Generalized shared-parameter models and missingness at random

Peer-reviewed author version

CREEMERS, An; HENS, Niel; AERTS, Marc; MOLENBERGHS, Geert; VERBEKE, Geert & Kenward, Michael G. (2011) Generalized shared-parameter models and missingness at random. In: STATISTICAL MODELLING, 11(4). p. 279-310.

DOI: 10.1177/1471082X1001100401
Handle: http://hdl.handle.net/1942/12220

Generalized Shared-parameter Models and Missingness at Random

An Creemers[1] **Niel Hens**[1,2] **Marc Aerts**[1] **Geert Molenberghs**[1,3]

Geert Verbeke[3,1] **Michael G. Kenward**[4]

[1] *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium*

[2] *Centre for Health Economics Research and Modeling Infectious Diseases (CHERMID), Centre for the Evaluation of Vaccination (WHO Collaborating Centre), Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium*

[3] *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

[4] *Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E7HT, UK*

Summary

When data are incomplete, models are often cataloged according to one of three modeling frameworks to which they belong: selection models (SeM), pattern-mixture models (PMM), and shared-parameter models (SPM). At the same time, the missing data mechanism is conventionally classified as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Whereas MCAR is naturally simple, in the sense that measurement and missingness processes are fully independent, and hence easy to describe in every modeling framework, this is less the case for MAR. The conventional definition (Rubin 1976) is cast in the SeM framework. Molenberghs *et al* (1998) provided a characterization for PMM. In this paper, MAR is characterized, for the first time, for the SPM framework too, based upon a general, appealing definition of a broad class of SPM. This result is important from a conceptual point of view. A specific sub-family, satisfying the MAR condition, is studied in some detail. Particular implications for non-monotone missingness as well as for time-ordered, longitudinal data subject to dropout are studied. In particular, it is indicated how SPM can be constrained such that dropout at a given point in time can depend on current and past, but not on future measurements, in analogy to the result by Kenward, Molenberghs, and Thijs [16] for the PMM family. While a natural requirement, it is less easily imposed in the PMM and SPM frameworks than in the SeM case. Some of the models proposed are illustrated using a clinical trial in toenail dermatophyte onychomycosis.

Keywords and phrases: Available-case missing value restrictions; Ignorability; Missing at random counterpart; Missing completely at random; Missing non-future dependent restrictions; Non-future dependence; Pattern-mixture model; Selection 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 [32] provided a formal framework for the field of incomplete data by introducing the important taxonomy of missing data mechanisms, consisting of *missing completely at random* (MCAR), *missing at random* (MAR), and *missing not at random* (MNAR). An MCAR mechanism potentially depends on observed covariates, but neither on observed nor unobserved outcomes. An MAR mechanism depends on the observed outcomes and perhaps also on the covariates, but not further on unobserved measurements. Finally, when an MNAR mechanism is operating, missingness does depend on unobserved measurements, maybe in addition to dependencies on covariates and/or on observed outcomes.

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 ith subject's outcomes, denoted $\boldsymbol{Y}_i$, and vector of missingness indicators, written $\boldsymbol{R}_i$, is factored as the marginal outcome distribution and the conditional distribution of $\boldsymbol{R}_i$ given $\boldsymbol{Y}_i$. 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 $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$ processes. An important version of such a model further asserts that, conditional on the latent variables, $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$ exhibit no further dependence. Rubin [32] 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. These conditions are sufficient, not necessary. All of these concepts will be formalized in Section 2 and amplified with the need arising in subsequent sections.

The concept of MAR has typically been framed within the SeM framework, while [25] 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. [23] 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* [15]. These authors focused on the SeM and PMM frameworks.

In this paper, 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* [23]. 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, Molenberghs, and Thijs [16], 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.

Our results are conceptual in nature, in the sense that we take no position as to whether either it is natural, for a particular application, to assume that missingness is MAR or does not depend on future observations. Rather, we ensure that the modeller is able to consider such mechanisms within the SPM framework, in analogy with the SEM and PMM frameworks. That said, in many situations, one would want to avoid missingness to further depend on future observations, given past ones.

The remainder of the paper is organized as follows. Notation and formal concepts, used throughout the paper, are detailed in Section 2. The background results regarding MAR counterparts to MNAR models, necessary in what follows, are reviewed in Section 3. Section 4 defines a general class of SPM models, within which MAR is then characterized. A particularly appealing set of MAR-type SPM, satisfying the characterization, is presented. It is also shown that there exist models of the SPM type that do not belong to this particular family. Implications for non-monotone missingness and longitudinal data with dropout, where time-ordering is important, are the subject of Sections 5 and 6, respectively. A set of clinical-trial data, is introduced and analyzed in Section 7.

2 Notation and Concepts

Let the random variable Y_{ij} denote the response of interest, for the ith study subject, designed to be measured at occasions t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$. 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 $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$. In addition, define a vector of missingness indicators $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{in_i})'$ with $R_{ij} = 0$ if Y_{ij} is observed and 1 otherwise. In the specific case of dropout, $\boldsymbol{R}_i$ can usefully be replaced by the dropout indicator

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

Note that the concept of dropout refers to time-ordered variables, such as in longitudinal studies. For a complete sequence, $\boldsymbol{R}_i = \boldsymbol{0}$ and/or $D_i = n_i$. It is customary to split the vector $\boldsymbol{Y}_i$ into observed ($\boldsymbol{Y}_i^o$) and missing ($\boldsymbol{Y}_i^m$) components, respectively. When $\boldsymbol{R}_i$ is conditioned up, $\boldsymbol{Y}_i^o$ and $\boldsymbol{Y}_i^m$ 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 $f(\boldsymbol{y}_i, \boldsymbol{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi})$, where the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describe the measurement and missingness processes, respectively. Covariates are assumed to be measured and grouped in a vector $\boldsymbol{x}_i$ 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 [32, 22]:

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

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; [18, 19]) using the reversed factorization

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

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* [42, 40, 41, 35, 11, 20]. One then might assume there exists a vector of random effects $\boldsymbol{b}_i$, conditional upon which the measurement and dropout processes are independent. This *shared-parameter model* (SPM) is formulated by way of the following factorization

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

and hence

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

Here, $\boldsymbol{b}_i$ 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 4.

3 Every MNAR Model Has an MAR Counterpart

In this section, based on the argument of Molenberghs *et al* [23], we restate that for every MNAR model fitted to a set of data, there is a unique MAR counterpart providing exactly the same fit to the data. We will sketch these results in view of their transposition in what follows to the general SPM case. 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.

In the first step, fit an MNAR model to the observed data, with likelihood:

$$L = \prod_i \int f(\boldsymbol{y_i}^o, \boldsymbol{y_i}^m, \boldsymbol{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) d\boldsymbol{y_i}^m. \tag{5}$$

Using hats for estimated parameters, express the full density in PMM form:

$$f(\boldsymbol{y_i}^o, \boldsymbol{y_i}^m | \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) f(\boldsymbol{r}_i | \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) = f(\boldsymbol{y_i}^o | \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) f(\boldsymbol{r}_i | \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) f(\boldsymbol{y_i}^m | \boldsymbol{y_i}^o, \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}). \tag{6}$$

A similar reformulation can be considered for an SPM, as will be shown in the next section. Molenberghs *et al* [23] formally that the fit remains the same after such a substition, which led them to state that 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 determine $h(\boldsymbol{y_i}^m | \boldsymbol{y_i}^o)$. This means that, 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. [25] have shown that, for the case of dropout, the so-called *available case missing value restrictions* (ACMV) provide a practical computational scheme. Molenberghs *et al* [23] discuss computational schemes and provide a number of illustrations, with particular emphasis on contingency tables subject to both monotone and non-monotone missingness.

When applying these ideas, computational issues will arise. There are various options available. For example, in a pattern-mixture context, some authors have made use of multiple imputation. The same could be envisaged for shared-parameter models. Admittedly, there are alternatives, however, and the choice among this will often be a pragmatic one.

4 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 [38, 44, 12, 29].

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 [42], Wu and Bailey [40], and Wu and Bailey [41]. Wu and Bailey [40] proposed such a model for what they termed informative right censoring. For a continuous response, Wu and Carroll [42] 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 [33] and DeGruttola and Tu [7] 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 [30] study the impact of random-effects misspecification in a shared parameter model. Beunckens *et al* [2] 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* [1] and Carpenter, Pocock, and Lamm [4]. Further references include Pawitan and Self [28], Taylor *et al* [34], Faucett and Thomas [10], Lavalley and DeGruttola [17], Hogan and Laird [13, 14], Wulfsohn and Tsiatis [43] and Xu and Zeger [45].

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 2, 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 [20], based on augmenting the joint density of $(\boldsymbol{y}_i, \boldsymbol{r}_i)$ with a vector of random effects $\boldsymbol{b}_i$:

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i, \boldsymbol{b}_i | \boldsymbol{\theta}, \psi, \boldsymbol{\xi}), \tag{7}$$

where $\boldsymbol{\xi}$ 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 (7), one can still consider the selection-model factorization:

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i, \boldsymbol{b}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{b}_i, \boldsymbol{\psi}) f(\boldsymbol{b}_i|\boldsymbol{\xi}) \tag{8}$$

and, likewise, the pattern-mixture model factorization:

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i, \boldsymbol{b}_i|\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\xi}) = f(\boldsymbol{y}_i|\boldsymbol{r}_i, \boldsymbol{b}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{b}_i, \boldsymbol{\psi}) f(\boldsymbol{b}_i|\boldsymbol{\xi}). \tag{9}$$

The notation is the same as in Section 2, with in addition $\boldsymbol{\xi}$ parameters describing the random-effects distribution. Little [20] refers to such decompositions as random-coefficient selection and pattern-mixture models, respectively. [21] and [46] present hybrid models with meaningful, identifiable parameters. Obviously, SeM (1) and PMM (2) follow by removing the random effects from (8) and (9), respectively or, at least, not having them in common between the models for $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$.

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

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i, \boldsymbol{b}_i|\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\xi}) = f(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\theta}) f(\boldsymbol{r}_i|\boldsymbol{b}_i, \boldsymbol{\psi}) f(\boldsymbol{b}_i|\boldsymbol{\xi}). \tag{10}$$

Model (10) 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 [11]. Note that, when $\boldsymbol{b}_i$ 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 $\boldsymbol{b}_i$ drives the entire process. Whilst holding on to the conditional-independence assumption, we will expand $\boldsymbol{b}_i$ to a set of latent structures, as in the following definition.

Definition 1 (A Generalized Shared-parameter Model Family.) *A general shared-parameter model is defined as one of the form*

$$f(\boldsymbol{y}_i^o|\boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{j}_i, \boldsymbol{\ell}_i) f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{k}_i, \boldsymbol{m}_i) f(\boldsymbol{r}_i|\boldsymbol{g}_i, \boldsymbol{j}_i, \boldsymbol{k}_i, \boldsymbol{q}_i), \tag{11}$$

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

In (11), parameters have been suppressed from notation. The same shorthand will be used in what follows, too. For convenience, write

$$\boldsymbol{b}_i = (\boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{j}_i, \boldsymbol{k}_i, \boldsymbol{\ell}_i, \boldsymbol{m}_i, \boldsymbol{q}_i). \tag{12}$$

Several remarks are in place. First, this is the most general random-effects model that can be considered in the sense that $\boldsymbol{g}_i$ is common to all three factors in (11), $\boldsymbol{h}_i$, $\boldsymbol{j}_i$, and $\boldsymbol{k}_i$ are shared between a pair of factors, and $\boldsymbol{\ell}_i$, $\boldsymbol{m}_i$, and $\boldsymbol{q}_i$ 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 (11) may come across as somewhat contrived. Our objective is not to postulate (11) 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 (11) assumes two different distributions for the outcome vector, *i.e.,* divorcing the observed from the missing components. This is not entirely the case because $\boldsymbol{g}_i$ and $\boldsymbol{h}_i$ still tie both factors together. The impact of $\boldsymbol{j}_i$, $\boldsymbol{k}_i$, $\boldsymbol{\ell}_i$, and $\boldsymbol{m}_i$ 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 (11) 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 (10) follows by removing all random effects but $\boldsymbol{g}_i$.

Definition (11) 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 or the PMM

characterization of MAR.

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

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

where integration over $\boldsymbol{b}_i$ is shorthand for integration over all component vectors making up $\boldsymbol{b}_i$, listed in (12), 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 $\boldsymbol{m}_i$ are not identifiable; they are included for completeness only.

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

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

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 (11):*

$$f(\boldsymbol{y}_i^o|\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{m}_i)f(\boldsymbol{r}_i|\boldsymbol{j}_i,\boldsymbol{q}_i), \tag{13}$$

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

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

Theorem 1 (A Class of MAR-based SPM Models.) *The shared-parameter model (13) is missing at random.*

We have not addressed necessity thus far. To this effect, we need to derive general expressions for MAR in the PMM case, respectively. First, for the left hand side:

$$\begin{aligned} & f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{r}_i) \\ &= \frac{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i,\boldsymbol{q}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}{\int\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i,\boldsymbol{q}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i\,d\boldsymbol{y}_i^m} \\ &= \frac{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i,\boldsymbol{q}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i,\boldsymbol{q}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}. \end{aligned} \tag{14}$$

Second, for the right hand side, consider:

$$\begin{aligned} & f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o) \\ &= \frac{\int\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i,\boldsymbol{q}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i\,d\boldsymbol{r}_i}{\int\int\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i,\boldsymbol{q}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i\,d\boldsymbol{y}_i^m\,d\boldsymbol{r}_i} \\ &= \frac{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i,\boldsymbol{\ell}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}. \end{aligned} \tag{15}$$

Equating (14) and (15) and, for brevity, integrating over random effects that occur in one component only, produces the general conditions, laid out in the next theorem.

Theorem 2 (Characterization of MAR in the Generalized Shared-parameter Family.) *A member of the general SPM family (11) is MAR if and only if*

$$\frac{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i)f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{j}_i)f(\boldsymbol{r}_i|\boldsymbol{g}_i,\boldsymbol{j}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}$$

$$= \frac{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i,\boldsymbol{h}_i) f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i) f(\boldsymbol{b}_i)\, d\boldsymbol{b}_i}{f(\boldsymbol{y}_i^o)}. \qquad (16)$$

Evidently, again assuming that $\boldsymbol{g}_i$, $\boldsymbol{h}_i$, and $\boldsymbol{k}_i$ cancel, reduces (16) to a tautological statement, showing that (13) satisfies Theorem 2.

There are situations where (16) is satisfied, without the triplet $(\boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{k}_i)$ vanishing, but these will necessarily be more *ad hoc* and less intuitively appealing than these laid out in Theorem 1. 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 (16) 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 $\boldsymbol{\ell}_i$, $\boldsymbol{m}_i$, and $\boldsymbol{q}_i$. From the four remaining random-effects, retain only $\boldsymbol{j}_i$ and $\boldsymbol{k}_i$, 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 (16) can be simplified to:*

$$\left(\sum_j \pi^1_{y_1|j}\pi_j\right) \cdot \left(\sum_{j,k} \pi^1_{y_1|j}\pi^2_{y_2|y_1k}\pi_{r=0|jk}\pi_j\pi_k\right)$$

$$= \left(\sum_{j,k} \pi^1_{y_1|j}\pi_{r=0|jk}\pi_j\pi_k\right) \cdot \left(\sum_{j,k} \pi^1_{y_1|j}\pi^2_{y_2|y_1k}\pi_j\pi_k\right), \qquad (17)$$

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 (17) 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 (17) produces two equations,* i.e.,*there are two constraints to be satisfied. For the first equation, in* $(Y_1, Y_2) = (0, 0)$*, choose* $x = \pi^2_{0|01}$ *as the parameter to be determined. This means that (17) is a linear equation in* x*. Clearly, setting* $\pi^2_{0|00} = \pi^2_{0|01}$ *solves the equation, based on two observations. First, a constant factor* $\pi^2_{y2|y1}$ *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 $\boldsymbol{g}_i$ is present; and (2) where only $\boldsymbol{h}_i$ and $\boldsymbol{j}_i$ are present. In these two cases, as well as in Example 1, a single random effect describes $\pi^2_{y_2|y_1}$.. 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.

Example 2 (MAR Example Violating Definition 1.) *Retain the setting of Example 1, but now with the pair of random effects* $\boldsymbol{h}_i$ *and* $\boldsymbol{k}_i$ *present. This particular choice leads to a different simplification of (16):*

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

$$= \left(\sum_{h,k} \pi^1_{y_1|h}\pi_{r=0|k}\pi_h\pi_k\right) \cdot \left(\sum_{h,k} \pi^1_{y_1|h}\pi^2_{y_2|y_1hk}\pi_h\pi_k\right). \qquad (18)$$

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 (18) for $(Y_1, Y_2) = (0, 0)$ *and* $(1, 0)$*. Concentrating on the first of these, and singling out* $\pi^2_{0|011}$ *as the parameter to identify from the others, it follows that*

$$\pi^2_{0|011} = \frac{ab - de}{df - ac}, \qquad (19)$$

with

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

The derivations for $(Y_1, Y_2) = (1, 0)$ *is entirely similar and leads to (19) with the first conditioning argument '1' rather than '0'. A numerical example is provided in Table 1, establishing that the random effects* $\boldsymbol{b}_i$ *and* $\boldsymbol{k}_i$ *do influence the distribution of* Y_2*, given* Y_1*, in the dropout pattern.*

Finally, the characterization of Theorem 2 allows us to construct an MAR counterpart to an arbitrary SPM of the form (11). It is necessary to (1) retain the fit of the model to the observed data, while (2) ensuring that (16) 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 $\boldsymbol{g}_i$, $\boldsymbol{h}_i$, and $\boldsymbol{k}_i$, where fitted parameters are plugged into the densities.

Theorem 3 (An MAR Counterpart to a Generalized SPM.) *The MAR counterpart, to an arbitrary general SPM of the type (11) is found by replacing* $f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{k}_i, \boldsymbol{m}_i)$ *with*

$$h(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{m}_i) = \int_{\boldsymbol{g}_i}\int_{\boldsymbol{h}_i}\int_{\boldsymbol{k}_i} f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{k}_i, \boldsymbol{m}_i)d\boldsymbol{g}_i d\boldsymbol{h}_i d\boldsymbol{k}_i \qquad (20)$$

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 (16), instead of $f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{k}_i, \boldsymbol{m}_i)$, of Theorem 2, 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 (20) becomes summation.

5 Non-monotone Missing Data

The characterization of MAR in the SPM family, the formulation of specific family (13), and the construction of an MAR counterpart to a general SPM, are all independent of the missing-data patterns that occur in a given study. Specifically, these results apply to monotone and non-monotone patterns alike. The same is true for earlier work on the definition of MAR in the PMM framework [25], and the MAR counterpart to a general model in the SeM and PMM families [23], both reviewed in Section 3.

Nevertheless, it is useful to realize that the purely probabilistic concept behind MAR, as intended by Rubin [32], is not necessarily the same as the pragmatic view taken by the statistical modeler. Indeed, consider the simple data setting of Examples 1 and 2. From a modeler's perspective, MAR might usefully mean that missingness depends on Y_1 but not on Y_2, whereas from a probabilistic point of view, it merely means that Y_1 and Y_2 influence missingness among the completers, whereas missingness is determined by Y_1 only among the incomplete observations.

This apparent discrepancy is resolved by noting that the modeler voluntarily restricts oneself to a practically meaningful sub-class of probabilistic MAR mechanisms. This notwithstanding, Molenberghs *et al* [23] provide examples of MAR mechanisms, often as so-called MAR counterparts to MNAR models, that would hardly be considered purely on grounds of appeal to the modeler. Also Molenberghs *et al* [25] provide such an example, taking the form of a 2×2 contingency table subject to non-monotone missingness.

In the same spirit, for SPM's, the modeler may consider an SPM of MAR type, *i.e.,* a model fulfilling the characterization of Theorem 2, practically unnatural. In this respect, the sub-class of Definition 2 may or may not be deemed a practically more appealing choice.

When measurements are collected longitudinally, then typically time is prominently present in model formulation, and most model choices will be judged in the light of their (un)desirable time-related implications. The presence of 'time' also provides a vehicle to make the set of measurements that influence missingness vary from pattern to pattern, while retaining an intuitively appealing flavor. For example, when missingness is confined to dropout, a natural restriction is to allow only past

measurements influence dropout, an MAR mechanism. Evidently, the set of past measurements is not static, but itself a function of the time point at which dropout occurs.

These and other *proper-time-dependence* considerations are the subject of the next section.

6 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 [9]: (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. [9], 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, Molenberghs, and Thijs [16] 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 6.1, and then present a similar taxonomy for the SPM in Section 6.2.

6.1 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 $(t = 1, \ldots, n)$. Thijs *et al* [36] 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* [36]: complete-case missing value restrictions [18], where information is borrowed from the completers only, available-case missing values, equivalent to MAR (Molenberghs *et al* [25]; see also Section 5), 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

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

Pattern-mixture models now take the form:

$$\begin{aligned} & f(y_1, \cdots, y_n, r = t) \\ &= f(y_1, \cdots, y_n | r = t) f(r = t) \\ &= f_t(y_1, \cdots, y_n) f(r = t) \\ &= f_t(y_1, \cdots, y_t) f_t(y_{t+1} | y_1, \cdots, y_t) f_t(y_{t+2}, \cdots, y_n | y_1, \cdots, y_{t+1}) f(r = t), \end{aligned} \tag{21}$$

where $f_t(y_1, \cdots, y_n) = f(y_1, \cdots, y_n | r = t)$. The first three factors in (21) 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*

$$f(r = t|y_1, \cdots, y_n) = f(r = t|y_1, \cdots, y_{t+1}). \tag{22}$$

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 Restrictions (NFMV).) *Within the PMM framework, we define non-future dependent missing value restrictions as follows:*

$$f(y_t|y_1, \cdots, y_{t-1}, r = j) = f(y_t|y_1, \cdots, y_{t-1}, r \geq t - 1), \tag{23}$$

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:

$$f(y_{t+1}|y_1, \cdots, y_t, r = t). \tag{24}$$

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, (23) 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 (24) of the same functional form as (23) 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, Molenberghs, and Thijs [16], establishes the equivalence between NFD and NFMV, showing the NFMV restrictions correspond to NFD, just as ACMV corresponds to MAR.

Theorem 4 (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 (23) 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 (24). Kenward, Molenberghs, and Thijs [16] offered a tractable, sampling-based implementation and applied it to the analysis of a set of data.

6.2 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 (22) 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 2 was derived from the standard MAR definition, it immediately follows that a characterization of NFD-SPM is as follows.

Theorem 5 (Non-future Dependent Shared-parameter Models.) *A member of the general SPM family (11) is NFD if and only if*

$$\frac{\int f(\boldsymbol{y}_i^{pc}|\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{j}_i)f(\boldsymbol{y}_i^{f}|\boldsymbol{y}_i^{pc},\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i)f(r_i|\boldsymbol{g}_i,\boldsymbol{j}_i,\boldsymbol{k}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}{\int f(\boldsymbol{y}_i^{pc}|\boldsymbol{g}_i,\boldsymbol{j}_i)f(r_i|\boldsymbol{g}_i,\boldsymbol{j}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}$$

$$= \frac{\int f(\boldsymbol{y}_i^{pc}|\boldsymbol{g}_i,\boldsymbol{h}_i)f(\boldsymbol{y}_i^{f}|\boldsymbol{y}_i^{pc},\boldsymbol{g}_i,\boldsymbol{h}_i)f(\boldsymbol{b}_i)\,d\boldsymbol{b}_i}{f(\boldsymbol{y}_i^{pc})}, \qquad (25)$$

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

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

Likewise, the sub-class (13) 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 (11):*

$$f(\boldsymbol{y}_i^{pc}|\boldsymbol{j}_i, \boldsymbol{\ell}_i) f(\boldsymbol{y}_i^{f}|\boldsymbol{y}_i^{pc}, \boldsymbol{m}_i) f(r_i|\boldsymbol{j}_i, \boldsymbol{q}_i), \tag{26}$$

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

The key assumption here is that all information about the missing data is contained in the observed data, given which no further information is needed from neither the missing data mechanism nor the random effects. 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 1.

7 The Toenail Data

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 [6]. TDO is a common toenail infection, difficult to treat, affecting more than 2 out of 100 persons [31]. 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 reduces our sample under consideration to 146 and 148 subjects, in group A and group B, respectively.

Figure 2 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 [39]. 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 3. Due to a variety of reasons, the outcome has been measured at all 7 scheduled time points, for only 224 (76%) out of the 294 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.

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.

7.1 Continuous Unaffected Nail Length

Consider a general model of the form (11), with random effects confined to $\boldsymbol{g}_i$, i.e., common to all three components. For the measurement model, assume a linear mixed model [39], with general form:

$$\boldsymbol{Y}_i|\boldsymbol{g}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{g}_i, \Sigma_i), \tag{27}$$

$$\boldsymbol{g}_i \sim N(0, D). \tag{28}$$

Based on (27) and (28), the so-called marginal model can be derived

$$\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i). \tag{29}$$

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 (27) as

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

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

$$\begin{aligned} \boldsymbol{Y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i \;\sim\; & N\Big[(X_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1} X_i^o)\boldsymbol{\beta} + \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\boldsymbol{y}_i^o + (Z_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1} Z_i^o)\boldsymbol{g}_i, \\ & \Sigma_i^{mm} - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\Sigma_i^{om}\Big]. \end{aligned} \tag{30}$$

Now, (30) 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 (29), now applied to (30), we obtain:

$$\begin{aligned} \boldsymbol{Y}_i^m|\boldsymbol{y}_i^o \;\sim\; & N\Big[(X_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1} X_i^o)\boldsymbol{\beta} + \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\boldsymbol{y}_i^o, \\ & (Z_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1} Z_i^o)D(Z_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1} Z_i^o)' \\ & + \Sigma_i^{mm} - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\Sigma_i^{om}\Big]. \end{aligned} \tag{31}$$

Hence, (31) is the MAR counterpart to (30). For the unaffected nail length, we choose for (27)–(28):

$$E(Y_{ij}|g_i, T_i, t_j, \boldsymbol{\beta}) = \beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j, \tag{32}$$

$g_i \sim N(0, d)$, 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 ($i = 1, \ldots, 298$). Finally, t_j is the time at which the jth measurement is taken ($j = 1, \ldots, 7$).

Given these choices, (30) and (31) simplify to

$$\boldsymbol{Y}_i^m|\boldsymbol{y}_i^o, g_i \;\sim\; N(X_i\boldsymbol{\beta} + Z_i^m g_i, \sigma^2 I_i), \tag{33}$$

$$\boldsymbol{Y}_i^m|\boldsymbol{y}_i^o \;\sim\; N(X_i\boldsymbol{\beta}, dJ_i + \sigma^2 I_i), \tag{34}$$

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 (11). The sequence $\boldsymbol{r}_i$ 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:

$$\text{logit}\left[P(R_{ij} = 1 | R_{i,j-1} = 0, g_i, T_i, t_j, \boldsymbol{\gamma})\right] = \gamma_0 + \gamma_{01} g_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j, \qquad (35)$$

($j > 1$), 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, $\gamma_{01} g_i \sim N(0, \gamma_{01}^2 d)$.

The model specified by (32) and (35) 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 4. 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 [39].

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 (33). When deemed less plausible, the fully general structure (30) can be implemented.

7.2 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 sometimes missing final point in time. The data are displayed in Table 5. By way of illustration, we will assume a single dichotomous random effect, of the $\boldsymbol{g}_i$ type. This imposes a latent-class structure. Decompose the cell probabilities as:

$$\pi_{gi_1i_2rt} = \pi_g \pi_{i_1|g} \pi_{i_2|i_1gt} \pi_{r|g}, \tag{36}$$

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 (36) are modeled as:

$$\begin{aligned}
\pi_g &= \frac{e^{\alpha g}}{1 + e^{\alpha}}, \\
\pi_{i_1|g} &= \frac{e^{(\beta_0 + \beta_1 g) i_1}}{1 + e^{\beta_0 + \beta_1 g}}, \qquad (37) \\
\pi_{i_2|i_1gt} &= \frac{e^{(\gamma_0 + \gamma_1 i_1 + \gamma_2 g + \gamma_3 i_1 g + \gamma_4 t) i_2}}{1 + e^{\gamma_0 + \gamma_1 i_1 + \gamma_2 g + \gamma_3 i_1 g + \gamma_4 t}}, \qquad (38) \\
\pi_{r|g} &= \frac{e^{(\delta_0 + \delta_1 g) r}}{1 + e^{\delta_0 + \delta_1 g}}.
\end{aligned}$$

In Model 'Bin1', we will set $\beta_1 = 0$ in (37) for reasons of identifiability. In Model 'Bin2', $\gamma_2 = \gamma_3 = 0$ in (38). 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 5. 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 (36) by their estimates, followed by summing over g. The MAR counterpart is obtained as $\pi_{gi_1i_2rt} = \pi_g \pi_{i_1|g} \widetilde{\pi}_{i_2|i_1t} \pi_{r|g}$, where

$$\widetilde{\pi}_{i_2|i_1t} = \sum_g \pi_g \pi_{i_2|i_1gt}.$$

Parameter estimation by both maximum likelihood, as well as the EM algorithm [8] is particularly easy. For direct likelihood, the log-likelihood function takes the form

$$\ell = \sum_{i_1,i_2,t} Z_{i_1 i_2, r=1,t} \ln\left(\sum_g \pi_g \pi_{i_1|g} \pi_{i_2|i_1 g t} \pi_{r=1|g}\right) + \sum_{i_1,t} Z_{i_1, r=0,t} \ln\left(\sum_g \pi_g \pi_{i_1|g} \pi_{r=0|g}\right), \quad (39)$$

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 (39) to a standard numerical optimizer.

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

$$\begin{aligned}
\ell^* &= \sum_{g,i_1,i_2,r,t} Z^*_{g i_1 i_2 r t} \ln\left(\pi_g \pi_{i_1|g} \pi_{i_2|i_1 g t} \pi_{r|g}\right) \\
&= \sum_g Z^*_{g++++} \ln(\pi_g) + \sum_{g,i_1} Z^*_{g i_1 +++} \ln\left(\pi_{i_1|g}\right) \\
&\quad + \sum_{g,i_1,i_2,t} Z^*_{g i_1 i_2 + t} \ln\left(\pi_{i_2|i_1 g t}\right) + \sum_{g,r} Z^*_{g++r+} \ln\left(\pi_{r|g}\right). \quad (40)
\end{aligned}$$

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:

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

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 (40).

8 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* [25]. 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* [23] 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, we present 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, Molenberghs, and Thijs [16] translated this requirement to the PMM family, this had not yet been done for the SPM. As a third result, we characterize so-called *non-future dependent* mechanisms within the SPM family.

While our results 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. [27].

It might appear counterintuitive that the issues arising from incompleteness are further compounded by allowing for a whole collection of random effects. While this adds a great deal of flexibility, thereby enabling proper characterization of MAR, it needs to be limited. In practice, substantive considerations would be critical in reducing the number of these, leaving a more manageable set which would form the basis for sensitivity analyses.

Finally, the results of this paper open avenues for sensitivity analysis regarding substantive conclusions with respect to missingness [24, 5]. Thanks to the results in this and previous papers, and the ensuing classification of model families versus missing data mechanisms (Figure 1), 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.

Acknowledgement

We gratefully acknowledge support from Belgian IUAP/PAI network "Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data".

References

[1] 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.

[2] Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2007). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics* **63**, 000–000.

[3] Beunckens, C., Sotto, C., Molenberghs, G.,, and Verbeke, G. (2007). An integrated sensitivity analysis of the Slovenian Public Opinion Survey data. *Submitted for publication*.

[4] 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.

[5] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M.G. (2009). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal* **00**, 000–000.

[6] 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.

[7] DeGruttola, V. and Tu, X.M. (1994). Modelling progression of CD4 lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.

[8] 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.

[9] Diggle, P.J., and Kenward, M.G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–93.

[10] 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.

[11] Follmann, D., and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168.

[12] Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480.

[13] 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.

[14] 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.

[15] 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**, 830–858.

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

[17] LaValley, M.P. and DeGruttola, V. (1996). Models for empirical Bayes estimators of longitudinal CD4 counts. *Statistics in Medicine* **15**, 2289–2305.

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

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

[20] Little, R.J.A. (1995). Modelling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–121.

[21] Little, R.J.A. (2009). Selection and pattern-mixture models. In: *Advances in Longitudinal Data Analysis,* G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs (eds.). London: CRC/Chapman & Hall, pp. 409–431.

[22] Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

[23] Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M.G. (2007). Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B* **00**, 000–000.

[24] Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies.* Chichester: John Wiley & Sons.

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

[26] 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**, 445–464.

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

[28] Pawitan, Y. and Self, S. (1993). Modeling disease marker processes in AIDS. *Journal of the American Statistical Association* **88**, 719–726.

[29] 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.

[30] Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2007). Shared parameter models under random-effects misspecification. *Biometrika* **94**, 000–000.

[31] 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.

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

[33] Schluchter, M.D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11**, 1861–1870.

[34] 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.

[35] 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**, 367–383.

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

[37] 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**, 425–438.

[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.

[39] Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer‘.

[40] 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**, 337–346.

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

[42] 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**, 175–188.

[43] Wulfsohn, M.S. and Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330–339.

[44] Xu, J. and Zeger, S.L. (2001a). The evaluation of multiple surrogate endpoints. *Biometrics* **57**, 81–87.

[45] Xu, J. and Zeger, S.L. (2001b). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387.

[46] Yuan, Y. and Little, R.J. (2008). Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics*, **64**, 000–000.

Table 1: *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 (11), but not to the specific MAR sub-class (13).*

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, given $k=0$	$\rho_0 = \pi_{0\|0}$	0.4500	$1-\rho_0 = \pi_{1\|0}$	0.5500
R, given $k=1$	$\rho_1 = \pi_{0\|1}$	0.8000	$1-\rho_1 = \pi_{1\|1}$	0.2000
Y_1, given $h=0$	$\pi^1_{0\|0}$	0.3000	$\pi^1_{1\|0}$	0.7000
Y_1, given $h=1$	$\pi^1_{0\|1}$	0.2000	$\pi^1_{1\|1}$	0.8000
Y_2, given $Y_1=0$, $h=0$, and $k=0$	$\pi^2_{0\|000}$	0.1500	$\phi^2_{1\|000}$	0.8500
Y_2, given $Y_1=0$, $h=0$, and $k=1$	$\pi^2_{0\|001}$	0.2500	$\pi^2_{1\|001}$	0.7500
Y_2, given $Y_1=0$, $h=1$, and $k=0$	$\pi^2_{0\|010}$	0.3500	$\pi^2_{1\|010}$	0.6500
Y_2, given $Y_1=0$, $h=1$, and $k=1$	$\pi^2_{0\|011}$	0.2857	$\pi^2_{1\|011}$	0.7143
Y_2, given $Y_1=1$, $h=0$, and $k=0$	$\pi^2_{0\|100}$	0.2000	$\pi^2_{1\|100}$	0.8000
Y_2, given $Y_1=1$, $h=0$, and $k=1$	$\pi^2_{0\|101}$	0.3000	$\pi^2_{1\|101}$	0.7000
Y_2, given $Y_1=1$, $h=1$, and $k=0$	$\pi^2_{0\|110}$	0.4000	$\pi^2_{1\|110}$	0.6000
Y_2, given $Y_1=1$, $h=1$, and $k=1$	$\pi^2_{0\|111}$	0.3625	$\pi^2_{1\|111}$	0.6375

Table 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%

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

	Group A		Group B	
# Obs.	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%

Table 4: *Toenail Data. Continuous, longitudinal unaffected-nail-length outcome. Parameter estimates (standard errors) for the model specified by (32) and (35).*

	Unaffected nail length		Dropout	
Effect	Parameter	Estimate (s.e.)	Parameter	Estimate (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)

Table 5: *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

$$
\begin{array}{lcccccccc}
\text{SeM} & : & \text{MCAR} & \subset & \text{MAR} & \subset & \text{NFD} & \subset & \text{general MNAR} \\
 & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\
\text{PMM} & : & \text{MCAR} & \subset & \text{ACMV} & \subset & \text{NFMV} & \subset & \text{general MNAR} \\
 & & & & & \cap & \neq & \subset & \\
 & & & & & & \text{interior} & & \\
 & & \updownarrow & & \updownarrow & & \updownarrow & & \updownarrow \\
\text{SPM} & : & \text{MCAR} & \subset & \text{Theorem 2} & \subset & \text{Theorem 5} & \subset & \text{general MNAR} \\
 & & & & \cup & & \cup & & \\
 & & & & \text{Definition 2} & \subset & \text{Definition 5} & &
\end{array}
$$

Figure 1: *Subset-elationships 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.*

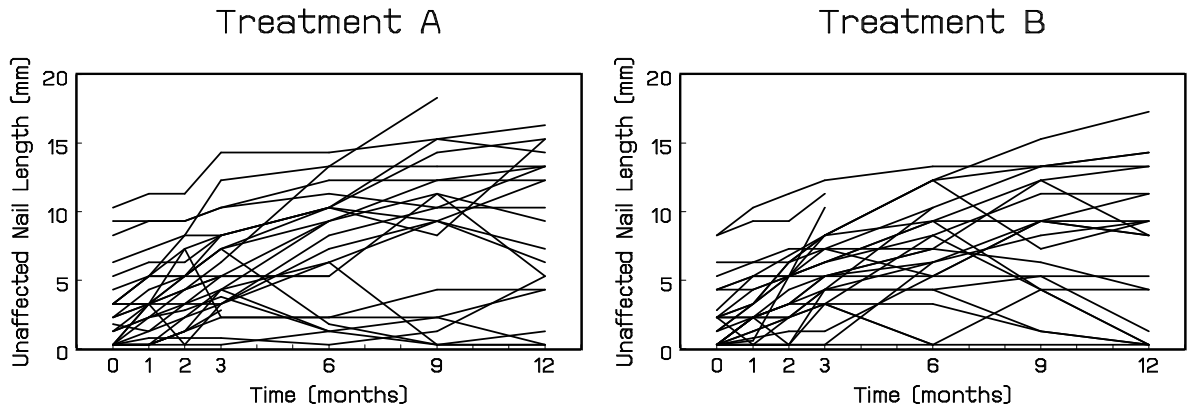

Figure 2: *Toenail Data. Individual profiles of 30 randomly selected subjects in each of the treatment groups in the toenail experiment.*

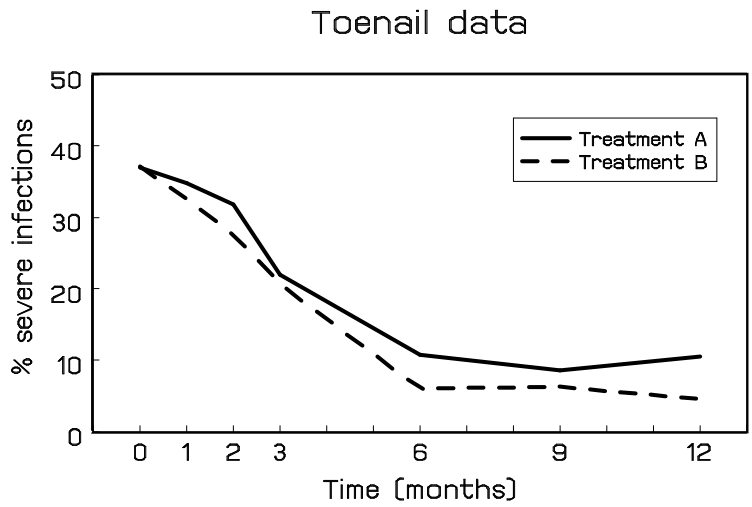

Figure 3: *Toenail Data. Evolution of the observed percentage of severe toenail infections in the two treatment groups separately.*

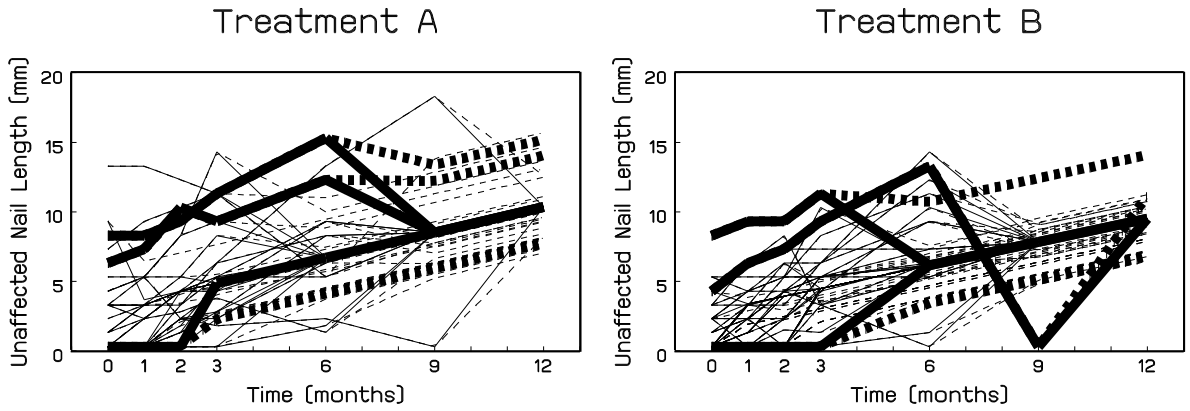

Figure 4: *Toenail Data. Individual profiles of subjects with incomplete data, for each treatment arm, extended using MNAR Model (32) (dashed line) and using the model's MAR counterpart (solid line). In each group, three subjects are highlighted.*