 \eta (1 - \varphi) + \pi_{0|1}^1 \rho_0 (1 - \eta) \varphi + \pi_{0|1}^1 \rho_1 (1 - \eta) (1 - \varphi), \\ e &= \pi_{0|0}^1 \pi_{000}^2 \eta \varphi + \pi_{0|0}^1 \pi_{001}^2 \eta (1 - \varphi) + \pi_{0|1}^1 \pi_{010}^2 (1 - \eta) \varphi, \\ f &= \pi_{0|1}^1 (1 - \eta) (1 - \varphi). \end{aligned}$$

The derivations for $(Y_1, Y_2) = (1, 0)$ is entirely similar and leads to (70) with the first conditioning argument ‘1’ rather than ‘0’. A numerical example is provided in Table 4, establishing that the random effects \mathbf{h}_i and \mathbf{k}_i do influence the distribution of $Y_2|Y_1$, in the dropout pattern.

Finally, the characterization of Theorem 3 allows us to construct an MAR counterpart to an arbitrary SPM of the form (62). It is necessary to (1) retain the fit of the model to the observed data, while (2) ensuring that (67) hold. This is easily done by *a-posteriori* integrating the shared random effects out of the densities describing the unobserved measurements, given the observed ones. Here, integration takes place over the densities of \mathbf{g}_i , \mathbf{h}_i , and \mathbf{k}_i , where fitted parameters are plugged into the densities.

Theorem 4 (An MAR Counterpart to a General SPM.) *The MAR counterpart, to an arbitrary general SPM of the type (62) is found by replacing $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i)$ with*

$$h(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) = \int_{\mathbf{g}_i} \int_{\mathbf{h}_i} \int_{\mathbf{k}_i} f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) d\mathbf{g}_i d\mathbf{h}_i d\mathbf{k}_i \quad (71)$$

Tabela 4: *Bivariate binary outcome with the first component fully observed and the second component partially missing. The missing data mechanism is MAR. The model belongs to general SPM family (62), but not to the specific MAR sub-class (64).*

Effect	'Failure (0)'		'Success (1)'	
Random h effect	$\eta = \pi_{h=0}$	0.3000	$1 - \eta = \pi_{h=1}$	0.7000
Random k effect	$\varphi = \pi_{k=0}$	0.4000	$1 - \varphi = \pi_{k=1}$	0.6000
$R k = 0$	$\rho_0 = \pi_{0 0}$	0.4500	$1 - \rho_0 = \pi_{1 0}$	0.5500
$R k = 1$	$\rho_1 = \pi_{0 1}$	0.8000	$1 - \rho_1 = \pi_{1 1}$	0.2000
$Y_1 h = 0$	$\pi_{0 0}^1$	0.3000	$\pi_{1 0}^1$	0.7000
$Y_1 h = 1$	$\pi_{0 1}^1$	0.2000	$\pi_{1 1}^1$	0.8000
$Y_2 Y_1 = 0, h = 0, \& k = 0$	$\pi_{0 000}^2$	0.1500	$\phi_{1 000}^2$	0.8500
$Y_2 Y_1 = 0, h = 0, \& k = 1$	$\pi_{0 001}^2$	0.2500	$\pi_{1 001}^2$	0.7500
$Y_2 Y_1 = 0, h = 1, \& k = 0$	$\pi_{0 010}^2$	0.3500	$\pi_{1 010}^2$	0.6500
$Y_2 Y_1 = 0, h = 1, \& k = 1$	$\pi_{0 011}^2$	0.2857	$\pi_{1 011}^2$	0.7143
$Y_2 Y_1 = 1, h = 0, \& k = 0$	$\pi_{0 100}^2$	0.2000	$\pi_{1 100}^2$	0.8000
$Y_2 Y_1 = 1, h = 0, \& k = 1$	$\pi_{0 101}^2$	0.3000	$\pi_{1 101}^2$	0.7000
$Y_2 Y_1 = 1, h = 1, \& k = 0$	$\pi_{0 110}^2$	0.4000	$\pi_{1 110}^2$	0.6000
$Y_2 Y_1 = 1, h = 1, \& k = 1$	$\pi_{0 111}^2$	0.3625	$\pi_{1 111}^2$	0.6375

First, it is clear that this marginalization is merely describing the model-based prediction of the unobserved outcomes, given the observed ones. Hence, the choice for $h(\cdot)$ does not alter the fit. Second, observe that using $h(\cdot)$ in (67), instead of $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i)$, of Theorem 3, reduces the equation to a trivial identity, and hence the second condition is also satisfied.

For categorical random effects, such as in Examples 1 and 2, the integral in (71) becomes summation.

7 Longitudinal Data With Dropout: Non-future Dependence

When measurements are taken longitudinally, it is good practice to ensure that the implied time dependencies are logical from a substantive standpoint. For example, in a variety of contexts, such as growth, regression functions over time may be constrained to non-decreasing forms.

Let us turn to the nature of the missingness mechanism. Throughout the section, assume that missingness is confined to dropout. From a SeM perspective, one often classifies missing data mechanisms as (Diggle, and Kenward,

1994): (1) independent of outcomes; (2) dependent on previous measurements only; (3) dependent on the current and perhaps previous measurements only; (4) fully arbitrary, *i.e.*, where missingness can depend on previous, current, and future measurements. Evidently, (1) is MCAR, (2) is MAR, and (4) is MNAR, without restrictions. (Diggle, and Kenward, 1994), for example, did not consider (4) but restricted MNAR to mechanism (3) only. While this is very restrictive, it is also extremely appealing since it prevents dropout at a given point in time to depend on future measurements; these are termed non-future dependent in the next section.

Clearly, the concepts of the previous paragraph are very natural by virtue of framing them in the SeM. Kenward *et al* (2003) underscored that the situation is less clear in the PMM family and then translated the mechanisms from the SeM to the PMM framework. We will review these in Section 7.1, and then present a similar taxonomy for the SPM in Section 7.2.

7.1 Non-future Dependence in the PMM Framework

Since we are restricting attention to monotone missingness, we can easily indicate a drop-out pattern by the numbers of observations made. In this sense, pattern t collects all individuals with the first t measurements taken ($t = 1, \dots, n$). Thijs *et al* (2002) constructed a general identifying-restrictions framework in which the distribution of the $(t + 1)$ th measurement, given the earlier measurements, in pattern t , y_{t+1} say, is set equal to a linear combination of the corresponding distributions in patterns $t + 1$ to n . Since this family is characterized by the use of observable distributions to identify the unobservable ones, we term it the ‘interior’ family of identifying-restrictions. Three members of this family are studied in detail by Thijs *et al* (2002): complete-case missing value restrictions (Little, 1993), where information is borrowed from the completers only, available-case missing values, equivalent to MAR (Molenberghs *et al* (1998)), for which a particular linear combination needs to be considered, and neighboring-case missing value restrictions, where information is borrowed from the closest available pattern.

The equivalence of available-case missing values and MAR is important in that it enables us to make a clear connection between the selection and pattern-mixture frameworks. By implication, the other members of the interior family are of MNAR type, while at the same time there do exist MNAR type restrictions that are not captured by this family.

We will now characterize missing-data mechanisms that prevent missingness from depending on future unobserved measurements. To this effect, it is useful to consider the SeM and PMM factorizations for the specific context of longitudinal data. Let $r = t \leq n$ be the number of measurements actually observed. The selection model factorization for this context is given by

$$f(y_1, \dots, y_n, r = t) = f(y_1, \dots, y_n) f(r = t | y_1, \dots, y_n).$$

Pattern-mixture models now take the form:

$$\begin{aligned}
& f(y_1, \dots, y_n, r = t) \\
&= f(y_1, \dots, y_n | r = t) f(r = t) \\
&= f_t(y_1, \dots, y_n) f(r = t) \\
&= f_t(y_1, \dots, y_t) f_t(y_{t+1} | y_1, \dots, y_t) f_t(y_{t+2}, \dots, y_n | y_1, \dots, y_{t+1}) f(r = t) \tag{72}
\end{aligned}$$

where $f_t(y_1, \dots, y_n) = f(y_1, \dots, y_n | r = t)$. The first three factors in (72) are referred to as the distributions of past, present, and future measurements, respectively. Only the first and the fourth factors are identifiable from the data.

Definition 3 (Non-future Dependence (NFD).) *In the SeM context, we can formulate missing non-future dependent as*

$$f(r = t | y_1, \dots, y_n) = f(r = t | y_1, \dots, y_{t+1}). \tag{73}$$

Note that MAR is a special case of missing non-future dependent, which in turn is a sub-class of MNAR.

Definition 4 (Non-future Dependent Missing Value (NFMV).) *Within the PMM framework, we define non-future dependent missing value restrictions as follows:*

$$f(y_t | y_1, \dots, y_{t-1}, r = j) = f(y_t | y_1, \dots, y_{t-1}, r \geq t - 1), \tag{74}$$

for all $t \geq 2$ and all $j < t - 1$.

Non-future missing values is not a comprehensive set of restrictions, but rather leaves one conditional distribution per incomplete pattern unidentified:

$$f(y_{t+1} | y_1, \dots, y_t, r = t). \tag{75}$$

In other words, the distribution of the ‘current’ unobserved measurement, given the previous ones, is unconstrained. This implies that the NFMV class contains members outside of the interior family, where every restriction takes the form of a linear combination of observable distributions. Conversely, (74) excludes such mechanisms as complete-case missing values and neighboring-case missing values, showing that there are members of the interior family that are not of non-future missing values type. Finally, choosing (75) of the same functional form as (74) establishes available-case missing values as a member of the intersection of the interior and non-future missing values families. The latter is particularly important since it shows, because of the equivalence of ACMV and MAR, that MAR belongs to both families.

The following theorem, the proof of which is to be found in Kenward *et al* (2003), establishes the equivalence between NFD and NFMV, showing the NFMV restrictions correspond to NFD, just as ACMV corresponds to MAR.

Theorem 5 (Equivalence Between NFD and NFMV.) *For longitudinal data with drop-outs, missing non-future dependence is equivalent to non-future missing values.*

A consequence of using (74) is that the joint distribution will not typically have a simple analytical representation. This is to be understood in the sense that covariate effects would not necessarily be linear on an appropriate scale. However, this is not to say there is no analytical form. Moreover, it does not have to be a major disadvantage, provided the resulting distribution is empirically reasonable. Such a requirement may help guide the choice for (75). Kenward *et al* (2003) offered a tractable, sampling-based implementation and applied it to the analysis of a set of data.

7.2 Non-future Dependence in the SPM Framework

It is now particularly easy to derive a general characterization of non-future dependent SPM. First, note that (73) in Definition 3 can be seen as a longitudinal dropout-based definition of MAR, “one component shifted to the right,” i.e., where y_{t+1} , in spite of its missingness, is also allowed to influence missingness. Given that Theorem 3 was derived from the standard MAR definition, it immediately follows that a characterization of NFD-SPM is as follows.

Theorem 6 (Non-future Dependent Shared-parameter Models.) *A member of the general SPM family (62) is NFD if and only if*

$$\begin{aligned} & \frac{\int f(\mathbf{y}_i^{pc} | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i) f(\mathbf{y}_i^f | \mathbf{y}_i^{pc}, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i) f(r_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^{pc} | \mathbf{g}_i, \mathbf{j}_i) f(r_i | \mathbf{g}_i, \mathbf{j}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \\ &= \frac{\int f(\mathbf{y}_i^{pc} | \mathbf{g}_i, \mathbf{h}_i) f(\mathbf{y}_i^f | \mathbf{y}_i^{pc}, \mathbf{g}_i, \mathbf{h}_i) f(\mathbf{b}_i) d\mathbf{b}_i}{f(\mathbf{y}_i^{pc})}, \end{aligned} \quad (76)$$

where $\mathbf{y}_i^{pc} = (y_1, \dots, y_{t+1})'$ and $\mathbf{y}_i^f = (y_{t+2}, \dots, y_n)'$.

Note that the subscript ‘pc’ refers to ‘previous and current,’ while ‘f’ refers to ‘future.’

Likewise, the sub-class (64) of Definition 2 can be ‘shifted’ to yield an NFD version.

Definition 5 (A NFD Sub-class of SPM Models.) *Define a sub-class of shared-parameter model (62):*

$$f(\mathbf{y}_i^{pc} | \mathbf{j}_i, \boldsymbol{\ell}_i) f(\mathbf{y}_i^f | \mathbf{y}_i^{pc}, \mathbf{m}_i) f(r_i | \mathbf{j}_i, \mathbf{q}_i), \quad (77)$$

where \mathbf{j}_i , $\boldsymbol{\ell}_i$, \mathbf{m}_i , and \mathbf{q}_i are independent random-effects vectors.

With similar logic as before, Definition 5 offers a class of missing-data mechanism that belongs to the NFD family. The relationship between the various mechanisms in the three families is depicted in Figure 3.

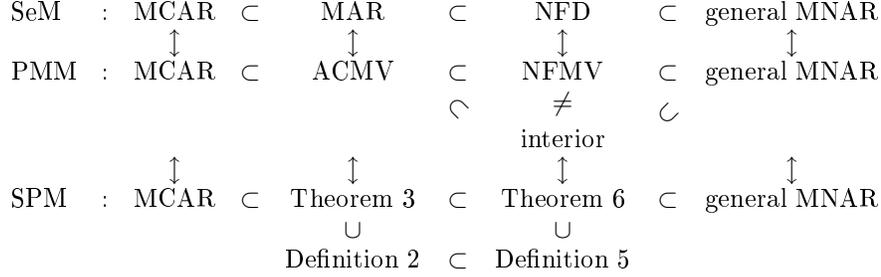


Figure 3: *Subset-relationships between nested families within the selection model (SeM), pattern-mixture model (PMM), and shared-parameter model (SPM) families. MCAR: missing completely at random; MAR: missing at random; MNAR: missing not at random; NFD: non-future dependence; ACMV: available-case missing values; NFMV: non-future missing values. The vertical two-headed arrows indicate equivalence between mechanisms across model families.*

8 Analysis of The Slovenian Public Opinion Survey

8.1 The BRD Models

Baker *et al* (1992) proposed a log-linear based family of models for the four-way classification of both outcomes, together with their respective missingness indicators: $\nu_{10,jk} = \nu_{11,jk}\beta_{jk}$, $\nu_{01,jk} = \nu_{11,jk}\alpha_{jk}$, and $\nu_{00,jk} = \nu_{11,jk}\alpha_{jk}\beta_{jk}\gamma$, with

$$\alpha_{jk} = \frac{\phi_{01|jk}}{\phi_{11|jk}}, \quad \beta_{jk} = \frac{\phi_{10|jk}}{\phi_{11|jk}}, \quad \gamma = \frac{\phi_{11|jk}\phi_{00|jk}}{\phi_{10|jk}\phi_{01|jk}}.$$

Furthermore $\nu_{r_1 r_2, jk}$ is the model for the four cells, indexed by j and k , in pattern (r_1, r_2) , where $(r_1, r_2) = (1, 1)$ corresponds to completers, etc.

The α (β) parameters describe missingness in the independence (attendance) question, and γ captures the interaction between both. The subscripts are missing from γ since Baker *et al* (1992) have shown that this quantity is independent of j and k in every identifiable model. These authors considered nine models, based on setting α_{jk} and β_{jk} constant in one or more indices, and enumerated using the ‘BRD’ abbreviation:

$$\begin{array}{lll}
\text{BRD1} & : & (\alpha, \beta) \\
\text{BRD2} & : & (\alpha, \beta_j) \\
\text{BRD3} & : & (\alpha_k, \beta) \\
\text{BRD4} & : & (\alpha, \beta_k) \\
\text{BRD5} & : & (\alpha_j, \beta) \\
\text{BRD6} & : & (\alpha_j, \beta_j) \\
\text{BRD7} & : & (\alpha_k, \beta_k) \\
\text{BRD8} & : & (\alpha_j, \beta_k) \\
\text{BRD9} & : & (\alpha_k, \beta_j).
\end{array}$$

Interpretation is straightforward, for example, BRD1 is MCAR, and in BRD4 missingness in the first variable is constant, while missingness in the second variable depends on its value. BRD6–BRD9 saturate the observed data degrees

Tabela 5: (Part I). Analysis of the Slovenian Public Opinion Survey, restricted to the independence and attendance questions. The observed data are shown, as well as the fit of models BRD1, BRD2, BRD7, and BRD9, and their MAR counterparts, to the observed data. The contingency tables' rows (columns) correspond to 'yes' vs. 'no' on the independence (attendance) question. The four tables in each row correspond to: (i) people responding to both questions; (ii) people responding to independence only; (iii) people responding to attendance only; (iv) people responding to neither question.

Observed data &					
fit of BRD7, BRD7(MAR), BRD9, and BRD9(MAR) to incomplete data					
1439	78	159	144	54	136
16	16	32			
Fit of BRD1 and BRD1(MAR) to incomplete data					
1381.6	101.7	182.9	179.7	18.3	136.0
24.2	41.4	8.1			
Fit of BRD2 and BRD2(MAR) to incomplete data					
1402.2	108.9	159.0	181.2	16.8	136.0
15.6	22.3	32.0			

of freedom, while the lower numbered ones leave room for a non-trivial model fit to the observed data.

8.2 Analysis of the Slovenian Public Opinion Data

The ideas developed in this paper can be illustrated easily by means of 4 models from the BRD family, fitted to the independence and attendance outcomes, i.e., collapsing Table 1. We select models BRD1, BRD2, BRD7, and BRD9. Model BRD1 assumes missingness to be MCAR. All others are of the MNAR type. Model BRD2 has 7 free parameters, and hence does not saturate the observed data degrees of freedom, while models BRD7 and BRD9 saturate the 8 data degrees of freedom. The collapsed data, together with the model fits, are displayed in Table 5. Each of the four models is doubled up with its MAR counterpart.

Table 5 presents, apart from the raw data, for each of the models and its MAR counterpart, the fit to the observed and the hypothetical complete data. The fits of models BRD7, BRD9, and their MAR counterparts to the observed data, coincide with the observed data. As the theory states, every MNAR model and its MAR counterpart produce exactly the same fit to the observed data, which is therefore also seen for BRD1 and BRD2. However, while Models BRD1

and BRD1(MAR) coincide in their fit to the hypothetical complete data, this is not the case for the other three models. The reason is clear: since model BRD1 belongs to the MAR family from the start, its counterpart BRD1(MAR) will not produce any difference, but merely copies the fit of BRD1 to the unobserved data, given the observed ones. Finally, while BRD7 and BRD9 produce a different fit to the complete data, BRD7(MAR) and BRD9(MAR) coincide. This is because the fits of BRD7 and BRD9 coincide with respect to their fit to the observed data, and indeed, due to their saturation, coincide with the observed data as such. This fit is the sole basis for the models' MAR extensions. It is noteworthy that, while BRD7, BRD9, and BRD7(MAR)≡BRD9(MAR) all saturate the observed data degrees of freedom, their complete-data fits are dramatically different.

Let us return to the implications of our results for the primary estimand θ , the proportion of people voting YES by simultaneously being in favor of independence and deciding to take part in the vote. Rubin *et al* (1995) considered, apart from simple models such as complete case analysis ($\hat{\theta} = 0.928$) and available case analyses ($\hat{\theta} = 0.929$), both ignorable models ($\hat{\theta} = 0.892$ when based on the two main questions and $\hat{\theta} = 0.883$ when using the secession question as an auxiliary variable) and a non-ignorable one ($\hat{\theta} = 0.782$). Since the value of the plebiscite was $\theta_{\text{pleb}} = 0.885$, an important benchmark obtained four weeks after the SPO, they concluded the MAR was preferable. Molenberghs *et al* (2001) supplemented these analysis with a so-called pessimistic-optimistic interval, obtained from replacing the incomplete data with NO and YES, respectively, and obtained: $\theta \in [0.694, 0.904]$. Further, they considered all nine BRD models, producing a range for θ from 0.741 to 0.892. Ultimately, these authors devised a method to consider overspecified models, in which point estimates are replaced by interval estimates, so-called *intervals of ignorance*.

Let us consider the results obtained from fitting each of the nine BRD models. Molenberghs *et al* (2001) presented a summary table but unfortunately there was a small computational error that had to be corrected, for which reason the corrected results are reproduced here (Table 6). BRD1 produces $\hat{\theta} = 0.892$, exactly the same estimate as the first MAR estimate obtained by Rubin *et al* (1995). This should not come as a surprise, since both BRD1 and Rubin's model assume MAR and use information from the two main questions. Before continuing with the models' interpretation, it is necessary to assess their fit. Conducting likelihood ratio tests for BRD1 versus the ones with 7 parameters, BRD2–BRD5, and then in turn for BRD2–BRD5 versus the saturated modes BRD6–BRD9, suggests the lower numbered models do not fit well, leaving us with BRD6–BRD9. The impression might be generated that the poor model fit of BRD1 might be seen as evidence for discarding the MAR-based value 0.892. However, studying the MAR values from each of the models BRD1(MAR)–BRD9(MAR), as displayed in the last column of Table 6, it is clear that this value is remarkably stable and hence a value of $\hat{\theta} = 0.892$, based on the four counterparts BRD6(MAR)–BRD9(MAR), is a sensible choice after all. Thus,

a main contribution resulting from considering the counterparts in this particular example, is the provision of a solid basis for the MAR-based estimate. Obviously, since Models BRD6(MAR)–BRD9(MAR) are exactly the same and exhibit a perfect fit, the corresponding probabilities $\hat{\theta}_{\text{MAR}}$ are exactly equal too. In this particular case, even though BRD2(MAR)–BRD5(MAR) differ among each other, the probability of being in favor of independence and attending the plebiscite is constant across these four models. This is a mere coincidence, since all three other cell probabilities are different, but only slightly so. For example, the probability of being in favour of independence combined with not attending ranges over 0.066–0.0685 across these four models.

We have made the following two-stage use of Models BRD6(MAR)–BRD9(MAR). At the first stage, in a conventional way, the fully saturated model is selected as the only adequate description of the observed data. At the second stage, these models are transformed into their MAR counterpart, from which inferences are drawn. As such, the MAR counterpart usefully supplements the original models BRD6–BRD9 and provide one further, important scenario to model the incomplete data. In principle, the same exercise can be conducted when the additional secession variable would be used.

9 Analysis of the Onychomycosis Trial

We will first analyze the entire longitudinal profile of continuous outcomes (unaffected nail length), and then switch to the binary outcome (severity of infection) and confine attention to the first and last time points.

9.1 Continuous Unaffected Nail Length

Consider a general model of the form (62), with random effects confined to \mathbf{g}_i , i.e., common to all three components. For the measurement model, assume a linear mixed model (Verbeke, and Molenberghs, 2000), with general form:

$$\mathbf{Y}_i | \mathbf{g}_i \sim N(X_i \boldsymbol{\beta} + Z_i \mathbf{g}_i, \Sigma_i), \quad (78)$$

$$\mathbf{g}_i \sim N(0, D). \quad (79)$$

Based on (78) and (79), the so-called marginal model can be derived

$$\mathbf{Y}_i \sim N(X_i \boldsymbol{\beta}, Z_i D Z_i' + \Sigma_i). \quad (80)$$

To compute the model's prediction for the unobserved data, given the observed measurements, the corresponding density needs to be derived. To this end, first decompose the mean and variance in (78) as

$$\begin{pmatrix} \mathbf{Y}_i^o \\ \mathbf{Y}_i^m \end{pmatrix} | \mathbf{g}_i \sim N \left[\begin{pmatrix} X_i^o \\ X_i^m \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} Z_i^o \\ Z_i^m \end{pmatrix} \mathbf{g}_i, \begin{pmatrix} \Sigma_i^{oo} & \Sigma_i^{om} \\ \Sigma_i^{mo} & \Sigma_i^{mm} \end{pmatrix} \right].$$

This expression can easily be used to construct the conditional density:

$$\begin{aligned} \mathbf{Y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i \sim N & \left[(X_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} X_i^o) \boldsymbol{\beta} + \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{y}_i^o \right. \\ & \left. + (Z_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} Z_i^o) \mathbf{g}_i, \Sigma_i^{mm} - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \Sigma_i^{om} \right] \end{aligned} \quad (81)$$

Now, (81) corresponds to the model as formulated, and will typically be of the MNAR type. To derive the MAR counterpart, we need to integrate over the random effect. With similar logic that leads to (80), now applied to (81), we obtain:

$$\begin{aligned} \mathbf{Y}_i^m | \mathbf{y}_i^o \sim N & \left[(X_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} X_i^o) \boldsymbol{\beta} + \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{y}_i^o, \right. \\ & (Z_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} Z_i^o) D (Z_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} Z_i^o)' \\ & \left. + \Sigma_i^{mm} - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \Sigma_i^{om} \right]. \end{aligned} \quad (82)$$

Hence, (82) is the MAR counterpart to (81). For the unaffected nail length, we choose for (78)–(79):

$$E(Y_{ij} | g_i, T_i, t_j, \boldsymbol{\beta}) = \beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j, \quad (83)$$

$g_i \sim N(0, d)$, and $\Sigma_i = \sigma^2 I_7$, where I_7 is a 7×7 identity matrix. Further, $T_i = 0$ if patient i received standard treatment and 1 for experimental therapy ($i = 1, \dots, 298$). Finally, t_j is the time at which the j th measurement is taken ($j = 1, \dots, 7$).

Given these choices, (81) and (82) simplify to

$$\mathbf{Y}_i^m | \mathbf{y}_i^o, g_i \sim N(X_i \boldsymbol{\beta} + Z_i^m g_i, \sigma^2 I_i), \quad (84)$$

$$\mathbf{Y}_i^m | \mathbf{y}_i^o \sim N(X_i \boldsymbol{\beta}, d J_i + \sigma^2 I_i), \quad (85)$$

with I_i an identity matrix and J_i a matrix of ones, with dimensions equal to the number of missing measurements for subject i . Especially owing to the conditional independence assumption, the simplification is dramatic.

Next, let us formulate a model for the missingness mechanism in (62). The sequence \mathbf{r}_i can take one of two forms in our case. Either, it is a length-7 vector of ones, for a completely observed subject, or it is a sequence of k ones followed by a sole zero $1 \leq k \leq 6$, for someone dropping out. Note that k is 1 at least, since for everyone the initial measurement has been observed. It is convenient to assume a logistic regression of the form:

$$\text{logit} [P(R_{ij} = 1 | R_{i,j-1} = 0, g_i, T_i, t_j, \boldsymbol{\gamma})] = \gamma_0 + \gamma_{01} g_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j, \quad (86)$$

($j > 1$), where γ_{01} is a scale factor for the shared random effect in the missingness model; forcing the variance in the measurement and dropout indicator sequences to be equal would make no sense. As a result, $\gamma_{01} g_i \sim N(0, \gamma_{01}^2 d)$.

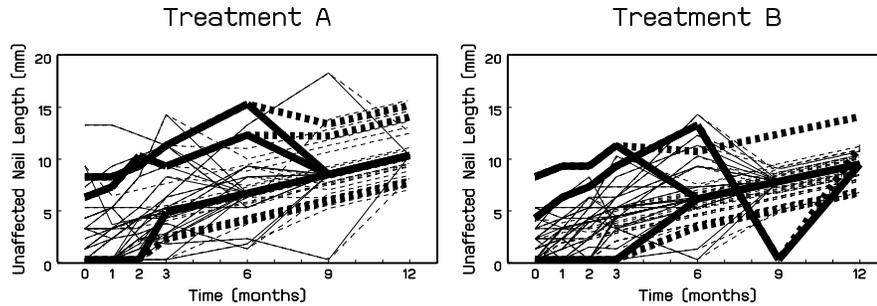


Figura 4: *Toenail Data*. Individual profiles of subjects with incomplete data, for each treatment arm, extended using MNAR Model (83) (dashed line) and using the model's MAR counterpart (solid line). In each group, three subjects are highlighted.

The model specified by (83) and (86) can easily be fitted using, for example, the SAS procedure NLMIXED, details about which are provided in the Appendix.

Parameter estimates and standard errors are displayed in Table 7. It is noteworthy that the scale factor γ_{01} is estimated to be negative, even though it is not significant. While we should not overly stress its importance, there is some indication that a higher subject-specific profile of unaffected nail length corresponds with a lower dropout probability, which is not surprising. The magnitude of the scale factor allows us to 'translate' the subject-specific effect from the continuous outcome scale, expressed in mm, to the unitless logit scale on which the probability of missingness is described. Note that the random-intercept variance is highly significant among unaffected nail length outcomes; the same is not true for the dropout model, with $p = 0.2487$, using a 50 : 50 mixture of a χ_0^2 and χ_1^2 distribution (Verbeke, and Molenberghs, 2000).

Figure 4 displays the incomplete profiles, extended beyond the time of dropout, using prediction based on: (1) the original model (dashed lines); (2) the MAR counterpart (solid lines). Within each of the treatment arms, three profiles are highlighted. The MAR counterpart reduces all predictions to the same profile, whereas the MNAR model predicts different evolutions for different subjects, implied by the presence of the random effect. The simple MAR-based prediction structure follows directly from the conditional independence assumption, present in (84). When deemed less plausible, the fully general structure (81) can be implemented.

9.2 Dichotomous Severity of Infection

Let us turn attention to the binary severity of infection outcome, for the pair of time points formed by the always recorded initial measurement and the some-

times missing final point in time. The data are displayed in Table 8. By way of illustration, we will assume a single dichotomous random effect, of the \mathbf{g}_i type. This imposes a latent-class structure. Decompose the cell probabilities as:

$$\pi_{g i_1 i_2 r t} = \pi_g \pi_{i_1|g} \pi_{i_2|i_1 g t} \pi_{r|g}, \quad (87)$$

with $g = 0, 1$ indicating the latent class, $i_1, i_2 = 0, 1$ non-severe *versus* severe infection at the first and last occasions, respectively, $r = 0, 1$ referring to the dropouts *versus* completers groups, and $t = 0, 1$ denoting standard *versus* experimental treatment arm. The probability factors on the right hand side of (87) are modeled as:

$$\pi_g = \frac{e^{\alpha g}}{1 + e^{\alpha}},$$

$$\pi_{i_1|g} = \frac{e^{(\beta_0 + \beta_1 g) i_1}}{1 + e^{\beta_0 + \beta_1 g}}, \quad (88)$$

$$\pi_{i_2|i_1 g t} = \frac{e^{(\gamma_0 + \gamma_1 i_1 + \gamma_2 g + \gamma_3 i_1 g + \gamma_4 t) i_2}}{1 + e^{\gamma_0 + \gamma_1 i_1 + \gamma_2 g + \gamma_3 i_1 g + \gamma_4 t}}, \quad (89)$$

$$\pi_{r|g} = \frac{e^{(\delta_0 + \delta_1 g) r}}{1 + e^{\delta_0 + \delta_1 g}}.$$

In Model ‘Bin1’, we will set $\beta_1 = 0$ in (88) for reasons of identifiability. In Model ‘Bin2’, $\gamma_2 = \gamma_3 = 0$ in (89). This implies the latter model is of the MAR type, and hence its MAR counterpart will equal the original model. Fitted counts are presented in Table 8. For the dropout group, both the fit to the pair of observed counts and the prediction of the underlying unobserved two-by-two table is given. Note that the MAR counterpart preserves the distribution of the first outcome, within each treatment and dropout group; the difference between original model and MAR counterpart is confined to the distribution of the second outcome, given the first one. The fits of the models is obtained by replacing all quantities in (87) by their estimates, followed by summing over g . The MAR counterpart is obtained as $\pi_{g i_1 i_2 r t} = \pi_g \pi_{i_1|g} \tilde{\pi}_{i_2|i_1 t} \pi_{r|g}$, where

$$\tilde{\pi}_{i_2|i_1 t} = \sum_g \pi_g \pi_{i_2|i_1 g t}.$$

Parameter estimation by both maximum likelihood, as well as the EM algorithm (Dempster, Laird, and Rubin, 1977) is particularly easy. For direct likelihood, the log-likelihood function takes the form

$$\begin{aligned} \ell = & \sum_{i_1, i_2, t} Z_{i_1 i_2, r=1, t} \ln \left(\sum_g \pi_g \pi_{i_1|g} \pi_{i_2|i_1 g t} \pi_{r=1|g} \right) \\ & + \sum_{i_1, t} Z_{i_1, r=0, t} \ln \left(\sum_g \pi_g \pi_{i_1|g} \pi_{r=0|g} \right), \end{aligned} \quad (90)$$

where $Z_{i_1 i_2, r=1, t}$ and $Z_{i_1, r=0, t}$ are the observed-data counts, with obvious notation. Maximization then proceeds by feeding (90) to a standard numerical optimizer.

The complete-data log-likelihood, needed for the EM algorithm, takes the form:

$$\begin{aligned} \ell^* &= \sum_{g, i_1, i_2, r, t} Z_{g i_1 i_2 r t}^* \ln(\pi_g \pi_{i_1|g} \pi_{i_2|i_1 g t} \pi_r |g) \\ &= \sum_g Z_{g++++}^* \ln(\pi_g) + \sum_{g, i_1} Z_{g i_1 +++}^* \ln(\pi_{i_1|g}) \\ &\quad + \sum_{g, i_1, i_2, t} Z_{g i_1 i_2 + t}^* \ln(\pi_{i_2|i_1 g t}) + \sum_{g, r} Z_{g++r+}^* \ln(\pi_r |g). \end{aligned} \quad (91)$$

Here, $Z_{g i_1 i_2 r t}^*$ is the (hypothetical) count in bivariate severity category (i_1, i_2) , in missingness group r , treatment arm t , and allocated to latent class g . A plus in lieu of a subscript indicates summation over the corresponding index. To proceed, the expected values of the complete-data sufficient statistics need to be computed. Thanks to the multinomial structure of ℓ^* , this is straightforward and hence the E step consists of:

$$\begin{aligned} E(Z_{g++++}^*) &= \pi_g Z_{++++}, \\ E(Z_{g i_1 +++}^*) &= \pi_g \pi_{i_1|g} Z_{i_1 +++}, \\ E(Z_{g i_1 i_2 + t}^*) &= \pi_g Z_{i_1 i_2, r=1, t} + \pi_g \pi_{i_2|i_1 g t} Z_{i_1+, r=0, t}, \\ E(Z_{g++r+}^*) &= \pi_g \pi_r |g Z_{++r+}. \end{aligned}$$

Finally, the M step takes the form of four separate logistic regressions, in the α , β , γ , and δ parameters, respectively, i.e., for each of the four terms in (91).

10 Concluding Remarks

Incomplete data are governed by a number of taxonomies and classification systems, two of which were of relevance here. A first one is concerned with the type of missing data mechanism (MCAR, MAR, and MNAR), whereas a second one classifies joint models for the outcome and missing data processes as belonging to the SeM, PMM, and SPM model families. Since MCAR merely comes ‘down to independence between both processes, perhaps conditional on fixed covariates, it takes a trivial form regardless of the model family. Whereas MAR has been defined in an SeM fashion, it has been characterized in a PMM way and studied further for the specific context of longitudinal data by Molenberghs *et al* (1998). Characterizing MAR in the SPM family is less straightforward and, to our knowledge, had not formally been done before. As a first result, we have

provided such a characterization in this paper, after defining a very general class of SPM that encompasses many earlier, specific instances. Since the characterization, in its full generality, may be somewhat awkward to work with, a more restrictive but appealing sub-class of SPM, satisfying MAR, has been proposed too.

Molenberghs *et al* (2007) established that every MNAR model fitted to a particular set of data can be replaced by a unique MAR counterpart, i.e., a model producing exactly the same fit to the observed data but where the prediction of the unobserved outcomes given the observed ones is of the MAR type. While their result is general, they focused on the SeM and PMM frameworks. As a second result, Creemers *et al* (2008) presented a generic format of this counterpart for the SPM family.

Apart from considerations on the basis of taxonomy, particular design aspects may be used to further focus one's model choices. For example, in a longitudinal study subject to dropout, one will often cast missingness mechanisms in terms of previous, current, and future measurements, rather than simply in terms of observed and unobserved measurements. There is a subtle distinction. While previous and observed measurements are synonymous in such a case, the unobserved measurements are further sub-divided into current and future measurements. Substantively, it is usually conceivable to assume that dropout is driven by the current, perhaps unobserved measurement, but it will not always be sensible to let dropout depend on future measurements. Constraining a SeM to this effect is particularly straightforward, but this is less trivial for the other two families. While Kenward *et al* (2003) translated this requirement to the PMM family, this had not yet been done for the SPM. As a third result reviewed here, Creemers *et al* (2008) characterize so-called *non-future dependent* mechanisms within the SPM family.

While the results reviewed in this paper are predominantly of a conceptual nature, a number of them have been illustrated, for enhanced insight, using both a continuous and a binary outcome from a two-armed clinical trial in toenail dermatophyte onychomycosis. In the continuous case, a linear mixed model was combined with logistic regression contributions for dropout. In the binary case, a dichotomous random effect was assumed, i.e., a latent class, reducing the analysis to one of incompletely observed contingency tables. Evidently, within each of the analyses done, a wider variety of model specifications can be entertained. Moreover, the ideas developed in this paper are generic and one could, for example, consider generalized linear mixed models for the entire binary profile, etc. (Molenberghs, and Verbeke, 2005).

Finally, the results of this paper open avenues for sensitivity analysis regarding substantive conclusions with respect to missingness (Molenberghs and Kenward, 2007). Thanks to the results in this and previous papers, and the ensuing classification of model families versus missing data mechanisms (Figure 3), one could, for example, select an insightful set models across families and mechanisms, perhaps supplementing MNAR models with their MAR counterparts, and then assess formally or informally how key conclusions change when

ranging over models.

Agradecimentos

The author gratefully acknowledges support from IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

Referências

- Baker, S.G. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics* **51**, 1042-52.
- Baker, S.G., Rosenberger, W.F., and DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine* **11**, 643-57.
- Best, N.G., Spiegelhalter, D.J., Thomas, A., and Brayne, C.E.G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society, Series A* **159**, 323-342.
- Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2007). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics* **63**, 000-000.
- Carpenter, J., Pocock, S., and Lamm, C.J. (2002). Coping with missing data in clinical trials: a model based approach applied to asthma trials. *Statistics in Medicine* **21**, 1043-1066.
- Chen, T., and Fienberg, S.E. (1974). Two-dimensional contingency tables with both completely and partially cross-classified data. *Biometrics* **30**, 629-42.
- Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M.G. (2008). Shared-parameter models and missingness at random. *Submitted for publication*.
- De Backer, M., De Keyser, P., De Vroey, C., and Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/day—a double-blind comparative trial. *British Journal of Dermatology* **134**, 16-17.
- DeGruttola, V. and Tu, X.M. (1994). Modelling progression of CD4 lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003-1014.
- Dempster, A.P., Laird, N.M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1-38.
- Diggle, P.J., and Kenward, M.G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49-93.

- Faucett, C.L. and Thomas, D.C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.
- Follmann, D., and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–68.
- Gill RD, van der Laan MJ, Robins JM. (1997). Coarsening at random: characterizations, conjectures and counterexamples. In: *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, DY Lin and TR Fleming (eds), pp. 255–294. New York: Springer. In: *Proceedings of the First Seattle Symposium on Survival Analysis*, pp. 255–294.
- Glynn, R.J., Laird, N.M., and Rubin, D.B. (1986). Selection modelling versus mixture modelling with non-ignorable nonresponse. In: *Drawing Inferences from Self Selected Samples*, Ed. H. Wainer, pp. 115–142. New York: Springer-Verlag.
- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.
- Hogan, J.W. and Laird, N.M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239–258.
- Hogan, J.W. and Laird, N.M. (1998). Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. *Statistical Methods in Medical Research* **7**, 28–48.
- Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G., and Kenward, M.G. (2006). The nature of sensitivity in missing not at random models. *Computational Statistics and Data Analysis* **50**, 830–58.
- Jansen, I., Molenberghs, G., Aerts, M., Thijs, H., and Van Steen, K. (2003). A Local influence approach applied to binary data from a psychiatric study. *Biometrics* **59**, 410–419.
- Kenward, M.G. (1998). Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine* **17**, 2723–32.
- Kenward, M.G., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53–71.
- Laird, N.M. (1994). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 84.
- LaValley, M.P. and DeGruttola, V. (1996). Models for empirical Bayes estimators of longitudinal CD4 counts. *Statistics in Medicine* **15**, 2289–2305.
- Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–34.
- Little, R.J.A. (1994a). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–83.

- Little, R.J.A. (1994b). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 78.
- Little, R.J.A. (1995). Modelling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112-21.
- Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: Wiley.
- Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, 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., Kenward, M.G., and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics* **50**, 15-29.
- Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley.
- Molenberghs, G., Kenward, M.G., and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with non-random dropout. *Biometrika* **84**, 33-44.
- Molenberghs, G., Michiels, B., Kenward, M.G., and Diggle, P.J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica* **52**, 153-61.
- Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M.G., Mallinckrodt, C., and Carroll, R.J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics* **5**, 445-64.
- Molenberghs, G., and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
- Pawitan, Y. and Self, S. (1993). Modeling disease marker processes in AIDS. *Journal of the American Statistical Association* **88**, 719–726.
- Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., and Buyse, M. (2002). Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. *Biometrical Journal* **44**, 921–935.
- Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2007). Shared parameter models under random-effects misspecification. *Biometrika* **94**, 000–000.
- Roberts, D.T. (1992). Prevalence of dermatophyte onychomycosis in the United Kingdom: Results of an omnibus survey. *British Journal of Dermatology* **126 Suppl. 39**, 23–27.
- Robins, J.M., Rotnitzky, A., and Scharfstein, D.O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321-39.

- Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106-21.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581-92.
- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Rubin, D.B.. (1994). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 80-82.
- Rubin, D.B., Stern, H.S., and Vehovar, V. (1995). Handling "don't know" survey responses: the case of the Slovenian plebiscite. *Journal of the American Statistical Association* **90**, 822-8.
- Schafer, J.L. and Graham, J.W. (2002). Missing data: our view of the state of the art. *Psychological Methods* **2**, 147-177.
- Schluchter, M.D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11**, 1861-1870.
- Taylor, J.M.G., Cumberland, W.G., and Sy, J.P. (1994). A stochastic model for analysis of longitudinal AIDS data. *Journal of the American Statistical Association* **89**, 727-736.
- TenHave, T.R., Kunselman, A.R., Pulkstenis, E.P., and Landis, J.R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54**, 367-83.
- Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3**, 245-65.
- Troxel, A.B., Harrington, D.P., and Lipsitz, S.R. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Appl. Statist.* **47**, 425-38.
- Tsiatis, A.A., DeGruttola, V., and Wulfsohn, M.S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90**, 27-37.
- Van Steen, K., Molenberghs, G., Verbeke, G., and Thijs, H. (2001). A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling: An International Journal* **1**, 125-142.
- Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M.G. (2001). Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics* **57**, 7-14.

- Wu, M.C., and Bailey, K.R. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* **7**, 337-46.
- Wu, M.C., and Bailey, K.R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939-55.
- Wu, M.C., and Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics* **44**, 175-88.
- Wulfsohn, M.S. and Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330-339.
- Xu, J. and Zeger, S.L. (2001a). The evaluation of multiple surrogate endpoints. *Biometrics* **57**, 81-87.
- Xu, J. and Zeger, S.L. (2001b). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375-387.

Tabela 5: (Part II). Analysis of the Slovenian Public Opinion Survey, restricted to the independence and attendance questions. The fit of models BRD1, BRD2, BRD7, and BRD9, and their MAR counterparts, to the hypothetical complete data is shown. The contingency tables' rows (columns) correspond to 'yes' vs. 'no' on the independence (attendance) question. The four tables in each row correspond to: (i) people responding to both questions; (ii) people responding to independence only; (iii) people responding to attendance only; (iv) people responding to neither question.

Fit of BRD1 and BRD1(MAR) to complete data							
1381.6	101.7	170.4	12.5	176.6	13.0	121.3	9.0
24.2	41.4	3.0	5.1	3.1	5.3	2.1	3.6
Fit of BRD2 to complete data							
1402.2	108.9	147.5	11.5	179.2	13.9	105.0	8.2
15.6	22.3	13.2	18.8	2.0	2.9	9.4	13.4
Fit of BRD2(MAR) to complete data							
1402.2	108.9	147.7	11.3	177.9	12.5	121.2	9.3
15.6	22.3	13.3	18.7	3.3	4.3	2.3	3.2
Fit of BRD7 to complete data							
1439	78	3.2	155.8	142.4	44.8	0.4	112.5
16	16	0.0	32.0	1.6	9.2	0.0	23.1
Fit of BRD9 to complete data							
1439	78	150.8	8.2	142.4	44.8	66.8	21.0
16	16	16.0	16.0	1.6	9.2	7.1	41.1
Fit of BRD7(MAR) and BRD9(MAR) to complete data							
1439	78	148.1	10.9	141.5	38.4	121.3	9.0
16	18	11.8	20.2	2.5	15.6	2.1	3.6

Tabela 6: *Analysis of the Slovenian Public Opinion Survey, restricted to the independence and attendance questions. Summaries on each of the Models BRD1–BRD9 are presented.*

Model	Structure	d.f.	loglik	$\hat{\theta}$	C.I.	$\hat{\theta}_{\text{MAR}}$
BRD1	(α, β)	6	-2495.29	0.892	[0.878;0.906]	0.8920
BRD2	(α, β_j)	7	-2467.43	0.884	[0.869;0.900]	0.8915
BRD3	(α_k, β)	7	-2463.10	0.881	[0.866;0.897]	0.8915
BRD4	(α, β_k)	7	-2467.43	0.765	[0.674;0.856]	0.8915
BRD5	(α_j, β)	7	-2463.10	0.844	[0.806;0.882]	0.8915
BRD6	(α_j, β_j)	8	-2431.06	0.819	[0.788;0.849]	0.8919
BRD7	(α_k, β_k)	8	-2431.06	0.764	[0.697;0.832]	0.8919
BRD8	(α_j, β_k)	8	-2431.06	0.741	[0.657;0.826]	0.8919
BRD9	(α_k, β_j)	8	-2431.06	0.867	[0.851;0.884]	0.8919

Tabela 7: *Toenail Data. Continuous, longitudinal unaffected-nail-length outcome. Parameter estimates (standard errors) for the model specified by (83) and (86).*

Effect	Unaffected nail length		Dropout	
	Par.	Est. (s.e.)	Par.	Est. (s.e.)
Mean structure parameters				
Intercept	β_0	2.510 (0.247)	γ_0	-3.127 (0.282)
Treatment	β_1	0.255 (0.347)	γ_1	-0.538 (0.436)
Time	β_2	0.558 (0.023)	γ_2	0.035 (0.041)
Treatment-by-time	β_3	0.048 (0.031)	γ_3	0.040 (0.061)
Variance-covariance structure parameters				
Residual variance	σ^2	6.937(0.248)		
Scale factor			γ_{01}	-0.076 (0.057)
Rand. int. variance	τ^2	6.507 (0.630)	$\gamma_{01}^2 \tau^2$	0.038 (0.056)

Tabela 8: *Toenail Data. Bivariate binary severity index at first and last time points. The observed data are shown, as well as the fit of Models ‘Bin1’ and ‘Bin2’, together with their corresponding counterparts. Both the fit to the observed data as well as to the hypothetical complete data are shown.*

Standard treatment				Experimental treatment					
Completers		Dropouts		Completers		Dropouts			
Observed data									
77	5		10	79	3		11		
42	9		3	42	3		6		
Fit of Model ‘Bin1’									
76.85	5.66	9.04	0.34	9.38	81.21	2.43	9.36	0.15	9.51
40.60	7.99	4.62	0.90	5.52	45.62	3.63	5.19	0.41	5.60
Fit of Model ‘Bin1(MAR)’									
77.12	5.39	8.77	0.61	9.38	81.32	2.32	9.24	0.26	9.51
40.61	7.98	4.62	0.91	5.52	45.63	3.63	5.18	0.41	5.59
Fit of Model ‘Bin2’≡‘Bin2(MAR)’									
75.86	5.58	9.72	0.72	10.44	80.16	2.40	10.27	0.31	10.58
41.50	8.15	3.74	0.73	4.47	46.61	3.72	4.20	0.34	4.53