

Published in final edited form as:

*Comput Stat Data Anal.* 2010 March 1; 54(3): 790–801. doi:10.1016/j.csda.2009.01.016.

## Impact of non-normal random effects on inference by multiple imputation: A simulation assessment

Recai M. Yucel<sup>a,\*</sup> and Hakan Demirtas<sup>b</sup>

<sup>a</sup>Department of Epidemiology and Biostatistics, School of Public Health, University at Albany, SUNY, One University Place Room 139, Rensselaer, NY 12144, United States

<sup>b</sup>Department of Epidemiology and Biostatistics (MC923), University of Illinois at Chicago, 1603 West Taylor Street, Chicago, IL 60612, United States

### Abstract

Multivariate extensions of well-known linear mixed-effects models have been increasingly utilized in inference by multiple imputation in the analysis of multilevel incomplete data. The normality assumption for the underlying error terms and random effects plays a crucial role in simulating the posterior predictive distribution from which the multiple imputations are drawn. The plausibility of this normality assumption on the subject-specific random effects is assessed. Specifically, the performance of multiple imputation created under a multivariate linear mixed-effects model is investigated on a diverse set of incomplete data sets simulated under varying distributional characteristics. Under moderate amounts of missing data, the simulation study confirms that the underlying model leads to a well-calibrated procedure with negligible biases and actual coverage rates close to nominal rates in estimates of the regression coefficients. Estimation quality of the random-effect variance and association measures, however, are negatively affected from both the misspecification of the random-effect distribution and number of incompletely-observed variables. Some of the adverse impacts include lower coverage rates and increased biases.

### 1. Introduction

The existence of missing data is inevitable in many settings. Among them are multilevel applications where missing values are arbitrarily seen in the observational units nested within naturally occurring groups, such as patients nested within hospitals. As most statistical methods for analyzing data under such circumstances are not designed to handle arbitrary missingness (some model fitting techniques can handle longitudinal data with drop-out under missing at random), in general, missing values are recognized as a complicating factor to the analyses. Indeed, identifying appropriate analytical tools to draw the desired inferences in the presence of missing data is often a daunting task because such tools need to account for the reasons for missingness (i.e. missingness mechanism) as well as the key features of the multilevel design such as levels of variations. Ignoring these essential characteristics may result in misleading results such as biased parameter estimates (e.g. underestimated variances) and reduced statistical power (Rubin, 1976, 1987; Schafer, 1997).

One of the most popular methods in conducting inference from incomplete data applications is multiple imputation (MI) (Rubin, 1987). MI is a simulation-based inferential tool operating on  $M > 1$  “completed” data sets, where the missing values are replaced by random draws from

their respective predictive distributions (e.g. posterior predictive distribution of missing data). These  $M$  versions of completed data are then analyzed by standard complete-data methods and the results are combined into a single inferential statement using rules to yield estimates, standard errors and  $p$ -values that formally incorporate the missing data uncertainty into the modeling process (Rubin, 1987). The key ideas and the advantages of MI are given by Rubin (1987) and Schafer (1997).

The crucial step in MI is filling in the missing data by drawing from the posterior predictive distribution of the missing data (i.e. the conditional distribution of the missing data, given the observed data and the unknown parameters). This typically involves positing a parametric model for the data and using it to derive this conditional distribution. In multilevel data applications, multivariate extensions of the mixed-effects models based on normality have often been perceived as a natural assumption because (1) it reflects the design features; and (2) the conditional distribution of the missing data given the observed data is easily tractable (Schafer and Yucel, 2002). There are two software products implementing the MI-based inference in multilevel data applications. The first one is freeware and available as a R (R Development Core Team, 2007) package called `pan` (Schafer and Yucel, 2002), and the second one is a macro called `mimacro` (Carpenter and Goldstein, 2004) which operates within a commercially available software package `MLwiN` (Rasbash et al., 2006). Because these are extensively used, it is important to examine how good their performance is when the underlying distributional assumption on the random-effects is not met. To the best of our knowledge, despite their growing use in the analyses of multilevel incomplete data, the plausibility and impact of the normality assumption of the random-effects have not been thoroughly evaluated. An evaluation using simulated examples that represent deviations from the standard normality assumption, such as skewness, non-zero peakedness, heavy tails, and flatness of the density that clearly violate the characteristics of normality, is provided in the current study. Performance of MI under varying distributional assumptions and sample sizes is also examined. Throughout the computations, R package `pan` (Schafer and Yucel, 2002) was used.

Section 2 gives a brief overview of the model used to impute the missing values for continuous or nearly-continuous variables and provides a discussion of the crucial assumptions of MI inference. Section 3 then summarizes how multiple imputations are created under a multivariate generalization of mixed-effects models. Section 4 outlines and discusses the findings of the simulations. Finally, Section 5 concludes the manuscript with a discussion of the implications of these findings.

## 2. Models used in MI inference

### 2.1. Imputation model

Below, a brief review on the imputation model commonly used to jointly impute missing values for a set of continuous random variables under a multivariate linear mixed-effects model is given. This model assumes two levels in the data collection, e.g. repeated measurements from individuals or secondary sampling units (US residents) within primary sampling units (US Census blocks). Allowing higher levels of nesting is possible by incorporating the corresponding random-effects in the imputation model (1) below, which has been implemented in `MLwiN` (Browne, 2003, ch. 17; Carpenter and Goldstein, 2004). We focus on the two-level model because it simplifies the assessment of the impact of the departures from normality and it is the underlying model of the R package `pan`.

Let  $y_i$  denote an  $n_i \times r$  matrix of multivariate responses for sample unit  $i$ ,  $i = 1, 2, \dots, m$ , where each row of  $y_i$  is a joint realization of a set of  $r$  variables subject to missing values. In longitudinal studies, for example, the rows correspond to the measurements of an individual at each point in time, and in clustered data, the rows correspond to the individuals within a

cluster. From the perspective of imputing missing values, the variables subject to missingness are included in  $y_i$ , regardless of being classified as a predictor (covariate) or a response. Often there exists auxiliary information that should be included for the reasons of post-imputation analyses or enriching the reasons for missingness, and we will denote this auxiliary information by  $X_i$  for the  $i$ th unit or cluster. Multivariate extension of a linear mixed-effects model with conventional Gaussian error structures on the residuals and random-effects is used as the imputation model for the missing values in  $y_i$ :

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i, \quad (1)$$

where  $X_i$  ( $n_i \times p$ ) and  $Z_i$  ( $n_i \times q$ ) are known covariate matrices,  $\beta$  ( $p \times r$ ) is a matrix of regression coefficients (fixed-effects) common to all units, and  $b_i$  ( $q \times r$ ) is a matrix of random coefficients, exhibiting the deviations of cluster  $i$  from the overall mean structure. In a fully parametric mixed-effects model framework, a normal probability distribution is often imposed on these deviations:  $\text{vec}(b_i) \sim N(0, \Psi)$  (here the  $\text{vec}$  operator vectorizes a matrix by stacking its columns). Model (1) assumes that the rows of the residuals matrix  $\varepsilon_i$  are independently distributed as  $N(0, \Sigma)$ , or alternatively,  $\text{vec}(\varepsilon_i) \sim N(0, \Sigma \otimes I_{n_i})$ .

Following the standard notation in the missing data literature, the complete data  $Y$  is comprised of the observed data  $Y_{obs}$  and  $Y_{mis}$ :  $Y = (Y_{obs}, Y_{mis})$ . Note that  $Y_{obs} = (y_{1(obs)}, y_{2(obs)}, \dots, y_{m(obs)})$  and  $Y_{mis} = (y_{1(mis)}, y_{2(mis)}, \dots, y_{m(mis)})$  denote all the observed and missing responses. Further,  $\theta$  includes all the unknown parameters  $\beta, \Sigma, \Psi$ .

## 2.2. Models for missingness mechanism

Explicit or implicit assumptions are made about the missing-data mechanisms in all missing-data methods. To set the notation for the discussion of missingness mechanisms, let  $R$  denote the set of missing-value indicators; note that  $R$  has the same dimension as  $Y$ , and it is always observed. Each element of  $R$  takes the value of 0 or 1 depending on whether the corresponding element in  $Y$  is missing or observed, respectively. Similar to  $Y$ ,  $R$  can be seen as a random variable; and the conditional distribution of  $R$  given  $Y$  depends on a set of parameters, say  $\gamma$ . The small letters  $r$  and  $y_{obs}$  denote the realized values of  $R$  and  $Y_{obs}$ .

Most tools available to the practitioners of MI (via software packages such as SAS PROC MIXED, R packages *norm*, *cat*, *mix* and *pan*) assume *missing at random* (MAR) as the missingness mechanism. This assumption simply means that the missingness probability may depend on the observed data but not on the missing data over the conditional distribution of  $R$  given  $Y_{obs}$ . More formally,  $P(R|Y_{obs}, Y_{mis}, \gamma) = P(R|Y_{obs}, \gamma)$ . Some misconceptions among practitioners exist due to the name MAR. The missing values do not occur at random under MAR; when they do, the mechanism is, in fact, *missing completely at random* (MCAR). Under MCAR, the missingness probabilities are independent from both  $Y_{obs}$  and  $Y_{mis}$ :  $P(R|Y_{obs}, Y_{mis}, \gamma) = P(R|\gamma)$ . When the missingness probabilities depend on  $Y_{mis}$ , the missingness mechanism is called *missing not at random* (MNAR). Under MNAR, one must posit a model for the complete data as well as for  $R$ . These models are usually very sensitive to the model assumptions (see detailed discussion by Schafer and Graham (2002) for more information). Another important concept is the “ignorability” of the missingness mechanism, and it is often seen as an implied condition once MAR is assumed. Specifically, ignorability occurs when the mechanism is MAR and the parameters  $\gamma$  and  $\theta$  are distinct:  $f(Y_{obs}, R|\theta, \gamma) = f(Y_{obs}|\theta)f(R|\gamma)$  (see Little and Rubin (2002) and Schafer (1997) for more details).

### 3. Multiple imputation under a multivariate linear mixed-effects model (MLMM)

A brief overview of MI and how it is conducted under a multivariate linear mixed-effects model (MLMM) for a two-level incomplete data is given (see Schafer and Yucel (2002) for computational details). MI is a Monte Carlo simulation-based technique where the missing values are replaced by a moderate number ( $M > 1$  and usually less than 10) of simulated values. How these simulated values are drawn relate to distributional assumptions and the missingness mechanism (Schafer and Yucel, 2002; Browne, 2003, ch. 17; Carpenter and Goldstein, 2004).

The principle idea of MI is that it treats the missing data as an explicit source of random variability which is incorporated into computation of standard errors. The process of creating imputations, analyzing the imputed data sets and combining the results is a Monte Carlo version of averaging the statistical results over the predictive distribution of the missing data,  $\int P(\theta | Y_{obs}, Y_{mis}) P(Y_{mis} | Y_{obs}) dY_{mis}$ . In applications with a moderate level of missingness, reasonably well-calibrated results can be obtained with  $M \leq 10$ . Once the imputations have been created, the  $M$  completed data sets may be analyzed with complete-data techniques, resulting in  $M$  set of estimates and associated standard errors, which are then combined by rules described by Rubin (1987).

The role of the imputation model in multilevel data such as those arising in longitudinal or clustered designs, is to preserve the relationships among variables (subject to missingness or not) and correlations among the observations from the same subject or cluster. Often the meaning or interpretation of its parameters is not crucial. The key importance of the imputation model is in its functionality to predict, simulate the missing observations and to capture, as much as possible, reasons for the missingness (to make the MAR plausible). MLMM given in (1), for these reasons, has found a common use among imputers. While other models of a similar nature are also possible, we focus on model (1) as it is the only one implemented in software products (R package pan and MLwiN).

Multiple imputations under model (1) are  $M$  independent draws  $Y_{mis}^{(1)}, \dots, Y_{mis}^{(M)}$  from a posterior predictive distribution for the missing data,  $P(Y_{mis} | Y_{obs}) = \int P(Y_{mis} | Y_{obs}, \theta) P(\theta | Y_{obs}) d\theta$  where  $P(\theta | Y_{obs})$  is the observed-data posterior density, which is proportional to the product of a prior density  $\pi(\theta)$  and the observed-data likelihood function  $L(\theta | Y_{obs}) = \int L(\theta | Y) dY_{mis}$ . As in most missing-data problems with arbitrary missingness patterns, the posterior distribution  $P(Y_{mis} | Y_{obs})$  cannot be simulated directly. Markov Chain Monte Carlo (MCMC) techniques facilitate the creation of random draws of  $Y_{mis}$  from  $P(Y_{mis} | Y_{obs})$ . In MCMC, one generates a sequence of dependent random variates whose distribution converges to the desired target — in this case the posterior distribution of missing data.

An MCMC procedure called the Gibbs sampler is the underlying iterative simulation algorithm used to draw the missing values from  $P(Y_{mis} | Y_{obs})$ . Specifically, suppose the current versions of the unknown parameters  $\theta^{(t)} = (\beta^{(t)}, \Sigma^{(t)}, \Psi^{(t)})$  and missing data  $Y_{mis}^{(t)}$  are updated in three steps: (1)  $b_i^{(t+1)} \sim P(b_i | Y_{obs}, Y_{mis}^{(t)}, \theta^{(t)})$ , independently for  $i = 1, \dots, m$ ; (2)  $\theta^{(t+1)} \sim P(\theta | Y_{obs}, Y_{mis}^{(t)}, B^{(t+1)})$ ; and finally, (3)  $y_{i(mis)}^{(t+1)} \sim P(y_{i(mis)} | Y_{obs}, B^{(t+1)}, \theta^{(t+1)})$ , for  $i = 1, \dots, m$ , where  $B = (b_1, \dots, b_m)$ . Given the starting values  $\theta^{(0)}$  and  $Y_{mis}^{(0)}$ , these steps define one cycle of an MCMC procedure, or the Gibbs sampler. Executing the cycle repeatedly creates sequences  $\{\theta^{(1)}, \theta^{(2)}, \dots\}$  and  $\{Y_{mis}^{(1)}, Y_{mis}^{(2)}, \dots\}$  whose limiting distributions are  $P(\theta | Y_{obs})$  and  $P(Y_{mis} | Y_{obs})$ , respectively.

To reflect the uncertainty in the unknown parameters  $\theta$ , the Bayesian paradigm has often been used in the preceding MCMC to create multiply imputed data sets. Under the model given in (1), independent inverted Wishart priors are used on the variance components:  $\Sigma^{-1}, \Psi^{-1}: \Sigma^{-1} \sim W(v_1, \Lambda_1)$  and  $\Psi^{-1} \sim W(v_2, \Lambda_2)$ , where  $W(v, \Lambda)$  denotes a Wishart variate with  $v > 0$  degrees of freedom and mean  $v\Lambda > 0$ . It is helpful to regard  $v_1^{-1}\Lambda_1^{-1}$  and  $v_2^{-1}\Lambda_2^{-1}$  as prior guesses for  $\Sigma$  and  $\Psi$  with confidence equivalent to  $v_1$  and  $v_2$  degrees of freedom, respectively. For  $\beta$ , an improper prior uniform ‘density’ over  $\mathcal{R}^{Pr}$  is often employed. In practice, determination of these hyperparameters can be difficult. Here, we used the R package **mlmmm** (Yucel, 2007) which computes maximum likelihood estimates of model (1) parameters to assign as the initial guesses for  $\Psi$  and  $\Sigma$  and as the initial values  $\beta^{(0)}, \Sigma^{(0)}$  and  $\Psi^{(0)}$  for the Gibbs sampler.

Computational details of the algorithm derived under these priors, Gaussian assumptions on the  $b_i, \varepsilon_i$ , and model (1) are given in detail by Schafer and Yucel (2002). An important question is how to monitor convergence of the underlying Gibbs sampler. As noted by many, including Gelman et al. (2003), and Schafer (1997), there are two general issues with the convergence: (1) whether the simulations are representative of the predictive distribution of missing data  $P(Y_{mis}|Y_{obs})$ , and (2) whether the successive draws are significantly correlated. These two issues, individually or collectively, can seriously damage the overall quality of inferences in terms of accuracy and representativeness of the target distribution. Simple investigation of time series plots and autocorrelation function plots is often indicative of the interdependency of the successive draws. These plots, along with mode-finding algorithms such as EM, can be used to assess convergence behavior and to determine the number of iterations needed to achieve independent samples from the desired distributions  $P(Y_{mis}|Y_{obs})$  and  $P(\theta|Y_{obs})$ . Running multiple independent chains from common starting values and checking if they converge to the same  $P(Y_{mis}|Y_{obs})$  is also a powerful and practical tool to monitor the convergence. We provide more specific details on Section 4.1.5.

Studies in the multiple imputation literature indicate that inference by MI is quite robust to departures from model assumptions used to derive the posterior predictive distribution of missing data  $P(Y_{mis}|Y_{obs})$  (Schafer, 1997; Demirtas et al., 2008). Under Gaussian assumptions on the random effects,  $b_i$ , and the error terms,  $\varepsilon_i$ , MLMM-based MI uses simple conjugacy to derive the conditional posterior distributions of  $\theta$ , random-effects and missing data which are all either normal or Wishart-based distributions (Schafer and Yucel, 2002). These conditionals constitute the steps of the Gibbs sampler given above, leading to the posterior predictive distribution of missing data. How sensitive is the inference by multiple imputation to the Gaussian assumption on the random effects? This question begs several sub-questions: what are the consequences of imputing under MLMM when the true distribution has heavier tails than normal, or when it is not symmetric and so on. The real data often do not conform with normality, and many procedures of statistical practice use the normality as an approximation. It is impossible to tailor the analytical solutions and computational algorithms suitable to tackle the underlying missing-data problem. Therefore, many of the consumers of multiple imputation rely on the normality assumption when imputing continuous multilevel data.

The next section assesses the viability of the Gaussian assumption in extensive simulation scenarios representing a diverse set of distributional violations of the assumed normality of the random-effects. We devise a simulation study primarily to generate the conditions under which the incomplete data were realized and where the imputation approach relies on the incompatible normality assumption. This strategy provides a feasible way of evaluation of the performance of the MLMM imputation with normal random-effects (Demirtas, 2004, 2005).

## 4. A simulation study

The characteristics of the MLMM-based MI when the real data-generating mechanism employs non-normal random effects are evaluated via the simulation study. Simulation design consists of the following steps: data generation, imposing a missingness mechanism, MI inference under a hypothetical analyst's model, parameter estimation, and evaluation of this estimation.

### 4.1. Simulation design

**4.1.1. Data generation**—A bivariate linear mixed-effects model was used to generate the complete data:

$$[y_{ij1} y_{ij2}]^T = [\beta_1 \beta_2]^T + [b_{i1} b_{i2}]^T + [\varepsilon_{ij1} \varepsilon_{ij2}]^T, \quad (2)$$

where  $i = 1, 2, \dots, m$ , and  $j = 1, 2, \dots, n_i$ . The number of clusters ( $m$ ) and observations per cluster ( $n_i$ ) are varied so that large and small sample behavior were observed. We allowed  $m$  to take values of 100, 50 and 20;  $n_i$  to take values of 5 and 10 (the  $n_i$  are assumed to be same across the clusters).

The error term  $\varepsilon_i$  are assumed to be distributed as  $N(0, \Sigma)$  and our simulations used values of  $\sigma_{11} = 0.9$ ,  $\sigma_{12} = 0.3$  and  $\sigma_{22} = 0.4$ . Subject-specific random-effects  $b_i = [b_{i1} b_{i2}]^T$  were imposed, varying distributions with means zero to investigate the performance of MI-based inferences under MLMM with non-normal random-effects. Results pertaining to varying distributions of the random effects are summarized in the Section 4.2. Similar results are also seen with non-normal distributions on  $\varepsilon_i$  and are available upon request from the authors.

- **Multivariate normal distribution:** This is the default assumption under MLMM imputation, and was primarily used for comparative reasons. Random variates, say  $b$ , were drawn from  $N_2(0, \Sigma_b)$ :

$$f(b|\Sigma_b) = (2\pi)^{-1} |\Sigma_b|^{-1/2} \exp \left( -\frac{1}{2} b^T \Sigma_b^{-1} b \right).$$

Elements of  $\Sigma_b$  were set as:  $\sigma_{b_1}^2 = 3$ ,  $\sigma_{b_2}^2 = 1.5$  and  $\sigma_{b_1, b_2} = 0.25$ . These values were also used in the following distributions.

- **Multivariate t distribution:** Multivariate  $t$  distribution with  $v = 4$  was used to observe the behavior of MI under MLMM when the underlying distribution of the random-effects has heavier tails than normal (see Fig. 1). Here the random variates follow the density function of the form:

$$f(b|\Sigma_b) = \frac{\Gamma((v+d)/2)}{\Gamma(v/2) v^{d/2} \pi^{d/2}} |\Sigma_b|^{-1/2} \left( 1 + \frac{1}{v} b^T \Sigma_b^{-1} b \right)^{-(v+d)/2}.$$

- **Multivariate Laplace distribution:** Multivariate Laplace distribution or double exponential distribution was considered to examine the behavior of MI inferences under distributions with heavier tails than normal (see Fig. 2). The density function is given by,

$$f(b|\Sigma_b, \gamma) = \frac{\gamma \Gamma(1)}{2\pi \Gamma(2/\gamma)} |\Sigma_b|^{-1/2} \exp \left( -(b^T \Sigma_b^{-1} b)^{\gamma/2} \right).$$



- **Multivariate uniform distribution:** Random bi-variables from  $U_d(0, 1)$  were generated with  $\Sigma_b$  to assess the behavior under a flat density.

**4.1.2. Imposing missingness mechanism**—Throughout the simulation study, we considered two general cases: (1) missingness is confined to covariates; and (2) arbitrary missingness is present in both response and covariates of the analyst's model which is described in Section 4.1.3. To study the first case, we imposed missing values on the covariate  $Y_2$  under the missingness mechanism MAR defined by

$$\log \frac{P(r_{Y_2}=1|Y_1)}{1 - P(r_{Y_2}=1|Y_1)} = 1.3 + 1.1y_1,$$

where  $r_{Y_2}$  is the missingness indicator for  $Y_2$ . Conditioning on completely-observed  $y_1$  makes the MAR assumption (Little and Rubin, 2002) as assumed by MLMM imputation, more plausible. The model stated above results in approximately 45% missingness in  $Y_2$  values. Tables 1 and 2 in the results section pertain to case (1).

Missing values were imposed under two independent intercept-only regression models leading to MCAR mechanisms (stronger assumption than MAR) to study the arbitrary missingness on both covariates and responses:

$$\log \frac{P(r_{Y_1}=1)}{1 - P(r_{Y_1}=1)} = -0.2, \quad \log \frac{P(r_{Y_2}=1)}{1 - P(r_{Y_2}=1)} = 0.2.$$

Under these specifications, rates of missingness in  $Y_1$  and  $Y_2$  are, on average, 45% and 55%. Tables 3 and 4 in the results section pertain the case (2).

**4.1.3. Estimation of the Analyst's model**—To mimic the practice of MI, the model given in (3) is assumed to underlie the substantive goal of the analysis, i.e. the analyst's model. Hence, we performed MI inference on the parameters of this model and evaluated the estimation quality using the criteria given in Section 4.1.4.

$$y_{1,ij} = \beta_0 + \beta_1 y_{2,ij} + b_i + \varepsilon_{ij}, \quad (3)$$

where  $b_i$  and  $\varepsilon_i$  are the random errors for level-1 and level-2 observational units, and assumed to be  $N(0, \sigma_b^2)$  and  $N(0, \sigma_\varepsilon^2)$ . In a typical analysis, the parameters of interest are  $\beta = (\beta_0, \beta_1)$ ,  $\sigma_b$ . For each of the imputed datasets created under MLMM model (1) via the R package pan, these parameters were estimated using the R package called lme4 (Bates and Sarkar, 2007). In Section 4.2, the performance of MI under MLMM in estimating these unknown quantities is summarized.

One of the important issues in MI inference is the compatibility of the MI model with the analyst's model as well as the data-generation mechanism. MI practitioners are typically advised to include auxiliary variables either to account for the causes of missingness or to improve the missing data procedure. Collins et al. (2001) discuss this subject extensively in cross-sectional settings and present a simulation study assessing the benefits and costs of inclusive and restrictive MI strategies. Theoretical aspects of the compatibility of the MI model with the analyst's model was earlier explored by Meng (1994), who also coined the term "uncongeniality" (i.e. incompatibility of MI and analyst's model). In general, to prevent

inferential biases, the MI models should be kept general enough to imply any particular analyst's model.

**4.1.4. Evaluation criteria**—Multiple imputations were created using the R package *pan*. The simulation procedure (data generation, imposing missing values, creating MIs, finding the estimates of the analyst's model (3) for each of the imputed datasets, and combining these estimates using Rubin's rules) was repeated 1000 times to fully describe the behavior in a repetitive sampling environment as well as to reduce the simulation error as much as possible. Several quantities were used to measure the accuracy as well as the efficiency of parameter estimates under MLMM-based MI when the distributional assumptions are violated. Briefly, these quantities are (assuming the unknown parameter is  $\theta$ ):

- **Coverage rate (CR):** The percentage of times that the true parameter value is covered in the 95% confidence interval. Here the true parameter value is the average parameter estimate across the simulations before the missing values are imposed. If a procedure is working well, the actual coverage should be close to the nominal rate of 95% in our study. However, it is important to evaluate coverage with the other measures because high variances can lead to higher CRs. The performance of the procedure is regarded to be poor if its coverage drops below 90% (Collins et al., 2001). If the procedure results in CRs that are close to 100% or below 85%, extra caution should be taken when using that procedure.
- **Root-mean-square error (RMSE):** Because nonresponse or missing values have undesirable effect on the variances, it is important to evaluate this adverse effect. An integrated measure of bias and variance is used, evaluating  $\hat{\theta}$  in terms of combined accuracy and precision.  $RMSE(\hat{\theta})$  is defined as  $\sqrt{E(\hat{\theta} - \theta)^2}$ .
- **Average width of confidence interval (AW):** The distance between the average lower and upper confidence interval limits across 1000 confidence intervals. A high CR along with narrow, calibrated confidence intervals translates into greater accuracy and higher power.

**4.1.5. Notes on computation**—As noted in Section 3, it is important to monitor the convergence of the Gibbs sampler in the R package *pan*, the computational engine for producing the multiple imputations under model MLMM (1). To assess the convergence, the Gibbs sampler was run for 10,000 cycles under mild prior distributions; we set

$\nu_1=2, \Lambda_1^{-1}=2\hat{\Sigma}, \nu_2=2$ , and  $\Lambda_2^{-1}=2\hat{\Psi}$  where  $\hat{\Sigma}$  and  $\hat{\Psi}$  were computed by the R package *mlmmm* (Yucel, 2007). We assessed time-series plots and sample autocorrelations for elements of  $\beta$ ,  $\Sigma$  and  $\Psi$  for several of the simulations. This exercise suggested that, even in the extreme cases of  $m = 100, n_i = 10$  and  $m = 20, n_i = 5$ , several thousand cycles were needed for the dependence between successive Gibbs draws to die out. To be quite conservative for the convergence, each of the simulations in all scenarios used a burn-in period of 10,000 cycles and saved the simulated values of missing data at 11,000, 12,000, ..., 20,000 as multiple imputations. This required about 45 min on a 2.6 GHz Intel Core Duo workstation for each of the simulation scenarios presented in Tables 1–4. These multiply imputed datasets were used to estimate the model given in (3) using the R package *lme4* (Bates and Sarkar, 2007), leading to 10 sets of  $(\hat{\beta}, \widehat{s.e.}(\hat{\beta})), (\hat{\sigma}_b, \widehat{s.e.}(\hat{\sigma}_b))$  which were then combined using the MI combining rules defined by Rubin (1987).

## 4.2. Results

The simulation results focus on the performance of the MLMM-based MI on the parameter estimates of the analyst's model given by (3) (i.e.  $\hat{\beta}_0, \hat{\beta}_1$  and  $\hat{\sigma}_b$ ). Results show that the residual



variance term is mostly unaffected by the current model misspecifications, and are, therefore, not reported here. Tables 1–4 present the results from scenarios that mimic a missing covariate ( $y_2$  missing on average 45% under MAR) and an arbitrary missingness (both  $y_1$  and  $y_2$  missing on average 45% and 55%, respectively, under MCAR). The first column of Tables 1–4 denotes the respective distributions of the random-effects and error terms used to simulate data and sample size ( $m, n_i$ ). The second column denotes the parameter of interest (PI):  $\beta_0, \beta_1$ , and  $\sigma_b$ . The next column contains the true value (TV) of the PI. The columns 4–7 state the evaluation measures for each of the unique combination of the first three column values. These are evaluation measures under the MI inference (first entry) and complete-case only or case deletion (second entry). For example, when both the random-effects and residuals are assumed to be distributed as normal with  $m = 100, n_i = 10$ , the true value for  $\beta_0$  is calculated to be 2.713, as the average of  $\beta_0$  across the 1000 simulated data before missing values are imposed. The average MI-estimate across the 1000 simulations is  $-2.694$  and the average complete-case-only (restricting the analysis to completely-observed cases) estimate is  $-1.398$ . Corresponding coverage rates are 93% and 8.7%, respectively. Either method of handling missing data uses the same model-fitting technique as implemented by the R package lme4 (Bates and Sarkar, 2007).

Tables 1 and 2 show that, across the board, the MI inferences lead to acceptable parameter estimation and coverage, indicating a good performance of the MLM-based MI procedure. We specifically observe estimates with minimal biases and CIs with excellent coverage rates, even when the distribution assumption is clearly violated and when the sample sizes are small. The performance of the MI is far superior to the unprincipled method of simple case-deletion, which leads to significant biases as well as dismal coverage rates (in some cases it is as dramatic as 0%). It is also noted that in some conditions the performance of the MI procedure is not as satisfactory. When the underlying distribution of the random-effects has heavier tails than normal, specifically a  $t$  distribution (see the bottom panel in Table 1), we observe weak performance (e.g. higher RMSEs, lower CRs and larger AWs) in the estimation of  $\sigma_b$ . However, in most conditions, regardless of the distributional assumption of the random-effects, the MI procedure results in well-calibrated estimates with minimal bias and coverage rates that are around the nominal rate of 95%. This consistent performance across the distributions and the varying sample sizes indicate that, under the MLM-based MI inference, large and small sample properties of the maximum likelihood estimates are satisfied (Demidenko, 2004). Examples of these properties include unbiasedness, smaller uncertainty measures, and decreased RMSE for increased sample size. In the mixed-effects framework, we follow the convention of large  $m$  and  $n_i$  for “large sample” while noting that the large  $m$  is more important for the large sample properties to hold. For more discussion see Demidenko (2004).

Overall performance of the MLM-based MI is weaker when both variables are subject to missingness under the non-normal random-effects. We note bigger biases, relatively lower CRs, higher RMSEs and larger AWs in Tables 3 and 4 compared to Tables 1 and 2. This simply means that under arbitrary patterns of missing data, the inferences are more vulnerable to the misspecification of the random-effect distribution as well as to the small sample sizes. The estimates describing the associations and the spread of the random-effect distribution are most impacted. There is little discernible difference in the overall performance in the estimation of  $\beta_0$  between the two missing-data scenarios (univariate versus arbitrary). When both response and covariate are subject to missingness, even though the estimation of  $\beta_1$  is still impacted by the relatively higher RMSE, AW, the CR and small biases indicate that the inference on the fixed-effects remains insensitive to the non-normality of the random-effects and to the small sample sizes. This essentially implies that the MI inference leads to estimates with good frequentist properties even when the joint aspects of data are not well-estimated due to arbitrary patterns of missing data. However, when the distribution of random-effects is not normal, the

validity of the MI inferences on the random-effect variance is highly questionable in terms of bias, CR or RMSE.

Recall that data generation scenario (given by model (2)) assumes a weak correlation between  $Y_1$  and  $Y_2$ . To assess whether this assumption has any discernible impact on the performance measures, a similar simulation experiment to the one presented above under three different conditions for weak, moderate and high correlations among  $Y_1$  and  $Y_2$  was also conducted. All scenarios are in agreement on well-calibrated MI inference under MLMM. Higher correlations between  $Y_1$  and  $Y_2$  translate into the improved accuracy measures of performance, such as reduced AWs, higher rates of coverage, and lower RMSE. However, it should be noted that even when  $Y_1$  and  $Y_2$  are weakly correlated, these performance measures still indicate inferences with desirable properties (e.g. unbiasedness, low average width, low RMSE, close to 95% coverage rates). The effect of the correlations has little impact on the overall performance measures (Table 5).

## 5. Discussion

The overall goal of our work was to assess the impact of the normality assumption of the random effects on the MI-based inferences in multilevel incomplete data applications. In problems where the missing values are confined to a small number of variables, or when there are sources for auxiliary information that can be used as predictors in the imputation model, the assumption of normality does not seem to matter on the final inferences, even when the random-effects are wrongly assumed to be normal. However, when the missingness is arbitrary in many variables and the rates of missingness are high (or when the sample sizes are small), the validity of MI-based estimation is more adversely affected under the similar non-normal random effects. For incorrectly specified distributions, this negative effect is more clearly seen in parameters describing the relationships and/or the distribution of the random-effects. This is somewhat expected, as very little useful data might be available to estimate the “relatedness” aspects of the distribution. Our simulation study also suggests that taking no action about the missing data can lead to inferences that are far more misleading, regardless of correct or incorrect specification of distributional assumptions.

For the purposes of assessing the sensitivity of MLMM-based MI to distributional assumptions on the random-effects, a limited number of distinct distributions was considered. Therefore, the simulation study does not claim to be generalizable, as there are many other scenarios that can be potentially encountered in real data applications. However, we believe that our simulation study is valuable in the sense that it sheds some light on the effects of making wrong assumptions on the distribution of the random effects (or any distribution on the subject-specific effects when there is not much data support for estimation) in the imputation model.

MI-based procedures, when cautiously used, produce reasonably well-calibrated inference as long as the missingness is not widely spread across the data. However, it is important for the practitioners of missing data methods not to use the MLMM-based software blindly. Similar to the general practice of model-fitting and the associated checks for assessing the goodness-of-fit, inference by multiple imputation should always be performed in conjunction with the tools necessary to assess the plausibility of the underlying assumptions. What are these tools? While they are not as widely implemented as the imputation procedures, there are still well-established techniques for this purpose. Posterior predictive checks, as suggested by Gelman et al. (1996, 2003), can be employed to assess the plausibility of the imputation model. As a direct check for the quality of the assumed model in creating imputations, one can devise a calibration strategy to make the imputations better represent the structure of the data and possibly the varying structure of the clusters. Post-imputation checks and calibration-based methods are presented in detail by Gelman et al. (2005) and Yucel et al. (2008).

The results suggest a number of directions for future research. The first direction is the imputation models with less restrictive distributional assumptions such as those based on marginal models. These models can be particularly useful in small sample scenarios, especially when the missingness is seen indiscriminately in the response or covariates. A second potentially fruitful direction is to explore the ad-hoc methods of imputation, e.g. predictive mean matching applied to multilevel settings. These methods usually make minimal assumptions, but may lead to underestimation of the standard errors in the multilevel settings. However, with improvements to adjust the standard errors, the pay-off might be considerable when there is a reasonable hesitation to use a fully parametric approach for multiple imputation. Finally, we believe that the most important direction is the dissemination of any computational development and/or improvements by means of statistical software.

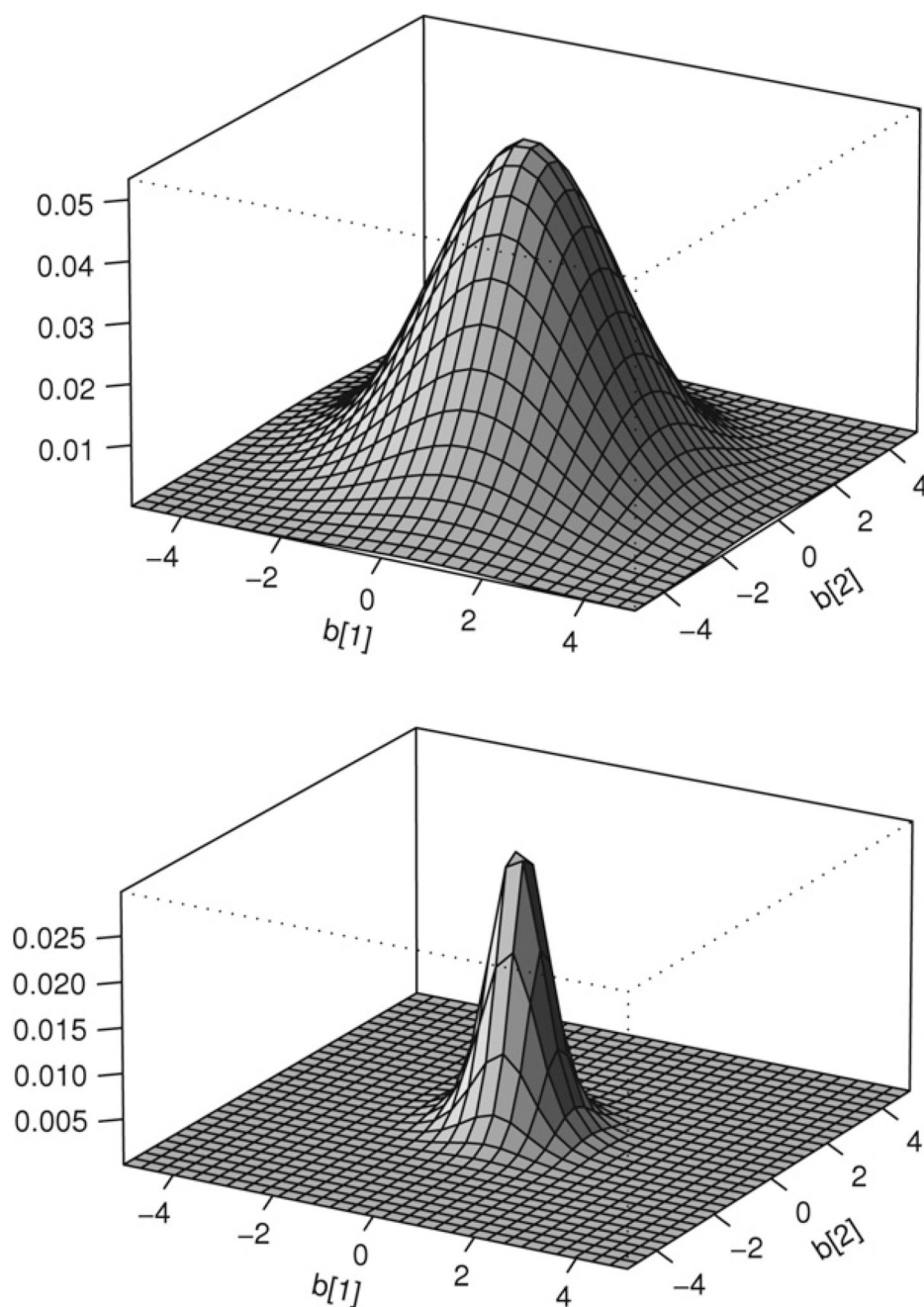
## Acknowledgments

The authors thank the associate reviewer and the reviewer for helpful comments which greatly improved our manuscript. First author's research was partially supported by the National Science Foundation.

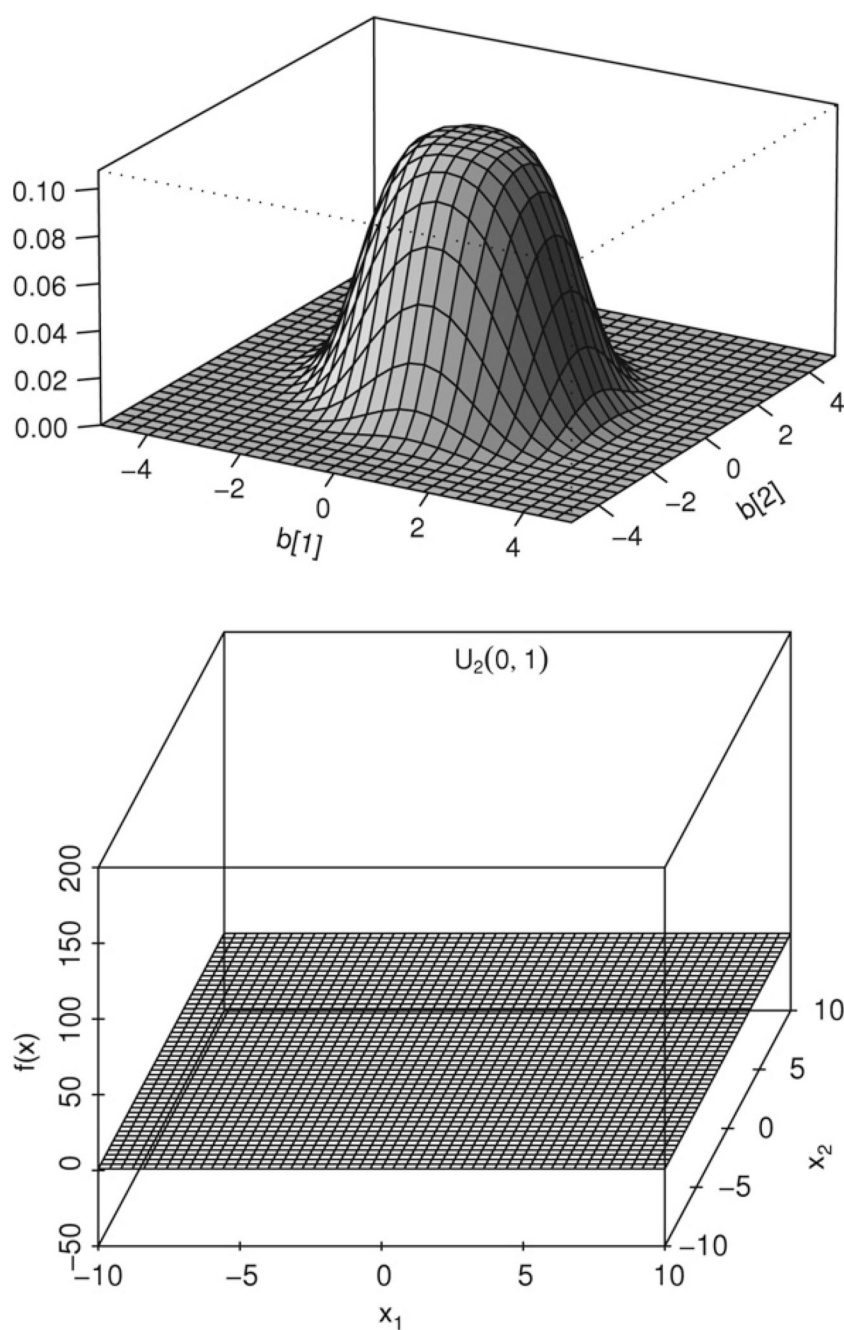
## References

- Bates D, Sarkar D. lme4: Linear mixed-effects models using s4 classes. 2007
- Browne, WJ. MCMC estimation in MLwiN (version 2.0). London, UK: Institute of Education, University of London; 2003.
- Carpenter J, Goldstein H. Multiple imputation in MLwiN. Multilevel modelling newsletter 4. 2004
- Collins L, Schafer J, Kam C-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* 2001;6:330–351. [PubMed: 11778676]
- Demidenko, E. *Mixed Models: Theory and Applications*. New York, NY: J. Wiley & Sons; 2004.
- Demirtas H. Simulation-driven inferences for multiply imputed longitudinal datasets. *Statistica Neerlandica* 2004;58:466–482.
- Demirtas H. Multiple imputation under Bayesianly smoothed pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* 2005;24:2345–2363. [PubMed: 15977286]
- Demirtas H, Freels S, Yucel R. Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: A simulation assessment. *Journal of Statistical Computation and Simulation* 2008;78:69–84.
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian Data Analysis*. second ed.. London: Chapman & Hall Ltd.; 2003.
- Gelman A, Mechelen I, Verbeke G, Heitjan D, Meulders M. Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics* 2005;61:74–85. [PubMed: 15737080]
- Gelman A, Meng X, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 1996;6:733–807.
- Little, R.; Rubin, D. *Statistical Analysis with Missing Data*. second ed.. New York, NY: J. Wiley & Sons; 2002.
- Meng X. Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* 1994;10:538–573.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R foundation for statistical Computing. Vienna, Australia: 2007. ISBN: 3-900051-07-0, URL: <http://www.R-project.org>
- Rasbash, J.; Steel, F.; Browne, W.; Prosser, B. *MLwin User's Manual*. Centre for Multilevel Modelling. Bristol, UK: 2006.
- Rubin DB. Inference and missing data. *Biometrika* 1976;63:581–590.
- Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons; 1987.
- Schafer, J. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997.
- Schafer J, Graham J. Missing data: Our view of the state of the art. *Psychological Methods* 2002;7:147–177. [PubMed: 12090408]

- Schafer J, Yucel R. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* 2002;11:421–442.
- Yucel, R. R mlmmm package: Fitting multivariate linear mixed-effects models with missing values. 2007. URL <http://www.R-project.org>
- Yucel R, He Y, Zaslavsky A. Using calibration to improve rounding in imputation. *The American Statistician* 2008;62:125–129.



**Fig. 1.**  
Bivariate normal and  $t$  (with scale parameter 1 and 10 degrees of freedom) distributions.



**Fig. 2.**  
Bivariate Laplace and uniform distributions.



**Table 1**

Summary of the results under random-effects following a multivariate normal and  $t_4$ . Missingness rate is (approximately) 45% on  $y_2$ . For each parameter of interest (PI), MI results and complete-case only analysis are given in the first and second rows, respectively.  $\beta_0$ ,  $\beta_1$  and  $\sigma_b^2$  are the parameters of model (3). For the acronyms, see page 12.

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma), N(\beta, \Psi)$ (100, 10)	$\beta_0$	-2.713	-2.694	0.292	0.930	1.110
			-1.398	1.335	0.087	0.785
	$\beta_1$	0.365	0.360	0.077	0.931	0.299
			0.161	0.212	0.034	0.192
	$\sigma_b$	2.103	2.102	0.372	0.946	1.519
			1.030	1.096	0.390	0.814
$N(0, \Sigma), N(\beta, \Psi)$ (20, 10)	$\beta_0$	-2.715	-2.610	0.608	0.931	2.490
			-1.411	1.405	0.217	1.752
	$\beta_1$	0.365	0.338	0.161	0.938	0.660
			0.167	0.242	0.517	0.428
	$\sigma_b$	2.115	2.103	0.835	0.928	4.112
			1.040	1.191	0.353	2.277
$N(0, \Sigma), N(\beta, \Psi)$ (50, 5)	$\beta_0$	-2.336	-2.279	0.481	0.944	1.987
			-0.989	1.390	0.038	1.287
	$\beta_1$	0.250	0.236	0.138	0.937	0.570
			0.072	0.199	0.450	0.331
	$\sigma_b$	1.820	1.836	0.525	0.939	2.145
			0.876	0.991	0.542	1.207
$N(0, \Sigma), N(\beta, \Psi)$ (20, 5)	$\beta_0$	-2.341	-2.222	0.753	0.935	3.094
			-1.014	1.463	0.315	2.047
	$\beta_1$	0.254	0.230	0.203	0.938	0.874
			0.081	0.235	0.690	0.529
	$\sigma_b$	1.876	1.888	0.817	0.932	4.082
			0.910	1.096	0.405	2.334

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma), t_4(\beta, \Psi)$ (100, 10)	$\beta_0$	-2.684	-2.675	0.323	0.944	1.251
			-1.184	1.666	0.001	0.893
	$\beta_1$	0.361	0.361	0.073	0.949	0.294
			0.167	0.203	0.060	0.191
	$\sigma_b$	5.539	5.466	6.167	0.821	3.926
			2.327	18.143	0.164	1.826
$N(0, \Sigma), t_4(\beta, \Psi)$ (20, 10)	$\beta_0$	-2.668	-2.552	0.848	0.927	2.866
			-1.162	1.838	0.234	2.040
	$\beta_1$	0.367	0.338	0.165	0.925	0.645
			0.177	0.237	0.539	0.425
	$\sigma_b$	6.397	6.355	9.189	0.885	12.855
			2.986	27.686	0.385	7.558
$N(0, \Sigma), t_4(\beta, \Psi)$ (50, 5)	$\beta_0$	-2.333	-2.275	0.578	0.944	2.174
			-0.730	1.794	0.052	1.433
	$\beta_1$	0.249	0.241	0.140	0.934	0.568
			0.080	0.196	0.454	0.330
	$\sigma_b$	4.951	4.879	6.497	0.850	5.600
			2.014	11.870	0.252	2.660
$N(0, \Sigma), t_4(\beta, \Psi)$ (20, 5)	$\beta_0$	-2.387	-2.182	0.950	0.920	3.388
			-0.743	2.023	0.303	2.332
	$\beta_1$	0.261	0.228	0.217	0.920	0.841
			0.096	0.228	0.692	0.528
	$\sigma_b$	6.182	5.528	8.432	0.867	11.539
			2.410	17.556	0.170	7.191

Table 2

Summary of the results under random-effects following a multivariate Laplace and Uniform. Missingness rate is (approximately) 45% on  $y_2$ . Format is the same as Table 1.

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (100, 10)	$\beta_0$	-2.694	-2.670	0.279	0.952	1.108
			-1.382	1.331	0.000	0.786
	$\beta_1$	0.362	0.355	0.074	0.947	0.299
			0.158	0.212	0.038	0.192
	$\sigma_b$	2.109	2.114	0.324	0.964	1.531
			1.054	1.073	0.399	0.837
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (20, 10)	$\beta_0$	-2.716	-2.618	0.626	0.938	2.468
			-1.432	1.390	0.253	1.779
	$\beta_1$	0.368	0.342	0.166	0.917	0.653
			0.172	0.239	0.537	0.432
	$\sigma_b$	2.138	2.139	0.785	0.960	4.136
			1.098	1.154	0.379	2.414
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (50, 5)	$\beta_0$	-2.335	-2.258	0.478	0.951	1.950
			-0.988	1.389	0.034	1.288
	$\beta_1$	0.251	0.232	0.135	0.943	0.555
			0.074	0.198	0.450	0.330
	$\sigma_b$	1.828	1.823	0.442	0.957	2.131
			0.899	0.972	0.353	1.243
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (20, 5)	$\beta_0$	-2.334	-2.184	0.696	0.951	3.041
			-1.006	1.443	0.317	2.046
	$\beta_1$	0.258	0.223	0.194	0.952	0.862
			0.082	0.229	0.406	0.529
	$\sigma_b$	1.882	1.889	0.727	0.953	4.042
			0.940	1.065	0.210	2.403

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma)$ , Uniform( $\beta$ ) (100, 10)	$\beta_0$	0.284	0.285	0.053	0.942	0.206
			0.461	0.185	0.377	0.200
	$\beta_1$	0.429	0.428	0.051	0.953	0.199
			0.373	0.073	0.166	0.178
	$\sigma_b$	0.139	0.139	0.029	0.967	0.129
			0.111	0.039	0.000	0.115
$N(0, \Sigma)$ , Uniform( $\beta$ ) (20, 10)	$\beta_0$	0.282	0.287	0.116	0.956	0.463
			0.460	0.211	0.614	0.444
	$\beta_1$	0.434	0.427	0.113	0.942	0.446
			0.377	0.119	0.200	0.398
	$\sigma_b$	0.143	0.143	0.068	0.922	0.340
			0.112	0.070	0.002	0.278
$N(0, \Sigma)$ , Uniform( $\beta$ ) (50, 5)	$\beta_0$	0.295	0.300	0.094	0.945	0.357
			0.476	0.203	0.487	0.352
	$\beta_1$	0.402	0.396	0.101	0.946	0.396
			0.347	0.107	0.299	0.351
	$\sigma_b$	0.138	0.139	0.062	0.901	0.241
			0.111	0.065	0.000	0.206
$N(0, \Sigma)$ , Uniform( $\beta$ ) (20, 5)	$\beta_0$	0.305	0.315	0.147	0.944	0.563
			0.480	0.228	0.337	0.554
	$\beta_1$	0.397	0.382	0.158	0.949	0.627
			0.347	0.156	0.415	0.557
	$\sigma_b$	0.130	0.131	0.090	0.852	0.382
			0.106	0.088	0.003	0.314

**Table 3**

Summary of the results under random-effects following a multivariate normal and  $t_4$ . Missingness rates are around 45% and 55% on  $y_1$  and  $y_2$ , respectively. Format is the same as Table 1.

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma), N(\beta, \Psi)$ (100, 10)	$\beta_0$	-2.697	-2.765	0.346	0.946	1.373
			-1.226	1.497	0.191	0.938
	$\beta_1$	0.364	0.388	0.094	0.940	0.375
			0.096	0.276	0.015	0.226
	$\sigma_b$	2.111	2.198	0.445	0.953	1.795
			0.983	1.149	0.187	0.844
$N(0, \Sigma), N(\beta, \Psi)$ (20, 10)	$\beta_0$	-2.715	-2.672	0.700	0.954	3.053
			-1.228	1.608	0.265	2.123
	$\beta_1$	0.366	0.357	0.189	0.946	0.819
			0.097	0.311	0.246	0.511
	$\sigma_b$	2.171	2.294	1.028	0.922	4.860
			1.011	1.280	0.140	2.392
$N(0, \Sigma), N(\beta, \Psi)$ (50, 5)	$\beta_0$	-2.322	-2.364	0.605	0.947	2.546
			-0.900	1.495	0.120	1.677
	$\beta_1$	0.247	0.268	0.168	0.943	0.719
			0.027	0.249	0.040	0.409
	$\sigma_b$	1.844	1.946	0.663	0.936	2.831
			0.905	1.004	0.133	1.426
$N(0, \Sigma), N(\beta, \Psi)$ (20, 5)	$\beta_0$	-2.372	-2.287	0.809	0.961	3.877
			-0.969	1.603	0.434	2.673
	$\beta_1$	0.258	0.244	0.218	0.968	1.082
			0.041	0.289	0.182	0.656
	$\sigma_b$	1.851	1.902	0.915	0.931	4.943
			0.908	1.124	0.241	2.670

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma), t_4(\beta, \Psi)$ (100, 10)	$\beta_0$	-2.699	-2.769	0.415	0.925	1.527
			-1.017	1.782	0.001	1.108
	$\beta_1$	0.362	0.386	0.097	0.927	0.378
			0.098	0.274	0.028	0.231
	$\sigma_b$	5.737	5.331	8.630	0.256	4.475
			2.392	5.861	0.022	2.138
$N(0, \Sigma), t_4(\beta, \Psi)$ (20, 10)	$\beta_0$	-2.735	-2.687	0.901	0.944	3.453
			-1.060	1.885	0.172	2.487
	$\beta_1$	0.370	0.362	0.187	0.942	0.827
			0.107	0.311	0.185	0.524
	$\sigma_b$	5.660	5.666	10.057	0.503	12.986
			2.338	5.805	0.213	6.682
$N(0, \Sigma), t_4(\beta, \Psi)$ (50, 5)	$\beta_0$	-2.307	-2.363	0.697	0.945	2.671
			-0.680	1.834	0.137	1.908
	$\beta_1$	0.247	0.270	0.170	0.939	0.697
			0.038	0.242	0.179	0.411
	$\sigma_b$	4.989	4.359	7.845	0.361	5.999
			2.022	4.880	0.167	3.167
$N(0, \Sigma), t_4(\beta, \Psi)$ (20, 5)	$\beta_0$	-2.384	-2.281	1.050	0.947	4.320
			-0.714	2.040	0.117	3.184
	$\beta_1$	0.264	0.246	0.234	0.935	1.087
			0.048	0.296	0.063	0.673
	$\sigma_b$	8.665	5.587	25.985	0.294	14.317
			3.145	25.892	0.172	12.429



**Table 4**

Summary of the results under random-effects distributed as Laplace and Uniform. Missingness rates are around 45% and 55% on  $y_1$  and  $y_2$ , respectively. Format is the same as Table 1.

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (100, 10)	$\beta_0$	-2.697	-2.765	0.346	0.946	1.373
			-1.226	1.497	0.008	0.938
	$\beta_1$	0.364	0.388	0.094	0.940	0.375
			0.096	0.276	0.015	0.226
	$\sigma_b$	2.111	2.198	0.445	0.953	1.795
			0.983	1.149	0.287	0.844
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (20, 10)	$\beta_0$	-2.715	-2.672	0.700	0.954	3.053
			-1.228	1.608	0.265	2.123
	$\beta_1$	0.366	0.357	0.189	0.946	0.819
			0.097	0.311	0.346	0.511
	$\sigma_b$	2.171	2.294	1.028	0.922	4.860
			1.011	1.280	0.340	2.392
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (50, 5)	$\beta_0$	-2.322	-2.364	0.605	0.947	2.546
			-0.900	1.495	0.120	1.677
	$\beta_1$	0.247	0.268	0.168	0.943	0.719
			0.027	0.249	0.240	0.409
	$\sigma_b$	1.844	1.946	0.663	0.936	2.831
			0.905	1.004	0.333	1.426
$N(0, \Sigma)$ , Laplace( $\beta, \Psi$ ) (20, 5)	$\beta_0$	-2.372	-2.287	0.809	0.961	3.877
			-0.969	1.603	0.434	2.673
	$\beta_1$	0.258	0.244	0.218	0.968	1.082
			0.041	0.289	0.182	0.656
	$\sigma_b$	1.851	1.902	0.915	0.931	4.943
			0.908	1.124	0.241	2.670

Distribution ( $m, n_i$ )	PI	TV	AE	RMSE	CR	AW
$N(0, \Sigma)$ , Uniform( $\beta$ ) (100, 10)	$\beta_0$	-2.699	-2.769	0.415	0.925	1.527
			-1.017	1.782	0.001	1.108
	$\beta_1$	0.362	0.386	0.097	0.927	0.378
			0.098	0.274	0.028	0.231
	$\sigma_b$	5.737	5.331	8.630	0.256	4.475
$N(0, \Sigma)$ , Uniform( $\beta$ ) (20, 10)	$\beta_0$	-2.735	-2.687	0.901	0.944	3.453
			-1.060	1.885	0.272	2.487
	$\beta_1$	0.370	0.362	0.187	0.942	0.827
			0.107	0.311	0.185	0.524
	$\sigma_b$	5.660	5.666	10.057	0.503	12.986
$N(0, \Sigma)$ , Uniform( $\beta$ ) (50, 5)	$\beta_0$	-2.307	-2.363	0.697	0.945	2.671
			-0.680	1.834	0.137	1.908
	$\beta_1$	0.247	0.270	0.170	0.939	0.697
			0.038	0.242	0.179	0.411
	$\sigma_b$	4.989	4.359	7.845	0.361	5.999
$N(0, \Sigma)$ , Uniform( $\beta$ ) (20, 5)	$\beta_0$	-2.384	-2.281	1.050	0.947	4.320
			-0.714	2.040	0.217	3.184
	$\beta_1$	0.264	0.246	0.234	0.935	1.087
			0.048	0.296	0.163	0.673
	$\sigma_b$	8.665	3.145	25.892	0.372	12.429
			5.587	25.985	0.294	14.317

Table 5

Results of performance measures under weak, moderate and high correlation among  $Y$ -variables of model (1), sample size is assumed to 100 clusters with 10 units each. Normal random-effects are assumed. All the other simulation conditions are same, with the exceptions of  $\Sigma$  and  $\Psi$  which are set to values to lead to correlations stated below.

$\text{Corr}(Y_1, Y_2)$	Parameter interest	True value	Avg. est.	RMSE	Cvg. rate	Avg. width
0.20	$\beta_0$	-2.713	-2.694	0.292	0.930	1.110
	$\beta_1$	0.365	0.360	0.077	0.931	0.299
	$\sigma_b$	2.103	2.102	0.372	0.946	1.519
0.50	$\beta_0$	-1.535	1.541	0.187	0.946	0.836
	$\beta_1$	0.265	0.268	0.059	0.944	0.240
	$\sigma_b$	2.831	2.719	0.372	0.946	1.305
0.80	$\beta_0$	-5.021	-5.015	0.158	0.950	0.637
	$\beta_1$	1.069	1.066	0.023	0.954	0.095
	$\sigma_b$	1.541	1.545	0.249	0.952	1.011