A Note on MAR, Identifying Restrictions, Model Comparison, and Sensitivity Analysis in Pattern Mixture Models With and Without Covariates for Incomplete Data

Chenguang Wang

Division of Biostatistics, Center for Devices and Radiological Health, FDA, Silver Spring, Maryland 20993

email: chenguang.wang@fda.hhs.gov

and

Michael J. Daniels

Department of Statistics,University of Florida, Gainesville, FL 32611

email: mdaniels@stat.ufl.edu

SUMMARY: Pattern mixture modeling is a popular approach for handling incomplete longitudinal data. Such models are not identifiable by construction. Identifying restrictions are one approach to mixture model identification (Little, 1995; Little and Wang, 1996; Thijs et al., 2002; Kenward et al., 2003; Daniels and Hogan, 2008) and are a natural starting point for missing not at random sensitivity analysis (Thijs et al., 2002; Daniels and Hogan, 2008). However, when the pattern specific models are multivariate normal, identifying restrictions corresponding to missing at random may not exist. Furthermore, identification strategies can be problematic in models with covariates (e.g. baseline covariates with time-invariant coefficients). In this paper, we explore conditions necessary for identifying restrictions that result in missing at random (MAR) to exist under a multivariate normality assumption and strategies for identifying sensitivity parameters for sensitivity analysis or for a fully Bayesian analysis with informative priors. In addition, we propose alternative modeling and sensitivity analysis strategies under a less restrictive assumption for the distribution of the observed response data. We adopt the deviance information criterion for model comparison and perform a simulation study to evaluate the performances of the different modeling approaches. We also apply the methods to a longitudinal clinical trial. Problems caused by baseline covariates

with time-invariant coefficients are investigated and an alternative identifying restriction based on residuals is proposed as a solution.

KEY WORDS: Missing at random; Non-future dependence; Deviance information criterion.

1. Introduction

For analyzing longitudinal studies with informative missingness, popular modeling frameworks include pattern mixture models, selection models and shared parameter models, which differ in the way the joint distribution of the outcome and missing data process are factorized (for a comprehensive review, see Little, 1995; Hogan and Laird, 1997; Kenward and Molenberghs, 1999; Molenberghs and Kenward, 2007; Daniels and Hogan, 2008). In this paper, we concern ourselves with pattern mixture models with monotone missingness (i.e., drop-out). For pattern mixture models with non-monotone (i.e., intermittent) missingness (details go beyond the scope of this paper), one approach is to partition the missing data and allow one (or more) or the partitions to be ignored given the other partition(s) (Harel and Schafer, 2009; Wang et al., 2010).

  It is well known that pattern-mixture models are not identified: the observed data does not provide enough information to identify the distributions for incomplete patterns. The use of identifying restrictions that equate the inestimable parameters to functions of estimable parameters is an approach to resolve the problem (Little, 1995; Little and Wang, 1996; Thijs et al., 2002; Kenward et al., 2003; Daniels and Hogan, 2008). Common identifying restrictions include complete case missing value (CCMV) constraints and available case missing value (ACMV) constraints. Molenberghs et al. (1998) proved that for discrete time points and monotone missingness, the ACMV constraint is equivalent to missing at random (MAR), as defined by Rubin (1976) and Little and Rubin (1987). A key and attractive feature of identifying restrictions is that they do not impact the fit of the model to the observed data. Understanding (identifying) restrictions that lead to MAR is an important first step for sensitivity analysis under missing not at random (MNAR) (Scharfstein et al., 2003; Zhang and Heitjan, 2006; Daniels and Hogan, 2008). In particular, MAR provides a good starting

point for sensitivity analyses and sensitivity analyses are essential for inference on incomplete data (Scharfstein et al., 1999; Vansteelandt et al., 2006; Daniels and Hogan, 2008).

The normality of response data (if appropriate) for pattern mixture models is desirable as it easily allows incorporation of baseline covariates and introduction of sensitivity parameters (for MNAR analysis) that have convenient interpretations as deviations of means and variances from MAR (Daniels and Hogan, 2008). However, multivariate normality within patterns can be overly restrictive. We explore such issues in this paper.

One criticism of mixture models is that they often induce missing data mechanisms that depend on the future (Kenward et al., 2003). We explore such non-future dependence in our context here and show how mixture models that have such missing data mechanisms have fewer sensitivity parameters.

In Section 2, we show conditions under which MAR exists and does not exist when the full-data response is assumed multivariate normal within each missing pattern. In Section 3 and Section 4 in the same setting, we explore sensitivity analysis strategies under MNAR and under non-future dependent MNAR respectively. In Section 5, we propose a sensitivity analysis approach where only the *observed* data within pattern are assumed multivariate normal. In Section 6, we apply the frameworks described in previous sections to a randomized clinical trial for estimating the effectiveness of recombinant growth hormone for increasing muscle strength in the elderly and propose a criterion to compare the fit of different models to the observed data. The behavior of this criterion and the model frameworks proposed in Sections **??**-5 are assessed by simulation in Section 7. In Section 8, we show that in the presence of baseline covariates with time-invariant coefficients, standard identifying restrictions cause over-identification of the baseline covariate effects and we propose a remedy. We provide conclusions and discussion in Section 9.

2. Existence of MAR under Multivariate Normality within Pattern

Let $\boldsymbol{Y}$ be a J-dimensional longitudinal response vector with components scheduled to be measured at time points t_j $(j \in \{1, \ldots, J\})$; this is the full data response. Without loss of generality, we assume Y_1 is always observed. Let $S = s$ denote the number of observed responses $(s = 1, 2, \ldots, J)$ corresponding to the follow up time t_s. Let $\overline{Y}_j$ denote the historical response vector $(Y_1, Y_2, \ldots, Y_j)$. Finally, we define $p_s(\cdot) = p(\cdot|S = s)$.

We show that MAR does not necessarily exist when it is assumed that

$$\boldsymbol{Y}|S = s \sim \mathrm{N}(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)}) \text{ for all } s. \tag{1}$$

To see this, we introduce some further notation. Let

$$\boldsymbol{\mu}^{(s)}(j) = \mathrm{E}(\overline{Y}_j|S = s) = \begin{bmatrix} \boldsymbol{\mu}_1^{(s)}(j) \\ \mu_2^{(s)}(j) \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}^{(s)}(j) = \mathrm{Var}(\overline{Y}_j|S = s) = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{(s)}(j) & \boldsymbol{\Sigma}_{12}^{(s)}(j) \\ \boldsymbol{\Sigma}_{21}^{(s)}(j) & \Sigma_{22}^{(s)}(j) \end{bmatrix}$$

where $\boldsymbol{\mu}_1^{(s)}(j) = \mathrm{E}(\overline{Y}_{j-1}|S = s)$, $\mu_2^{(s)}(j) = \mathrm{E}(Y_j|S = s)$, $\boldsymbol{\Sigma}_{11}^{(s)}(j) = \mathrm{Var}(\overline{Y}_{j-1}|S = s)$, $\Sigma_{22}^{(s)}(j) = \mathrm{Var}(Y_j|S = s)$, $\boldsymbol{\Sigma}_{12}^{(s)}(j) = \mathrm{Cov}(\overline{Y}_{j-1}, Y_j|S = s)$ and $\boldsymbol{\Sigma}_{21}^{(s)}(j)$ is the transpose of $\boldsymbol{\Sigma}_{12}^{(s)}(j)$.

LEMMA 1: *For monotone dropout, under the model given in (1), define*

$$\kappa_1^{(s)}(j) = \boldsymbol{\Sigma}_{21}^{(s)}(j) \left(\Sigma_{11}^{(s)}(j)\right)^{-1}$$

$$\kappa_2^{(s)}(j) = \mu_2^{(s)}(j) - \kappa_1^{(s)}(j)\boldsymbol{\mu}_1^{(s)}(j)$$

$$\kappa_3^{(s)}(j) = \Sigma_{22}^{(s)}(j) - \boldsymbol{\Sigma}_{21}^{(s)}(j) \left(\boldsymbol{\Sigma}_{11}^{(s)}(j)\right)^{-1} \boldsymbol{\Sigma}_{12}^{(s)}(j).$$

The condition that for a given j, the conditional distributions $p_s(y_j|\overline{y}_{j-1})$ are identical for all s is equivalent to $\kappa_1^{(s)}(j)$, $\kappa_2^{(s)}(j)$ and $\kappa_3^{(s)}(j)$ being constant in s.

Proof. The proof is trivial since

$$Y_j|\overline{Y}_{j-1}, S = s \sim \mathrm{N}\left(\kappa_2^{(s)}(j) - \kappa_1^{(s)}(j)\overline{Y}_{j-1}, \kappa_3^{(s)}(j)\right).$$

In other words, if the condition in Lemma 1 is satisfied, then there exists a conditional distribution $p_{\geq s}(y_j | \overline{y}_{j-1})$ such that $p_s(y_j | \overline{y}_{j-1}) = p_{\geq s}(y_j | \overline{y}_{j-1})$ for all $s \geq j$. We now state a theorem that gives the restrictions on the model given in (1) for MAR to exist. Note that the proofs of rest of the theorems and corollaries in this and the subsequent section can be found in Web Appendix A.

THEOREM 1: *For pattern mixture models with monotone dropout, under the model given in (1), identification via MAR constraints exists if and only if $\boldsymbol{\mu}^{(s)}$ and $\boldsymbol{\Sigma}^{(s)}$ satisfy Lemma 1 for $s \geq j$ and $1 < j < J$.*

So, a default approach for continuous Y, assuming the full data response is multivariate normal within pattern, does not allow an MAR restriction (unless the restrictions in Theorem 1 are imposed).

We now examine the corresponding missing data mechanism (MDM), $S|\boldsymbol{Y}$. We use "$\simeq$" to denote equality in distribution.

COROLLARY 1: *For pattern mixture models of the form (1) with monotone dropout, MAR holds if and only if $S|\boldsymbol{Y} \simeq S|Y_1$.*

Thus, the implicit MDM is very restrictive and does not depend on the entire history, $\overline{Y}_s$. We now show connections to missing completely at random (MCAR) and other common identifying restrictions.

COROLLARY 2: *For pattern mixture models of the form (1) with monotone dropout, MCAR is equivalent to MAR if $p_s(y_1) = p(y_1)$ for all s.*

COROLLARY 3: *For pattern mixture models of the form (1) with monotone dropout,*

MAR constraints are identical to complete case missing value (CCMV) and nearest-neighbor constraints (NCMV).

The results in this section were all based on specifying the mixture model in (1) and demonstrate that MAR only exists under the fairly strict conditions given in Theorem 1.

3. Sequential Model Specification and Sensitivity Analysis under MAR

Due to the structure of $\boldsymbol{\mu}^{(s)}$ and $\boldsymbol{\Sigma}^{(s)}$ under MAR constraints as outlined in Section 2, we propose to follow the approach in Daniels and Hogan (2008, Chapter 8) and specify distributions of observed Y within pattern as:

$$
\begin{aligned}
p_s(y_1) &\sim \mathrm{N}(\mu_1^{(s)}, \sigma_1^{(s)}) & 1 \leq s \leq J \\
p_s(y_j|\overline{y}_{j-1}) &\sim \mathrm{N}(\mu_{j|j^-}^{(\geq j)}, \sigma_{j|j^-}^{(\geq j)}) & 2 \leq j \leq s \leq J
\end{aligned}
\tag{2}
$$

where $j^- = \{1, 2, \ldots, j-1\}$. Note that by construction, we assume $p_s(y_j|\overline{y}_{j-1})$ are identical for all $j \leq s \leq J$. Consequently, we have $p_s(y_j|\overline{y}_{j-1}) = p(y_j|\overline{y}_{j-1}, S \geq s)$, denoted as $p_{\geq s}(y_j|\overline{y}_{j-1})$.

COROLLARY 4: *For pattern mixture models of the form (1) with monotone dropout, identification via MAR constraints exists if and only if the observed data can be modeled as (2).*

Corollary 4 implies that under the multivariate normality assumption in (1) and the MAR assumption, a sequential specification as in (2) always exists.

We provide some details for MAR in model (1) (which implies the specification in (2) as stated in Corollary 4) next. Distributions for missing data (which are not identified) are specified as:

$$
p_s(y_j|\overline{y}_{j-1}) \sim \mathrm{N}(\mu_{j|j^-}^{(j)}, \sigma_{j|j^-}^{(j)}) \qquad 1 \leq s < j \leq J.
$$

The conditional mean structure of $\mu_{j|j^-}^{(\geq j)}$ and $\mu_{j|j^-}^{(j)}$ is parameterized as follows:

$$\mu_{j|j^-}^{(\geq j)} = \beta_0^{(\geq j)} + \sum_{l=1}^{j-1} \beta_l^{(\geq j)} y_l$$

$$\mu_{j|j^-}^{(j)} = \beta_0^{(j)} + \sum_{l=1}^{j-1} \beta_l^{(j)} y_l.$$

To identify the full-data model, the MAR constraints require that

$$p_k(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1})$$

for $k < j$, which implies that $\mu_{j|j^-}^{(j)} = \mu_{j|j^-}^{(\geq j)}$ and $\sigma_{j|j^-}^{(j)} = \sigma_{j|j^-}^{(\geq j)}$ for $2 \leq j \leq J$. Since the equality of the conditional means need to hold for all $\boldsymbol{Y}$, this further implies that the MAR assumption requires that $\beta_l^{(j)} = \beta_l^{(\geq j)}, 0 \leq l < j \leq J$.

The motivation of the proposed sequential model is to allow a straightforward extension of the MAR specification to a large class of MNAR models indexed by parameters measuring departures from MAR, as well as the attraction of doing sensitivity analysis on means and/or variances in normal models.

For example, we can let

$$\beta_l^{(j)} = \Delta_l^{(j)} + \beta_l^{(\geq j)} \quad \text{and} \quad \log \sigma_{j|j^-}^{(j)} = \Delta_\sigma^{(j)} + \log \sigma_{j|j^-}^{(\geq j)}$$

for all $j > 1$ and $0 \leq l < j$. Sensitivity analysis can be done on these Δ parameters that capture the information about the missing data mechanism (see Web Appendix B for the impact of the Δ parameters on the MDM). For example, in a Bayesian framework, we may assign informative priors elicited from experts to these sensitivity parameters Δ. Note in general we may have separate $\Delta_l^{(j)}$ and $\Delta_\sigma^{(j)}$ for each pattern s $(s \leq j)$, but in practice it is necessary to limit the dimensionality of these (Daniels and Hogan, 2008). Indeed, we could make $\Delta_l^{(j)}$ and $\Delta_\sigma^{(j)}$ independent of j to further reduce the number of sensitivity parameters.

In general the MDM depends on $\overline{Y}_J$, i.e. MNAR, in presence of the the Δ parameters. However, one might want hazard at time t_s to only depend on $\overline{Y}_{s+1}$, in which case we need

to have different distributions and assumptions on $[Y_j|\overline{Y}_{j-1}, S = k]$ for $k < j - 1$ and $j > 2$, as shown in the next section.

4. Non-future Dependence and Sensitivity Analysis under Multivariate Normality within Pattern

Non-future dependence assumes that missingness only depends on observed data and the current missing value, i.e.

$$[S = s|\boldsymbol{Y}] \simeq [S = s|\overline{Y}_{s+1}],$$

and can be viewed as a special case of MNAR and an extension of MAR (Kenward et al., 2003). Kenward et al. (2003) showed that non-future dependence holds if and only if for each $j \geq 3$ and $k < j - 1$,

$$p_k(y_j|\overline{y}_{j-1}) = p_{\geq j-1}(y_j|\overline{y}_{j-1}).$$

An approach to implement non-future dependence within the framework of Section 3 is as follows. We model the observed data as in (2). For the conditional distribution of the current missing data (Y_{s+1}), we assume that

$$p_s(y_{s+1}|\overline{y}_s) \sim \mathrm{N}\left(\beta_0^{(\geq s+1)} + \Delta_0^{(s+1)} + \sum_{l=1}^{s}(\beta_l^{(\geq s+1)} + \Delta_l^{(s+1)})y_l, e^{\Delta_\sigma^{(s+1)}}\sigma_{s|s^-}^{(\geq s+1)}\right) \quad 2 \leq s < J$$

and for the conditional distribution of the future missing data $(Y_{s+2}, \ldots, Y_J)$, we assume that

$$p_s(y_j|\overline{y}_{j-1}) = p_{\geq j-1}(y_j|\overline{y}_{j-1}) \quad 2 \leq s < j - 1 \leq J - 1,$$

where

$$p_{\geq j-1}(y_j|\overline{y}_{j-1}) = \frac{p(S = j - 1)}{p(S \geq j - 1)}p_{j-1}(y_j|\overline{y}_{j-1}) + \frac{p(S \geq j)}{p(S \geq j - 1)}p_{\geq j}(y_j|\overline{y}_{j-1}).$$

Note that by this approach, although the model for future missing data is a mixture of normals, the sensitivity parameters are kept the same as in Section 3 ($\Delta_l^{(j)}$ and $\Delta_\sigma^{(j)}$, $j = 2, \ldots, J$ and $l = 0, \ldots, j - 1$). In addition, this significantly reduces the number of potential sensitivity parameters. For J-dimensional longitudinal data, the total number of

sensitivity parameters, $(2J^3 + 3J^2 + J)/6 - J$ is reduced to $(J^2 + 3J - 4)/2$; for J=3 (6), from 11 (85) to 7 (25). Further reduction is typically needed. See the data example in Section 6 as an illustration. If all of the remaining sensitivity parameters are set to zero, we have $p_s(y_{s+1}|\overline{y}_s) = p_{\geq s+1}(y_{s+1}|\overline{y}_s)$ for $2 \leq s < J$ and $p_s(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1})$ for $2 \leq s < j - 1 \leq J - 1$, which implies $p_s(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1})$ for all $s < j$, i.e. MAR.

5. MAR and Sensitivity Analysis with Multivariate Normality on the Observed-data Response

If we assume multivariate normality only on observed data response, $\boldsymbol{Y}_{\text{obs}}|S$ instead of the full data response, $\boldsymbol{Y}|S$, we can weaken the restrictions on $p_s(y_j|\overline{y}_{j-1})$ for $s \geq j$ and allow the MDM to incorporate all observed data under MAR (cf. Corollary 1).

For example, we may specify distributions $\boldsymbol{Y}_{\text{obs}}|S$ as follows:

$$p_s(y_1) \sim \text{N}(\mu_1^{(s)}, \sigma_1^{(s)}) \quad 1 \leq s \leq J$$

$$p_s(y_j|\overline{y}_{j-1}) \sim \text{N}(\mu_{j|j-}^{(s)}, \sigma_{j|j-}^{(s)}) \quad 2 \leq j \leq s \leq J$$

where

$$\mu_{j|j-}^{(s)} = \beta_{j,0}^{(s)} + \sum_{l=1}^{j-1} \beta_{j,l}^{(s)} Y_l.$$

To identify the full-data model, recall the MAR constraints imply that

$$p_s(y_j|\overline{y}_{j-1}) = p_{\geq j}(y_j|\overline{y}_{j-1}) = \sum_{k=j}^{J} \frac{P(S = k)}{P(S \geq j)} p_k(y_j|\overline{y}_{j-1}) \tag{3}$$

for $s < j$, which are mixture of normals. For sensitivity analysis in this setting of mixture of normals, we propose to introduce sensitivity parameters Δ_μ (location) and Δ_σ (scale) such that for $s < j$

$$p_s(y_j|\overline{y}_{j-1}) = e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} p_k\left(\frac{y_j - \Delta_\mu^{(j)} - (1 - e^{\Delta_\sigma^{(j)}})\mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}}\bigg|\overline{y}_{j-1}\right) \tag{4}$$

where $\varpi_{j,k} = \frac{P(S=k)}{P(S \geq j)}$. The rationale for this parameterization is that each $p_k(\cdot|\overline{y}_{j-1})$ in the summation will have mean $\Delta_\mu^{(j)} + \mu_{j|j-}^{(k)}$ and variance $e^{2\Delta_\sigma^{(j)}} \sigma_{j|j-1}^{(k)}$. To reduce the dimension of

the sensitivity parameters, we could make $\Delta_\mu^{(j)}$ and $\Delta_\sigma^{(j)}$ common for all j (namely Δ_μ and Δ_σ).

In this set up, we have

$$\mu_{j|j-}^{(s),\mathrm{MNAR}} = \Delta_\mu^{(j)} + \sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)}$$

and

$$\sigma_{j|j-}^{(s),\mathrm{MNAR}} = e^{2\Delta_\sigma^{(j)}} \left\{ \sum_{k=j}^{J} \varpi_{j,k} \left(\sigma_{j|j-1}^{(k)} + (\mu_{j|j-1}^{(k)})^2 \right) - \left(\sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2 \right\} + (1 - e^{2\Delta_\sigma^{(j)}}) \mathscr{M}$$

where

$$\mathscr{M} = \sum_{k=j}^{J} \varpi_{j,k} (\mu_{j|j-}^{(k)})^2 - \left(\sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)} \right)^2$$

(see Web Appendix C for details). Note that $\mathscr{M}$ does not depend on $\sigma_{j|j-1}^{(k)}$ for $k = j, \ldots, J$.

Under an MAR assumption (3), for $[Y_j | \overline{Y}_{j-1}, S = s]$, we have

$$\mu_{j|j-}^{(s),\mathrm{MAR}} = \sum_{k=j}^{J} \varpi_{j,k} \mu_{j|j-}^{(k)}$$

and

$$\sigma_{j|j-}^{(s),\mathrm{MAR}} = \sum_{k=j}^{J} \varpi_{j,k} \left(\sigma_{j|j-1}^{(k)} + (\mu_{j|j-1}^{(k)})^2 \right) - (\mu_{j|j-}^{(s),\mathrm{MAR}})^2.$$

Therefore, under MNAR assumption (4), the two sensitivity parameters control the departure of the mean and variance from MAR in the following way,

$$\mu_{j|j-}^{(s),\mathrm{MNAR}} = \Delta_\mu^{(j)} + \mu_{j|j-}^{(s),\mathrm{MAR}} \qquad \text{and} \qquad \sigma_{j|j-}^{(s),\mathrm{MNAR}} = e^{2\Delta_\sigma^{(j)}} \sigma_{j|j-}^{(s),\mathrm{MAR}} + (1 - e^{2\Delta_\sigma^{(j)}}) \mathscr{M},$$

with $\Delta_\mu^{(j)}$ being a location parameter and $\Delta_\sigma^{(j)}$ being a scale parameter. The MNAR class allows MAR when $\Delta_\mu^{(j)} = \Delta_\sigma^{(j)} = 0$ for all $j \geq 2$.

By assuming non-future dependence, we obtain

$$p_s(y_j | \overline{y}_{j-1}) = p_{\geq j-1}(y_j | \overline{y}_{j-1}) = \frac{p(S = j - 1)}{p(S \geq j - 1)} e^{-\Delta_\sigma^{(j)}} \sum_{k=j}^{J} \varpi_{j,k} p_k \left(\frac{y_j - \Delta_\mu^{(j)} - (1 - e^{\Delta_\sigma^{(j)}}) \mu_{j|j-}^{(k)}}{e^{\Delta_\sigma^{(j)}}} \middle| \overline{y}_{j-1} \right)$$

$$+ \sum_{k=j}^{J} \frac{P(S = k)}{p(S \geq j - 1)} p_k(y_j | \overline{y}_{j-1}) \quad 2 \leq s < j - 1 \leq J - 1,$$

for the future data and (4) for the current data ($j = s + 1$). The number of sensitivity

parameters in this setup is reduced from $J(J-1)$ by $(J-2)(J-1)$ to $2(J-1)$; so, for $J=3$ (6), from 6 (30) to 2 (20). Further reductions are illustrated in Section 6.

6. Example: Growth Hormone Study

We analyze a longitudinal clinical trial using the framework from Sections 4 and 5 that assume multivariate normality for the full-data response within pattern (MVN) or multivariate normality for the observed data response within pattern (OMVN). We assume non-future dependence for the missing data mechanism to minimize the number of sensitivity parameters.

The growth hormone (GH) trial was a randomized clinical trial conducted to estimate the effectiveness of recombinant human growth hormone therapy for increasing muscle strength in the elderly. The trial had four treatment arms: placebo (P), growth hormone only (G), exercise plus placebo (EP), and exercise plus growth hormone (EG). Muscle strength, here mean quadriceps strength (QS), measured as the maximum foot-pounds of torque that can be exerted against resistance provided by a mechanical device, was measured at baseline, 6 months and 12 months. There were 161 participants enrolled on this study, but only (roughly) 75% of them completed the 12 month follow up. Researchers believed that dropout was related to the unobserved strength measures at the dropout times.

For illustration, we confine our attention to the two arms using exercise: exercise plus growth hormone (EG) and exercise plus placebo (EP). Table 1 contains the observed data for the two arms.

[Table 1 about here.]

Let (Y_1, Y_2, Y_3) denote the full-data response corresponding to baseline, 6 months, and 12 months. Let Z be the treatment indicator (1 = EG, 0 = EP). Our goal is to draw inference about the mean difference of QS between the two treatment arms at month 12.

That is, the treatment effect $\theta = E(Y_3|Z = 1) - E(Y_3|Z = 0)$. In the full-data model for each treatment under non-future dependence, there are seven sensitivity parameters for the MVN model: $\{\Delta_0^{(2)}, \Delta_1^{(2)}, \Delta_0^{(3)}, \Delta_1^{(3)}, \Delta_2^{(3)}, \Delta_\sigma^{(2)}, \Delta_\sigma^{(3)}\}$, and four sensitivity parameters for OMVN model: $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}, \Delta_\sigma^{(2)}, \Delta_\sigma^{(3)}\}$; see Web Appendix D for details on the models. For the MNAR analysis, we reduced the number of sensitivity parameters as follows:

- $\Delta_\sigma^{(2)}$ and $\Delta_\sigma^{(3)}$ do not appear in the posterior distribution of $E(Y_3|Z)$ for $Z = 0, 1$, and thus are not necessary for inference on θ.

- We restrict to MNAR departures from MAR in terms of the intercept terms by assuming $\Delta_1^{(2)} = \Delta_1^{(3)} = \Delta_2^{(3)} \equiv 0$.

- We assume the sensitivity parameters are identical between treatments.

This reduces the set of sensitivity parameters to $\{\Delta_0^{(2)}, \Delta_0^{(3)}\}$ for MVN model and $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}\}$ for the OMVN model.

There are a variety of ways to specify priors for the sensitivity parameters $\Delta_0^{(2)}$ and $\Delta_0^{(3)}$,

$$\Delta_0^{(2)} = E(Y_2|Y_1, S = 1) - E(Y_2|Y_1, S \geq 2)$$

$$\Delta_0^{(3)} = E(Y_3|Y_2, Y_1, S = 2) - E(Y_3|Y_2, Y_1, S = 3).$$

Both represent the difference of conditional means between the observed and unobserved responses. $\Delta_\mu^{(2)}$ and $\Delta_\mu^{(3)}$ have (roughly) the same interpretation as $\Delta_0^{(2)}$ and $\Delta_0^{(3)}$.

Based on discussion with investigators, we made the assumption that dropouts do worse than completers; thus, we restrict the Δ's to be less than or equal to zero. To do a fully Bayesian analysis to fairly characterize the uncertainty associated with the missing data mechanism, we assume a uniform prior for the Δ's as a default choice. Subject matter considerations gave an upper bound of zero for the uniform distributions. We set the lower bound using the variability of the observed data as follows. We estimate the residual variances of $Y_2|Y_1$ and $Y_3|Y_2, Y_1$ using the observed data; we denote these by $\tau_{2|1}$ and $\tau_{3|2,1}$ respectively. We use the square root of these estimates as the lower bounds. In particular, we specify the

priors for $\{\Delta_0^{(2)}, \Delta_0^{(3)}\}$ as well as $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}\}$ as $\mathrm{Unif}(\mathscr{D}(\boldsymbol{\tau}))$, where

$$\mathscr{D}(\boldsymbol{\tau}) = \left[-\tau_{2|1}^{-1/2}, 0\right] \times \left[-\tau_{3|2,1}^{-1/2}, 0\right]. \tag{5}$$

Based on the estimates $\tau_{2|1}^{-1/2} = 18$ and $\tau_{3|2,1}^{-1/2} = 12$, the priors are $[-18, 0] \times [-12, 0]$ for $\{\Delta_0^{(2)}, \Delta_0^{(3)}\}$ and for $\{\Delta_\mu^{(2)}, \Delta_\mu^{(3)}\}$. For the other parameters in the full-data model, we assign $N(0, 10^6)$ for mean parameters (μ, β) and $Unif(0, 100)$ for variance parameters $(\sigma^{1/2})$.

We fit the model using `WinBUGS`, with multiple chains of $25,000$ iterations and 4000 burn-in. Convergence was checked by examining trace plots of the multiple chains.

The results of the MVN and OMVN models are given in Table 2. Under MNAR, the posterior mean (posterior standard deviation) of the difference in quadriceps strength at 12 months between the two treatment arms was 4.0 (8.9) and 4.4 (10) for the MVN and OMVN models. Under MAR the differences were 5.4 (8.8) and 5.8 (9.9) for the MVN and OMVN models, respectively. The smaller differences under MNAR were due to quadriceps strength at 12 months being lower under MNAR due to the assumption that dropouts do worse than completers.

To compare the fit of models OMVN and MVN, we used the deviance information criterion that is based on the observed data likelihood (DIC_O),

$$\mathrm{DIC}_O = -4E\{\log L(\theta|\boldsymbol{y}_{\mathrm{obs}}, S)\} + 2\log L\{E(\theta|\boldsymbol{y}_{\mathrm{obs}}, S)|\boldsymbol{y}_{\mathrm{obs}}, S\}. \tag{6}$$

The results favored the OMVN model (Table 2). The behavior of DIC_O for model selection in the setting of incomplete data is explored in Section 7. Note that given the OMVN or MVN specification for the observed data, the fit (as measured by the DIC_O) was equivalent across the different missingness mechanisms as it should be since the observed data provide no information about this. We conclude that the treatment difference, θ was not significantly different from zero.

[Table 2 about here.]

To assess the sensitivity of the informative priors $\text{Unif}(\mathscr{D}(\boldsymbol{\tau}))$, we evaluated several scenarios by making the priors more or less informative by modifying the range (see Table 3). Specifically, we considered (1) $\mathscr{D}(\boldsymbol{\tau}) = [-20, 0] \times [-20, 0]$, (2) $\mathscr{D}(\boldsymbol{\tau}) = [-9, 0] \times [-6, 0]$, (3) $\mathscr{D}(\boldsymbol{\tau}) = [-20, 0] \times [-6, 0]$ and (4) $\mathscr{D}(\boldsymbol{\tau}) = [-9, 0] \times [-20, 0]$; note that the lower bounds of -9 and -6 are $-\tau_{2|1}^{-1/2}/2$ and $-\tau_{3|2,1}^{-1/2}/2$, respectively. Based on subject matter considerations, we assumed the difference of the conditional means between the observed and unobserved responses did not exceed 20 foot pounds of torque. None of the scenarios considered resulted in a significant mean treatment difference at month 12.

[Table 3 about here.]

7. Model Comparison and Simulations

Although the specification in Section 3-4 arguably offers a simpler sensitivity analysis than that in Section 5, the former may not fit the observed data as well (as indicated by DIC_O in Section 6). In addition, if the conclusions between OMVN and MVN were substantially different (unlike in the example here), we would need to decide which model to use for inference.

Spiegelhalter et al. (2002) developed the deviance information criterion (DIC) for Bayesian model comparison. In the setting of incomplete data, many alternative representations of DIC have been proposed (Celeux et al., 2006; Daniels and Hogan, 2008). The underlying complication is the fact that we do not observe the full data. Recommendations in Celeux et al. (2006) were based on missing data that were actually latent variables in random effects and mixture models, not potentially observable missing data as is our focus. Daniels and Hogan (2008) recommended DIC_O (6) on intuitive grounds but did no exploration of its operating characteristics. The following simulation study explores the behavior of DIC_O, as well as the behavior of MVN and OMVN models.

We simulated observed data from the MVN and OMVN models. Parameters estimated by fitting the MVN and OMVN models to the observed EG arm data from the growth hormone study were the basis for the simulation (see Table 4).

[Table 4 about here.]

For each sample size, 50, 150 and 300, we simulated 500 datasets for MVN and OMVN observed data models. The performance of the different approaches were assessed using both the DIC_O and the mean squared error (MSE) of $E(Y_j) : j = 2, 3$; see Table 5. Note that we computed the "true" $E(Y_2)$ and $E(Y_3)$ for MNAR cases using the uniform prior with $\mathscr{D}(\boldsymbol{\tau}) = [-18, 0] \times [-12, 0]$ from Section 6. For reference, the MSE's associated with the true data generating model are bolded.

The simulations favored the correctly specified model with respect to both DIC_O and MSE. As the sample size was increased, the MSE's decreased and the difference between the correct and the incorrect models increased. The MSE for $E[Y_j] : j = 2, 3$ provides information on the fit to the full data response model (both the observed and missing data), not just the observed data model; however, all that can be checked from the observed data is the fit to the observed data response model. Based on this (limited) simulation, we conclude that DIC_O is a reasonable criterion to choose between different incomplete data models and has more power when the sample size is moderate or large. Sensitivity analysis and all inferences can then be conducted on the model that fits the observed data the best. The simulation result also showed that when the MVN specification is true, OMVN is competitive, but not vice versa. This is reasonable because OMVN is the conservative choice here given that MVN is nested within it.

To compare the robustness of the MVN and OMVN models when neither is correct, we considered selection models where the distribution of $\boldsymbol{Y}$ is heavy-tailed or asymmetric. Specifically, we simulated response data from a multivariate t (MVT) distribution $\boldsymbol{Y} \sim$

$\mathrm{MVT}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, df)$ with df being the degrees of freedom, and a multivariate skew-normal (MSKN) distribution $\boldsymbol{Y} \sim \mathrm{MSKN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \omega)$ with ω being the skewness factor (Azzalini and Valle, 1996). See Web Appendix E for details of data generation for these cases.

The average DIC_O are reported in Table 6. Note that MAR and MNAR cases correspond to different observed data models in this simulation setting. Therefore, we report DIC_O for MAR and MNAR separately.

From the simulation, we can see that MVN model is more likely to be selected by DIC_O when sample size is small. When sample size is moderate or large, OMVN model fits the observed data better for all the cases considered and offers more robustness (in particular for the heavy-tailed scenario). OMVN appears to be the better default choice for medium to large sample sizes.

[Table 5 about here.]

[Table 6 about here.]

8. ACMV Restrictions and Multivariate Normality with Baseline Covariates

In this section, we show that common identifying restrictions over-identify estimable parameters in the presence of baseline covariates with time invariant coefficients and offer a solution.

Consider the situation when $\boldsymbol{Y} = (Y_1, Y_2)$ is a bivariate normal response ($J = 2$) with missing data only in Y_2, i.e. $S = 1$ or 2. Assume there are baseline covariates X with time invariant coefficients α. We model $p(S)$ and $p(\boldsymbol{Y}|S)$ as follows:

$$S|X \sim \mathrm{Bern}(\phi(X))$$

$$\boldsymbol{Y}|S = s \sim \mathrm{N}(\boldsymbol{\mu}^{(s)}(X), \boldsymbol{\Sigma}^{(s)}) \qquad s = 1, 2$$

where

$$\boldsymbol{\mu}^{(s)} = \begin{bmatrix} \mu_1^{(s)} + X\alpha^{(s)} \\ \mu_2^{(s)} + X\alpha^{(s)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{(s)} = \begin{bmatrix} \sigma_{11}^{(s)} & \sigma_{12}^{(s)} \\ \sigma_{21}^{(s)} & \sigma_{22}^{(s)} \end{bmatrix}.$$

MAR (ACMV) implies the following restriction

$$[Y_2|Y_1, S = 1] \simeq [Y_2|Y_1, S = 2].$$

This implies that conditional means, $E(Y_2|Y_1, X, S = s)$ for $s = 1, 2$, are equal, i.e.

$$\mu_2^{(1)} + X\alpha^{(1)} + \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}(Y_1 - \mu_1^{(1)} - X\alpha^{(1)}) = \mu_2^{(2)} + X\alpha^{(2)} + \frac{\sigma_{21}^{(2)}}{\sigma_{11}^{(2)}}(Y_1 - \mu_1^{(2)} - X\alpha^{(2)}). \quad (7)$$

For (7) to hold for all Y_1 and X, we need that

$$\alpha^{(1)} = \alpha^{(2)}.$$

However, both $\alpha^{(1)}$ and $\alpha^{(2)}$ are already identified by the observed data Y_1. Thus the ACMV

(MAR) restriction affects the model fit to the observed data. This is against the principle of

applying identifying restrictions (Little and Wang, 1996; Daniels and Wang, 2009).

To resolve the over-identification issue , we propose to apply MAR constraints on residuals

instead of directly on the responses. In the bivariate case, the corresponding restriction is

$$\left[Y_2 - X\alpha^{(1)}|Y_1 - X\alpha^{(1)}, X, S = 1\right] \simeq \left[Y_2 - X\alpha^{(2)}|Y_1 - X\alpha^{(2)}, X, S = 2\right]. \quad (8)$$

Since the conditional distributions

$$\left[Y_2 - X\alpha^{(s)}|Y_1 - X\alpha^{(s)}, X, S = s\right] \sim N\left(\mu_2^{(s)} + \frac{\sigma_{21}^{(s)}}{\sigma_{11}^{(s)}}(Y_1 - \mu_1^{(s)}), \sigma_{22}^{(s)} - \frac{(\sigma_{21}^{(s)})^2}{\sigma_{11}^{(s)}}\right)$$

are independent of $\alpha^{(s)}$ for $s = 1, 2$, the restriction (8) places no constraints on $\alpha^{(s)}$, thus

avoiding over-identification.

The MDM corresponding to the ACMV(MAR) on the residuals is given by

$$\log \frac{P(S = 1|\mathbf{Y}, X)}{P(S = 2|\mathbf{Y}, X)} = \log \frac{\phi(X)}{1 - \phi(X)} - \frac{1}{2\sigma^*}\Bigg\{(1 - B)^2 X(\alpha^{(2)}\alpha^{(2)T} - \alpha^{(1)}\alpha^{(1)T})X^T$$

$$- 2(1 - B)(Y_2 - \Delta(Y_1))X(\alpha^{(2)} - \alpha^{(1)})\Bigg\} - \frac{1}{2}\log \frac{\sigma_{11}^{(2)}}{\sigma_{11}^{(1)}}$$

$$- \frac{(Y_1 - X\alpha^{(2)} - \mu_1^{(2)})^2}{2(\sigma_{11}^{(2)})^2} + \frac{(Y_1 - X\alpha^{(1)} - \mu_1^{(1)})^2}{2(\sigma_{11}^{(1)})^2},$$

where $\sigma^* = \sigma_{22}^{(2)} - \frac{(\sigma_{21}^{(2)})^2}{\sigma_{11}^{(2)}}$, $B = \frac{\sigma_{21}^{(1)}}{\sigma_{11}^{(1)}}$ and $\Delta(Y_1) = \mu_2^{(2)} + \frac{\sigma_{21}^{(2)}}{\sigma_{11}^{(2)}}(Y_1 - \mu_1^{(2)})$. Hence by assuming MAR on the residuals, we have the MDM being a quadratic form of Y_1, but independent of Y_2 if and only if $\alpha^{(2)} = \alpha^{(1)}$. In other words, assumption (8) implies MAR if and only if $\alpha^{(2)} = \alpha^{(1)}$. So in general, MAR on residuals does not imply that missingness in Y_2 is MAR. However, it is an identifying restriction that does not impact the fit of the model to the observed data. CCMV and NCMV restrictions can be applied similarly to the residuals.

Remark: In general, $\mu_l^{(s)}$ can be replaced by $\mu_{il}^{(s)}$ if there are subject-specific covariates with time varying coefficients.

The ACMV (MAR) on the residuals restriction can be applied to the multivariate case and results in a similar MDM. Detailed discussion is provided in Web Appendix F.

9. Summary

Most pattern mixture models allow the missingness to be MNAR, with MAR as a unique point in the parameter space. The magnitude of departure from MAR can be quantified via a set of sensitivity parameters. For MNAR analysis, it is critical to find scientifically meaningful and dimensionally tractable sensitivity parameters. For this purpose, (multivariate) normal distributions are often found attractive since the MNAR departure from MAR can be parsimoniously defined by deviations in the mean and (co-)variance.

However, a simple pattern mixture model based on multivariate normality for the full data response within patterns does not allow MAR without special restrictions that themselves, induce a very restrictive missing data mechanism. We have explored this fully and proposed alternatives based on multivariate normality for the observed data response within patterns. In both these contexts, we proposed strategies for specifying sensitivity parameters.

The proposed modeling and sensitivity analysis approaches based on within-pattern multivariate normality for the full data response or the observed data response may lead to

contradicting study conclusions, due to the different (unverifiable) missing data mechanism assumptions and the different observed data model. For the latter, we proposed to use the deviance information criterion that is based on the observed data likelihood for model comparison and selection and showed via simulations that it appears to perform well. Sensitivity analysis and inferences can then be based upon the model that fits the observed data the best.

In addition, we showed that when introducing baseline covariates with time invariant coefficients, standard identifying restrictions result in over-identification of the model. This is against the principle of applying identifying restriction in that they should not affect the model fit to the observed data. We proposed a simple alternative set of restrictions based on residuals that can be used as an 'identification' starting point for an analysis using mixture models.

In the growth hormone study data example, we showed how to reduce the number of sensitivity parameters in practice and a default way to construct informative priors for sensitivity parameters based on limited knowledge about the missingness. In particular, all the values in the range $\mathscr{D}$ were weighted equally via a uniform distribution. If there is additional external information from expert opinion or historical data, informative priors may be used to incorporate such information (for example, see Ibrahim and Chen, 2000; Wang et al., 2010). Finally, an important consideration in sensitivity analysis and constructing informative priors is that they should avoid extrapolating missing values outside of a reasonable range (e.g., in the growth hormone trial example, we would not want to impute a negative quadriceps strength).

10. Supplementary Materials

Web Appendices and Tables referenced in Sections 2, 3, 5, 6, 7, and 8 are available under the Paper Information link at the Biometrics website http://www.biometrics.tibs.org.

References

Azzalini, A. and Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83,** 715–726.

Celeux, G., Forbes, F., Robert, C., and Titterington, D. (2006). Deviance information criteria for missing data models. *Bayesian Analysis* **1,** 651–674.

Daniels, M. and Hogan, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* Chapman & Hall/CRC.

Daniels, M. and Wang, C. (2009). Discussion of "Missing Data in longitudinal studies: A review" by Ibrahim and Molenberghs. *TEST* **18,** 51–58.

Harel, O. and Schafer, J. (2009). Partial and latent ignorability in missing-data problems. *Biometrika* **96,** 37.

Hogan, J. and Laird, N. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16,** 259–272.

Ibrahim, J. and Chen, M. (2000). Power prior distributions for regression models. *Statistical Science* **15,** 46–60.

Kenward, M. and Molenberghs, G. (1999). Parametric Models for Incomplete Continuous and Categorical Longitudinal Data. *Statistical Methods in Medical Research* **8,** 51.

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

Little, R. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90,**.

Little, R. and Rubin, D. (1987). *Statistical Analysis with Missing Data.* Wiley.

Little, R. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52,** 98–111.

Molenberghs, G. and Kenward, M. (2007). *Missing Data in Clinical Studies.* Wiley.

Molenberghs, G., Michiels, B., Kenward, M., and Diggle, P. (1998). Monotone Missing Data and Pattern-Mixture Models. *Statistica Neerlandica* **52,** 153–161.

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

Scharfstein, D., Daniels, M., and Robins, J. (2003). Incorporating Prior Beliefs about Selection Bias into the Analysis of Randomized Trials with Missing Outcomes. *Biostatistics* **4,** 495.

Scharfstein, D., Rotnitzky, A., and Robins, J. (1999). Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *Journal of the American Statistical Association* **94,** 1096–1146.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 583–639.

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3,** 245.

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

Wang, C., Daniels, M., Scharfstein, D., and Land, S. (2010). A Bayesian shrinkage model for incomplete longitudinal binary data with application to the breast cancer prevention trial. *Journal of the American Statistical Association (in press)* .

Zhang, J. and Heitjan, D. (2006). A simple local sensitivity analysis tool for nonignorable coarsening: application to dependent censoring. *Biometrics* **62,** 1260–1268.

Table 1
Growth Hormone Study: Sample mean (standard deviation) stratified by dropout pattern.

Treatment	Dropout Pattern	Number of Participants	Month 0	6	12
EG	1	12	58(26)		
	2	4	57(15)	68(26)	
	3	22	78(24)	90(32)	88(32)
	All	38	69(25)	87(32)	88(32)
EP	1	7	65(32)		
	2	2	87(52)	86(51)	
	3	31	65(24)	81(25)	72(21)
	All	40	66(26)	82(26)	72(21)

Table 2
Growth Hormone Study: Posterior mean (standard deviation) stratified by treatment.

Treatment	Month	Observed Data	MVN MAR	MVN MNAR	OMVN MAR	OMVN MNAR
EG	0	69(7.3)	69(4.9)	69(4.9)	69(4.9)	69(4.9)
	6	87(16)	81(6.8)	78(7.1)	82(7.7)	79(8.0)
	12	88(6.8)	78(7.2)	76(7.5)	79(7.8)	76(8.0)
EP	0	66(9.9)	66(6.0)	66(6.0)	66(6.0)	66(6.0)
	6	82(18)	82(5.9)	80(6.0)	81(8.2)	80(8.3)
	12	72(3.8)	73(4.9)	72(5.0)	73(6.1)	71(6.1)
Difference at 12 mos.		15.9(7.8)	5.4(8.8)	5.8(9.9)	4.0(8.9)	4.4(10)
DIC_O			1786.6		1779.8	

Table 3
Growth Hormone Study: MNAR Sensitivity Analysis

$\mathscr{D}(\boldsymbol{\tau})$			Difference at 12 mos.	
			MVN	OMVN
$[-20, 0]$	$\times$	$[-20, 0]$	3.5(9.0)	3.9(10.1)
$[-9, 0]$	$\times$	$[-6, 0]$	4.7(8.8)	5.1(9.9)
$[-20, 0]$	$\times$	$[-6, 0]$	4.2(8.9)	4.6(10.0)
$[-9, 0]$	$\times$	$[-20, 0]$	4.1(8.9)	4.5(10.1)

Table 4

Simulation Scenario: Parameters based on the observed EG arm from the growth hormone study.

$\phi_i = P(S = i)$		OMVN		MVN	
ϕ_1	0.3	$\beta_{2,0}^{(2)}$	-28	$\beta_0^{(\geq 2)}$	15
ϕ_2	0.2	$\beta_{2,1}^{(2)}$	2	$\beta_1^{(\geq 2)}$	1
ϕ_3	0.5	$\beta_{2,0}^{(3)}$	18	$\beta_0^{(\geq 3)}$	-5
Baseline		$\beta_{2,1}^{(3)}$	1	$\beta_1^{(\geq 3)}$	0.5
$\mu_1^{(1)}$	58	$\beta_{3,0}^{(3)}$	-5	$\beta_2^{(\geq 3)}$	0.7
$\mu_1^{(2)}$	56	$\beta_{3,1}^{(3)}$	0.5	$\sigma_{2|2-}^{(\geq 2)}$	4
$\mu_1^{(3)}$	78	$\beta_{3,2}^{(3)}$	0.7	$\sigma_{3|3-}^{(\geq 3)}$	2.3
$\sigma_1^{(1)}$	12.3	$\sigma_{2|2-}^{(2)}$	9		
$\sigma_1^{(2)}$	16	$\sigma_{2|2-}^{(3)}$	6.3		
$\sigma_1^{(3)}$	4	$\sigma_{3|3-}^{(3)}$	4		

Table 5

Simulation Results: Comparison of MSE and performance of DIC_O in MVN and OMVN models. The columns correspond to the models fit and the rows to the scenario under which the data was generated. Bold indicates the MSE results for the true model. Rate in favor corresponds to the percentage of simulated datasets where the model (either MVN or OMVN) had a smaller DIC_O.

	Parameter	MVN		OMVN	
		MAR	MNAR	MAR	MNAR
MVN, Sample Size 50					
MAR	$E(Y_2)$	**2.25**	11.6	3.32	10.9
	$E(Y_3)$	**7.40**	23.3	7.94	24.1
MNAR	$E(Y_2)$	8.72	**3.47**	11.5	4.45
	$E(Y_3)$	19.0	**9.82**	18.7	10.1
DIC_O (rate in favor)		610.3(78.0%)		613.6(22.0%)	
MVN, Sample Size 150					
MAR	$E(Y_2)$	**0.77**	8.82	1.09	7.79
	$E(Y_3)$	**2.26**	16.9	2.53	16.6
MNAR	$E(Y_2)$	7.75	**1.22**	9.43	1.55
	$E(Y_3)$	12.8	**2.98**	13.7	3.27
DIC_O (rate in favor)		1780.3(73.3%)		1783.9 (26.7%)	
MVN, Sample Size 500					
MAR	$E(Y_2)$	**0.26**	7.70	0.45	6.43
	$E(Y_3)$	**0.93**	14.9	0.91	14.2
MNAR	$E(Y_2)$	7.55	**0.41**	9.19	0.59
	$E(Y_3)$	11.2	**1.09**	12.0	1.12
DIC_O (rate in favor)		5874.7(75.2%)		5880.7 (24.8%)	
OMVN, Sample Size 50					
MAR	$E(Y_2)$	3.65	4.37	**2.85**	9.06
	$E(Y_3)$	11.9	22.9	**11.7**	27.9
MNAR	$E(Y_2)$	18.3	4.50	12.1	**3.80**
	$E(Y_3)$	27.8	13.1	23.0	**13.4**
DIC_O (rate in favor)		676.9(0.9%)		649.4 (99.1%)	
OMVN, Sample Size 150					
MAR	$E(Y_2)$	2.45	2.61	**0.96**	7.53
	$E(Y_3)$	3.35	11.5	**3.45**	17.7
MNAR	$E(Y_2)$	17.3	2.74	9.32	**1.31**
	$E(Y_3)$	21.5	3.91	15.0	**3.99**
DIC_O (rate in favor)		1970.5(0.0%)		1883.5 (100.0%)	
OMVN, Sample Size 500					
MAR	$E(Y_2)$	2.10	2.00	**0.35**	6.56
	$E(Y_3)$	1.13	10.3	**1.14**	16.3
MNAR	$E(Y_2)$	16.8	2.22	8.81	**0.50**
	$E(Y_3)$	17.5	1.43	11.3	**1.26**
DIC_O (rate in favor)		6515.6(0.0%)		6213.1 (100.0%)	

Table 6
Simulation Results: Assessing fit of MVN and OMVN models using DIC_O under heavy-tailed and skewed data generating mechanisms. Values in the table are DIC_O (rate in favor). Rate in favor corresponds to the percentage of simulated datasets where the model (either MVN or OMVN) had a smaller DIC_O.

Sample Size	MVN	OMVN
MVT, MAR		
50	781.8(49.2%)	780.1 (50.8%)
150	2344.8(26.0%)	2338.4 (74.0%)
500	7864.4(5.0%)	7844.4 (95.0%)
MVT, MNAR		
50	790.8(54.0%)	790.1 (46.0%)
150	2381.6(38.8%)	2377.7 (61.3%)
500	7951.4(12.5%)	7938.9 (87.5%)
MSKN, MAR		
50	623.1(72.9%)	625.2 (27.1%)
150	1845.2(61.3%)	1846.0 (38.7%)
500	6119.6(26.5%)	6115.3 (73.5%)
MSKN, MNAR		
50	632.5(75.4%)	635.0 (24.6%)
150	1872.4(63.7 %)	1873.9 (36.3%)
500	6211.9(41.2 %)	6210.2 (58.8%)