Arbitrariness of models for augmented and coarse data, with emphasis on incomplete data and random effects models

DOI: 10.1177/1471082X0901000403
Handle: http://hdl.handle.net/1942/11484

Arbitrariness of Models for Augmented and Coarse Data, With Emphasis on Incomplete-data and Random-effects Models

Geert Verbeke[1,2] **Geert Molenberghs**[2,1]

Interuniversity Institute for Biostatistics and statistical Bioinformatics

[1] *Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

[2] *Universiteit Hasselt, B-3590 Diepenbeek, Belgium*

Abstract

Statistical models often extend beyond the data available. First, in coarse data, what is actually observed is less detailed than what might ideally be observed, owing to, for example, incompleteness, censoring, grouping, or a combination thereof. Second, in augmented data settings, the observed data are hypothetically but conveniently supplemented with such structures as random effects, latent variables, latent classes, or component membership in mixture distributions. The two settings together will be referred to as *enriched data*. Reasons for modeling enriched data encompass mathematical and computational convenience, advantages in interpretation, and substantive plausibility of such constructions. It is generally known that models for enriched data combine evidence coming from empirical data with unverifiable model components, resting entirely on assumptions. While a result that is fairly generally true, it has accute consequences in the case of enriched data. This notwithstanding, knowledge about and insight into this issue is somewhat scattered. We provide a unified framework for enriched data and show, generally on the one hand and with focus on the cases of incomplete-data models and random-effects models on the other hand, that to any given model an entire class of models can be assigned, with all of its members producing the same fit to the observed data but arbitrary regarding the unobservable parts of the enriched data. The concepts developed are illustrated using a clinical trial in toenail dermatophyte onychomycosis and a developmental toxicity study conducted in mice.

Some Keywords: Compound-symmetry; Empirical Bayes; Enriched data; Exponential random effects; Gamma random effects; Linear mixed model; Missing at random; Missing completely at random; Non-future dependence; Pattern-mixture model; Selection model; Shared-parameter model.

1 Introduction

Augmented data, in the sense of supplementing the observed data with latent, unobserved structures, is common throughout statistics. Examples include models for incompletely observed data, describing observed and unobserved outcomes alike, random-effects models, latent class models, latent variable

models, censored survival data, etc. Heitjan (Heitjan and Rubin, 1991; Zhang and Heitjan, 2007) has unified some of these settings in a concept called *coarsening*, broadly referring to the fact that the observed data are coarser than the hypothetically conceived data structures, while models target the latter. It is obvious that models for such augmented structures are identifiable only by virtue of making sometimes strong but always partially unverifiable modeling assumptions. These settings taken together and from now on termed *enriched data*, will be treated in a unified way, such that important, common features can be singled out and studied. Having said this, there is a subtle distinction between both concepts. In the coarse-data setting, it is understood that a part of the data would ideally be observed but are not in practice (e.g., actual survival time after censoring, outcomes after dropout, etc.). Augmented data refers rather to the addition of useful but artificial constructs to the data setting, such as random effects, latent classes, latent variables. These are assumed to always be fully unobservable. Such augmentations then allow to drastically simplify model development. A key example of coarsening is incomplete data, whereas a model with random effects is a paradigm for augmented data.

Quite a bit of work to explore and address issues arising from enrichment has been done for the context of incomplete data. In that setting, an often made assumption is that the missing data are missing at random (MAR), meaning that the unobserved data provide no further information about the mechanism governing missingness, given the observed outcomes (Rubin, 1976; Molenberghs and Kenward, 2007). More general mechanisms are then termed missing not at random (MNAR). Apart from classifying the missingness mechanisms, one often employs a taxonomy for joint models that simultaneously describe the measurement and missingness mechanisms: (1) In a *selection model* (SeM), the joint distribution of the ith subject's outcomes, denoted $\boldsymbol{Y}_i$, and vector of missingness indicators, written $\boldsymbol{R}_i$, is factored as the marginal outcome distribution and the conditional distribution of $\boldsymbol{R}_i$ given $\boldsymbol{Y}_i$; (2) A *pattern-mixture model* (PMM) approach starts from the reverse factorization; (3) In a *shared-parameter model* (SPM), a set of latent variables, latent classes, and/or random effects is posited, that are then assumed to drive both the $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$ processes, in the sense that conditional on the latent variables, $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$ exhibit no further dependence.

Especially during the last decade, work has been done to increase understanding of the issues arising

from coarsening with incomplete data. This has led to a vast body of so-called sensitivity analyses (Molenberghs and Kenward, 2007). Moreover, Molenberghs *et al* (2008) and Creemers *et al* (2008) have established that for any MNAR model, there exists a counterpart of MAR type, with exactly the same fit to the observed data. Whereas the former authors presented the result in generic incomplete-data terms and then focused on SeM and PMM, the latter paper is dedicated to the SPM setting. Obviously, such work needs a characterization of MAR in each of the three frameworks. The original definition of MAR (Rubin, 1976) had been given in SeM terms and Molenberghs *et al* (1998) had provided a characterization for PMM. The SPM characterization of MAR can be found in Creemers *et al* (2008).

In this paper, we focus on to the general enriched-data case and establish that always a part of the model is fully unidentifiable from the observed data. The evident consequence is that the identification of such a part can come from assumptions only and points at the same time to the considerable risk for conclusions to be sensitive to such assumptions, and ultimately to the need for conducting sensitivity analyses. It implies that such non-identified parts can be replaced arbitrarily, without altering the fit to the observed data but with potentially grave consequences for inferences and substantive conclusions. Put simply, one's inferential conclusions may strongly depend on such unverifiable portions of the model. The result is also presented in full generality. Two specific cases are treated in more detail. First, we bring together and unify results from Beunckens *et al* (2007b) and Creemers *et al* (2008), dealing with missing data and, in particular, shared-parameter models. Second, we study in detail the linear mixed model as a key paradigm within the mixed-model family.

The remainder of the paper is organized as follows. Two sets of data, a clinical trial and a developmental toxicity study, are introduced in Section 2, and analyzed in Section 7. Notation and basic concepts are introduced in Section 3. Our general result, regarding data-enriched structures is given in Section 4. In Section 5, the focus is on incomplete data, with particular emphasis on so-called shared-parameter results, bringing together and unifying earlier results (Beunckens *et al*, 2007b; Creemers *et al*, 2008). Section 6 focuses on the specific but insightful and ubiquitous linear mixed-effects model (Verbeke and Molenberghs 2000), with particular attention to exchangeable, compound-symmetric data.

2 Motiving Data Sets

2.1 A Clinical Trial in Onychomycosis

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

In total, 2×189 patients, distributed over 36 centers, were randomized. Subjects were followed during 12 weeks (3 months) of treatment and followed further, up to a total of 48 weeks (12 months). Measurements were taken at baseline, every month during treatment, and every 3 months afterwards, resulting in a maximum of 7 measurements per subject. At the first occasion, the treating physician indicates one of the affected toenails as the target nail, the nail which will be followed over time. We will restrict our analyses to only those patients for which the target nail was one of the two big toenails. This 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). Figure 1 shows the observed profiles of 30 randomly selected subjects from treatment group A and treatment group B, respectively. In Table 1, the amount of missingness is brought to the forefront, by listing the number of repeated measures available per subject, for each of the two treatment arms separately. A linear mixed model will be considered, in which enrichment arises through the inclusion of random effects.

2.2 A Developmental Toxicity Study

This developmental toxicity study investigates the dose-response relationship in mice of the potentially hazardous chemical compound di(2-ethylhexyl)phthalate (DEHP), used in vacuum pumps (Windholz, 1983) and as plasticizers for numerous plastic devices made of polyvinyl chloride. DEHP provides the finished plastic products with desirable flexibility and clarity (Shiota, Chou, and Nishimura, 1980). It has been well documented that small quantities of phthalic acid esters, of which DEHP is an instance, may leak out of polyvinyl chloride plastic containers in the presence of food, milk, blood, or various solvents. Due to their ubiquitous distribution and presence in human and animal tissues, considerable concern has developed as to the possible toxic effects of the phthalic acid esters (Autian, 1973). The developmental toxicity study, conducted in timed-pregnant mice during the period of major organogenesis and described by Tyl *et al* (1988), has attracted much interest in the toxicity of DEHP. The doses selected for the study were 0, 0.025, 0.05, 0.1, and 0.15%, corresponding to a DEHP consumption of 0, 44, 91, 191, and 292 mg/kg/day, respectively. The dams were sacrificed, slightly prior to normal delivery, and the status of uterine implantation sites recorded. A total of 1082 live fetuses were dissected from the uterus, anesthetized, and examined for external, visceral, and skeletal malformations, as well as for body weight. Our focus will be on the continuous weight outcome. Evidently, fetuses are clustered within mothers; hence the implied association needs to be accommodated in the analysis. When done through random effects, data enrichment arises. Summary data are presented in Table 2. Table 2 makes clear, when the number of viable fetuses (litter size) is compared to the number of implants, that there is a substantial amount of depletion and that it, not surprisingly, increases with dose.

3 Concepts and Notation

Let the random variable Y_{ij} denote the response of interest, for the ith study subject, designed to be measured at occasions t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$. Independence across subjects is assumed. The outcomes can conveniently be grouped into a vector $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$. In addition, define a vector of missingness indicators $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{in_i})'$ with $R_{ij} = 1$ if Y_{ij} is observed and 0 otherwise.

In principle, one would like to consider the density of the full data $f(\boldsymbol{y}_i, \boldsymbol{r}_i|\boldsymbol{\theta}, \boldsymbol{\psi})$, where the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ describe the measurement and missingness processes, respectively. Covariates are assumed to be measured and grouped in a vector $\boldsymbol{x}_i$ but, throughout, are suppressed from notation.

We will now sketch the modeling frameworks (Section 3), present the definition of MAR in each one of them (Section 5.1), and then establish that every MNAR model can be doubled up with a MAR counterpart that preserves the fit to the observed data (Section 5.2).

The full density function can be factored in different ways, each leading to a different framework, already briefly mentioned in the introduction.

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

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

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

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

The *shared-parameter model* (Wu, and Carroll, 1988; Wu, and Bailey, 1988, 1989; TenHave *et al*, 1998; Follmann, and Wu, 1995; Little, 1995) assumes a vector of random effects $\boldsymbol{b}_i$, shared between both processes, conditional upon which the measurement and missingness processes are independent, and often taking the form of random effects with a specific parametric distribution. This *shared-parameter model* (SPM) is formulated by way of the following factorization

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

and hence

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

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 $\boldsymbol{b_i}$ drives the entire process, one can expand $\boldsymbol{b_i}$ to a set of latent structures.

4 General Result About Counterparts in Enriched-data Structures

Let us state our general result. We need slightly more general notation. Assume data $\boldsymbol{Z}_i$ for an independent unit $i = 1, \ldots, N$ are augmented with $\boldsymbol{c}_i$. The $\boldsymbol{c}_i$ 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 shared-parameter framework, that will be considered in the next section.

Assume a joint model of the generic form $f(\boldsymbol{z}_i, \boldsymbol{c}_i|\boldsymbol{\theta}, \boldsymbol{\psi})$, where covariates have been suppressed for notational simplicity. Consider the factorizations:

$$f(\boldsymbol{z}_i, \boldsymbol{c}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = f(\boldsymbol{z}_i|\boldsymbol{c}_i, \boldsymbol{\theta}) f(\boldsymbol{c}_i|\boldsymbol{\psi}), \tag{5}$$

$$= f(\boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) f(\boldsymbol{c}_i|\boldsymbol{z}_i, \boldsymbol{\theta}, \boldsymbol{\psi}). \tag{6}$$

Borrowing terminology from the hierarchical-models context, such as mixed models, which are further given specific consideration in Section 6, every factor in both (5) and (6) can usefully be given a name. The left hand side is the *joint model*. Let us turn to the right hand sides. The first factor in (5) is the *hierarchical model* and the second one is the *prior density* for the enriched data. The first factor in (6) 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 (5)–(6) and the mixed-model setting. The link with incomplete data follows by setting $\boldsymbol{c}_i \equiv \boldsymbol{y}_i^m$ and $\boldsymbol{z}_i = (\boldsymbol{y}_i^o, \boldsymbol{r}_i)$. Hence, again, we are naturally led to the PMM framework. In PMM factorization (16) the marginal model is further factored, but this is immaterial. The key is the third factor on the right hand side of (16), i.e., the second factor in (6).

These considerations immediately establish the following theorem.

Theorem 1 (A Family of Counterparts to a Given Model for Enriched Data.) *Let us assume that data $\boldsymbol{z}_i$ are enriched with $\boldsymbol{c}_i$. Then, any model (5) formulated for and fitted to such data, can be replaced by an infinite family of models, all retaining the fit to the observed data as achieved by the original model. This is done by preserving the marginal model $f(\boldsymbol{z}_i|\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}})$ and replacing the*

posterior density $f(\boldsymbol{c}_i|\boldsymbol{z}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}})$ *by an arbitrary conditional density*

$$f(\boldsymbol{d}_i|\boldsymbol{z}_i, \boldsymbol{\gamma}). \tag{7}$$

Here, $\boldsymbol{d}_i$ rather than $\boldsymbol{c}_i$ is used to indicate that there need not be any connection between the original and substituted enriched data. Also, the new density (7) can be parameterized by a completely new parameter $\boldsymbol{\gamma}$.

5 Incomplete Data

In Section 3, we introduced standard concepts regarding incomplete data. We will now move beyond this by enlarging the family of shared-parameter models, then zoom in on missingness at random and study the impact of our general result for this particular case.

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

$$f(\boldsymbol{y}_i^o|\boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{j}_i, \boldsymbol{\ell}_i) f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{k}_i, \boldsymbol{m}_i) f(r_i|\boldsymbol{g}_i, \boldsymbol{j}_i, \boldsymbol{k}_i\, \boldsymbol{n}_i), \tag{8}$$

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

Here, $\boldsymbol{y}_i^o$ ($\boldsymbol{y}_i^m$) refers to the observed (missing) components for subject i. While fixed effects are allowed to accompany each of the random effects, they are suppressed from notation.

Several remarks are in place. First, this is the most general random-effects model that can be considered in the sense that $\boldsymbol{g}_i$ is common to all three factors in (8), $\boldsymbol{h}_i$, $\boldsymbol{j}_i$, and $\boldsymbol{k}_i$ are shared between pairs of factors, and $\boldsymbol{\ell}_i$, $\boldsymbol{m}_i$, and $\boldsymbol{n}_i$ are restricted to a single factor. Depending on the application, one may choose to either retain all random effects or to omit some. For example, $\boldsymbol{j}_i$ is present in the first factor but not in the second, with the reverse holding for $\boldsymbol{k}_i$. Retaining these is useful when it is deemed plausible that, at the time of dropout, the process governing the outcome is sufficiently altered so as to modulate the effects of $\boldsymbol{g}_i$ and $\boldsymbol{h}_i$, which are common to both. Note also that $\boldsymbol{m}_i$ is never identifiable from data but is introduced as the basis for sensitivity analysis. It

will then be useful to have a perspective on the implications of such simplifications, preferably also in terms of the missing-data mechanism operating. This is why it is useful to establish conditions under which MAR operates on the one hand, and missingness does not depend on future, unobserved measurements in a longitudinal context on the other hand. Second, in full generality, model (8) may come across as somewhat contrived. The objective of formulating Definition 1 is not to postulate (8) as a model for 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 may seem that (8) assumes two completely different distributions for the outcome vector, i.e., divorcing the observed from the missing components. This is not entirely the case because $\boldsymbol{g}_i$ and $\boldsymbol{h}_i$ still tie both factors together. The impact of $\boldsymbol{j}_i$, $\boldsymbol{k}_i$, $\boldsymbol{\ell}_i$, and $\boldsymbol{m}_i$ is to modify one's latent process in terms of missingness. In other words, the most general model assumes that observed and missing components are governed in part by common processes and partly by separate processes. Third, in principle, we could expand (8) 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, a conventional SPM formulation follows by removing all random effects but $\boldsymbol{g}_i$. For convenience, write

$$\boldsymbol{b_i} = (\boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{j}_i, \boldsymbol{k}_i, \boldsymbol{\ell}_i, \boldsymbol{m}_i, \boldsymbol{n}_i). \tag{9}$$

5.1 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

$$f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{\psi}) = f(\boldsymbol{r}_i|\boldsymbol{y}_i^o, \boldsymbol{\psi}), \tag{10}$$

i.e., the missing-data mechanism depends on the observed outcomes but, given these, not further on the unobserved ones. In the MNAR case, missingness depends on the unobserved outcomes $\boldsymbol{y}_i^m$, regardless of the observed outcomes and the covariates.

Molenberghs *et al* (1998, 2008), among others, formulated MAR in the PMM setting:

Theorem 2 (Missingness at Random in the Pattern-mixture Framework.) *In the PMM framework, the missing-data mechanism is MAR if and only if*

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

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. Put differently, prediction of the unobserved outcomes can be done merely using the observed ones with no further information coming from the missing-data mechanism. Note that, owing to this result, MAR can be formulated in terms of R given Y, but also in terms of Y given R.

Creemers *et al* (2008) characterized MAR in the SPM framework:

Theorem 3 (Characterization of MAR in the General Shared-parameter Family.) *A member of the general SPM family (8) is MAR if and only if*

$$\begin{aligned}\frac{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{j}_i) f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i, \boldsymbol{h}_i, \boldsymbol{k}_i) f(\boldsymbol{r}_i|\boldsymbol{g}_i, \boldsymbol{j}_i, \boldsymbol{k}_i) f(\boldsymbol{b_i})\, d\boldsymbol{b_i}}{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i, \boldsymbol{j}_i) f(\boldsymbol{r}_i|\boldsymbol{g}_i, \boldsymbol{j}_i) f(\boldsymbol{b_i})\, d\boldsymbol{b_i}} \\ = \quad \frac{\int f(\boldsymbol{y}_i^o|\boldsymbol{g}_i, \boldsymbol{h}_i) f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i, \boldsymbol{h}_i) f(\boldsymbol{b_i})\, d\boldsymbol{b_i}}{f(\boldsymbol{y}_i^o)}.\end{aligned} \tag{12}$$

Note that the random effects $\boldsymbol{\ell}_i$, $\boldsymbol{m}_i$, and $\boldsymbol{n}_i$, pertaining to a single factor only, are suppressed from notation but are allowed to be present. Clearly, this result is not as intuitive as the SeM and PMM versions and, as such, the above result has little immediate data-analytic value. Therefore, fortunately, these authors also showed that the following family satisfies the MAR property:

Definition 2 (A Sub-class of SPM Models.) *Define a sub-class of shared-parameter model (8):*

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

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

At the same time, they established that there are members of the SPM family satisfying Theorem 3 but that are not of the (13) type.

5.2 Every MNAR Model Has Got an MAR Counterpart

In this section, based on the argument of Molenberghs *et al* (2008), we restate that for every MNAR model fitted to a set of data, there is a unique MAR counterpart providing exactly the same fit to the data. Whereas these authors confined attention to the missing-data setting, in the next section we will provide a much more general result, pertaining to all data-enriched structures.

The concept of model fit should be understood as being measured using such conventional methods as deviance measures and, of course, in as far as the observed data are concerned. The following steps are involved: (1) fitting an MNAR model to the data; (2) reformulating the fitted model in PMM form; (3) replacing the density or distribution of the unobserved measurements given the observed ones and given a particular response pattern by its MAR counterpart; (4) establishing that such an MAR counterpart uniquely exists.

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

$$L = \prod_i \int f({\boldsymbol{y}_i}^o, {\boldsymbol{y}_i}^m, \boldsymbol{r}_i | \boldsymbol{\theta}, \boldsymbol{\psi}) d{\boldsymbol{y}_i}^m. \tag{14}$$

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

$$f({\boldsymbol{y}_i}^o, {\boldsymbol{y}_i}^m, \boldsymbol{r}_i | \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) = f({\boldsymbol{y}_i}^o, {\boldsymbol{y}_i}^m | \widehat{\boldsymbol{\theta}}) f(\boldsymbol{r}_i | {\boldsymbol{y}_i}^o, {\boldsymbol{y}_i}^m, \widehat{\boldsymbol{\psi}}). \tag{15}$$

To undertake the second step, full density (15) can be re-expressed in PMM form as:

$$f({\boldsymbol{y}_i}^o, {\boldsymbol{y}_i}^m | \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) f(\boldsymbol{r}_i | \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) = f({\boldsymbol{y}_i}^o | \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) f(\boldsymbol{r}_i | \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) f({\boldsymbol{y}_i}^m | {\boldsymbol{y}_i}^o, \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}). \tag{16}$$

Note that the final term on the right hand side of (16), $f({\boldsymbol{y}_i}^m | {\boldsymbol{y}_i}^o, \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}})$, is not identified from the observed data. In this case, it is determined solely from modeling assumptions. Within the PMM framework, identifying restrictions have to be considered (Little, 1994; Molenberghs *et al*, 1998; Kenward, Molenberghs, and Thijs, 2003).

The third step requires replacing this factor by the appropriate MAR counterpart. Now, using Theorem 2, it is clear that $f({\boldsymbol{y}_i}^m | {\boldsymbol{y}_i}^o, \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}})$ needs to be replaced with

$$f^*({\boldsymbol{y}_i}^m | {\boldsymbol{y}_i}^o, \boldsymbol{r}_i) = f^*({\boldsymbol{y}_i}^m | {\boldsymbol{y}_i}^o) = f({\boldsymbol{y}_i}^m | {\boldsymbol{y}_i}^o, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}), \tag{17}$$

where the $f^*(\cdot)$ notation is used for shorthand purposes. Note that the density in (17) follows from the SeM-type marginal density of the complete data vector. Sometimes, therefore, it may be more convenient to replace the notation ${\boldsymbol{y}_i}^o$ and ${\boldsymbol{y}_i}^m$ by one that explicitly indicates which components are observed and missing in pattern $\boldsymbol{r}_i$ under consideration:

$$f^*({\boldsymbol{y}_i}^m|{\boldsymbol{y}_i}^o,\boldsymbol{r}_i) = f^*({\boldsymbol{y}_i}^m|{\boldsymbol{y}_i}^o) = f[(y_{ij})_{r_j=0}|(y_{ij})_{r_j=1},\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}}]. \tag{18}$$

Thus, (18) provides a unique way of extending the model fit to the observed data, belonging to the MAR family. As stated before, the above construction does not lead to a member of a conventional parametric family. While this obviously implies limitations on its use, such is not dissimilar to the construction of some semi- and non-parametric estimators. Also, it helps to understand that an overall, definitive conclusion about the nature of the missing-data mechanism, 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). To show formally that the fit remains the same, Molenberghs *et al* (2008) considered the observed-data likelihood based on (14) and (16):

$$\begin{aligned}
\widehat{L} &= \prod_i \int f({\boldsymbol{y}_i}^o,{\boldsymbol{y}_i}^m|\widehat{\boldsymbol{\theta}}) f(\boldsymbol{r}_i|{\boldsymbol{y}_i}^o,{\boldsymbol{y}_i}^m,\widehat{\boldsymbol{\psi}}) d{\boldsymbol{y}_i}^m \qquad (19)\\
&= \prod_i \int f({\boldsymbol{y}_i}^o|\boldsymbol{r}_i,\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}}) f(\boldsymbol{r}_i|\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}}) f({\boldsymbol{y}_i}^m|{\boldsymbol{y}_i}^o,\boldsymbol{r}_i,\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}}) d{\boldsymbol{y}_i}^m\\
&= \prod_i f({\boldsymbol{y}_i}^o|\boldsymbol{r}_i,\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}}) f(\boldsymbol{r}_i|\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}})\\
&= \prod_i \int f({\boldsymbol{y}_i}^o|\boldsymbol{r}_i,\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}}) f(\boldsymbol{r}_i|\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\psi}}) f^*({\boldsymbol{y}_i}^m|{\boldsymbol{y}_i}^o) d{\boldsymbol{y}_i}^m. \qquad (20)
\end{aligned}$$

The above results show the following theorem:

Theorem 4 (MAR Counterpart to MNAR Models.) *Every fit to the observed data, obtained from fitting an MNAR model to a set of incomplete data, is exactly reproducible from an MAR decomposition.*

The key computational consequence is the need to determine $f^*({\boldsymbol{y}_i}^m|{\boldsymbol{y}_i}^o)$ in (17) or (18). This means that, for each pattern, the conditional density of the unobserved measurements given the observed ones needs to be extracted from the marginal distribution of the complete set of measurements. Molenberghs *et al* (1998) have shown that, for the special case of dropout (i.e., monotone

missingness), the so-called *available case missing value restrictions* (ACMV) provide a practical computational scheme.

The characterization of Theorem 3 allows us to construct an MAR counterpart to an arbitrary SPM of the form (8). It is necessary to (a) retain the fit of the model to the observed data, while (b) ensuring that (12) holds. This is easily done by *a-posteriori integrating* over the shared random effects in the densities describing the unobserved measurements, given the observed ones. Practically, integration takes place over the densities of $\boldsymbol{g}_i$, $\boldsymbol{h}_i$, and $\boldsymbol{k}_i$, where fitted parameters are plugged into the densities.

Theorem 5 (An MAR Counterpart to a General SPM.) *The MAR counterpart, to an arbitrary general SPM of the type (8) is found by replacing* $f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)$ *with*

$$f^*(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{m}_i)=\int_{\boldsymbol{g}_i}\int_{\boldsymbol{h}_i}\int_{\boldsymbol{k}_i} f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)f(\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i)d\boldsymbol{g}_i d\boldsymbol{h}_i d\boldsymbol{k}_i. \qquad (21)$$

First, it is clear that this marginalization is merely describing the model-based prediction of the unobserved outcomes, given the observed ones. Hence, the choice for $f^*(\cdot)$ does not alter the fit. Second, observe that using $f^*(\cdot)$ in (12), instead of $f(\boldsymbol{y}_i^m|\boldsymbol{y}_i^o,\boldsymbol{g}_i,\boldsymbol{h}_i,\boldsymbol{k}_i,\boldsymbol{m}_i)$, of Theorem 3, reduces the equation to a trivial identity, and hence the MAR condition is also satisfied. The importance of this result is that (21) provides an MAR scenario for the missing-data mechanism, consistent with the previously achieved model fit.

Some comments are in place. Note that our general result follows quite easily in this case by observing that any missing-data model can be recast as a full PMM, as in (16). This framework readily allows the construction of an MAR substitute (17), which renders Theorem 4 almost trivial, as verified in (19)–(20). Indeed, the key feature of PMM, distinguishing it from SeM, is that it happens to factor the joint distribution of observed measurements and missing-data indicators on the one hand and unobserved measurements on the other hand, in such a way that the conditional distribution of what is unobserved, given what is observed, is an explicit factor in the model. It is this particular conditional distribution that can be changed arbitrarily.

Note that the same feature is employed in Theorem 5, relative to the SPM. Indeed, also here the distribution of what is unobserved, given what is observed, is used. Three remarks are worth making.

First, the right hand side of (21) does not condition on $\boldsymbol{r}_i$, in spite of it being observed. Now, this absence is a key characteristic of SPM and therefore entirely logical.

Second, $f^*(\cdot)$ contains conditioning on random effects $\boldsymbol{m}_i$, in spite of them being unobserved. It is important to understand that the set of distributions that can be changed have to be of unobserved outcomes or structures, given other structures that can be an amalgamation of observed and unobserved structures. The only requirement is that such a factor needs to be placed into a chain of factors that properly factors the entire joint distribution.

Third, uniqueness results in the missing-data case come from the requirement that the counterpart is of MAR type. This can be relaxed by observing that, in (16), the factor $f({\boldsymbol{y}_i}^m|{\boldsymbol{y}_i}^o, \boldsymbol{r}_i, \widehat{\boldsymbol{\theta}}, \widehat{\psi})$ may be replaced by *any* valid density. This well-known result is: (1) placed in the broader context of enriched data; (2) also phrased in a shared-parameter context; (3) is illustrated in an insightful way.

6 Linear Mixed-effects Models

In Section 6.1, the linear mixed model will be considered for illustration. In Section 6.2, the special but important case of clustered data will be considered, with constant mean within clusters and compound-symmetry variance-covariance structure.

6.1 The Standard Linear Mixed-effects Model

Let us consider the linear mixed-effects model, in all components featuring in (5)–(6), and then apply Theorem 1 to replace the posterior density of the random effects, ordinarily normal, by two versions of the exponential density.

6.1.1 Standard Formulation of the Linear Mixed Model

Using notation as in Section 3, the fully hierarchically specified linear mixed-effects model takes the form (Verbeke and Molenberghs, 2000):

$$\boldsymbol{Y}_i|\boldsymbol{b_i} \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}, \Sigma_i), \tag{22}$$

$$\boldsymbol{b_i} \sim N(0, D), \tag{23}$$

where $\boldsymbol{\beta}$ is a vector of fixed effects, and X_i and Z_i are design matrices.

Based on (22) and (23), the marginal model and posterior distribution of the random effects can be derived (Searle, Casella, and McCulloch, 1996):

$$\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, V_i = Z_iDZ_i' + \Sigma_i), \tag{24}$$

$$\boldsymbol{b_i}|\boldsymbol{Y}_i \sim N[DZ_i'V_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}), (Z_i'\Sigma_i^{-1}Z_i + D^{-1})^{-1}]. \tag{25}$$

It is useful to also present the empirical Bayes predictions (Carlin and Louis, 1996; Verbeke and Molenberghs, 2000). For the random effects, these follow in a straightforward fashion as the mean of (25), i.e.,

$$\widehat{\boldsymbol{b_i}} = E(\boldsymbol{b_i}|\boldsymbol{Y}_i) = DZ_i'V_i^{-1}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}). \tag{26}$$

For the prediction of outcome $\boldsymbol{Y}_i$, the value in (26) is plugged into the mean of the hierarchical model (22):

$$\widehat{\boldsymbol{Y}_i} = (Z_iDZ_i') \cdot V_i^{-1}\boldsymbol{y}_i + (\Sigma_i) \cdot V_i^{-1}X_i\boldsymbol{\beta}, \tag{27}$$

the familiar "weighted average" of the observed outcomes $\boldsymbol{y}_i$ and the marginal mean $X_i\boldsymbol{\beta}$.

6.1.2 A First Normal-exponential Version of the Linear Mixed Model

To illustrate the arbitrariness of the posterior density, brought forward by Theorem 1 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 gamma random effects, where each outcome component Y_{ij} is paired with a gamma random effect g_{ij}. The conventional density for a gamma variable ϕ is

$$f(\phi) = [\beta_*^{\alpha_*}\Gamma(\alpha_*)]^{-1}\,\phi^{\alpha_*-1}e^{-\phi/\beta_*}, \tag{28}$$

with $\alpha_*, \beta_* \geq 0$ parameters. For convenience, let us set $\alpha_* = 1$ and $\delta = 1/\beta_*$ in (28), producing

$$f(\phi) = \delta e^{-\phi\delta}, \tag{29}$$

which is the exponential density special case of the gamma family. Clearly, the mean of ϕ then is $E(\phi) = \delta^{-1}$. Note that the choice for an exponential distribution here is not aimed at proposing a viable model for data analysis. The choice is made to conveniently illustrate Theorem 1, in such

a way that reasonably tractable closed-form solutions can be obtained, at the same time allowing for choice within the exponential framework. Indeed, the choice to be made next can be juxtaposed with the one of Section 6.1.3.

Our first choice is completed by choosing a conditional density of the form (29) for $\phi = g_{ij}$, with $\delta = \gamma_j y_{ij}$, where γ_j is an unspecified parameter. The marginal model (24) is retained and coupled with the posterior:

$$f(\boldsymbol{g}_i|\boldsymbol{y}_i) = \prod_{j=1}^{n_i} \gamma_j y_{ij} e^{-g_{ij}\gamma_j y_{ij}}. \tag{30}$$

The joint density of $\boldsymbol{y}_i$ and $\boldsymbol{g}_i$ obviously follows as the product of the density corresponding to (24) and density (30), and hence, after some algebra, the hierarchical model and prior can be seen to take the forms:

$$f(\boldsymbol{g}_i) = \left(\prod_{j=1}^{n_i} \gamma_j\right) e^{\boldsymbol{\mu}_i'\boldsymbol{\theta}_i + \frac{1}{2}\boldsymbol{\theta}_i' V_i \boldsymbol{\theta}_i} M_{n_i}(\boldsymbol{\mu}_i + V_i\boldsymbol{\theta}_i, V_i), \tag{31}$$

$$f(\boldsymbol{y}_i|\boldsymbol{g}_i) = \frac{\left(\prod_{j=1}^{n_i} y_{ij}\right) e^{\boldsymbol{\theta}_i'(\boldsymbol{y}_i - \boldsymbol{\mu}_i)} e^{-\frac{1}{2}\left[(\boldsymbol{y}_i-\boldsymbol{\mu}_i)' V_i^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_i) + \boldsymbol{\theta}_i' V_i \boldsymbol{\theta}_i\right]}}{(2\pi)^{n_i/2}|V_i|^{1/2} M_{n_i}(\boldsymbol{\mu}_i + V_i\boldsymbol{\theta}_i, V_i)}, \tag{32}$$

where $\boldsymbol{\mu}_i = X_i\boldsymbol{\beta}$, $\boldsymbol{\theta}_i$ has components $\theta_{ij} = -g_{ij}\gamma_j$, and $M_n(\boldsymbol{k}, V) = E(Y_1 \ldots Y_n; \boldsymbol{k}, V)$, i.e., the sole nth order moment, relative to a normal distribution with mean $\boldsymbol{k}$ and variance V, where each component occurs exactly once. From Willink (2005) it follows that a simple recursive relationship can be used, based on the concept of Hermite polynomials, to calculate such moments:

$$M_n(\boldsymbol{k}, V) = k_n M_{n-1}(\boldsymbol{k}, V) + \sum_{j=1}^{n-1} v_{jn} M_{1\ldots j-1, j+1\ldots n-1}(\boldsymbol{k}, V),$$

where the last term is an $(n-2)$th order moment, with both the jth and nth components left out; k_j is the jth element of the vector $\boldsymbol{k}$ and v_{jn} is the (j,n)th entry of the matrix V.

The empirical Bayes predictions take the form:

$$\widehat{g_{ij}} = 1/(\gamma_j y_{ij}), \tag{33}$$

$$\widehat{\boldsymbol{y}_i} = \frac{\boldsymbol{P}_{n_i}(\boldsymbol{\mu}_i - V_i\boldsymbol{z}_i, V_i)}{M_{n_i}(\boldsymbol{\mu}_i - V_i\boldsymbol{z}_i, V_i)}, \tag{34}$$

where $\boldsymbol{P}_{n_i}(\boldsymbol{\mu}_i - V_i\boldsymbol{z}_i, V_i)$ is an n_i-dimensional vector with components defined by:

$$P_{nj}(\boldsymbol{k}, V_i) = E(Y_1 \ldots Y_{i,j-1} Y_{ij}^2 Y_{i,j+1} \ldots Y_n; \boldsymbol{k}, V). \tag{35}$$

Also here, the following recursive relationship is useful to calculate the components of (35) (Willink, 2005):

$$\begin{aligned} P_{nj}(\boldsymbol{k}, V) &= k_j M_n(\boldsymbol{k}, V) + \sum_{k \neq j} v_{jk} E(Y_1 \dots Y_{i,j-1} Y_{ij}^2 Y_{i,j+1} \dots Y_{i,k-1} Y_{i,k+1} \dots Y_n) \\ &\quad + v_{jj} E(Y_1 \dots Y_{i,j-1} Y_{i,j+1} \dots Y_n). \end{aligned}$$

Finally, $\boldsymbol{z}_i$ is a vector with components $z_{ij} = 1/y_{ij}$.

There is an obvious consequence resulting from these developments regarding the meaning of model parameters. In specifying the original hierarchical model (22)–(23), the parameters $\boldsymbol{\beta}$, Σ_i, and D in general, but D in particular, are part of a hierarchical specification. Since (24)–(25) taken together are equivalent to the original pair of equations, one might argue that 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 (22)–(23) 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 (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 (24). This is clear when considering the model composed of (24) and, for example, either (30) or (36). 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 (24) 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 (24), which is the only channel by which the data convey information. The model merely *appears* interpretable at a hierarchical, or enriched, level since (25) contains these, and only these parameters.

Note that the choice $\delta = \gamma_j y_{ij}$ is pragmatic, in the sense that δ should be non-negative. This is fine for a data set where the outcomes are sufficiently bounded away from zero, such as body length. However, it may be deemed less elegant, in which case it may make sense to square or exponentiate y_{ij}, motivating the following, alternative formulation.

Should one adhere to a Bayesian interpretation of the original model then $\boldsymbol{b_i} \sim N(\mathbf{0}, D)$ is a conventional prior distribution, and arbitrariness pertains to the posterior distribution. While conventionally uncommon to specify the posterior first and then work back to the prior, it does help to understand that there is an observable and an unobservable part of the joint distribution. Also, it opens avenues for sensitivity analysis, as we will discuss further in Section 8.

6.1.3 A Second Normal-exponential Version of the Linear Mixed Model

Let us consider an alternative choice for (29): $\delta = e^{\gamma_j y_{ij}}$. Straightforward algebra, thereby making use of the identity:

$$\prod_{j=1}^{n_i} e^{-q_{ij}e^{\gamma_j y_j}} = \sum_{m_1=0}^{\infty} \cdots \sum_{m_{n_i}=0}^{\infty} \frac{(-q_{i1})^{m_1} \cdots (-q_{in_i})^{m_{n_i}}}{m_1! \cdots m_{n_i}!} e^{m_1\gamma_1 y_{i1} + \cdots + m_{n_i}\gamma_{n_i} y_{in_i}},$$

leads to the following model equations, that are in the same order and with the same notation as in the first normal-exponential case:

$$f(\boldsymbol{q}_i|\boldsymbol{y}_i) = \prod_{j=1}^{n_i} e^{\gamma_j y_{ij}} e^{-q_{ij}e^{\gamma_j y_{ij}}}, \tag{36}$$

$$f(\boldsymbol{q}_i) = \sum_{\boldsymbol{m}} \left(\prod_{j=1}^{n_i} \frac{(-q_{ij})^{m_j}}{m_j!} \right) e^{\boldsymbol{\mu}_i'\boldsymbol{\lambda}_m + \frac{1}{2}\boldsymbol{\lambda}_m' V_i \boldsymbol{\lambda}_m}, \tag{37}$$

$$f(\boldsymbol{y}_i|\boldsymbol{q}_i) = \frac{\prod_{j=1}^{n_i} e^{\gamma_j y_{ij}} e^{-q_{ij}e^{\gamma_j y_{ij}}} e^{-\boldsymbol{\mu}_i'\boldsymbol{\lambda}_m - \frac{1}{2}\left[(\boldsymbol{y}_i - \boldsymbol{\mu}_i)' V_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\lambda}_m' V_i \boldsymbol{\lambda}_m)\right]}}{(2\pi)^{n_i/2} |V_i|^{1/2} \sum_{\boldsymbol{m}} \left(\prod_{j=1}^{n_i} \frac{(-q_{ij})^{m_j}}{m_j!} \right)}, \tag{38}$$

$$\widehat{q_{ij}} = e^{-\gamma_j y_{ij}}, \tag{39}$$

$$\widehat{\boldsymbol{y}_i} = \frac{\sum_{\boldsymbol{m}} \left[\prod_{j=1}^{n_i} \frac{(-e^{-\gamma_j y_{ij}})^{m_j}}{m_j!} \right] e^{\boldsymbol{\mu}_i'\boldsymbol{\lambda}_m + \frac{1}{2}\boldsymbol{\lambda}_m' V_i \boldsymbol{\lambda}_m} (\boldsymbol{\mu}_i + V_i \boldsymbol{\lambda}_m)}{\sum_{\boldsymbol{m}} \left[\prod_{j=1}^{n_i} \frac{(-e^{-\gamma_j y_{ij}})^{m_j}}{m_j!} \right] e^{\boldsymbol{\mu}_i'\boldsymbol{\lambda}_m + \frac{1}{2}\boldsymbol{\lambda}_m' V_i \boldsymbol{\lambda}_m}}, \tag{40}$$

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

6.2 Exchangeable Data With Compound-symmetry Covariance

Let us now consider the special but enlightening case of exchangeable, compound-symmetry data, in the sense that all members of a cluster have the same mean μ_i and the variance-covariance matrix

is of a compound symmetry structure $V_i = \sigma^2 I_{n_i} + dJ_{n_i}$, where I_{n_i} is an n_i-dimensional identity matrix and J_{n_i} is an $n_i \times n_i$ matrix consisting of ones. We will simply refer to this setting as the "exchangeable" one.

Let us, for each of the three model formulations in Section 6, present the six model equations considered there, for the special case of interest here.

6.2.1 The Standard Linear Mixed-effects Model

Let $\mathbf{1}_{n_i}$ be a length n_i vector of ones and denote by $\overline{y_i}$ the average of the components of the outcome vector $\boldsymbol{y}_i$. Further, the following expressions are useful:

$$V_i^{-1} = \frac{1}{\sigma^2}\left(I_{n_i} - \frac{d}{dn_i + \sigma^2} J_{n_i}\right), \qquad |V_i| = \sigma^{2n_i} + n_i \sigma^{2(n_i-1)} d.$$

The exchangeable versions of (23)–(27) are:

$$\begin{aligned}
\boldsymbol{Y}_i | b_i &\sim N(\mathbf{1}_{n_i}\mu_i + \mathbf{1}_{n_i} b_i, \sigma^2 I_{n_i}), && (41)\\
b_i &\sim N(0, d), && (42)\\
\boldsymbol{Y}_i &\sim N(\mathbf{1}_{n_i}\mu_i, V_i = \sigma^2 I_{n_i} + dJ_{n_i}), && (43)\\
b_i | \boldsymbol{Y}_i &\sim N\left[\frac{n_i d}{\sigma^2 + n_i d}(\overline{y_i} - \mu_i), \frac{\sigma^2}{\sigma^2 + n_i d} d\right], && (44)\\
\widehat{b_i} &= \frac{n_i d}{\sigma^2 + n_i d}(\overline{y_i} - \mu_i), && (45)\\
\widehat{\boldsymbol{Y}_i} &= \frac{n_i d \overline{y_i} + \sigma^2 \mu_i}{\sigma^2 + n_i d} \cdot \mathbf{1}_{n_i}. && (46)
\end{aligned}$$

6.2.2 A First Normal-exponential Version of the Linear Mixed Model

It now makes sense to assume, like in Section 6.1.2, that there is a single, exponentially distributed, random effect. This alters the model from Section 6.2.1 a bit, in addition to obvious simplification. This means that (43) will be coupled with

$$f(g_i | \boldsymbol{y}_i) = \gamma \overline{y_i} e^{-g_i \gamma \overline{y_i}}. \qquad (47)$$

We obtain the following sequence of model equations:

$$f(g_i) = \gamma e^{-g_i\mu_i\gamma+\frac{1}{2}\frac{g_i^2\gamma^2}{n_i}(\sigma^2+n_id)}\left[\frac{n_i\mu_i - g_i\gamma(\sigma^2+n_id)}{n_i}\right], \tag{48}$$

$$f(\boldsymbol{y}_i|g_i) = \frac{n_i\overline{y_i}e^{-\frac{1}{2}\left[\frac{1}{\sigma^2}(\boldsymbol{y}_i-\mathbf{1}_{n_i}\overline{y_i})'(\boldsymbol{y}_i-\mathbf{1}_{n_i}\overline{y_i})+\frac{n_i}{\sigma^2+n_id}(\overline{y_i}-\mu_i)^2\right]-g_i\gamma(\overline{y_i}-\mu_i)}}{(2\pi)^{n_i/2}|V_i|^{1/2}e^{\frac{1}{2}\frac{g_i^2\gamma^2}{n_i}(\sigma^2+n_id)}[n_i\mu_i - g_i\gamma(\sigma^2+n_id)]} \tag{49}$$

$$\widehat{g_i} = 1/(\gamma\overline{y_i}), \tag{50}$$

$$\widehat{\boldsymbol{y}_i} = \frac{\left\{\left[n_i\mu_i - \frac{1}{\overline{y_i}}(\sigma^2+n_id)\right]^2 + n_i(\sigma^2+n_id)\right\}\mathbf{1}_{n_i}}{n_i\left[n_i\mu_i - \frac{1}{\overline{y_i}}(\sigma^2+n_id)\right]}. \tag{51}$$

6.2.3 A Second Normal-exponential Version of the Linear Mixed Model

Now, (42) will be coupled with

$$f(q_i|\boldsymbol{y}_i) = e^{\gamma\overline{y_i}}e^{-q_ie^{\gamma\overline{y_i}}}. \tag{52}$$

This then produces the following sequence of model equations:

$$f(q_i) = \sum_{m=0}^{\infty}\frac{(-q_i)^m}{m!}e^{\mu_i\gamma(m+1)+\frac{1}{2}\frac{\gamma^2(m+1)^2}{n_i}(\sigma^2+n_id)}, \tag{53}$$

$$f(\boldsymbol{y}_i|q_i) = \frac{e^{-\frac{1}{2}\left[\frac{1}{\sigma^2}(\boldsymbol{y}_i-\mathbf{1}_{n_i}\overline{y_i})'(\boldsymbol{y}_i-\mathbf{1}_{n_i}\overline{y_i})+\frac{n_i}{\sigma^2+n_id}(\overline{y_i}-\mu_i)^2\right]+\gamma\overline{y_i}-q_ie^{\gamma\overline{y_i}}}}{(2\pi)^{n_i/2}|V_i|^{1/2}\sum_{m=0}^{\infty}\frac{(-q_i)^m}{m!}e^{\mu_i\gamma(m+1)+\frac{1}{2}\frac{\gamma^2(m+1)^2}{n_i}(\sigma^2+n_id)}}, \tag{54}$$

$$\widehat{q_i} = e^{-\gamma\overline{y_i}}, \tag{55}$$

$$\widehat{\boldsymbol{y}_i} = \frac{\sum_{m=0}^{\infty}\frac{(e^{-\gamma\overline{y_i}})^m}{m!}e^{\mu_i\gamma(m+1)+\frac{1}{2}\frac{\gamma^2(m+1)^2}{n_i}(\sigma^2+n_id)}\left[\mu_i+\frac{\gamma(m+1)}{n_i}(\sigma^2+n_id)\right]\mathbf{1}_{n_i}}{\sum_{m=0}^{\infty}\frac{(e^{-\gamma\overline{y_i}})^m}{m!}e^{\mu_i\gamma(m+1)+\frac{1}{2}\frac{\gamma^2(m+1)^2}{n_i}(\sigma^2+n_id)}}. \tag{56}$$

7 Data Analysis

7.1 Analysis of the Toenail Data

7.1.1 Focus on Random Effects

For the unaffected nail length, let us specify a linear mixed-effects model (22)–(23):

$$Y_{ij}|(b_{i0},b_{i1}) \sim N(\beta_0+b_{i0}+(\beta_1+b_{i1})t_j+\beta_2T_i+\beta_3T_it_j,\sigma^2), \tag{57}$$

$$\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_{00} & d_{01} \\ d_{10} & d_{11} \end{pmatrix}\right], \tag{58}$$

where $T_i = 0$ if patient i received standard treatment and 1 for experimental therapy ($i = 1, \ldots, 298$). Further, t_j is the time at which the jth measurement is taken ($j = 1, \ldots, 7$). Parameter estimates and standard errors, obtained through maximum likelihood (Verbeke and Molenberghs, 2000), are presented in Table 3.

We are now able to partially replace the model specified by (57)–(58) with the exponential-defined models. Let us choose, for illustration, the second exponential model of Section 6.1.3. This implies that the marginal model resulting from (57)–(58) is retained:

$$\boldsymbol{Y}_i \sim N[X_i(\beta_0, \beta_1, \beta_2, \beta_3)', \sigma^2 I_{n_i} + Z_i' D Z_i], \tag{59}$$

and coupled with (36). 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 second 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 (36), we have the freedom of specifying the parameters γ_j, because there is no information contained in the data. Indeed, they can be identified by additional assumptions only; they play the role of sensitivity parameters. We set them equal to $\gamma_j = 0.05$. Figure 2 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.

Of course, in theory, one could estimate the parameters γ_j, but the whole point here is that one can freely vary the parameters specific to the posterior distribution of the random effects, without affecting the marginal fit, i.e., without affecting what is verifiable directly from the data.

7.1.2 Focus on Missingness

Consider a general model of the form (8), with random effects confined to $\boldsymbol{g}_i$, i.e., common to all three components. For the measurement model, assume a linear mixed model (Verbeke and

Molenberghs, 2000), with general form:

$$\boldsymbol{Y}_i|\boldsymbol{g}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{g}_i, \Sigma_i), \tag{60}$$

$$\boldsymbol{g}_i \sim N(0, D). \tag{61}$$

Based on (60) and (61), the so-called marginal model can be derived

$$\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i). \tag{62}$$

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 (60) as

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

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

$$\begin{aligned}\boldsymbol{Y}_i^m|\boldsymbol{y}_i^o, \boldsymbol{g}_i \sim N\Big[&(X_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}X_i^o)\boldsymbol{\beta} + \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\boldsymbol{y}_i^o + (Z_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}Z_i^o)\boldsymbol{g}_i, \\ &\Sigma_i^{mm} - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\Sigma_i^{om}\Big].\end{aligned} \tag{63}$$

Now, (63) 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 (62), now applied to (63), we obtain:

$$\begin{aligned}\boldsymbol{Y}_i^m|\boldsymbol{y}_i^o \sim N\Big[&(X_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}X_i^o)\boldsymbol{\beta} + \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\boldsymbol{y}_i^o, \\ &(Z_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}Z_i^o)D(Z_i^m - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}Z_i^o)' \\ &+\Sigma_i^{mm} - \Sigma_i^{mo}\{\Sigma_i^{oo}\}^{-1}\Sigma_i^{om}\Big].\end{aligned} \tag{64}$$

Hence, (64) is the MAR counterpart to (63). For the unaffected nail length, we choose for (60)–(61):

$$E(Y_{ij}|g_i, T_i, t_j, \boldsymbol{\beta}) = \beta_0 + g_i + \beta_1 T_i + \beta_2 t_j + \beta_3 T_i t_j, \tag{65}$$

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

Given these choices, (63) and (64) simplify to

$$\boldsymbol{Y}_i^m|\boldsymbol{y}_i^o, g_i \sim N(X_i\boldsymbol{\beta} + Z_i^m g_i, \sigma^2 I_i), \tag{66}$$

$$\boldsymbol{Y}_i^m|\boldsymbol{y}_i^o \sim N(X_i\boldsymbol{\beta}, dJ_i + \sigma^2 I_i), \tag{67}$$

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

Next, let us formulate a model for the missingness mechanism in (8). The sequence $\boldsymbol{r}_i$ can take one of two forms in our case. Either, it is a length-7 vector of ones, for a completely observed subject, or it is a sequence of k ones followed by a sole zero $1 \leq k \leq 6$, for someone dropping out. Note that k is 1 at least, since for everyone the initial measurement has been observed. It is convenient to assume a logistic regression of the form:

$$\text{logit}\left[P(R_{ij} = 1|R_{i,j-1} = 0, g_i, T_i, t_j, \boldsymbol{\gamma})\right] = \gamma_0 + \gamma_{01} g_i + \gamma_1 T_i + \gamma_2 t_j + \gamma_3 T_i t_j, \tag{68}$$

($j > 1$), 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 (65) and (68) can easily be fitted using, for example, the SAS procedure NLMIXED.

Parameter estimates and standard errors are displayed in Table 4. It is noteworthy that the scale factor γ_{01} is estimated to be negative, even though it is not significant. While we should not overly stress its importance, there is some indication that a higher subject-specific profile of unaffected nail length corresponds with a lower dropout probability, which is not surprising. The magnitude of the scale factor allows us to 'translate' the subject-specific effect from the continuous outcome scale, expressed in mm, to the unitless logit scale on which the probability of missingness is described. Note that the random-intercept variance is highly significant among unaffected nail length outcomes; the same is not true for the dropout model, with $p = 0.2487$, using a $50 : 50$ mixture of a χ_0^2 and χ_1^2 distribution (Verbeke and Molenberghs, 2000).

Figure 3 displays the incomplete profiles, extended beyond the time of dropout, using prediction based on: (1) the original model (dashed lines); (2) the MAR counterpart (solid lines). Within each of the treatment arms, three profiles are highlighted. The MAR counterpart reduces all predictions to the same profile, whereas the MNAR model predicts different evolutions for different subjects, implied by the presence of the random effect. The simple MAR-based prediction structure follows directly from the conditional independence assumption, present in (66). When deemed less plausible, the fully general structure (63) can be implemented.

7.2 Analysis of the Developmental Toxicity Study

Let us consider the following hierarchically specified, exchangeable model for the DEHP data, introduced in Section 2.2:

$$Y_{ij}|b_i \sim N(\beta_0 + b_i + \beta_1 x_i, \sigma^2), \tag{69}$$

coupled with (42). Here x_i is rescaled dose, in the sense that the DEHP consumption doses of 0, 44, 91, 191, and 292 mg/kg/day are replaced by unit-interval standardized values 0.0000, 0.1507, 0.3116, 0.6541, and 1.0000, respectively. Parameter estimates and standard errors are presented in Table 5.

Following the developments in Section 6.2, Model (69)–(42), can be replaced by, for example, the models with exponential posterior distributions, described in Sections 6.1.2 and 6.1.3, respectively. This implies that the marginal model is retained, with

$$Y_{ij} \sim N(\beta_0 + \beta_1 x_i, \sigma^2 + d), \tag{70}$$

but with alternative posterior distributions, and hence EB estimates for the random effects and predictions, as presented by (55) and (56), respectively. The results are graphically depicted in Figure 4. For 11 selected clusters, spread over the various dose groups, the figure shows (1): observed average weight per cluster (2): the estimated marginal mean as given by (70); (3), (4), and (5): predictions following the normal, first, and second exponential models, respectively. We observe that, in line with the analysis of the toenail data, the exponential predictions lie closer to the marginal averages than is the case with the normal model.

8 Concluding Remarks

In this paper, we have unified the frameworks of coarse and augmented data. This brings together settings where the observed data are a less refined version of what might ideally be observed (incomplete, censored, and grouped data among others) and situations where the modeler posits fictitious structures (random effects, latent variables, latent classes, mixture components, etc.), for convenience and simplicity, for interpretation, because it is plausible from a substantive point of view, or a combination of these and other reasons. We have referred to this broad family of settings as *enriched data*.

The use of bringing out the commonality of otherwise quite dispersed models, designs, and data structures is that it allows to study, in a unified fashion, the impact of such structures, which extend beyond the available data. As we made clear, depending on the particular situation, there has been knowledge about such impacts, in particular also on ensuing inferences. For example, work has been done for decades in the incomplete data field to assess the sensitivity of, for example, deviations from the MAR assumption. Here, we reviewed and brought together results for the incomplete-data field. Precisely, we first defined MAR in all three commonly used modeling frameworks: selection models, pattern-mixture models, and shared-parameter models. We then showed that every posited MNAR model, regardless of the framework, can be paired with an MAR model that exhibits exactly the same fit to the observed data as the original, making them indistinguishable in terms of observed data alone. This work is based on results by Molenberghs *et al* (1998), Beunckens *et al* (2007b), and Creemers *et al* (2008).

Beyond the mere incomplete data setting, we have shown that every model for enriched-data settings can be factored as a product of two components: the first one, termed the marginal model, fully identifiable from the observed data; the second one, the conditional distribution of the enriched given the observed data, entirely arbitrary. In the missing data case, one could identify the second factor by requiring, for example, that it is of the MAR type. In the context of a conventional linear mixed model, we have illustrated the implications of the result by replacing the conditional distribution of the random effects given the data, i.e., the random effects' posterior, by two families of exponential distributions, special cases of the gamma family. The choice for exponential random effects was

made only for computational and illustrative convenience, rather than as a viable model for practice. Rather to the contrary, the developments underscore arbitrariness of part of the model's specification. For the specific case of exchangeable, compound-symmetric data, specific, useful expressions were derived.

While our results are predominantly of a conceptual nature, they have been illustrated, for enhanced insight, using longitudinal data from both a two-armed clinical trial in toenail dermatophyte onychomycosis and clustered data from a developmental toxicity study in mice. In the first case, a conventional linear mixed model was used as the basic model, whereas in the second case the specific clustered compound-symmetry version was considered. In both situations, the standard models were supplemented with the normal-exponential model. Evidently, similar illustrations could be given for all enriched-data settings. The specific case of incomplete data illustrated in Molenberghs *et al* (2008) and Creemers *et al* (2008).

The results of this paper open avenues for (1) better understanding of the dependence of one's inferences on non-verifiable model components and (2) developing sensitivity analysis tools regarding substantive conclusions with respect to data enrichment (Molenberghs and Kenward, 2007). Generally speaking, inferences relative to observed data only, such as fixed-effects parameter estimates (e.g., those in Tables 3–5, are unaffected by the choice of enrichment model. However, such aspects as empirical Bayes predictions in linear mixed models, or predictive distributions of unobserved measurements given observed ones, strongly rest on unverifiable modeling assumptions. This points to the need for sensitivity analysis. Rather than fitting a single model and putting blind belief in it, it is more reasonable to consider a discrete or continuous set of alternative model formulations, and assess how key inferences are vulnerable to choices made.

Without being exhaustive, we can point out several avenues for sensitivity analysis, some of which come from the incomplete-data literature (Molenberghs and Kenward, 2007). First, because it is clear in general for which component of the enriched-data model there is no information in the data, one could apply a number of operations to the posterior distribution, i.e., the model for the enriched data given the observed data. For example, one could let the posterior vary over a finite or infinite number of reasonable choices for the posterior, similar to the *interval of ignorance* ideas

of Molenberghs, Kenward, and Goetghebeur (2001) and Vansteelandt *et al* (2006). Second, one could identify the posterior using unverifiable, i.e., not driven by the data, but nevertheless explicit assumptions. One such route is in choosing the posterior of the MAR type in the incomplete-data setting. For different enriched-data settings, it has to be judged what is a reasonable identification. The advantage of making explicit such assumptions is that they then can be critiqued on substantive and/or mathematical grounds. For example, with random effects or other latent structures, one could express a preference for conjugate priors (Lee and Nelder, 1996, 2001, 2003) since, in the absence of identification, the convenience and appeal of conjugacy may be invoked. Third, one could focus on the impact of one or a few influential subjects on model-based conclusions, especially these pertaining to (almost) unidentified parameters.

Data sets and programs are available from the authors and through the journal's web pages.

Acknowledgment

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

References

Autian, J. (1973). Toxicity and health threats of phthalate esters: Review of the literature. *Environmental Health Perspectives* **4**, 3–26.

Beunckens, C., Sotto, C., Molenberghs, G., and Verbeke, G. (2007). 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.

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.

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

Heitjan, D.F. and Rubin, D.B. (1991). Ignorability and coarse data. *The Annals of Statistics*, **19**, 2244–2253.

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

Lee, Y. and Nelder, J.A. (1996) Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B* **58**, 619–678.

Lee, Y. and Nelder, J.A. (2001) Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions. *Biometrika*, **88**, 987–1006.

Lee, Y., and Nelder, J.A. (2003) Extended-REML estimators. *Journal of Applied Statistics*, **30**, 845–856.

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. (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.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M.G. (2008). Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B* **70**, 371–388

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

Molenberghs, G., Kenward, M.G., and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Applied Statistics* **50**, 15–29.

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

Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician* **61**, 1–6.

Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., and Buyse, M. (2002). Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. *Biometrical Journal* **44**, 921–935.

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.

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

Searle, S.R., Casella, G., and McCulloch, C.E. (1996). *Variance Components*. New York: John Wiley & Sons.

Shiota, K., Chou, M.J., and Nismimura, H. (1980). Embryotoxic effects of di-2-ethylhexyl phthalate (DEHP) and di-n-butyl phthalate (DBP) in mice. *Environmental Research* **22**, 245–253.

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.

Tyl, R.W., Price, C.J., Marr, M.C., and Kimmel, C.A. (1988). Developmental toxicity evaluation of dietary di(2-ethylhexyl)phthalate in Fischer 344 rats and CD-1 mice. *Fundamental and Applied Toxicology* **10**, 395–412.

Vansteelandt, S., Goetghebeur, E., Kenward, M.G., and Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, **16**, 953–979.

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.

Willink, R. (2005). Normal moments and Hermite polynomials. *Statistics & Probability Letters* **73**, 271–275.

Windholz, M. (1983). *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals (10th ed.)* Rahway, NJ: Merck and Co.

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.

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

	Group A		Group B	
# Obs.	N	%	N	%
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%

Table 2: *Developmental Toxicity Study (DEHP). Summary data by dose group.*

	# dams with		# live	average	
dose	implants	viable implants	fetuses	litter size	weight
0 mg/kg/day	30	30	330	13.2	0.9483
44 mg/kg/day	26	26	288	11.1	0.9592
91 mg/kg/day	26	26	277	10.7	0.8977
191 mg/kg/day	24	17	137	8.1	0.8509
292 mg/kg/day	25	9	50	5.6	0.6906

Table 3: *Toenail Data. (Unaffected nail length outcome). Parameter estimates (standard errors) for the model specified by (57) and (58).*

Effect	Parameter	Estimate	(Standard error)
Fixed effects:			
Intercept	β_0	2.46	(0.24)
Dose effect	β_1	0.59	(0.05)
Time effect	β_2	0.28	(0.34)
Dose by time interaction	β_3	0.04	(0.06)
Variance components:			
Random intercept variance	d_{00}	7.32	(0.70)
Random slope variance	d_{11}	0.22	(0.02)
Random effects covariance	d_{01}	-0.50	(0.10)
Residual variance	σ^2	3.15	(0.13)

Table 4: *Toenail Data. Continuous, longitudinal unaffected-nail-length outcome. Parameter estimates (standard errors) for the model specified by (65) and (68).*

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

Table 5: *Developmental Toxicity Study (DEHP). Parameter estimates (standard errors) for the model specified by (69) and (42).*

Effect	Parameter	Estimate	(Standard error)
Fixed effects:			
Intercept	β_0	0.9733	(0.0138)
Dose effect	β_1	-0.2563	(0.0327)
Variance components:			
Random intercept variance	d	0.0086	(0.0015)
Residual variance	σ^2	0.0195	(0.0009)

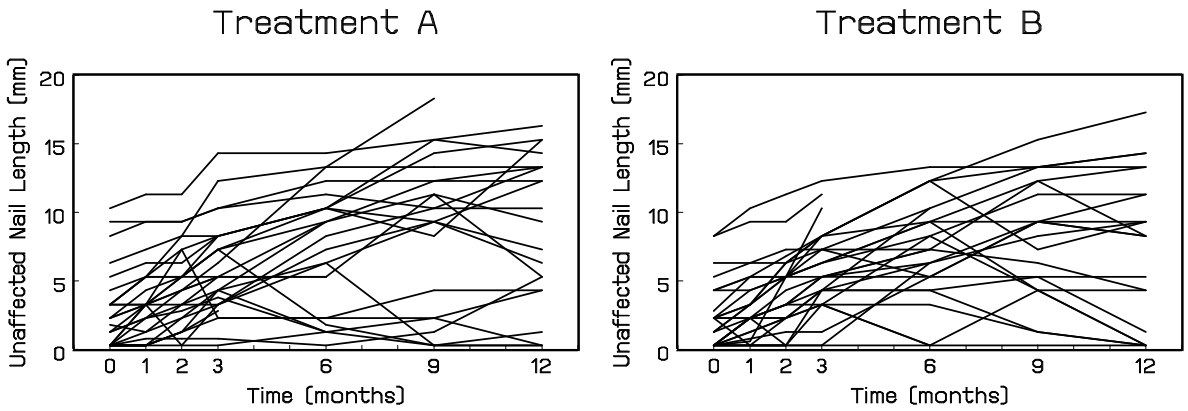

Figure 1: *Toenail Data. Individual profiles of 30 randomly selected subjects in each of the treatment groups in the toenail experiment.*

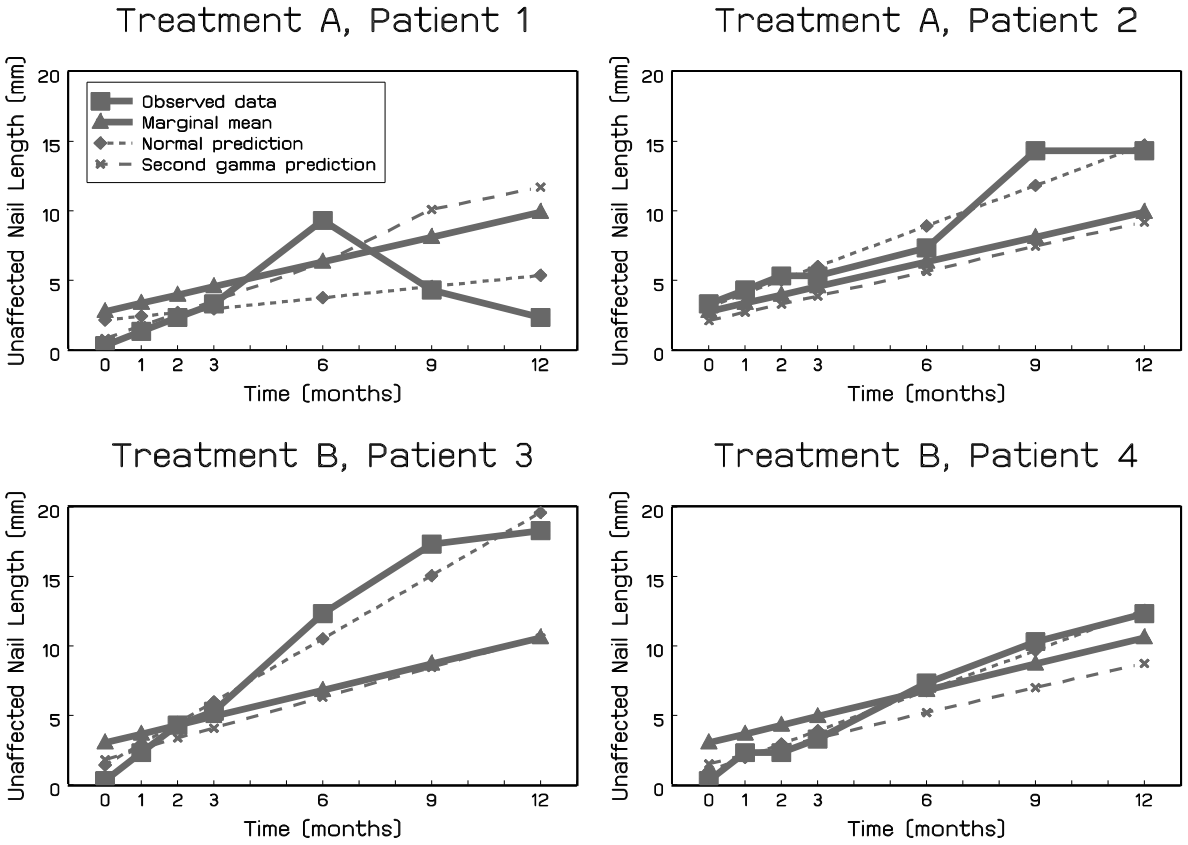

Figure 2: *Toenail data. For 4 selected subjects, two per treatment arm: (1): observed profile; (2) marginal mean profile (which solely depends on treatment); (3) prediction from the normal model (27); (4) prediction from the second exponential model (40).*

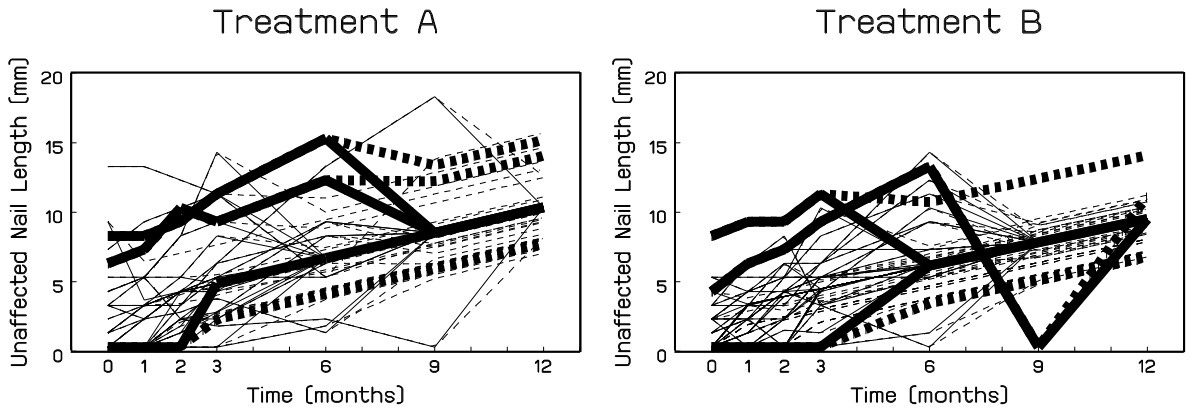

Figure 3: *Toenail Data. Individual profiles of subjects with incomplete data, for each treatment arm, extended using MNAR Model (65) (dashed line) and using the model's MAR counterpart (solid line). In each group, three subjects are highlighted.*

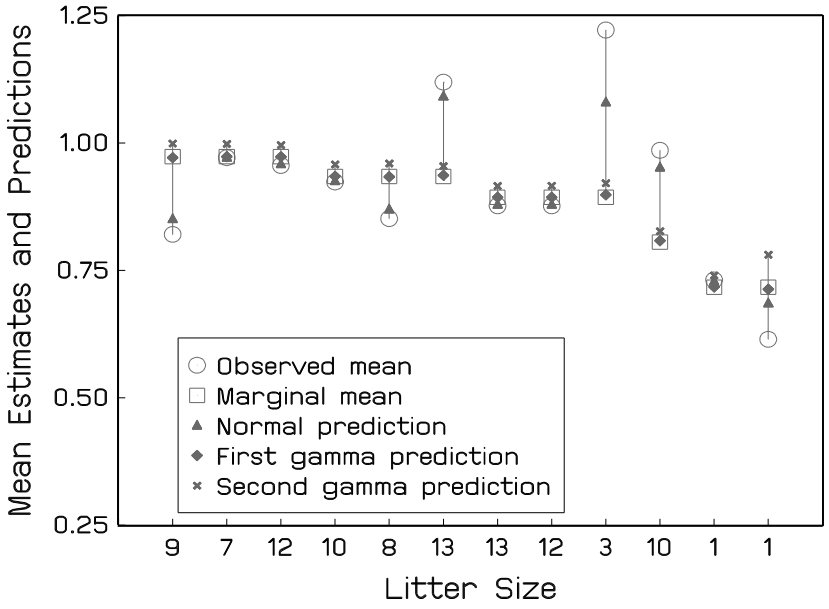

Figure 4: *Developmental Toxicity Study (DEHP). For 12 selected clusters from the control group (for which the size is shown in the x-axis): (1): observed average weight per cluster (2): the estimated marginal mean as given by (70); (3) prediction from the normal model (46); (4) prediction from the first exponential model (51); and (5): prediction from the second exponential model (56).*