

Incomplete Data in Clinical Studies: Analysis, Sensitivity, and Sensitivity Analysis

Peer-reviewed author version

MOLENBERGHS, Geert (2009) Incomplete Data in Clinical Studies: Analysis, Sensitivity, and Sensitivity Analysis. In: DRUG INFORMATION JOURNAL, 43(4). p. 409-429.

Handle: <http://hdl.handle.net/1942/9753>



Incomplete Data in Clinical Studies: Analysis, Sensitivity, and Sensitivity Analysis [Link](#)

Peer-reviewed author version

Made available by Hasselt University Library in [Document Server@UHasselt](#)

Reference (Published version):

Molenberghs, Geert(2009) Incomplete Data in Clinical Studies: Analysis, Sensitivity, and Sensitivity Analysis. In: DRUG INFORMATION JOURNAL, 43(4). p. 409-429

DOI: 10.1002/pst.1665

Handle: <http://hdl.handle.net/1942/9753>

Incomplete Data in Clinical Studies: Analysis, Sensitivity, and Sensitivity Analysis

Geert Molenberghs

*Center for Statistics, Universiteit Hasselt, B-3590 Diepenbeek, Belgium
Biostatistical Centre, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

Abstract

Statistical analysis often extends beyond the data available. This is especially true when data are incompletely recorded because both *ad hoc* as well as model-based approaches are rooted, not only in the observed data and the mechanism governing missingness, but also in the unobserved given the observed data. Other instances of this phenomenon include but are not limited to censored time-to-event data, random-effects models, and latent-class approaches. One needs to be aware of: (1) changes in results and intuition relative to complete-data analysis; (2) the assumptions under which such approaches are valid; (3) the sensitivities implied by departures; and (4) in response to these, what sensitivity analysis avenues are available. This paper provides a bird's eye perspective on these. Some of the developments are illustrated using data from a clinical trial in onychomycosis.

Some Key Words: Linear mixed model; Missing at random; Missing completely at random; Non-future dependence; Pattern-mixture model; Selection model; Shared-parameter model.

1 Introduction

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

During the same era, the *selection model* (SeM), *pattern-mixture model* (PMM), and *shared-parameter model* (SPM) frameworks have been established. In a selection model, the joint distribution of the i th subject's outcomes, denoted Y_i , and vector of missingness indicators, written R_i , is factored as the marginal outcome distribution and the conditional distribution of R_i given 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 steer both the Y_i

and R_i processes. An important version of such a model further asserts that, conditional on the latent variables, Y_i and R_i exhibit no further dependence. Rubin (1976) contributed the concept of *ignorability*, stating that under precise conditions, the missing data mechanism can be ignored when interest lies in inferences about the measurement process. Combined with regularity conditions, ignorability applies to MCAR and MAR combined, when likelihood or Bayesian inference routes are chosen, but the stricter MCAR condition is required for frequentist inferences to be generally valid. These concepts will be formalized in Section 3.

Traditionally, such simple methods as a complete case analysis or simple forms of imputation (e.g., last observation carried forward) have been in use. While they have the advantage of restoring balance and/or a rectangular data matrix, their implied and often severe biases and losses of efficiency have been properly documented (Molenberghs *et al*, 2004; Jansen *et al*, 2006; Molenberghs and Kenward, 2007) and should therefore be avoided. Because of a likelihood-based or Bayesian analysis' validity under MAR, as long as all observed data are included into the analysis, so-called direct likelihood analyses, their Bayesian counterparts, or multiple imputation (Rubin, 1987), are widely regarded as candidate primary analyses of a study. When semi-parametric inferences are desired, the methods proposed by Robins *et al* (1995, 1998) can be applied. Nevertheless, in spite of the flexibility and elegance brought by a direct-likelihood method, there are fundamental issues when selecting a model and assessing its fit to the observed data that do not occur with complete data (Molenberghs, Verbeke, and Beunckens, 2007), already in the MAR case and compounded further under MNAR. A number of such issues are taken up in Section 4.

The concept of MAR has typically been framed within the SeM framework, while Molenberghs *et al* (1998) provided a formulation in the PMM setting as well. Creemers *et al* (2008) studied MAR in the SPM family. The ensuing availability of operational MAR definitions across all frameworks, further studied in Section 5, is a strong asset for data analysis, especially in view of the following result.

Molenberghs *et al* (2007) showed that for every MNAR model, there is an MAR counterpart that produces exactly the same fit to the observed data. Hence, a given MNAR model and its MAR counterpart cannot be distinguished from one another based on observed data. These authors focused on the SeM and PMM frameworks, whereas Creemers *et al* (2008) established the corresponding result in the SPM family. Details can be found in Section 6.

In the context of longitudinal trials, the above MAR-based results ensure that missingness is allowed to depend on covariates and past outcomes, but neither on current nor future observations. A sensible extension towards MNAR would then allow, additionally, missingness to depend on the current, possible unobserved outcome. However, while formulating such a model in the SeM framework is natural (Diggle, and Kenward, 1994), it is less so in the PMM and SPM settings. Results by Kenward, Molenberghs, and Thijs (2003) and Creemers *et al* (2008), respectively, establish this so-called non-future dependence (NFD) property in these frameworks as well. The key aspects are summarized in Section 7.

Whereas the results of Section 6 pertain to incomplete data, this is but one setting where a model extends beyond the data available. Section 8 briefly discusses the more general result holding for any so-called coarse-data and/or data-augmented setting, including censoring, grouping, random-effects models, latent variable structures, and latent classes.

As is clear from what precedes, one cannot distinguish in a formal sense between MAR and MNAR;

Table 1: 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%

it *a fortiori* is simply difficult, if not impossible, to rule out than an MNAR mechanism be operating. It is then even more difficult to justify one the particular choice of MNAR model (Jansen *et al*, 2006). Without additional information, one can only distinguish between such models using their fit to the observed data, and so goodness-of-fit tools alone do not provide a relevant means of choosing between such models, naturally leading to sensitivity analysis, broadly defined as any instrument to assess the impact on statistical inferences from varying the, often untestable, assumptions in an MNAR model (Vach and Blettner, 1995; Copas and Li, 1997; Scharfstein, Rotnitzky, and Robins, 1999; Molenberghs and Kenward, 2007). The sensitivity analysis theme is taken up in Section 9. Let us now first introduce an illustrative case study.

2 A Clinical Trial in Onychomycosis

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

In total, 2×189 patients, distributed over 36 centers, were randomized. Subjects were followed during 12 weeks (3 months) of treatment and followed further, up to a total of 48 weeks (12 months). Measurements were taken at baseline, every month during treatment, and every 3 months afterwards, resulting in a maximum of 7 measurements per subject. At the first occasion, the treating physician indicates one of the affected toenails as the target nail, the nail which will be followed over time. We will restrict our analyses to only those patients for which the target nail was one of the two big toenails. This reduces our sample under consideration to 146 and 148 subjects, in group A and group B , respectively. Figure 1 shows the observed profiles of 30 randomly selected subjects from treatment group A and treatment group B , respectively.

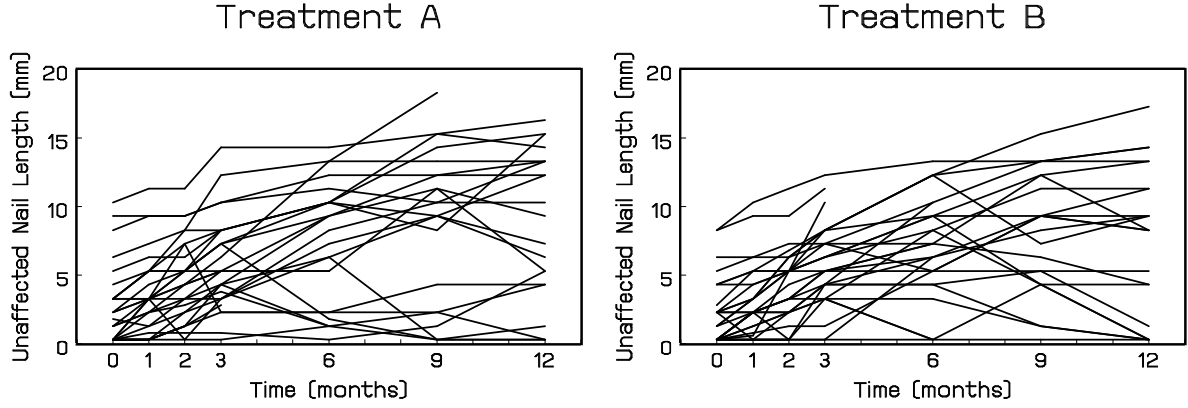


Figure 1: Toenail Data. Individual profiles of 30 randomly selected subjects in each of the treatment groups in the toenail experiment.

Table 2: Toenail Data. Number of available repeated measurements per subject, for each treatment arm separately.

# 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%

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). 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 percentage of severe infections decreased over time, and whether that evolution was different for the two 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 1. A graphical representation is given in Figure 2. Due to a variety of reasons, the outcome has been measured at all 7 scheduled time points, for only 224 (76%) out of the 298 participants. Table 2 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. Figure 1 shows the observed profiles of 30 randomly selected subjects from treatment group *A* and treatment group *B*, respectively.

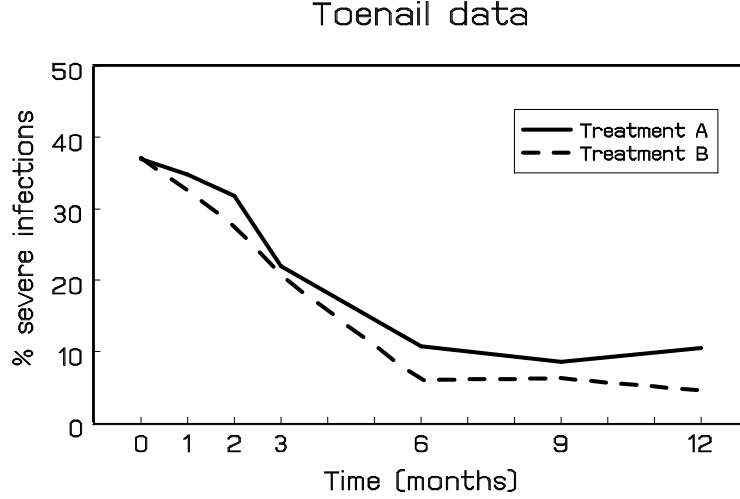


Figure 2: Toenail Data. Evolution of the observed percentage of severe toenail infections in the two treatment groups separately.

3 Modeling Frameworks

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} , $i = 1, \dots, N$, $j = 1, \dots, n_i$. The outcomes can conveniently be grouped into a vector $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})'$. In addition, define a vector of missingness indicators $\mathbf{R}_i = (R_{i1}, \dots, R_{in_i})'$ with $R_{ij} = 1$ if Y_{ij} is observed and 0 otherwise.

In 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 measured and grouped in a vector \mathbf{x}_i , though generally suppressed from notation. Let us formalize the frameworks, touched upon in the introduction (Little, 1995; Molenberghs and Kenward, 2007).

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

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

The first factor is the marginal density of the measurement process and the second one is the density of the missingness process, conditional on the outcomes. The *pattern-mixture models* (PMM) (Little, 1993, 1994) uses the reverse factorization

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

The conventional *shared-parameter model* (SPM) (Wu, and Carroll, 1988; Wu, and Bailey, 1988, 1989) assumes a vector of random effects \mathbf{b}_i , conditional upon which the measurement and dropout processes are independent:

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

and hence

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

Here, \mathbf{b}_i are shared parameters, often taking the form of random effects and following a specific parametric distribution. For our purposes, we will need a slightly more general SPM formulation, as presented by Creemers *et al* (2008). Indeed, while most formulations assume that a single, common set \mathbf{b}_i drives the entire process, one can expand \mathbf{b}_i to a set of latent structures:

$$f(\mathbf{y}_i^o | \mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{l}_i) f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) f(\mathbf{r}_i | \mathbf{g}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{n}_i), \quad (5)$$

where \mathbf{g}_i , \mathbf{h}_i , \mathbf{j}_i , \mathbf{k}_i , \mathbf{l}_i , \mathbf{m}_i , and \mathbf{n}_i are independent random-effects vectors, vectors of latent variables, etc. Obviously, a conventional SPM formulation follows by removing all random effects but \mathbf{g}_i . For convenience, write $\mathbf{b}_i = (\mathbf{g}_i, \mathbf{h}_i, \mathbf{j}_i, \mathbf{k}_i, \mathbf{l}_i, \mathbf{m}_i, \mathbf{n}_i)$.

4 Ignorability and the Likelihood: No Further Issues?

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

$$L_i = \int f(\mathbf{y}_i | \boldsymbol{\theta}) f(\mathbf{r}_i | \mathbf{y}_i^o, \mathbf{y}_i^m, \boldsymbol{\psi}) d\mathbf{y}_i^m. \quad (6)$$

In general, (6) does not simplify, but under MAR, we obtain:

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

Hence, likelihood and Bayesian inferences for the measurement model parameters $\boldsymbol{\theta}$ can be made without explicitly formulating the missing data mechanism, provided the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are distinct, meaning that their joint parameter space is the Cartesian product of the two component parameter spaces (Rubin, 1976). For Bayesian inferences, additionally the priors need to be independent (Little and Rubin, 2002). It is precisely this result which makes so-called direct-likelihood analyses, valid under MAR, viable candidates for the status of primary analysis in clinical trials and a variety of other settings (Molenberghs *et al*, 2004; Molenberghs and Kenward, 2007).

In spite of the appeal of ignorability for likelihood-based analysis of incomplete data under MAR, Molenberghs, Verbeke, and Beunckens (2008) have brought forward generic issues arising when fitting models to incomplete data: (i) the classical relationship between observed and expected features is convoluted since one observes the data only partially while the model describes all data; (ii) the independence of mean and variance parameters in a (multivariate) normal is lost, implying increased sensitivity, even under MAR; (iii) also the well-known agreement between the frequentist ordinary least squares (OLS) approach and maximum likelihood estimation methods for normal models is lost, as soon as the missing data mechanism is not of the MCAR type, with related results holding in the non-normal case; (iv) in a likelihood-based context, deviances and related information criteria cannot be used in the same vein as with complete data since they provide no information about a model's prediction of the unobserved data; and, in particular, (v) several models may saturate the observed-data degrees of freedom, while providing a different fit to the complete data, i.e., they only coincide in as far as they describe the observed data; as a consequence, different inferences may result from different saturated models, where 'saturation' is to be understood in terms of the observed but not the full data.

Based on these considerations, it follows that model assessment should always proceed in two steps. In the first step, the fit of a model to the *observed* data should be carefully assessed, while in the

second step the sensitivity of the conclusions to the *unobserved data given the observed data* should be addressed. Gelman *et al* (2005) proposed an approach to this effect, the essence of which is as follows. First, a model is fitted to the observed data. Under the fitted model, and assuming ignorable missingness, datasets simulated from the fitted model should ‘look similar’ to the actual data. Therefore, multiple sets of data are sampled from the fitted model, and compared to the dataset at hand. Because what one actually observes consists of, not only the actually observed outcome data, but also realizations of the missingness process, comparison with the simulated data would also require simulation from, hence full specification of, the missingness process. This added complexity is avoided by augmenting the observed outcomes with imputations drawn from the fitted model, conditional on the observed responses, and by comparing the so-obtained completed dataset with the multiple versions of simulated complete datasets. Such a comparison will usually be based on relevant summary characteristics such as time-specific averages or standard deviations. As suggested by Gelman *et al* (2005), this so-called data-augmentation step could be done multiple times, along multiple-imputation ideas from Rubin (1987).

5 Defining Missing at Random

The taxonomy of missing data mechanisms, introduced by Rubin (1976) and informally described in the introduction, is customarily formalized using the second factor on the right hand side of (1): A mechanism is MAR if $f(\mathbf{r}_i|\mathbf{y}_i, \psi) = f(\mathbf{r}_i|\mathbf{y}_i^o, \psi)$. In the MNAR case, missingness depends on the unobserved outcomes \mathbf{y}_i^m , regardless of the observed outcomes and the covariates.

Molenberghs *et al* (1998, 2007), among others, formulated MAR in the PMM setting:

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

This means that, in a given pattern, the conditional distribution of the unobserved components given the observed ones equals the corresponding distribution marginalized over the patterns. Note that, owing to this result, MAR can be formulated in terms of R given Y , but also in terms of Y given R . These authors also operationalized the definition by so-called identifying restrictions. For example, in a pattern where, say, three out of five measurements are obtained, the distribution of the fourth outcome given the first three is identified from patterns with either four or all measurements obtained. At the same time, the distribution of the fifth measurement given the earlier four is then identified from the completers, the only pattern with the fifth, and last, measurement observed. These restrictions are termed *available case missing value restrictions* (ACMV) by Molenberghs *et al* (1998).

Creemers *et al* (2008) characterized MAR in the SPM framework. Their general result, while providing necessary and sufficient conditions, is in terms of an integral equation and therefore is less intuitive and quite impractical. That is why we here restrict attention to a useful sub-family of SPM-based MAR models, defined by the following sub-class of model (5):

$$f(\mathbf{y}_i^o|\mathbf{j}_i, \ell_i)f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{m}_i)f(\mathbf{r}_i|\mathbf{j}_i, \mathbf{n}_i), \quad (9)$$

where \mathbf{j}_i , ℓ_i , \mathbf{m}_i , and \mathbf{n}_i are independent random-effects vectors. They are the general SPM models for which information about the missing outcomes stems is allowed to come from the observed outcomes but, given these, not further from shared random effects. Just like in PMM characterization (8), the appropriate conditional density of the missing components given the observed components is fully

unidentified. This is in contrast to the SeM framework, where the factorization is such that the data carry some information about all factors, which are then fully identified by making further restrictive modeling assumptions.

6 Every MNAR Model Has an MAR Counterpart

The correctness of an (MNAR) model fitted to incomplete data can be verified only in as far as it fits the *observed* data. Thus, evidence for or against MNAR can be provided solely within a particular, predefined parametric family, the plausibility of which cannot be verified in empirical terms alone. Hence, an omnibus assessment of MAR *versus* MNAR is not possible, since every MNAR model can be doubled up with a uniquely defined MAR counterpart, producing exactly the same fit as the original MNAR model, in the sense that it produces exactly the same predictions to the observed data as the original MNAR model, and depending on exactly the same parameter vector. Molenberghs *et al* (2007) showed that, while this so-called MAR counterpart generally does not belong to a conventional parametric family, its existence has important ramifications.

The construction of such a counterpart proceeds in four steps: (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 set of data, using the observed data likelihood:

$$L = \prod_i \int f(\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) d\mathbf{y}_i^m. \quad (10)$$

Upon denoting the obtained parameter estimates by $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\psi}}$ respectively, the fit to the hypothetical full data is

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

To undertake the second step, full density (11) can be re-expressed in PMM form as:

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

In line with what has been said before, the final term on the right hand side of (12), $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$, is not identified from the observed data but rests solely on model assumptions.

The third step requires replacing this factor by the appropriate MAR counterpart: $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{r}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$ needs to be replaced with

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

where the $h(\cdot)$ notation is used for shorthand purposes. The above construction does not lead to a member of a conventional parametric family. While this obviously implies limitations on its use, such is not dissimilar to the construction of some semi- and non-parametric estimators. Also, it helps understand that an overall, definitive conclusion about the nature of the missing data mechanism, solely based on the observed outcomes, is not possible, even though one can make progress if attention is confined to a given parametric family, in which one puts sufficiently strong prior belief (Jansen *et al*, 2006).

The key computational consequence is the need to determine $h(\mathbf{y}_i^m | \mathbf{y}_i^o)$ in (13). To this end, the aforementioned ACMV identifying restrictions can be used.

The characterization MAR in the SPM setting, owing to Creemers *et al* (2008) and discussed in Section 5, enables the construction of an MAR counterpart to an arbitrary SPM of the form (5). Practically, this is done by *a-posteriori integrating* over the shared random effects in the densities describing the unobserved measurements, given the observed ones, where integration takes place over the densities of \mathbf{g}_i , \mathbf{h}_i , and \mathbf{k}_i , with fitted parameters plugged in. Precisely, one replaces $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i)$ with

$$h(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{m}_i) = \int_{\mathbf{g}_i} \int_{\mathbf{h}_i} \int_{\mathbf{k}_i} f(\mathbf{y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i, \mathbf{h}_i, \mathbf{k}_i, \mathbf{m}_i) d\mathbf{g}_i d\mathbf{h}_i d\mathbf{k}_i \quad (14)$$

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.

7 Longitudinal Data With Dropout: Non-future Dependence

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

Let us turn to the nature of the missingness mechanism. Throughout the section, assume that missingness is confined to dropout. From a SeM perspective, one often classifies missing data mechanisms as (Diggle, and Kenward, 1994): (1) independent of outcomes; (2) dependent on previous measurements only; (3) dependent on the current and perhaps previous measurements only; (4) fully arbitrary, *i.e.*, where missingness can depend on previous, current, and future measurements. Evidently, (1) is MCAR, (2) is MAR, and (4) is fully unrestricted MNAR. The last category is not always desired. For example, (Diggle, and Kenward, 1994) did not consider (4) but restricted MNAR to mechanism (3). While restrictive, this is appealing since preventing dropout at a given point in time to depend on future measurements, *i.e.*, non-future dependent.

Clearly, the above ideas are easy to frame within the SeM family. Kenward, Molenberghs, and Thijs (2003) and Creemers *et al* (2008) underscored that the situation is less clear in the PMM and SPM families and then provided translations. These will be considered in Section 7.1, with illustrations offered in Section 7.2.

7.1 Non-future Dependence in the PMM and SPM Frameworks

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

neighboring-case missing value restrictions, where information is borrowed from the closest available pattern.

The equivalence of ACMV 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, for the SeM, PMM, and SPM frameworks in turn.

Starting with the SeM family, let $r = t \leq n$ be the number of measurements actually observed. The selection model factorization for this context is given by

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

One can now formulate missing non-future dependent as

$$f(r = t | y_1, \dots, y_T) = f(r = t | y_1, \dots, y_{t+1}). \quad (15)$$

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

Using the notational system of this section, pattern-mixture models take the form:

$$\begin{aligned} f(y_1, \dots, y_T, r = t) \\ = f_t(y_1, \dots, y_t) f_t(y_{t+1} | y_1, \dots, y_t) f_t(y_{t+2}, \dots, y_T | y_1, \dots, y_{t+1}) f(r = t), \end{aligned} \quad (16)$$

where $f_t(y_1, \dots, y_T) = f(y_1, \dots, y_T | r = t)$. The first three factors in (16) 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.

Within the PMM framework, define non-future dependent missing value restrictions as:

$$f(y_t | y_1, \dots, y_{t-1}, r = j) = f(y_t | y_1, \dots, y_{t-1}, r \geq t - 1), \quad (17)$$

for all $t \geq 2$ and all $j < t - 1$. Non-future missing values is not a comprehensive set of restrictions, but rather leaves one conditional distribution per incomplete pattern unidentified:

$$f(y_{t+1} | y_1, \dots, y_t, r = t). \quad (18)$$

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, (17) excludes such mechanisms like 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 (18) of the same functional form as (17) 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. Kenward, Molenberghs, and Thijs (2003) formally showed that NFD and NFMV are equivalent. Kenward, Molenberghs, and Thijs (2003) offered practical implementation strategies.

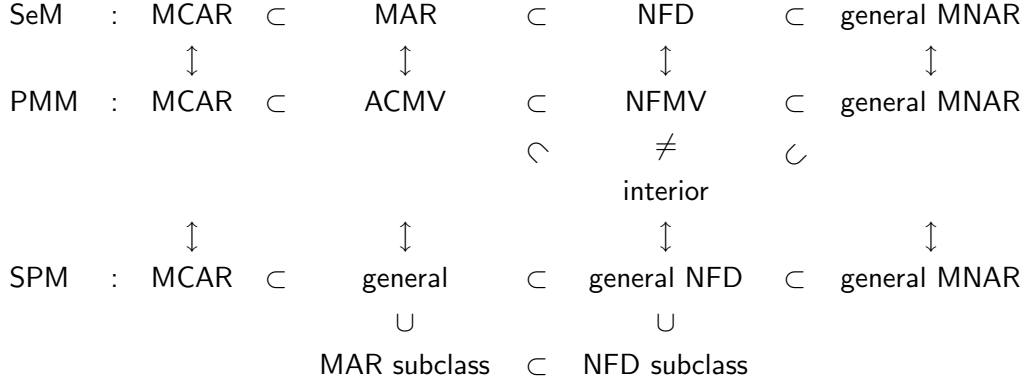


Figure 3: Relationship between nested families within the selection model (SeM), pattern-mixture model (PMM), and shared-parameter model (SPM) families. MCAR: missing completely at random; MAR: missing at random; MNAR: missing not at random; NFD: non-future dependence; ACMV: available-case missing values;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

Let us turn to SPM. First, note that (15) 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 the characterization of MAR by Creemers *et al* (2008) was derived from the standard MAR definition, it immediately follows that a characterization of NFD-SPM is, again, in terms of an integral equation. The interested reader will find details in Creemers *et al* (2008). Here, we restrict attention to the sub-class (9). Define a sub-class of shared-parameter model (5):

$$f(\mathbf{y}_i^{pc} | \mathbf{j}_i, \mathbf{l}_i) f(\mathbf{y}_i^f | \mathbf{y}_i^{pc}, \mathbf{m}_i) f(r_i | \mathbf{j}_i, \mathbf{n}_i), \quad (19)$$

where \mathbf{j}_i , \mathbf{l}_i , \mathbf{m}_i , and \mathbf{n}_i are independent random-effects vectors. The subscript ‘pc’ refers to ‘previous and current,’ while ‘f’ refers to ‘future.’ Hence, (19) offers a class of missing-data mechanisms that belongs to the NFD family. The relationship between the various mechanisms in the three families is depicted in Figure 3.

7.2 Analysis of the Toenail Data

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

7.2.1 Continuous Unaffected Nail Length

Consider a general model of the form (5), with random effects confined to \mathbf{g}_i , i.e., common to all three components. For the measurement model, assume a linear mixed model (Verbeke and Molenberghs, 2000), with general form:

$$\mathbf{Y}_i | \mathbf{g}_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{g}_i, \Sigma_i), \quad (20)$$

$$\mathbf{g}_i \sim N(0, D). \quad (21)$$

Based on (20) and (21), the so-called marginal model can be derived

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i D \mathbf{Z}_i' + \Sigma_i). \quad (22)$$

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 (20) as

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

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

$$\begin{aligned} \mathbf{Y}_i^m | \mathbf{y}_i^o, \mathbf{g}_i &\sim N \left[(\mathbf{X}_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{X}_i^o) \boldsymbol{\beta} + \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{y}_i^o + (\mathbf{Z}_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{Z}_i^o) \mathbf{g}_i, \right. \\ &\quad \left. \Sigma_i^{mm} - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \Sigma_i^{om} \right]. \end{aligned} \quad (23)$$

Now, (23) 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 (22), now applied to (23), produces:

$$\begin{aligned} \mathbf{Y}_i^m | \mathbf{y}_i^o &\sim N \left[(\mathbf{X}_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{X}_i^o) \boldsymbol{\beta} + \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{y}_i^o, \right. \\ &\quad (\mathbf{Z}_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{Z}_i^o) D (\mathbf{Z}_i^m - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \mathbf{Z}_i^o)' \\ &\quad \left. + \Sigma_i^{mm} - \Sigma_i^{mo} \{\Sigma_i^{oo}\}^{-1} \Sigma_i^{om} \right]. \end{aligned} \quad (24)$$

For the unaffected nail length, we choose for (20)–(21):

$$E(Y_{ij} | g_i, T_i, t_j, \boldsymbol{\beta}) = \beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j, \quad (25)$$

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

Given these choices, (23) and (24) simplify to

$$\mathbf{Y}_i^m | \mathbf{y}_i^o, g_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i^m g_i, \sigma^2 I_i), \quad (26)$$

$$\mathbf{Y}_i^m | \mathbf{y}_i^o \sim N(\mathbf{X}_i \boldsymbol{\beta}, d \mathbf{J}_i + \sigma^2 I_i), \quad (27)$$

with I_i an identity matrix and \mathbf{J}_i a matrix of ones, the dimensions of which are 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 (5). The sequence \mathbf{r}_i can take one of two forms in our case. Either, it is a length-7 vector of zeros, for a completely observed subject, or it is a sequence of k zeros followed by a sole one $1 \leq k \leq 6$, for someone dropping out. Note that k is 1 at least, since for everyone the initial measurement has been observed. It is convenient to assume a logistic regression of the form:

$$\text{logit}[P(R_{ij} = 1 | R_{i,j-1} = 0, g_i, T_i, t_j, \boldsymbol{\gamma})] = \gamma_0 + \gamma_{01} g_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j, \quad (28)$$

Table 3: Toenail Data. Continuous, longitudinal unaffected-nail-length outcome. Parameter estimates (standard errors) for the model specified by (25) and (28).

Effect	Unaffected nail length		Dropout	
	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)

($j > 1$), 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 (25) and (28) can easily be fitted using, for example, the SAS procedure NLMIXED. Details are provided in Creemers *et al* (2008).

Parameter estimates and standard errors are displayed in Table 3. It is noteworthy that the scale factor γ_{01} is estimated to be negative, even though it is not significant. While we should not overly stress its importance, there is some indication that a higher subject-specific profile of unaffected nail length corresponds with a lower dropout probability, which is not surprising. The magnitude of the scale factor allows us to ‘translate’ the subject-specific effect from the continuous outcome scale, expressed in mm, to the unitless logit scale on which the probability of missingness is described. Note that the random-intercept variance is highly significant among unaffected nail length outcomes; the same is not true for the dropout model, with $p = 0.2487$, using a 50 : 50 mixture of a χ_0^2 and χ_1^2 distribution (Verbeke and Molenberghs, 2000).

Figure 4 displays the incomplete profiles, extended beyond the time of dropout, using prediction based on: (1) the original model (dashed lines); (2) the MAR counterpart (solid lines). Within each of the treatment arms, three profiles are highlighted. The MAR counterpart reduces all predictions to the same profile, whereas the MNAR models 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 (26). When deemed less plausible, the fully general structure (23) can be implemented.

7.2.2 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 4. By way of illustration, we will assume a single dichotomous random effect,

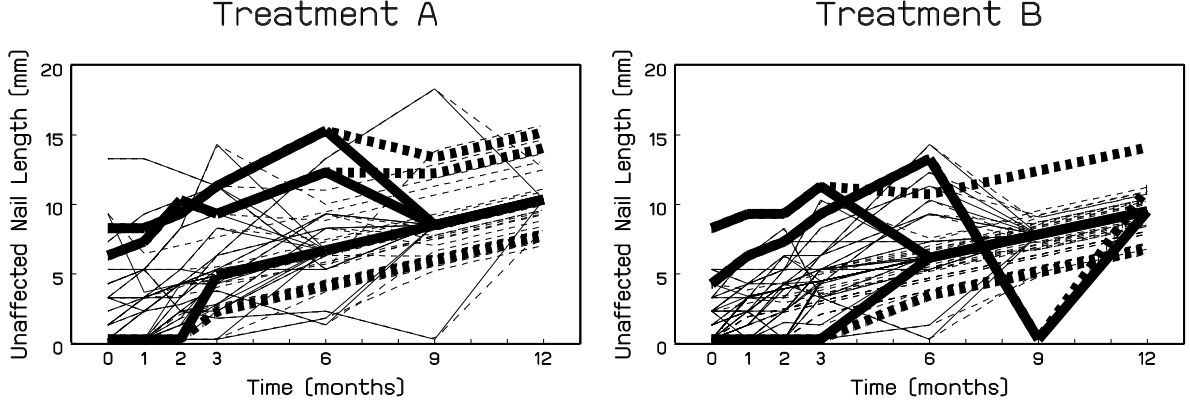


Figure 4: Toenail Data. Individual profiles of subjects with incomplete data, for each treatment arm, extended using MNAR Model (25) (dashed line) and using the model's MAR counterpart (solid line). In each group, three subjects are highlighted.

of the g_i type. This imposes a latent-class structure. Decompose the cell probabilities as:

$$\pi_{gi_1i_2rt} = \pi_g \pi_{i_1|g} \pi_{i_2|i_1gt} \pi_{r|g}, \quad (29)$$

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 (29) are modeled as:

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

$$\pi_{i_1|g} = \frac{e^{(\beta_0 + \beta_1 g)i_1}}{1 + e^{\beta_0 + \beta_1 g}}, \quad (30)$$

$$\pi_{i_2|i_1gt} = \frac{e^{(\gamma_0 + \gamma_1 i_1 + \gamma_2 g + \gamma_3 i_1 g + \gamma_4 t)i_2}}{1 + e^{\gamma_0 + \gamma_1 i_1 + \gamma_2 g + \gamma_3 i_1 g + \gamma_4 t}}, \quad (31)$$

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

In Model 'Bin1', we will set $\beta_1 = 0$ in (30) for reasons of identifiability. In Model 'Bin2', $\gamma_2 = \gamma_3 = 0$ in (31). 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 4. 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 (29) by their estimates, followed by summing over g . The MAR counterpart is obtained as $\pi_{gi_1i_2rt} = \pi_g \pi_{i_1|g} \tilde{\pi}_{i_2|i_1t} \pi_{r|g}$, where

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

Parameter estimation by both maximum likelihood, as well as the EM algorithm Dempster, Laird,

Table 4: *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 bodyguards. 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

and Rubin (1977) is particularly easy. For direct likelihood, the log-likelihood function takes the form

$$\ell = \sum_{i_1, i_2, t} Z_{i_1 i_2, r=1, t} \ln \left(\sum_g \pi_g \pi_{i_1|g} \pi_{i_2|i_1 g t} \pi_{r=1|g} \right) + \sum_{i_1, t} Z_{i_1, r=0, t} \ln \left(\sum_g \pi_g \pi_{i_1|g} \pi_{r=0|g} \right), \quad (32)$$

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 (32) to a standard numerical optimizer.

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

$$\begin{aligned} \ell^* &= \sum_{g, i_1, i_2, r, t} Z_{g i_1 i_2 r t}^* \ln \left(\pi_g \pi_{i_1|g} \pi_{i_2|i_1 g t} \pi_{r|g} \right) \\ &= \sum_g Z_{g++++}^* \ln(\pi_g) + \sum_{g, i_1} Z_{g i_1 +++}^* \ln(\pi_{i_1|g}) \\ &\quad + \sum_{g, i_1, i_2, t} Z_{g i_1 i_2 + t}^* \ln(\pi_{i_2|i_1 g t}) + \sum_{g, r} Z_{g++r+}^* \ln(\pi_{r|g}). \end{aligned} \quad (33)$$

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 . 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:

$$\begin{aligned}
E\left(Z_{g++++}^*\right) &= \pi_g Z_{++++}, \\
E\left(Z_{gi_1+++}^*\right) &= \pi_g \pi_{i_1|g} Z_{i_1+++} \\
E\left(Z_{gi_1i_2+t}^*\right) &= \pi_g Z_{i_1i_2,r=1,t} + \pi_g \pi_{i_2|i_1gt} Z_{i_1+,r=0,t}, \\
E\left(Z_{g++r+}^*\right) &= \pi_g \pi_{r|g} Z_{++r+}.
\end{aligned}$$

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 (33).

8 Data-enriched Structures

The results of Section 6 are not confined to incomplete data. Verbeke and Molenberghs (2008) show that the results hold for all coarsened-data and data-augmented settings. By coarsening, a term coined by Heitjan (Heitjan and Rubin, 1991; Zhang and Heitjan, 2007), one refers to the fact that the observed data are coarser than the hypothetically conceived data structures and for which models are built, and encompass incomplete, censored, grouped, and truncated data. Data augmentation refers to the introduction of unobservables such as random effects, latent variables, and latent classes. We capture both of these families under the common denominator *data enrichment*.

Assume data \mathbf{Z}_i for an independent unit $i = 1, \dots, N$ are augmented with \mathbf{c}_i . The \mathbf{c}_i can take any conventional enriched-data form. For example, the vector can refer to missing measurements, random effects, or perhaps a combination of both. An example of a setting where the latter situation arises naturally is the SPM framework. Assume further a joint model of the generic form $f(\mathbf{z}_i, \mathbf{c}_i | \boldsymbol{\theta}, \boldsymbol{\psi})$, where covariates have been suppressed for notational simplicity. Consider the factorizations:

$$f(\mathbf{z}_i, \mathbf{c}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{z}_i | \mathbf{c}_i, \boldsymbol{\theta}) f(\mathbf{c}_i | \boldsymbol{\psi}), \quad (34)$$

$$= f(\mathbf{z}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) f(\mathbf{c}_i | \mathbf{z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}). \quad (35)$$

Borrowing terminology from the hierarchical-models context, such as mixed models, every factor in both (34) and (35) can usefully be given a name. The left hand side is the *joint model*. Let us turn to the right hand side. The first factor in (34) is the *hierarchical model* and the second one is the *prior density* for the enriched data. The first factor in (35) may be termed the *marginal model*, whereas the second one is the *posterior density* of the enriched data. The above terminology makes clear the obvious link between (34)–(35) and the mixed-model setting. The link with incomplete data follows by setting $\mathbf{c}_i \equiv \mathbf{y}_i^m$ and $\mathbf{z}_i = (\mathbf{y}_i^o, \mathbf{r}_i)$. Hence, again, we are naturally led to the PMM framework. In PMM factorization (12), the marginal model is factored further, but this is immaterial. The key is the third factor on the right hand side of (12), i.e., the second factor in (35).

The practical implication is as follows. Assume that data \mathbf{z}_i are enriched with \mathbf{c}_i . Then, any model (34) formulated for and fitted to such data, can be replaced by an infinite family of models, all of which retain the fit to the observed data. This is done by preserving the marginal model $f(\mathbf{z}_i | \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$ and replacing the posterior density $f(\mathbf{c}_i | \mathbf{z}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\psi}})$ by an arbitrary density

$$f(\mathbf{d}_i | \mathbf{z}_i, \boldsymbol{\gamma}). \quad (36)$$

Here, d_i rather than c_i is used to indicate that there need not be any connection between the original and substituted enriched data. Also, the new density (36) can be parameterized by a completely new parameter γ .

8.1 Illustration: Linear Mixed-effects Models

Let us illustrate these ideas for the linear mixed model (Verbeke and Molenberghs, 2000), with notation as in Section 3, the fully hierarchically specified linear mixed-effects model takes the form (Verbeke and Molenberghs, 2000):

$$\mathbf{Y}_i | \mathbf{b}_i \sim N(X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i, \Sigma_i), \quad (37)$$

$$\mathbf{b}_i \sim N(0, D), \quad (38)$$

where $\boldsymbol{\beta}$ is a vector of fixed effects, and X_i and Z_i are design matrices. The marginal model and posterior distribution of the random effects are:

$$\mathbf{Y}_i \sim N(X_i \boldsymbol{\beta}, V_i = Z_i D Z_i' + \Sigma_i), \quad (39)$$

$$\mathbf{b}_i | \mathbf{Y}_i \sim N[D Z_i' V_i^{-1} (\mathbf{Y}_i - X_i \boldsymbol{\beta}), (Z_i' \Sigma_i^{-1} Z_i + D^{-1})^{-1}]. \quad (40)$$

Finally, the empirical Bayes predictions (Carlin and Louis, 1996; Verbeke and Molenberghs, 2000) and predicted values of the outcomes are:

$$\hat{\mathbf{b}}_i = E(\mathbf{b}_i | \mathbf{Y}_i) = D Z_i' V_i^{-1} (\mathbf{Y}_i - X_i \boldsymbol{\beta}), \quad (41)$$

$$\widehat{\mathbf{Y}}_i = (Z_i D Z_i') \cdot V_i^{-1} \mathbf{y}_i + (\Sigma_i) \cdot V_i^{-1} X_i \boldsymbol{\beta}, \quad (42)$$

the latter representing the familiar “weighted average” of the transformed observed outcomes $V_i^{-1} \mathbf{y}_i$ and the transformed marginal mean $V_i^{-1} X_i \boldsymbol{\beta}$.

To illustrate the arbitrariness brought forward by our results, and in this case referring to the posterior density of the random effects, let us replace the normally distributed random effects by a vector of n_i independent exponential random effects, where each outcome component Y_{ij} is paired with an exponential random effect g_{ij} . The conventional density for an exponential variable ϕ is

$$f(\phi) = \delta e^{-\phi \delta}. \quad (43)$$

Further, choose $\delta = e^{\gamma_j y_{ij}}$. Straightforward algebra leads to the following model equations:

$$f(\mathbf{q}_i | \mathbf{y}_i) = \prod_{j=1}^{n_i} e^{\gamma_j y_{ij}} e^{-q_{ij} e^{\gamma_j y_{ij}}}, \quad (44)$$

$$f(\mathbf{q}_i) = \sum_{\mathbf{m}} \left(\prod_{j=1}^{n_i} \frac{(-q_{ij})^{m_j}}{m_j!} \right) e^{\boldsymbol{\mu}_i' \boldsymbol{\lambda}_m + \frac{1}{2} \boldsymbol{\lambda}_m' V_i \boldsymbol{\lambda}_m}, \quad (45)$$

$$f(\mathbf{y}_i | \mathbf{q}_i) = \frac{\prod_{j=1}^{n_i} e^{\gamma_j y_{ij}} e^{-q_{ij} e^{\gamma_j y_{ij}}} e^{-\boldsymbol{\mu}_i' \boldsymbol{\lambda}_m - \frac{1}{2} [(\mathbf{y}_i - \boldsymbol{\mu}_i)' V_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\lambda}_m' V_i \boldsymbol{\lambda}_m]}}{(2\pi)^{n_i/2} |V_i|^{1/2} \sum_{\mathbf{m}} \left(\prod_{j=1}^{n_i} \frac{(-q_{ij})^{m_j}}{m_j!} \right)}, \quad (46)$$

$$\widehat{q_{ij}} = e^{-\gamma_j y_{ij}} \quad (47)$$

$$\widehat{y}_i = \frac{\sum_{\mathbf{m}} \left[\prod_{j=1}^{n_i} \frac{(-e^{-\gamma_j y_{ij}})^{m_j}}{m_j!} \right] e^{\boldsymbol{\mu}'_i \boldsymbol{\lambda}_m + \frac{1}{2} \boldsymbol{\lambda}'_m V_i \boldsymbol{\lambda}_m} (\boldsymbol{\mu}_i + V_i \boldsymbol{\lambda}_m)}{\sum_{\mathbf{m}} \left[\prod_{j=1}^{n_i} \frac{(-e^{-\gamma_j y_{ij}})^{m_j}}{m_j!} \right] e^{\boldsymbol{\mu}'_i \boldsymbol{\lambda}_m + \frac{1}{2} \boldsymbol{\lambda}'_m V_i \boldsymbol{\lambda}_m}}, \quad (48)$$

where \mathbf{m} ranges over all non-negative integer vectors $\mathbf{m} = (m_1, \dots, m_{n_i})$, and $\boldsymbol{\lambda}_m$ has components $\lambda_{mj} = (m_j + 1)\gamma_j$.

There is an obvious consequence to these developments regarding the meaning of model parameters. Indeed, in specifying the original hierarchical model (37)–(38), the parameters $\boldsymbol{\beta}$, Σ_i , and D in general, but D in particular, are part of a hierarchical specification. Since (39)–(40) taken together are equivalent to the original pair of equations, one might argue there still is the hierarchical interpretation. The difference now is that all three sets of parameters occur in each of the two models, whereas in the original specification (37)–(38) there is a separation between $\boldsymbol{\beta}$ and Σ_i on the one hand and D on the other hand. However, it has been argued by many (Verbeke and Molenberghs, 2000, 2003; Molenberghs and Verbeke, 2007) that there is a fundamental difference in parameter interpretation, even to the point of bearing on the inferences made, when one solely considers the marginal model (39). This is clear when considering the model composed of (39) and, for example, (44). Indeed, now all three parameters $\boldsymbol{\beta}$, Σ_i , and D feature in the marginal model only. The hierarchical parameters, γ_j in our particular instance, are completely separated from the marginal ones. This further implies that the so-called hierarchical parameter is estimable only because it also occurs in marginal model (39) for which, by definition, there is information in the data. Put differently, in the conventional hierarchical marginal model, all parameters are identifiable from marginal model (39), which is the only channel by which the data convey information. The model merely *appears* interpretable at a hierarchical, or enriched, level since (40) contains these, and only these parameters.

8.2 Analysis of the Toenail Data

For the unaffected nail length, and with notation as in Section 7.2, let us specify a linear mixed-effects model (37)–(38):

$$Y_{ij}|(b_{i0}, b_{i1}) \sim N(\beta_0 + b_{i0} + (\beta_1 + b_{i1})t_j + \beta_2 T_i + \beta_3 T_i t_j, \sigma^2), \quad (49)$$

$$\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_{00} & d_{01} \\ d_{10} & d_{11} \end{pmatrix} \right]. \quad (50)$$

Parameter estimates and standard errors are presented in Table 5.

We are now able to supplement the model specified by (49)–(50) with the exponentially defined models. Let us choose, for illustration, the exponential model. This implies that the marginal model resulting from (49)–(50) is retained:

$$\mathbf{Y}_i|(b_{i0}, b_{i1}) \sim N[X_i(\beta_0, \beta_1, \beta_2, \beta_3)' + Z_i(b_{i0}, b_{i1})', \sigma^2 I_{n_i} + Z_i' D Z_i], \quad (51)$$

and coupled with (44). Here, X_i and Z_i are the obvious $n_i \times 4$ and $n_i \times 2$ design matrices, respectively. Then, we can calculate empirical Bayes predictions under both the normal and the exponential model. These produce two different subject-specific profiles, in addition to the observed-data and marginal mean profiles. Note that, for the posterior density (44), we have the freedom to specify the parameters γ_j , since there is no information contained in the data. We set them equal to

Table 5: Toenail Data. (Unaffected nail length outcome). Parameter estimates (standard errors) for the model specified by (49) and (50).

Effect	Parameter	Estimate	(Standard error)
<i>Fixed effects:</i>			
Intercept	β_0	2.4562	(0.2400)
Dose effect	β_1	0.5932	(0.0450)
Time effect	β_2	0.2792	(0.3375)
Dose by time interaction	β_3	0.0350	(0.0627)
<i>Variance components:</i>			
Random intercept variance	d_{00}	7.3174	(0.6952)
Random slope variance	d_{11}	0.2239	(0.0231)
Random effects covariance	d_{01}	-0.4985	(0.0982)
Residual variance	σ^2	3.1508	(0.1245)

$\gamma_j = 0.05$. Figure 5 presents these four profiles for four selected subjects, two from each treatment arm, respectively. It is clear that the exponential choice produces predictions that lie much closer to the marginal mean profile and further away from the observed profile, than is the case with the normal random effects.

9 Sensitivity Analysis

In the previous sections, we have reviewed a number of issues the analyst ought to be aware of when dealing with incomplete data, arising from clinical or other studies. Many can be captured under the general denominator that models not only describe the observed data, but also “make statements” about the unobserved data given the observed ones. To further address this, a variety of sensitivity analysis routes have been proposed. For our purposes, one could informally define a sensitivity analysis as a way of exploring the impact of a model and/or selected observations on the inferences made when data are incomplete. However, the concept of sensitivity analysis is both older and broader.

Sensitivity analysis, generically defined as assessment of how scientific conclusions depend on model assumptions, influential observations and subjects, and the like, has a long history in statistics. Early instances include Cornfield’s work in the context of causal inference (Holland, 1986) and the study of the independent censoring assumption’s impact in time-to-event analyses, to which a large part of a joint US Air Force, National Cancer Institute, and Florida State University sponsored conference was devoted (Proschan and Serfling (1974), and several contributions therein, in particular by Fisher and Kanarek). A different strand is formed by input/output sensitivity in industrial applications (Haug, Choi, and Komkov, 1986).

Even when confining attention to the field of incomplete data, research is vast and disparate. This is

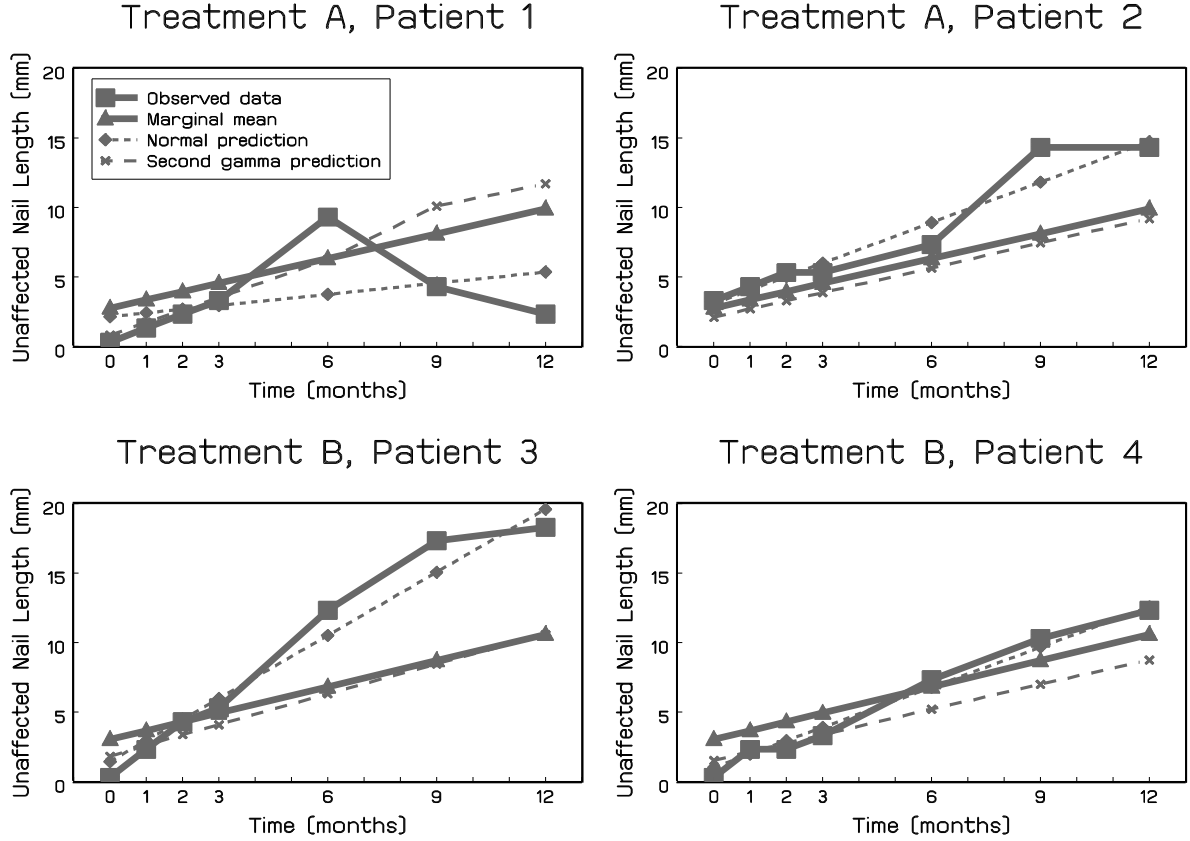


Figure 5: Toenail data. For 4 selected subjects, two per treatment arm: (1): observed profile; (2) marginal mean profile (which solely depends on treatment); (3) prediction from the normal model (42); (4) prediction from the exponential model.

not a negative point, rather it reflects broad awareness of the need for such sensitivity analysis. Earlier work on incomplete data was virtually exclusively focused on the formulation of ever more complex models. Both the pattern-mixture model framework (Little, 1993, 1994) and the shared-parameter framework (Wu, and Carroll, 1988; Wu, and Bailey, 1988, 1989) have provided useful vehicles for model formulation. In a pattern-mixture model, the outcome distribution is modeled conditional on the observed response pattern, as opposed to the selection model framework, used throughout this manuscript, where the unconditional outcome distribution is the centerpiece, sometimes supplemented with a model describing the non-response process, given the outcomes. In a shared-parameter model, the outcome and non-response processes are considered independent, *given* a set of common latent variables or random effects, which are assumed to drive both processes simultaneously. A particularly versatile research line is geared towards the formulation of semi-parametric approaches (Robins *et al*, 1995; Scharfstein, Rotnitzky, and Robins, 1999). Whereas in the parametric context one is often interested in quantifying the impact of model assumptions, the semi-parametric and non-parametric modelers aim at formulating models that have a high level of robustness against the impact of the missing data mechanism. A number of authors have aimed at quantifying the impact of one or a few observations on the substantive and missing data mechanism related conclusions (Copas and Li, 1997; Verbeke *et al*, 2001; Troxel *et al*, 1998).

A number of early references pointing to the aforementioned sensitivities and responses there to in-

clude Rosenbaum and Rubin (1983), Nordheim (1984), Little (1994b), Rubin (1994), Laird (1994), Vach and Blettner (1995), Fitzmaurice, Molenberghs, and Lipsitz (1995), Fitzmaurice, Molenberghs, and Lipsitz (1995), Molenberghs *et al* (1999), Kenward (1998), and Kenward and Molenberghs (1999). Rosenbaum and Rubin (1983) is a pivotal reference for its propensity-scores basis, a technique useful with incomplete data and beyond. A propensity score is, roughly, the probability of an observation being missing or an indication thereof. The method has been used as a basis for missing-data developments in general and sensitivity analysis in particular. For example, it is strongly connected to more recent inverse probability weighting methods, as well as to certain forms of multiple imputation (Rubin, 1987).

Apart from considering pattern-mixture models (PMM) for their own sake, they have been considered by way of a useful contrast to selection models, either (1) to answer the same scientific question, such as marginal treatment effect or time evolution, based on these two rather different modeling strategies, or (2) to gain additional insight by supplementing the selection model results with those from a PMM approach. Pattern-mixture models also have a special role in some multiple imputation based sensitivity analyses. Examples of PMM applications can be found in Cohen and Cohen (1983), Muthén, Kaplan, and Hollis (1987), Allison (1987), McArdle and Hamagani (1992), McArdle and Hamagani (1992), Little and Wang (1996), Little and Yau (1996), Hedeker and Gibbons (1997), Hedeker and Gibbons (1997), Hogan and Laird (1997), Ekholm and Skinner (1998), Molenberghs, Michiels, and Kenward (1998), Michiels, Molenberghs, and Lipsitz (1999), Verbeke, Lesaffre, and Spiessens (2001), Michiels *et al* (2002), Thijs *et al* (2002), and Rizopoulos, Verbeke, and Molenberghs (2008). Whereas the earlier references primarily focus on the use of the framework as such, the later ones emanate a gradual shift towards sensitivity analysis applications. Molenberghs *et al* (1998) and Kenward, Molenberghs, and Thijs (2003) studied the relationship between selection models and PMMs. The earlier paper presents the PMM's counterpart of MAR, whereas the later one states how pattern-mixture models can be constructed such that dropout does not depend on future points in time.

Turning to the shared-parameter (SPM) framework, one of its main advantages is that it can easily handle non-monotone missingness. Nevertheless, these models are based on very strong parametric assumptions, such as normality of the shared random effect(s). Of course, sensitivities abound in the selection and PMM frameworks as well, but the assumption of unobserved, random, or latent effects, further compounds the issue. Various authors have considered model extensions. An overview is given by Tsonaka, Verbeke, and Lesaffre (2007), who consider shared parameter models without any parametric assumptions for the shared parameters. A theoretical assessment of the sensitivity with respect to these parametric assumptions is presented in Rizopoulos, Verbeke, and Molenberghs (2008).

Beunckens *et al* (2007a) proposed a so-called *latent-class mixture model*, bringing together features of all three frameworks. Information from the location and evolution of the response profiles, a selection model concept, and from the dropout patterns, a pattern-mixture idea, is used simultaneously to define latent groups and variables, a shared-parameter feature. This brings several appealing features. First, one uses information in a more symmetric, elegant way. Second, apart from providing a more flexible modeling tool, there is room for use as a sensitivity analysis instrument. Third, a strong advantage over existing methods is the ability to classify subjects into latent groups. If done with due caution, it can enhance substantive knowledge and generate hypotheses. Fourth, while computational burden increases, fitting the proposed method is remarkably stable and acceptable in terms of computation time. Clearly, neither the proposed model nor any other alternative can

be seen as a tool to definitively test for MAR *versus* MNAR, as discussed earlier. This is why the method's use predominantly lies within the sensitivity analysis context. Such a sensitivity analysis is of use both when it modifies the results of a simpler analysis, for further scrutiny, as well as when it confirms these.

As stated earlier, a quite separate, extremely important line of research starts from a semi-parametric standpoint, as opposed to the parametric take on the problem that has prevailed throughout this chapter. Within this paradigm, weighted generalized estimating equations (WGEE), proposed by Robins *et al* (1995) and Robins and Rotnitzky (1995) play a central role. Rather than jointly modeling the outcome and missingness processes, the centerpiece is inverse probability weighting of a subject's contribution, where the weights are specified in terms of factors influencing missingness, such as covariates and observed outcomes. These ideas are developed in Robins *et al* (1998) and Scharfstein, Rotnitzky, and Robins (1999). Robins, Rotnitzky, and Scharfstein (2000) and Rotnitzky *et al* (2001) employ this modeling framework to conduct sensitivity analysis. They allow for the dropout mechanism to depend on potentially unobserved outcomes through the specification of a non-identifiable sensitivity parameter. An important special case for such a sensitivity parameter, τ say, is $\tau = 0$, which the authors term explainable censoring, which is essentially a sequential version of MAR. Conditional upon τ , key parameters, such as treatment effect, are identifiable. By varying τ , sensitivity can be assessed. As such, there is similarity between this approach and the interval of ignorance concept, touched upon in the second paragraph of the next section. There is a connection with pattern-mixture models too, in the sense that, for subjects with the same observed history until a given time $t - 1$, the distribution for those who drop at t for a given cause is related to the distribution of subjects who remain on study at time t .

Fortunately, it is often possible in problems of missing data, to bring in assumptions that are external to this study, in the sense of them being untestable from its data, but that are implied by the scientific body of knowledge surrounding the problem. An example is the so-called *exclusion restriction* in certain problems of causal inference. When such assumptions are brought in, the missing data distribution can become identifiable or, at least, the universe of possibilities may be reduced in size. In particular, such knowledge may provide external evidence against MAR. Key references include Angrist, Imbens, and Rubin (1996), Little and Yau (1996), and Frangakis and Rubin (2002). Their work is geared towards both study design and analysis methodology that can integrate such external knowledge.

Thus clearly, the field of sensitivity analysis, for incomplete data and beyond, is both blessed with a long and rich history and vibrantly alive. The amount of work in this field is vast. Classifying sensitivity analysis methods by means of a useful taxonomy is easier said than done. One could categorize according to the model family to which they are directed within which they are cast. Alternatively, one can distinguish between context-free techniques and methods that make use of substantive considerations. Some methods make simplifying assumptions and specific choices. For example, a number of sensitivity analysis tools are based upon considering a scalar or low-dimensional sensitivity parameter, often positioned within the original model at one of many possible locations. Such choices are entirely reasonable, and ought to be seen as a pragmatic compromise between the desire to explore sensitivity while keeping the ensuing analysis practically feasible and interpretable.

10 Concluding Remarks

In this paper, we have reviewed complexities that can and often will arise when data are incomplete or otherwise coarser than actually or counterfactually possible. It has been indicated that, already under MAR and/or ignorability, care needs to be taken since results stemming from complete-data analysis may no longer hold and intuition based there upon misleading. This notwithstanding, MAR plays a pivotal role and it is appealing to have it operationally defined in all three frameworks, selection models, pattern-mixture models, and shared-parameter models. At the same time, every MNAR model can be teamed up with an counterpart that is MAR and produces exactly the same fit to the observed data; this implies that in a fundamental sense it is not possible to distinguish between MAR and MNAR. Whereas this result holds for incomplete data, it holds more broadly for the entire family of data-enriched structures, encompassing coarsened and augmented data settings. Furthermore, in a longitudinal setting, it is in all three frameworks possible to constrain missing-data mechanisms to the non-future dependent subclass of MNAR, such that missingness depends on the current, possible unobserved, measurement, but not on future ones. All of this points to a great danger for inferences and hence conclusions to be based on unverifiable assumptions, leading to sensitivity analysis. Without going into a great amount of detail, a perspective on sensitivity analysis has been offered. Details can be found, for example, in Molenberghs and Kenward (2007) and Fitzmaurice *et al* (2008).

Acknowledgment

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

References

- Allison P.D. (1987). Estimation of linear models with incomplete data. *Sociology Methodology*, **17**, 71–103.
- Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, **91**, 444–455.
- Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2007a). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics* **63**, 000–000.
- Beunckens, C., Sotito, C., Molenberghs, G., and Verbeke, G. (2007b). An integrated sensitivity analysis of the Slovenian Public Opinion Survey data. *Submitted for publication*.
- Carlin, B.P. and Louis, T.A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman & Hall.
- Cohen, J. and Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Copas, J.B. and Li, H.G. (1997). Inference from non-random samples (with discussion). *Journal of the Royal Statistical Society, Series B*, **59**, 55–96.

- Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M.G. (2008). Shared-parameter models and missingness at random. *Submitted for publication*.
- De Backer, M., De Keyser, P., De Vroey, C., and Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/day—a double-blind comparative trial. *British Journal of Dermatology* **134**, 16–17.
- Dempster, A.P., Laird, N.M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Diggle, P.J., and Kenward, M.G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49–93.
- Ekholm, A. and Skinner, C. (1998). The muscatine children's obesity data reanalysed using pattern mixture models. *Applied Statistics*, **47**, 251–263.
- 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.
- Fitzmaurice, G., Davidian, M., Molenberghs, G., and Verbeke, G. (2008). *Longitudinal Data Analysis. Handbooks of Modern Statistical Methods*. New York: Chapman & Hall/CRC.
- 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.
- Frangakis, C.E. and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.
- Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D.F., and Meulders, M. (2005). Multiple imputation for model checking: completed-data plots with missing and latent data. *Biometrics*, **61**, 74–85.
- Haug, E.J., Choi, K.K., and Komkov, V. (1986). *Design of Sensitivity Analysis of Structural Systems*. Orlando, FL: Academic Press.
- Hedeker, D. and Gibbons, R.D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, **2**, 64–78.
- Heitjan, D.F. and Rubin, D.B. (1991). Ignorability and coarse data. *The Annals of Statistics*, **19**, 2244–2253.
- 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.
- Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945–960.
- 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 non-random dropout: an illustration of sensitivity. *Statistics in Medicine* **17**, 2723–32.
- 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.
- Kenward, M.G., Molenberghs, G., and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika* **90**, 53–71.
- Laird, N.M. (1994). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 84.
- 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**, 1112–1121.
- Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- Little, R.J.A. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, **52**, 98–111.
- Little, R.J.A. and Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics*, **52**, 1324–1333.
- McArdle, J.J. and Hamagani, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research*, **18**, 145–166.
- Michiels, B., Molenberghs, G., Bijmens, L., and Vangeneugden, T. (2002). Selection models and pattern-mixture models to analyze longitudinal quality of life data subject to dropout. *Statistics in Medicine*, **21**, 1023–1041.
- Michiels, B., Molenberghs, G., and Lipsitz, S.R. (1999). Selection models and pattern-mixture models for incomplete categorical data with covariates. *Biometrics*, **55**, 978–983.
- Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M.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.
- 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. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons.

- Molenberghs, G., Michiels, B., and Kenward, M.G. (1998). Pseudo-likelihood for combined selection and pattern-mixture models for missing data problems. *Biometrical Journal*, **40**, 557–572.
- 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., 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. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, **61**, 1–6.
- Molenberghs, G., Verbeke, G., and Beunckens, C. (2007). Formal and informal model selection with incomplete data. *Statistical Science*, **00**, 000–000.
- 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.
- Muthén, B., Kaplan, D., and Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, **52**, 431–462.
- Proschan, F. and Serfling, R.J. (1974). *Reliability and Biometry: Analysis of Lifelength*. Tallahassee, FL: Florida State University.
- Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random-effects misspecification. *Biometrika*, **95**, 63–74.
- Roberts, D.T. (1992). Prevalence of dermatophyte onychomycosis in the United Kingdom: Results of an omnibus survey. *British Journal of Dermatology* **126 Suppl. 39**, 23–27.
- Robins, J.M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, **90**, 122–129.
- Robins, J.M., Rotnitzky, A., and Scharfstein, D.O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–39.
- Robins, J.M., Rotnitzky, A., and Scharfstein, D.O. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, M.E. Halloran and D.A. Berry, eds. New York: Springer, pp. 1–94.
- Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–21.
- Rosenbaum, P.R. and Rubin, D.B. (1983). The central role of the propensity score method in observational studies for causal effects. *Biometrika*, **70**, 41–55.
- Rotnitzky, A., Scharfstein, D., Su, T.L., and Robins, J.M. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics*, **57**, 103–113.

- Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Rubin, D.B.. (1994). Discussion to Diggle, P.J. and Kenward, M.G.: Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 80–82.
- Scharfstein, D.O., Rotnitzky, A., and Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*, **94**, 1096–1146.
- Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3**, 245–265.
- 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.
- Tsonaka, R., Verbeke, G., and Lesaffre, E. (2007). A semi-parametric shared parameter model to handle non-monotone non-ignorable missingness. *Submitted for publication*.
- Vach, W. and Blettner, M. (1995). Logistic regresion with incompletely observed categorical covariates—investigating the sensitivity against violation of the missing at random assumption. *Statistics in Medicine*, **12**, 1315–1330.
- Verbeke, G., Lesaffre, E., and Spiessens, B. (2001). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, **35**, 419–434.
- Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**, 254–262.
- Verbeke, G. and Molenberghs, G. (2008). Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Submitted for publication*.
- Molenberghs, G., Verbeke, G., and Beunckens, C. (2007). Formal and informal model selection with incomplete data. *Statistical Science*, **00**, 000–000.
- Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M.G. (2001). Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics* **57**, 7–14.
- Wu, M.C., and Bailey, K.R. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* **7**, 337–346.
- 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.
- Zhang, J. and Heitjan, D.F. (2007). Impact of nonignorable coarsening on Bayesian inference. *Biostatistics*, **8**, 722–743.