

Sensitivity Analysis for Shared-Parameter Models for Incomplete
Longitudinal Outcomes

Peer-reviewed author version

CREEMERS, An; HENS, Niel; AERTS, Marc; MOLENBERGHS, Geert; VERBEKE, Geert & Kenward, Michael G. (2010) Sensitivity Analysis for Shared-Parameter Models for Incomplete Longitudinal Outcomes. In: BIOMETRICAL JOURNAL, 52(1). p. 111-125.

DOI: 10.1002/bimj.200800235

Handle: <http://hdl.handle.net/1942/10817>

A Sensitivity Analysis for Shared-parameter Models for Incomplete Longitudinal Outcomes

An Creemers¹, Niel Hens^{1,2}, Marc Aerts¹, Geert Molenberghs^{*1,3}, Geert Verbeke^{3,1}, and Michael G. Kenward⁴

¹ I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium

² Epidemiology and Social Medicine, Universiteit Antwerpen, B-2610 Antwerpen, Belgium

³ I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

⁴ Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

Received 1 November 2004, revised 30 November 2004, accepted 2 December 2004

Published online 3 December 2004

Summary

All models for incomplete data either explicitly make assumptions about aspects of the distribution of the unobserved outcomes, given the observed ones, or at least implicitly imply such. One consequence is that there routinely exist a whole class of models, coinciding in their description of the observed portion of the data but differing with respect to their “predictions” of what is unobserved. Within such a class, there always is a single model corresponding to so-called random missingness, in the sense that the mechanism governing missingness depends on covariates and observed outcomes, but given these not further on unobserved outcomes. We employ these results, in the context of so-called shared-parameter models where outcome and missingness models are connected by means of common latent variables or random effects, to devise a sensitivity analysis framework. Precisely, the impact of varying unverifiable assumptions about unobserved measurements on parameters of interest is studied. Apart from analytic considerations, the proposed methodology is applied to assess treatment effect in data from a clinical trial in toenail dermatophyte onychomycosis. While our focus is on longitudinal outcomes with incomplete outcome data, the ideas developed in this paper are of use whenever a shared-parameter model could be considered.

Key words: Missing at random, missing not at random, onychomycosis, pattern-mixture model, selection model.

1 Introduction

Missing data are common throughout empirical research, especially in longitudinal studies with human subjects. Simple methods like a complete case analysis and last observation carried forward (Molenberghs and Kenward, 2007) rest on strong, often unrealistic, assumptions. During the last decades, a multitude of advanced models, allowing for potentially complicated ways in which missingness is influenced by observed and unobserved measurements, have been formulated. To describe such methods, the taxonomy of missing data-mechanisms of Little and Rubin (1987) is used.

The missingness process is said to be missing completely at random (MCAR) if the missingness is independent of both unobserved and observed data. When, conditional on the observed data, the missingness is independent of the unobserved measurements, we call the process missing at random (MAR). If both MCAR and MAR are not valid, the process is said to be missing not at random (MNAR). In the likelihood framework, and when parameters describing the missingness process are independent of the parameters describing the measurement process, MCAR and MAR are ignorable, in the sense that incomplete data can be analyzed as they are, without the need for an explicit missingness model. Although applied a lot in

* Corresponding author: e-mail: geert.molenberghs@uhasselt.be, Phone: +32 11 26 8205, Fax: +32 11 26 8299

practice, rarely the assumption of ignorability holds and one might want to consider more general (MNAR) models.

Models are often cataloged according to one of the three modeling frameworks. In a *selection model* (SeM), the joint distribution of the measurement and missingness process is factored as the marginal measurement model and the conditional distribution of the missingness process given the measurement process. A *pattern mixture model* (PMM) starts from the reverse factorization: the marginal distribution of the missingness process and the conditional distribution of the measurement process given the missingness process. Finally, in a *shared-parameter model* (SPM), a set of latent variables, e.g., random effects, is assumed to drive both the measurement and missingness process. One conventionally assumes that, conditional on this set, no further dependency exists between missing and measurement processes, although this can be generalized (Little, 1995).

Even though MNAR models seem to be the proper answer to the need for more flexible models, one has to be extremely careful with interpreting evidence for or against MNAR using only the data under study. This is because for every MNAR model, there is an MAR counterpart with exactly the same fit to the observed data, but differing in the prediction of what is unobserved (Molenberghs *et al.*, 2008). Therefore, a compromise between putting too much faith in such an MNAR model on the one hand, or ignoring them altogether on the other hand, consists of making them a component of a sensitivity analysis. The latter can be defined as an analysis in which several statistical models are considered simultaneously (Molenberghs and Verbeke, 2005; Molenberghs and Kenward, 2007). One route forward is to fit a selected number of models that are all plausible, and check to what extent conclusions depend on the model. This provides an indication about the belief one can put into these conclusions.

Until now, sensitivity analysis is reasonably well-developed in the SeM and the PMM frameworks (Molenberghs and Verbeke, 2005; Molenberghs and Kenward, 2007). In this paper, and based on the SPM modeling ideas of Creemers *et al.* (2009), we will present a sensitivity analysis paradigm in the SPM framework, which is both intuitively appealing and relatively easy to fit. While Creemers *et al.* (2009) is devoted to the introduction of a general shared-parameter framework, thereby allowing for the missing data mechanism to be missing at random, here we introduce sensitivity analysis tools, based on this framework. Sensitivity analysis had not been considered by Creemers *et al.* (2009).

While our focus is on longitudinal outcomes with incomplete outcome data, the ideas developed in this paper are of use whenever a shared-parameter model could be considered.

The rest of the paper is organized as follows. In Section 2, the onychomycosis trial is introduced. After introducing notation and concepts in Section 3, a variety of SPM models for these data are considered in Section 4. The sensitivity analysis framework is then developed in Section 5.

2 The Toenail Dermatophyte Onychomycosis Trial

The data were obtained from a randomized, double-blind, parallel group, multi-center 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

Table 1: Toenail Data. Number of available repeated measurements per subject, for each treatment arm.

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

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.

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 *mm*. This outcome has been studied extensively in Verbeke and Molenberghs (2000). Owing 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 1 summarizes the number of repeated measurements available per subject, per treatment group. We observe that the occurrence of missingness is similar in both treatment groups for longer, but not for shorter sequences.

There might be reason to believe that the missing data might be missing at random or even missing completely at random. This underscores why it is useful to have available the missing at random mechanism even within the shared-parameter framework. Further, it is prudent not to *a priori* exclude a missing not at random mechanism, strongly pointing towards the need for sensitivity analysis.

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. The outcomes can be grouped into a vector $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})'$. Define a vector of missingness indicators $\mathbf{R}_i = (R_{i1}, \dots, R_{in_i})'$, where $R_{ij} = 1$ if Y_{ij} is observed, and 0 otherwise. Partition \mathbf{Y}_i into two sub-vectors: \mathbf{Y}_i^o , containing those Y_{ij} for which $R_{ij} = 1$, and \mathbf{Y}_i^m , grouping the remaining components.

Out of the three frameworks, informally introduced in the previous section, the focus here is on the SPM; in particular, we focus on the general version (GSPM), introduced by Creemers *et al.* (2009). 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 processes, respectively. Covariates are assumed to be fully measured and grouped in a vector \mathbf{x}_i but, throughout, these are suppressed from notation, for simplicity and without loss of generality. The GSPM assumes a set of random-effects vectors $\mathbf{b}_i = (\mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{l}_i, \mathbf{m}_i, \mathbf{q}_i)$, characterized by the components of the full-density factorization to which they apply, in the following way:

$$\begin{aligned}
 & f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{l}_i, \mathbf{m}_i, \mathbf{q}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) \\
 &= f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{l}_i, \boldsymbol{\theta}) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{q}_i, \boldsymbol{\psi}).
 \end{aligned} \tag{1}$$

A few comments are in place. First, this is the most general shared-parameter model that can be considered in the sense that \mathbf{g}_i is common to all three factors, \mathbf{h}_i , \mathbf{j}_i , and \mathbf{k}_i are shared between a pair of factors, and \mathbf{l}_i , \mathbf{m}_i , and \mathbf{q}_i are restricted to a single factor. Note that the random effect \mathbf{m}_i is not identifiable, because

it merely describes the missing data. The same holds for k_i because it is aliased with q_i , of which one is used twice, the other one in a single factor only. However, the occurrence of k_i in the middle factor does not separate it from q_i , because the middle factor is unidentifiable. The same applies to j_i and g_i , which are not separable. Consequently, they are of use only in the context of sensitivity analysis.

Depending on the application, one may choose to either retain all random effects or to omit some. A special case of this GSPM is the conventional SPM (Wu, and Carroll, 1988; Wu, and Bailey, 1988, 1989; TenHave *et al.*, 1998; Follmann, and Wu, 1995; Little, 1995), where only one vector of random effects is assumed, g_i , conditional upon which the measurement and dropout processes are independent:

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

It is useful to gauge the implications of such simplifications, preferably also in terms of the missing data mechanism operating. In particular, Creemers *et al.* (2009) established 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 GSPM (1) may come across as somewhat contrived. It is not the objective to postulate GSPM (1) 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 that general GSPM, represented by (1), assumes two different distributions for the outcome vector, *i.e.*, divorcing the observed from the missing components. This is not entirely the case because g_i and h_i still tie both factors together. The impact of j_i , k_i , ℓ_i , and 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, one could expand GSPM (1) 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 (2) follows by removing all random effects but g_i . Sixth, in agreement with Creemers *et al.* (2009) and as mentioned above, also an MAR sub-class of the GSPM can be defined.

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

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

whereas the PMM factorization (Little, 1993, 1994) takes the reverse form:

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

The missing data mechanism can be classified as MCAR, MAR, or MNAR, originally defined by Rubin (1976) for the SeM setting. More recently, they were generalized to the PMM (Molenberghs *et al.*, 1998) and the GSPM (Creemers *et al.*, 2009) context. Also, an expression in the SPM framework is possible, as given in what follows. Let us now describe each of the missingness mechanisms for every framework.

MCAR holds when the measurement and missingness processes are independent (conditional on possible covariates). In the aforementioned modeling frameworks, this means:

$$\begin{aligned} \text{SeM} &: f(\mathbf{r}_i | \mathbf{y}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i | \boldsymbol{\psi}), \\ \text{PMM} &: f(\mathbf{y}_i | \mathbf{r}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i | \boldsymbol{\theta}), \\ \text{SPM} &: f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{b}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i | \boldsymbol{\theta}) f(\mathbf{r}_i | \boldsymbol{\psi}), \\ \text{GSPM} &: f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \boldsymbol{\theta}) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{q}_i, \boldsymbol{\psi}) \\ &= f(\mathbf{y}_i^o | \mathbf{h}_i, \boldsymbol{\theta}) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{h}_i, \mathbf{m}_i, \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{q}_i, \boldsymbol{\psi}). \end{aligned}$$

Table 2: Overview: Sub-models of the general shared-parameter model considered.

GSPM _k	Model	MAR counterpart
GSPM ₁	$f(\mathbf{y}_i^o \mathbf{g}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{g}_i)f(\mathbf{r}_i \mathbf{g}_i)$	$f(\mathbf{y}_i^o)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i)$
GSPM ₂	$f(\mathbf{y}_i^o \mathbf{g}_i, \mathbf{h}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i)f(\mathbf{r}_i \mathbf{g}_i)$	$f(\mathbf{y}_i^o)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i)$
GSPM ₃	$f(\mathbf{y}_i^o \mathbf{g}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{k}_i)f(\mathbf{r}_i \mathbf{g}_i, \mathbf{k}_i)$	$f(\mathbf{y}_i^o)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i)$
GSPM ₄	$f(\mathbf{y}_i^o \mathbf{g}_i, \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{j}_i)f(\mathbf{r}_i \mathbf{g}_i, \mathbf{j}_i)$	$f(\mathbf{y}_i^o \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i \mathbf{j}_i)$
GSPM ₅	$f(\mathbf{y}_i^o \mathbf{h}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{h}_i, \mathbf{k}_i)f(\mathbf{r}_i \mathbf{k}_i)$	$f(\mathbf{y}_i^o)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i)$
GSPM ₆	$f(\mathbf{y}_i^o \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{k}_i)f(\mathbf{r}_i \mathbf{j}_i, \mathbf{k}_i)$	$f(\mathbf{y}_i^o \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i \mathbf{j}_i)$
GSPM ₇	$f(\mathbf{y}_i^o \mathbf{h}_i, \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{h}_i)f(\mathbf{r}_i \mathbf{j}_i)$	$f(\mathbf{y}_i^o \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i \mathbf{j}_i)$
GSPM ₈	$f(\mathbf{y}_i^o \mathbf{h}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{h}_i)f(\mathbf{r}_i)$	$f(\mathbf{y}_i^o)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i)$
GSPM ₉	$f(\mathbf{y}_i^o \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i \mathbf{j}_i)$	$f(\mathbf{y}_i^o \mathbf{j}_i)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i \mathbf{j}_i)$
GSPM ₁₀	$f(\mathbf{y}_i^o)f(\mathbf{y}_i^m \mathbf{y}_i^o, \mathbf{k}_i)f(\mathbf{r}_i \mathbf{k}_i)$	$f(\mathbf{y}_i^o)f(\mathbf{y}_i^m \mathbf{y}_i^o)f(\mathbf{r}_i)$

The mechanism is MAR when missingness depends on the observed outcomes and observed covariates, but not further on the unobserved outcomes. In the various frameworks, MAR applies if and only if

$$\begin{aligned}
\text{SeM} &: f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{\psi}), \\
\text{PMM} &: f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{r}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i^m|\mathbf{y}_i^o, \boldsymbol{\theta}), \\
\text{SPM} &: \int f(\mathbf{y}_i^o, \mathbf{r}_i, \mathbf{b}_i) \{f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{b}_i) - f(\mathbf{y}_i^m|\mathbf{y}_i^o)\} d\mathbf{b}_i = 0, \\
\text{GSPM} &: \frac{\int f(\mathbf{y}_i^o|\mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i)f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i)f(\mathbf{r}_i|\mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i)f(\mathbf{b}_i) d\mathbf{b}_i}{\int f(\mathbf{y}_i^o|\mathbf{g}_i, \mathbf{j}_i)f(\mathbf{r}_i|\mathbf{g}_i, \mathbf{j}_i)f(\mathbf{b}_i) d\mathbf{b}_i} \\
&= f(\mathbf{y}_i^o|\mathbf{g}_i, \mathbf{h}_i)f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i)f(\mathbf{b}_i) d\mathbf{b}_i f(\mathbf{y}_i^o).
\end{aligned} \tag{5}$$

The PMM model in (5) states that the unobserved outcome \mathbf{y}_i^m given the observed outcome \mathbf{y}_i^o does not depend on the missingness. GSPM (5) was obtained by Creemers *et al.* (2009). A proof of the result can be found in their paper. These authors define the following MAR sub-class:

$$f(\mathbf{y}_i^o|\mathbf{j}_i, \boldsymbol{\ell}_i)f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{m}_i)f(\mathbf{r}_i|\mathbf{j}_i, \mathbf{q}_i), \tag{6}$$

where the random effects $\mathbf{g}_i, \mathbf{h}_i$ and \mathbf{k}_i vanish. Although this subclass does not contain all MAR models, it has intuitive appeal. Examples of models that satisfy GSPM (5) without belonging to (6) are described in Creemers *et al.* (2009).

The above considerations can also be applied to the conventional SPM framework. The corresponding MAR sub-class containing only MAR models in this case becomes

$$f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{b}_i) = f(\mathbf{y}_i^m|\mathbf{y}_i^o). \tag{7}$$

A useful result in this context was obtained by Molenberghs *et al.* (2008). They state that every fit to the observed data, obtained from fitting a MNAR model to a set of incomplete data, is exactly reproducible from an MAR decomposition. Working with a GSPM, the MAR counterpart is found by replacing $f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i, \boldsymbol{\theta})$ in (1) with (Creemers *et al.*, 2009)

$$h(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{m}_i) = \int_{\mathbf{g}_i} \int_{\mathbf{h}_i} \int_{\mathbf{k}_i} f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) d\mathbf{g}_i d\mathbf{h}_i d\mathbf{k}_i. \tag{8}$$

It is clear that this marginalization only changes the predictions for the unobserved data and thus the choice of $h(\cdot)$ does not alter the fit of the model to the observed portion.

4 A Collection of SPM Models for the Onychomycosis Data

In this section, we will analyze the longitudinal profile of unaffected nail length using various sub-models of the Model (1), obtained by removing portions of the random-effects structure. Table 2 presents ten different forms, together with their MAR counterparts. In the face of such a collection, two views can be taken. First, a particular model can be selected for the sake of its interpretation. In the same way, one or a few models can be excluded for this purpose. For example, scenario 10 assumes that the observed measurements are independent, which may be deemed implausible for wide classes of repeated measures designs. Second, models can be considered jointly, for the sake of sensitivity analysis. We will predominantly take the latter view.

Let \mathbf{b}_i be the q -dimensional vector of all random effects considered in a certain scenario. Then, for the measurement model, we assume a linear mixed model (Verbeke and Molenberghs, 2000) with general form:

$$\begin{cases} \mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \\ \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i), \\ \mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N \text{ independent.} \end{cases} \quad (9)$$

Here, \mathbf{Z}_i and \mathbf{X}_i are $(n_i \times q)$ and $(n_i \times p)$ dimensional matrices of known covariates, $\boldsymbol{\beta}$ is a p -dimensional vector containing the fixed effects and $\boldsymbol{\varepsilon}_i$ is an n_i -dimensional vector of residual components. \mathbf{D} and $\boldsymbol{\Sigma}_i$ are $q \times q$ and $n_i \times n_i$ covariance matrices, respectively.

Under all scenarios, the mean structure includes effects of treatment (T_i), time (t_j), their interaction, as well as a random effect. Further, we assume the variance-covariance matrix $\boldsymbol{\Sigma}_i$ to be of the form $\sigma^2 \mathbf{I}_7$ with \mathbf{I}_7 the 7×7 identity matrix. A complete overview for mean and variance-covariance structures in the ten scenarios is presented in Table 3. Note that the identity choice for the residual matrix $\boldsymbol{\Sigma}_i$ is not very restrictive, because the marginal correlation structure is induced by the variance-covariance structure of the random-effects system, and the latter is potentially very rich, a signature feature of the GSPM. Note that, even with these assumptions, some parameters will not be fully identified. In cases where the user wants to generalize $\boldsymbol{\Sigma}_i$, there is no problem in doing so. Of course, one then has to check carefully whether the resulting model is still identifiable.

Furthermore, a model for the missingness mechanisms needs to be formulated. We allow the sequence \mathbf{r}_i to take one of two forms: either a length-7 vector of ones, for a completely observed subject, or a sequence of k ones followed by a sole zero, for someone dropping out at time k , $k = 1, \dots, 6$. $k \geq 1$, because the initial measurement has been observed for all participants. For the missingness mechanism, we assume a logistic regression with the same parametric structure as for the mean structure. The various forms are presented in Table 4. γ_g , γ_k , and γ_j are scale factors for the shared random effects in the missingness model, to avoid forced equality of the variance in the measurement and dropout model. As a result, $\gamma_g g_i \sim N(0, \gamma_g^2 d_g^2)$, $\gamma_k k_i \sim N(0, \gamma_k^2 d_k^2)$, and $\gamma_j j_i \sim N(0, \gamma_j^2 d_j^2)$.

Scenario 1 coincides with the conventional SPM (2) and is studied in detail in Creemers *et al.* (2009). The models were fitted using the SAS procedure NLMIXED. Parameter estimates (standard errors) for all scenarios are displayed in Table 5. Although not much difference can be observed in the estimates for the mean structure parameters, the variance-covariance parameter estimates change considerably over the different scenarios. The total variance, roughly 13.5, is spread out over the residual variance and the random-effects variances. The scale parameters are all estimated to be negative, although non-significantly so. Note that the scale parameter γ_k is never identifiable, because the random effect k_i is the link between the missingness process and the missing observations given the observed ones, which can never be observed. The same problem holds when both random effects g_i and j_i are included in the model; then the scale parameter γ_j is not identifiable. For these analyses, they were fixed to 1. However, this unidentifiability provides us with the opportunity to perform sensitivity analysis, as will be outlined in the next section. These parameters will then play the role of sensitivity parameters.

Table 3: *Toenail Data. Mean structures for the observed responses given the random effects, mean structures for the missing responses given the observed responses, and the random effects and the variance-covariance matrix for the random effects for the various scenarios. T_i is the treatment indicator for subject i and t_j is the time at which the j th measurement is taken.*

Scen.	$E[Y_{ij}^o \mathbf{b}_i]$	$E[Y_{ij}^m y_i^o, \mathbf{b}_i]$	BD
1	$\beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	d_g^2
2	$\beta_0 + g_i + h_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + g_i + h_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\begin{pmatrix} d_g^2 & 0 \\ 0 & d_h^2 \end{pmatrix}$
3	$\beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + g_i + k_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\begin{pmatrix} d_g^2 & 0 \\ 0 & d_k^2 \end{pmatrix}$
4	$\beta_0 + g_i + j_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\begin{pmatrix} d_g^2 & 0 \\ 0 & d_j^2 \end{pmatrix}$
5	$\beta_0 + h_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + h_i + k_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\begin{pmatrix} d_h^2 & 0 \\ 0 & d_k^2 \end{pmatrix}$
6	$\beta_0 + j_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + k_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\begin{pmatrix} d_j^2 & 0 \\ 0 & d_k^2 \end{pmatrix}$
7	$\beta_0 + h_i + j_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + h_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\begin{pmatrix} d_h^2 & 0 \\ 0 & d_j^2 \end{pmatrix}$
8	$\beta_0 + h_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + h_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	d_h^2
9	$\beta_0 + j_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	d_j^2
10	$\beta_0 + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	$\beta_0 + k_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j$	d_k^2

Table 4: *Toenail Data. Models for the missingness mechanism in the different scenarios.*

Scenario	$\text{logit}[P(R_{ij} = 1 R_{i,j-1} = 0, b_i, T_i, t_j, \gamma)]$
1	$\gamma_0 + \gamma_g g_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
2	$\gamma_0 + \gamma_g g_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
3	$\gamma_0 + \gamma_g g_i + \gamma_k k_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
4	$\gamma_0 + \gamma_g g_i + \gamma_j j_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
5	$\gamma_0 + \gamma_k k_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
6	$\gamma_0 + \gamma_k k_i + \gamma_j j_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
7	$\gamma_0 + \gamma_j j_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
8	$\gamma_0 + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
9	$\gamma_0 + \gamma_j j_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$
10	$\gamma_0 + \gamma_k k_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j$

At first sight, negative estimates for the sensitivity parameters may appear to be counter-intuitive. However, they merely imply that, with an increase of unaffected nail length, patients are less inclined to remain on study. It ought to be kept in mind that the active period of the compound is the first three months only. Thereafter, maintenance visits take place. Those patients fully cured are less likely to remain faithful to the trial schedule.

Using empirical Bayes estimates, predictions for the incomplete profiles are obtained. Figure 1 displays the incomplete profiles, extended beyond the time of dropout, using predictions based on the original model (dashed lines) and its MAR counterpart (solid lines) for 3 selected scenarios, #1, #3, and #10. In what follows, other scenarios will be discussed as well. The MAR counterpart reduces all predictions to the same profile, while in most cases the MNAR prediction yields different evolutions per subjects. This

is because the prediction of Y_i^m depends on y_i^o only, in case of the MAR model, whereas for MNAR also g_i , h_i , and k_i intervene. The fact that the various scenarios produce quite similar predictions enhances confidence in the conclusions reached about the evolution of toenail infection and the differences between both treatment arms. Furthermore, it is entirely possible to implement a more general variance-covariance structure. Again, one then has to check the identifiability of the resulting model.

For Scenarios 6, 9, and 10, MNAR and MAR predictions coincide. For Scenario 9 this is clearly due to the fact that the MNAR model is the same as its MAR counterpart (see Table 2). In Scenarios 6 and 10 this can be explained by the small (and highly non-significant) variance of the random effect k . While the fixed-effect estimates for Scenario 10 are somewhat different from the other scenarios, this is still minor in view of the precision estimates. Nevertheless, Scenario 10 is qualitatively different from the others, in the sense that the random effects within the observed sequence are assumed independent; this is rather unrealistic.

5 A Sensitivity Analysis in the Shared-parameter Framework

As mentioned in the introduction, blindly fitting a MNAR model is not recommendable, because of the strong, largely untestable assumptions they rest upon. This suggests further sensitivity analysis.

Particularly in the SeM framework, different methods that investigate the sensitivity of the results with respect to the model assumptions were proposed. For example Nordheim (1984); Little (1994); Rubin (1994); Laird (1994); Fitzmaurice, Molenberghs, and Lipsitz (1995); Molenberghs *et al.* (1999); Kenward (1998) and Kenward and Molenberghs (1999) considered useful, though relatively *ad hoc* methods to apply sensitivity analyses in the SeM framework. Van Steen *et al.* (2001) considered a local influence approach for ordinal data with dropout, which was generalized to the continuous case by Jansen *et al.* (2006). These authors pay attention, not only to the assumed distributional form of the model, but also to the impact one or a few influential subjects may have on the dropout and/or measurement model parameters. Hens *et al.* (2005) consider kernel weighted local influence measurements. Molenberghs, Kenward, and Goetghebeur (2001); Kenward, Goetghebeur, and Molenberghs (2001) work with the “interval of ignorance,” a sensitivity approach that takes into account two sources of uncertainty, i.e., resulting from finite sampling and incompleteness, respectively. Also, the PMM framework offers opportunities to perform a sensitivity analysis. Overviews are presented in Thijs *et al.* (2002) and Molenberghs *et al.* (2004).

In this section, we will propose a method for sensitivity analysis within the GSPM framework. This can, of course, take several forms. First, like in the previous section, a collection of sub-models can be considered and stability of inferences compared across them. Second, like will also be done here, sensitivity analysis can be placed within a particular model, making use of its non-identifiability. Technically, this means that a model actually corresponds to an infinite collection of models. For the sake of illustration, this will be done within Scenario 6.

As stated, some of the GSPM models in Section 4 were not fully identifiable. Consider, for example, Scenario 6, which only includes the random effects j and k . Because the second term in (1) ($f(y_i^m | y_i^o, g_i, h_i, k_i, m_i, \theta)$) is never observed, the sensitivity parameter γ_k is not identifiable. Only by fixing this γ_k , the *sensitivity parameter*, can the model be fitted. The principle is to let γ_k take different values and compare the inferences across fits. To this end, a grid in the sensitivity parameter space will be defined, and γ_k will take all values in this grid. For each of the values, the model will be fitted and then, based on this model, y_i^m will be imputed multiple times. This will be done using conventional multiple imputation (Little and Rubin, 2002; Molenberghs and Kenward, 2007). Details will be given in Section 5.1. Re-fitting the model using these imputations and summarizing the different inferences resulting from different imputations into a single set of inferences, purports to sensitivity analysis. In Section 5.1, an algorithmic outline for such a sensitivity analysis is proposed; in Section 5.2 results for this sensitivity analysis in Scenario 6 are summarized.

Table 5: Toenail Data. Parameter estimates and standard errors for the SPM fits.

Effect	Par.	Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5
<i>Mean Structure</i>						
Measurement						
intercept	β_0	2.510 (0.247)	2.510 (0.247)	2.510 (0.247)	2.510 (0.247)	2.517 (0.247)
treatment	β_1	0.255 (0.347)	0.255 (0.347)	0.255 (0.347)	0.255 (0.347)	0.249 (0.346)
time	β_2	0.558 (0.023)	0.558 (0.023)	0.558 (0.023)	0.558 (0.023)	0.561 (0.023)
treat x time	β_3	0.048 (0.031)	0.048 (0.031)	0.048 (0.031)	0.049 (0.031)	0.047 (0.031)
Dropout						
intercept	γ_0	-3.127 (0.282)	-3.109 (0.783)	-3.206 (0.864)	-3.208 (0.863)	-3.112 (0.281)
treatment	γ_1	-0.538 (0.436)	-0.545 (0.445)	-0.544 (0.445)	-0.544 (0.446)	-0.535 (0.436)
time	γ_2	0.035 (0.041)	0.039 (0.061)	0.039 (0.065)	0.039 (0.065)	0.033 (0.041)
treat x time	γ_3	0.040 (0.061)	0.039 (0.061)	0.039 (0.061)	0.039 (0.061)	0.041 (0.061)
<i>Variance-covariance structure</i>						
Measurement						
resid. var.	σ^2	6.937 (0.248)	6.938 (0.248)	6.937 (0.248)	6.937 (0.248)	6.940 (0.248)
rand. int. var.	d_g^2	6.507 (0.630)	1.256 (8.553)	6.507 (0.630)	6.366 (1.528)	
	d_h^2		5.249 (8.565)			6.486 (0.628)
	d_k^2			0.165 (1.707)		1.42E-6 (0.248)
	d_j^2				0.141 (1.394)	
Dropout						
scale factor	γ_g	-0.076 (0.057)	-0.409 (2.891)	-0.079 (0.066)	-0.103 (0.279)	
	γ_k			fixed:1		fixed:1
	γ_j				fixed:1	
rand. int. var.	$\gamma_g^2 d_g^2$	0.038 (0.056)	0.210 (1.557)	0.041 (0.068)	0.067 (0.351)	
Effect	Par.	Scen. 6	Scen. 7	Scen. 8	Scen. 9	Scen. 10
<i>Mean Structure</i>						
Measurement						
intercept	β_0	2.510 (0.247)	2.510 (0.247)	2.517 (0.247)	2.510 (0.247)	2.553 (0.179)
treatment	β_1	0.255 (0.347)	0.255 (0.347)	0.249 (0.346)	0.255 (0.347)	0.217 (0.251)
time	β_2	0.558 (0.023)	0.558 (0.023)	0.561 (0.0226)	0.558 (0.023)	0.575 (0.0307)
treat x time	β_3	0.048 (0.031)	0.048 (0.031)	0.049 (0.031)	0.048 (0.031)	0.036 (0.043)
Dropout						
intercept	γ_0	-3.206 (0.864)	-3.209 (0.780)	-3.112 (0.281)	-3.127 (0.282)	-3.112 (0.281)
treatment	γ_1	-0.544 (0.445)	-0.545 (0.445)	-0.535 (0.436)	-0.538 (0.436)	-0.535 (0.436)
time	γ_2	0.039 (0.065)	0.039 (0.061)	0.033 (0.041)	0.035 (0.041)	0.033 (0.041)
treat x time	γ_3	0.039 (0.061)	0.039 (0.061)	0.041 (0.061)	0.040 (0.061)	0.041 (0.061)
<i>Variance-covariance structure</i>						
Measurement						
resid. var.	σ^2	6.937 (0.248)	6.938 (0.248)	6.940 (0.248)	6.937 (0.248)	13.527 (0.444)
rand. int. var.	d_g^2					
	d_h^2		5.248 (1.861)	6.486 (0.628)		
	d_k^2	0.1647 (1.707)				1.561E-9 (2.04E-4)
	d_j^2	6.507 (0.630)	0.256 (8.514)		6.507 (0.630)	
Dropout						
scale factor	γ_k	fixed:1				fixed:1
	γ_j	-0.079 (0.066)	-0.408 (2.876)		-0.076 (0.057)	
rand. int. var.	$\gamma_j^2 d_j^2$	0.041 (0.068)	0.209 (1.549)		0.038 (0.056)	

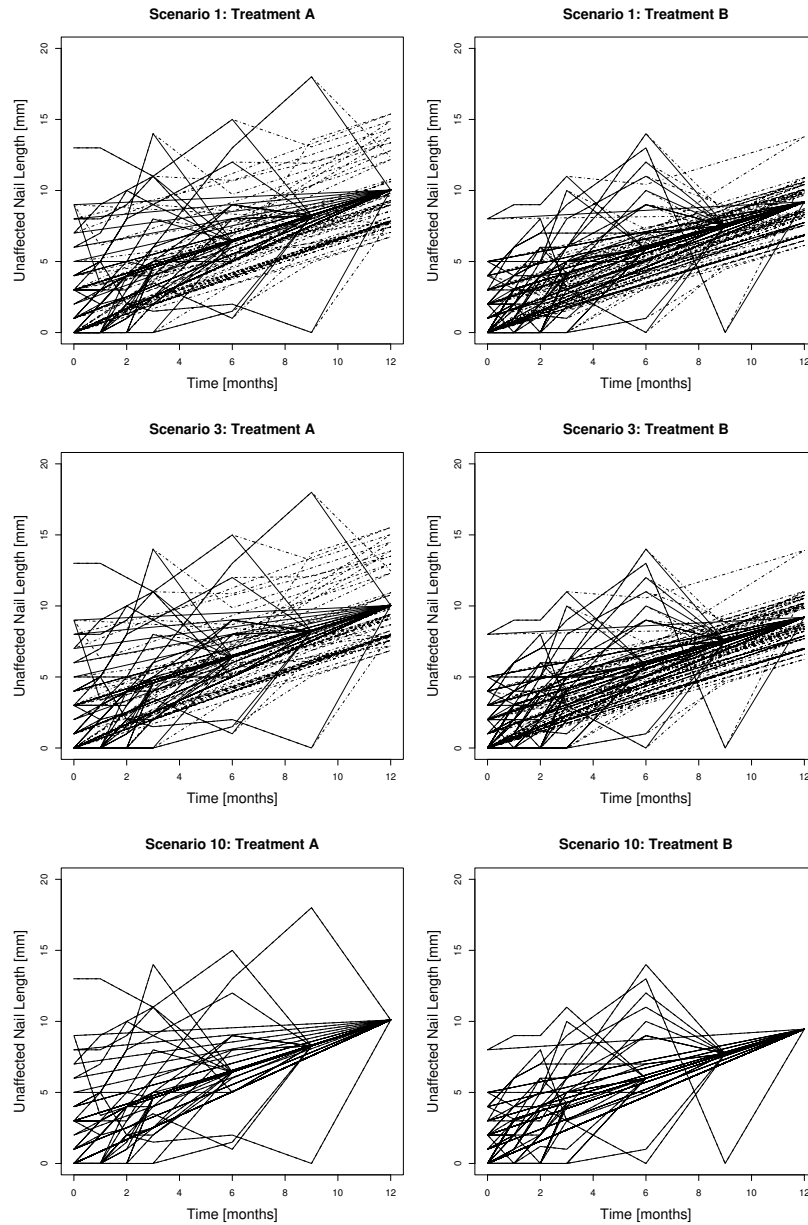


Fig. 1: Toenail Data. Individual profiles of subjects with incomplete data, for each treatment arm, extrapolated using MNAR model (dashed lines) and using the model's MAR counterpart (solid lines) for three of the models in Table 2.

5.1 The method step by step

For convenience, we will assume Scenario 6, but the sequence described in this section is applicable to all scenarios with at least 1 unidentifiable sensitivity parameter. The sensitivity parameter γ_k cannot be identified in this specific case and needs to be pre-specified. We define a range for the values of this sensitivity parameter, from -30 to 30 , with step length 1. Of course, should it be decided that a specific

sub-interval is of interest, one might decide to zoom in and apply rather a fine grid to it. This might occur, for example, when large values are deemed very implausible, because they would put much more emphasis on the missingness probability than on the missing observations. However, considering large values has the virtue of not *a priori* excluding scenarios. When (graphically) exploring the results, the user can decide to give more or less weight to certain regions of the sensitivity parameters, in accordance with prior belief or other scientific opinions held.

We perform the following steps for every fixed value in the range of the sensitivity parameter:

1. Fit the model using the fixed value for γ_k .
2. Calculate the conditional distribution $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{k}_i, \boldsymbol{\theta})$.
3. Using empirical Bayes estimates, simulate M values for \mathbf{y}_i^m , conditional on \mathbf{y}_i^o and the random effect \mathbf{k}_i . In this way, M different datasets are obtained.
4. On each of these (complete) datasets, fit a linear mixed model of the form

$$\begin{cases} Y_{ij} = \beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j + \varepsilon_{ij}, \\ g_i \sim N(0, d_g^2), \\ \varepsilon_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i), \\ g_1, \dots, g_N, \varepsilon_1, \dots, \varepsilon_N \text{ independent.} \end{cases} \quad (10)$$

Here, again, we assume $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_7$, with \mathbf{I}_7 the 7×7 identity matrix.

5. Use the results from these M datasets to construct a single set of inferences.

Step 3 can be seen as multiple imputation under MNAR and thus enables the use of conventional multiple imputation to perform step 5.

We will set $M = 5$. Such a small number is often satisfactory, as claimed by Little and Rubin (2002). However, for small to moderate datasets, considering larger number of imputations poses no problems.

Because $\gamma_k = 0$ corresponds to MAR, repeating this procedure for the whole grid results in a sensitivity analysis around MAR in the GSPM framework. If results are very sensitive to the value of γ_k , one should be very careful interpreting them.

5.2 Results for Scenario 6

The above procedure will now be applied on the onychomycosis dataset. We will focus on Scenario 6, even though others are possible, as well.

Scientific interest might focus on different parameters. First, one could be interested in the treatment effect on the last time point. Even though this effect is not captured by a single model parameter, using the imputed datasets, one can test this hypothesis. Second, all parameters in Model (10) could be of interest. We will test the significance of the treatment effect, the time effect, and the treatment \times time interaction. Next to this, we will also consider the variance parameters σ^2 and d_g^2 .

Figure 2 shows results for the test of a treatment effect at the last time point, i.e., the seventh measurement, taken at month 12. The left plot contains the values for the test statistic, while the right plot presents the corresponding p values. The points in both plots correspond to dataset-specific values. Per grid point, these 5 different values can be reduced to a single one using multiple imputation; they are indicated by the lines. The line with the overall p values is located above the 0.05 line, which indicates there is no significant treatment difference in unaffected nail length at month 7. For large negative values of γ_k , some dataset-specific p values are below 0.05. This indicates that ranging over γ_k implies some fluctuation in the p -values, but it never becomes significant.

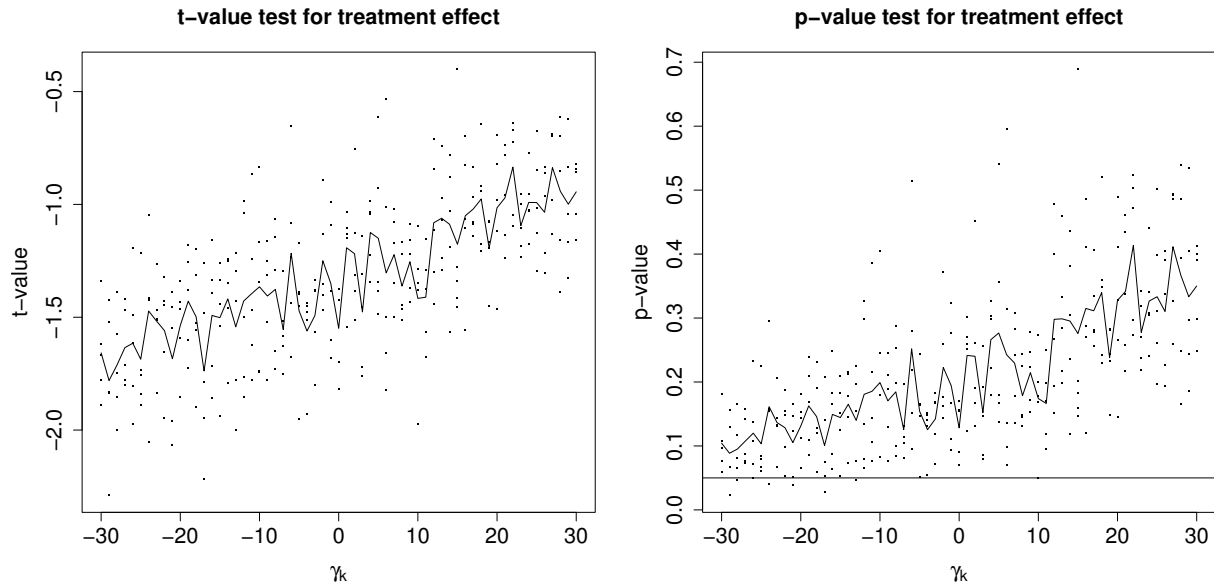


Fig. 2: Toenail Data. Results for sensitivity analysis applied under Scenario 6: t -test for difference between treatment effect at the last time point. (t -values (left) and p -values (right)). Points correspond to the dataset specific values (5 for every grid point), lines correspond to overall value.

In Figure 3, parameter estimates and p -values to test their significance are presented. The variance parameters σ^2 and dg^2 have p -values of order, respectively, 10^{-5} and 10^{-4} and thus are highly significant. Mean structure parameters β_1 (treatment effect) and β_3 (interaction effect) are not significant. Again, from the plot with p -values for β_3 , it turns out that varying γ_k controls the evidence for the effect of treatment, in as far as it is captured by the treatment-by-time interaction. It should be noted that, over the entire range of the sensitivity parameter, the p -values for β_3 remain consistent with no strong difference in evolution. Given that this is the main scientific parameter, this analysis strongly elevates the level of confidence with the conclusions reached. Further, a clear time effect (β_2) is observed (p values of the order 10^{-5}). Because parameter estimates for β_2 are increasing with increasing γ_k , the larger the sensitivity parameter, the bigger the time effect. Although the value of γ_k changes across models, the main result is qualitatively invariant. The inferences resulting from examining the p -values are only moderately sensitive to the choice of γ_k ; hence they can interpreted with confidence, even though the confidence statements issue a somewhat more subtle message.

6 Concluding Remarks

Missing data mechanisms (MCAR, MAR, and MNAR) and modeling frameworks (SeM, PMM, and (G)SPM) are familiar taxonomic entities in the modeler's toolkit. One should be cautious when fitting a MNAR model, in whatever framework, because they rest on strong, unverifiable assumptions. Therefore, a sensitivity analysis should be part of the data analysis when dealing with incompleteness. A lot of work has been done in the SeM framework, and some for PMM. The (G)SPM framework has not been considered thus far.

In this paper, we have proposed a sensitivity analysis tool for the GSPM framework. This is done by letting the values for unidentifiable sensitivity parameters take all values over a pre-specified grid. Based on empirical Bayes estimates, the missing values in the original dataset are imputed a number of times.

Using multiple imputation tools, these datasets are analyzed and inferences obtained. If results are very sensitive to the value of the sensitivity parameter, inferences should not be considered trustworthy.

While empirical Bayesian ideas are used, we have not presented a fully Bayesian sensitivity analysis. Arguably, such would be possible, but falls outside the scope of this manuscript.

Acknowledgements All authors gratefully acknowledge support from IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

The authors have declared no conflict of interest.

References

- Creemers, A., Hens, N., Aerts, M., Molenberghs, G., and Verbeke, G., Kenward, M.G. (2009). 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.
- Fitzmaurice, G.M., Molenberghs, G., and Lipsitz, S.R. (1995). Regression models for longitudinal binary responses with informative dropouts. *Journal of the Royal Statistical Society, Series B* **57**, 691–704.
- Follmann, D., and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168.
- Hens, N., Aerts, M., Molenberghs, G., Thijs, H., and Verbeke, G. (2005). Kernel weighted influence measures. *Computational Statistics and Data Analysis* **48**, 467–487.
- 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.
- Kenward, M.G. (1998). Selection models for repeated measurements with nonrandom dropout: an illustration of sensitivity. *Statistics in Medicine* **17**, 2723–2732.
- Kenward, M.G., Goetghebeur, E.J.T., Molenberghs, G. (2001). Sensitivity analysis of incomplete categorical data. *Statistical Modelling* **1**, 31–48.
- Kenward, M.G. and Molenberghs, G. (1999). Parametric models for incomplete continuous and categorical longitudinal studies data. *Statistical Methods in Medical Research* **8**, 51–83.
- Laird N.M. (1994). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics* **43**, 84.
- Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.
- Little, R.J.A. (1994a). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483.
- 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**, 1112–121.
- Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. Wiley, New York.
- Molenberghs, G., Beunckens, C., Sotito, 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. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. Wiley, Chichester.
- Molenberghs, G., Kenward, M.G., Goetghebeur, E. (2001). Simple fitting algorithms for incomplete categorical data. *Journal of the Royal Statistical Society, Series B* **59**, 401–414.

- Molenberghs, G., Michiels, B., Kenward, M.G., and Diggle, P.J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica* **52**, 153–161.
- Molenberghs, G., Goetghebeur, E.J.T., Lipsitz, S.R., Kenward, M.G. (1999). Non-random missingness in categorical data: strengths and limitations. *The American Statistician* **53**, 110–118.
- 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.
- Molenberghs, G., Thijs, H., Michiels, B., Verbeke, G., and Kenward, M.G. (2004). Pattern-mixture models. *Journal de la Société française de Statistique* **145**, 49–77.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.
- Nordheim, E.V. (1984). Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *Journal of the American Statistical Association* **79**, 772–780.
- 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.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Rubin, D.B. (1994). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics* **43**, 80–82.
- 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.
- Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., Curran, D. (2002). Strategies to fit pattern-mixture models.. *Biostatistics* **3**, 245–265.
- 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* **1**, 125–142.
- Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- 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.
- Wu, M.C., and Bailey, K.R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* **45**, 939–955.
- Wu, M.C., and 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.

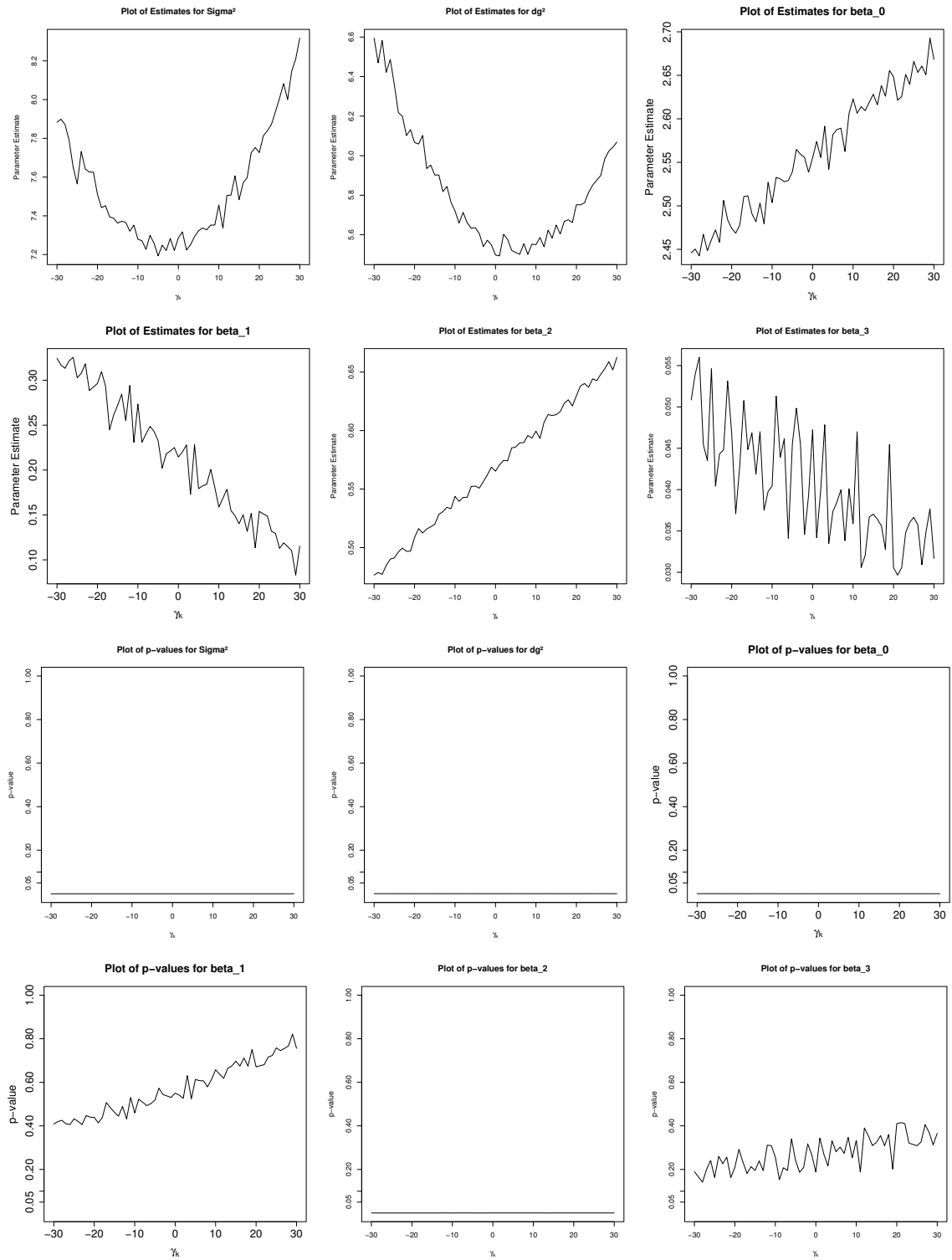


Fig. 3: Toenail Data. Results for sensitivity analysis applied to Scenario 6: parameter estimates (first and second row) and p-values (third and fourth row) for significance for the mean and variance parameters.