Multiple Imputation of Missing Covariate Values in Multilevel Models With Random Slopes: A

Cautionary Note

Simon Grund and Oliver Lüdtke

Leibniz Institute for Science and Mathematics Education;

Centre for International Student Assessment

Alexander Robitzsch

Federal Institute for Education Research, Innovation and Development of the Austrian School

System

Author Note

Abstract

Multiple imputation (MI) has become one of the main procedures used to treat missing data, but guidelines from the methodological literature are not easily transferred to multilevel research. For models including random slopes, proper MI can be difficult, especially when covariate values are partially missing. In the present paper, we discuss applications of MI in multilevel random coefficient models, theoretical challenges posed by slope variation, and current limitations of standard MI software. Our findings from three simulation studies suggest that (a) MI is able to recover most parameters but is currently not well suited to capture slope variation entirely when covariate values are missing, (b) MI offers reasonable estimates for most parameters even in smaller samples or when its assumptions are not met, and (c) listwise can be an alternative worth considering when preserving the slope variance is particularly important.

Keywords: missing data, multilevel, random slopes, multiple imputation, listwise deletion, covariate

Multiple Imputation of Missing Covariate Values in Multilevel Models With Random Slopes: A

Cautionary Note

Multilevel data are often found in psychological research. The complex pattern of

variability in such data allows the use of statistical models that can accommodate multiple sources

of variation. In recent years, multilevel models have become the standard tool for analyzing such

data structures (Raudenbush & Bryk, 2002). In the social sciences, missing data (MD) represent a

pervasive problem that has received considerable attention during the last two decades. There is

consensus in the methodological literature that methods such as multiple imputation (MI) are

much better suited for treating missing data than traditional approaches such as listwise or

pairwise deletion (Little & Rubin, 2002; Schafer & Graham, 2002).

Although several book-length treatises have familiarized applied researchers with modern

missing data methods (Allison, 2001; Enders, 2010; Graham, 2012; van Buuren, 2012), less has

been said about how to deal with missing values in multilevel research. Previous studies concerned

with missing data in multilevel modeling have consistently found that parameter estimates can be

seriously distorted if the multilevel structure is not taken into account in the imputation process

(Andridge, 2011; van Buuren, 2011). However, these studies have focused on random intercept

models, which assume that relations between variables do not vary across groups.

In the present article, we focused on random slope models. These models are frequently

used in organizational and educational research to investigate whether relations at Level 1 (e.g.,

students, employees) vary across Level 2 units (e.g., classes, working teams) or in longitudinal

research to assess different developmental trajectories across subjects. For example, Hochweber,

Hosenfeld, and Klieme (2014) investigated the relationship between students' mathematics

achievement and their math grades. Hussong et al. (2008) examined the effects of parents' alcohol

abuse on the development of children's internalizing behavior.

Using a multivariate mixed effects model and the software *pan*, we explored strategies for

dealing with missing data in models with random slopes. In three simulation studies, we

considered incomplete data on outcome variables, predictor variables, and both variables

simultaneously, as well as different sample properties and patterns of missing data.

Missing Data in Multilevel Research

The multivariate mixed effects model.   A few options for treating missing data in

multilevel models are available in standard statistical software. The pan package (Schafer & Zhao,

2014) has been recommended for MI of multilevel data (Enders, 2010; Graham, 2012) and is

easily accessible through the statistical software R (R Core Team, 2014). The statistical model

behind pan, which we will refer to as the *pan model*, is the multivariate mixed effects model as

presented by Schafer (1997). The pan model is capable of treating multilevel missing data but

may also be used to describe both the imputation and analysis of multilevel data. The model reads

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{b}_j + \mathbf{E}_j, \tag{1}$$

where $j = 1, \ldots, G$ denotes groups or other observational units at Level 2. Here, the response

matrix $\mathbf{Y}_j$ of group j is regressed on a design matrix $\mathbf{X}_j$ (containing intercept and predictor

values) with associated *fixed effects* $\boldsymbol{\beta}$ and a design matrix $\mathbf{Z}_j$ with associated group-specific

random effects $\mathbf{b}_j$. The random effects matrix $\mathbf{b}_j$ (with columns stacked) is assumed to follow a

normal distribution with mean zero and covariance matrix $\boldsymbol{\Psi}$ (iid. for all groups). Each row of the

error matrix $\mathbf{E}_j$ is assumed to follow a normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$ (iid. for all individuals). Note that in the pan model, $\boldsymbol{\Sigma}$ is missing the index j and is thus assumed to be the same for each group.

Suppose our dataset consists of two variables X and Y, both of which are Level 1 variables that have some variation at Level 2. In a special application of the pan model, we may want to estimate the regression of Y on X with varying coefficients across groups, that is, the random coefficient model (RC model). This model results if we write the outcome Y (e.g., students' math grades) on the left-hand side of Equation 1, and the covariate X (e.g., individual achievement) on the right-hand side, and allow for the intercepts and slopes to vary across groups. We will also call the RC model the *analyst's model* as it fits our supposed research question. Finally, we can express the parameters of the analyst's model in a single expression $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma})$ and write $f(Y|X, \boldsymbol{\theta}_X)$ in short for the RC model.

Missing data terminology.   The common classification of missing data mechanisms found in Rubin (1987) assumes a hypothetical complete data matrix, which is decomposed into observed and unobserved parts $\mathbf{Y} = (\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$ by an indicator matrix $\mathbf{R}$ denoting the missing data. If values are missing as a random sample of the hypothetical complete data, that is, $P(\mathbf{R}|\mathbf{Y}) = P(\mathbf{R})$, the data are missing completely at random (MCAR). If missingness depends on other variables but the data are MCAR with these partialled out, that is, $P(\mathbf{R}|\mathbf{Y}) = P(\mathbf{R}|\mathbf{Y}_{obs})$, the data are missing at random (MAR). These two missing data mechanisms are often called "ignorable". An ignorable missing data mechanism is highly beneficial for MI because all of the relevant information about the missing values is present in the dataset. This is in contrast to data that are missing *not* at random (MNAR) where missingness is additionally dependent on the missing part of the data, that is, $P(\mathbf{R}|\mathbf{Y}) = P(\mathbf{R}|\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$. For such "nonignorable"

missingness, a general approach to an analysis of missing data is not feasible and strong

assumptions have to be made about the missing data mechanism (Carpenter & Kenward, 2013).

Multiple imputation for multilevel models. Multiple imputation, as introduced by Rubin

(1987), is a convenient procedure for obtaining valid parameter estimates from partially

unobserved data that usually relies on the MAR assumption (i.e., the observed values provide

sufficient information about the missing data mechanism). Using MI, the researcher draws

independent random samples from the posterior predictive distribution of the missing values

given the observed data and a statistical model, thus generating a number of complete datasets to

use in further analyses. The final parameter estimates can be obtained according to the rules

described by Rubin (1987) simply by averaging over the parameter estimates from all imputed

datasets. Applying MI can be subtle and need not always be the most practical choice in

multilevel research (Peters et al., 2012; Twisk, de Boer, de Vente, & Heymans, 2013) because its

validity is subject to some further conditions.

First, with increasing variation and sample size at Level 2, it becomes necessary to include

the multilevel structure in the imputation model. Ignoring the multilevel structure using

single-level MI may result in biased parameter estimates (Taljaard, Donner, & Klar, 2008; van

Buuren, 2011). Second, the analyst's model has to be considered, and the imputation model must

be specified accordingly (Meng, 1994; Schafer, 2003). Broadly speaking, the imputation model

must account for the complexity of the desired analysis. If an imputation model is used that does

not include variables or parameters relevant to the analyst (e.g., slope variance), then the analysis

results will be biased. And third, the imputation model must incorporate relevant information

about the missing data process, that is, variables predictive of missing variables or of the

missingness itself (Carpenter & Kenward, 2013), to make the MAR assumption more plausible

(Collins, Schafer, & Kam, 2001). Satisfying these conditions can be cumbersome when varying slopes are of interest. However, little is known about how the quality of parameter estimates in multilevel modeling is affected if one of these conditions is not met.

Missing covariates in models with random slopes. When only the outcome variable contains missing values, MI for a random coefficient model is straightforward. The imputation model can be specified in pan by writing the outcome on the left-hand side of Equation 1, and the covariate with fixed and random effects on the right-hand side. This is the previously mentioned RC model, denoted $f(Y|X, \theta_X)$. The imputation model is then equivalent to the analyst's model.

Fewer guidelines are available if a covariate contains missing values. If the outcome is completely observed, then a *reversed* imputation model may be used. For this model, the covariate is written on the left-hand side of Equation 1 and the outcome on the right-hand side (with fixed and random effects). We will refer to this as the reversed RC model and denote it $f(X|Y, \theta_Y)$. This model assumes slope variation but does so by regressing X on Y, which might induce bias into the parameter estimation. So far, pan has been recommended only for missing covariates whose effect is fixed across groups (Schafer, 1997). Alternatively, for a *multivariate* imputation model, denoted $f(X, Y|\theta_0)$, both variables could be written on the left-hand side of Equation 1 with random intercepts for both variables. Slope variation is ignored in this model, but in contrast to the conditional models (i.e., reversed and regular RC) it is able to account for multivariate patterns of missing data. An additional description of these models can be found in Supplement A in the online supplemental materials. The supplemental online materials can be downloaded from `http://dx.doi.org/10.6084/m9.figshare.1206375`.

In three simulation studies, we assessed the performance of conditional and multivariate MI for random slope models. Study 1, Study 2 and Study 3 examined cases in which missing values

occurred on the outcome, the covariate, or both variables, respectively. Study 1 attempted to replicate findings of previous research on partially observed outcome variables. We expected both conditional MI and LD to provide approximately unbiased estimates if the outcome was MAR (Carpenter & Kenward, 2013; Little & Rubin, 2002). Study 2 focused on missing covariate data. We expected that the reversed model would recover most parameters of the RC model but that it might perform poorly for the slope variance. Listwise deletion was expected to provide biased estimates with MAR and MNAR data. Study 3 examined multivariate missing data. We expected that multivariate MI would underestimate the slope variance but would recover most other parameters. We expected the results for LD to be similar to the results from the second study.

Study 1

The first study compared the performance of LD, conditional MI, and multivariate MI when the only outcome had missing values. For conditional MI, both the analyst's model $f(Y|X, \boldsymbol{\theta}_X)$ and the imputation model $g(Y|X, \boldsymbol{\omega}_X)$ were RC models where $\boldsymbol{\omega}_X$ took on the same role as $\boldsymbol{\theta}_X$ but denoted a distinct set of model parameters. These models were equally complex and fit the clustered structure of the data. Multivariate MI was set up as described earlier, and LD was applied by restricting the analysis to complete cases only.

Simulation and Methods

Data generation and imposition of missing values.   Two standardized normal variables X and Y were simulated. Both varied at two levels as indicated by their intraclass correlations (ICCs) ρ_X and ρ_Y, respectively. The covariate X was simulated from its within- and between-group portions $X^W \sim N(0, 1 - \rho_X)$ and $X^B \sim N(0, \rho_X)$, respectively. Then Y was

simulated conditionally on X according to Equation 1 with fixed effects $\boldsymbol{\beta} = (\beta_0, \beta_1)$, where β_0

was zero due to standardization. The covariance matrix of random effects was $\boldsymbol{\Psi} = \begin{pmatrix} \psi_{11}^2 & 0 \\ 0 & \psi_{22}^2 \end{pmatrix}$.

Thus, the intercepts and slopes were uncorrelated. The Level 1 residual variance was $\boldsymbol{\Sigma} = \sigma^2$. The

variables in this study were parametrized by their ICC rather than their actual variance

components. Given the ICC and a slope variance ψ_{22}^2, the other variance components followed

(see Snijders & Bosker, 2012) as

$$\sigma^2 = (1 - \rho_Y) - \beta_1^2(1 - \rho_X) - \psi_{22}^2(1 - \rho_X)$$
$$\psi_{11}^2 = \rho_Y - \beta_1^2 \rho_X - \psi_{22}^2 \rho_X. \tag{2}$$

Missing values on Y were imposed using a linear model for the latent response variable R^*.

Values in Y were set to be missing if their respective $R^* > 0$ according to

$$R^* = \alpha + \lambda_1 X + \lambda_2 Y + \varepsilon_{R^*}, \tag{3}$$

where α is a value of the standard normal distribution according to a missing data probability

(e.g., $\alpha = -0.67$ for 25% missing data), and λ_1 and λ_2 are used to control the missing data

mechanism. The residuals were distributed normally with mean zero and variance

$$\sigma_{R^*}^2 = 1 - \lambda_1^2 - \lambda_2^2 - 2\lambda_1\lambda_2 \text{Cov}(X, Y). \tag{4}$$

Table 1 provides an overview of the conditions included in all three studies. The two ICCs were

set to be equal, that is, $\rho_X = \rho_Y = \rho$. In order for Y to be MCAR, we set $\lambda_1 = \lambda_2 = 0$, and for

MAR, we set $\lambda_1 = 0.5$ and $\lambda_2 = 0$. For Y to be MNAR, we chose equal values for λ_1 and λ_2 such

that the error variance in R^* was the same as in the MAR condition. Hence, with

$\text{Cov}(X, Y) = \beta_1 = 0.5$, we had $\lambda_1 = \lambda_2 = \sqrt{0.25/3} \approx 0.289$. The conditions were chosen to mimic

typical data in psychology and the behavioral sciences (Aguinis, Gottfredson, & Culpepper, 2013; Mathieu, Aguinis, Culpepper, & Chen, 2012; Murray & Blitstein, 2003).

In summary, each simulated setting was defined by the number of groups (G), the number of individuals within each group (N), the ICC of X and Y (ρ), the fixed slope (β_1), the slope variance (ψ_{22}^2), the proportion of missing data, and the missing data mechanism (including the missing data effects λ_1 and λ_2). Each setting was replicated 1,000 times.

Imputation and data analysis. The R package pan was used to impute missing values (Schafer & Zhao, 2014). We let pan perform 10,000 burn-in cycles before drawing one imputed dataset for every 200 cycles, leading to $M = 50$ imputed datasets and 20,000 cycles in total (see Graham, Olchowski, & Gilreath, 2007). Diagnostic plots regarding the convergence behavior of pan's Gibbs sampler are presented in Supplement B in the online supplemental materials.

Least-informative inverse-Wishart priors for $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ were chosen with $\boldsymbol{\Sigma} \sim W^{-1}(\mathbf{I}_1, 1)$ and $\boldsymbol{\Psi} \sim W^{-1}(\mathbf{I}_2, 2)$ for conditional MI, and $\boldsymbol{\Sigma} \sim W^{-1}(\mathbf{I}_2, 2)$ and $\boldsymbol{\Psi} \sim W^{-1}(\mathbf{I}_2, 2)$ for multivariate MI, where $\mathbf{I}_n$ denotes the identity matrix of size n. We fit the analyst's model to each imputed dataset using the R package lme4 (Bates, Maechler, Bolker, & Walker, 2013). The final parameter estimates were obtained according to Rubin's (1987) rules. We note that choosing least-informative priors implies a prior expectation of variances of .50, which might induce bias into small variance components. However, because non-informative priors are often desirable for MI, the same priors were used throughout the three studies. Possible alternative specifications of the prior distribution will be reviewed in the General Discussion. The computer code for running conditional and multivariate MI, with least-informative or alternative priors, is provided in Supplement C of the supplemental online materials.

Bias and the root-mean-square error (RMSE) were calculated for each condition and each

parameter. The bias is the mean difference between a parameter estimate $\hat{\theta}$ and its true value θ, and is crucial for statistical reasoning in general. The RMSE is the root of the mean squared difference between $\hat{\theta}$ and θ and represents both accuracy and precision (i.e., the variability) of an estimator. Thus, it is an important measure of practical utility.

Results and Discussion

Due to the large simulation design, only the most important findings will be reported. Furthermore, only results for 25% missing data will be reported as higher rates did not yield interesting results. The complete results for Study 1 are given in Supplement D in the online supplemental materials. Table 2 shows the results of the first study for samples that featured small variance components (i.e., ICC = .05, ψ_{22}^2 = .01) for MCAR and MAR data in smaller (N = 10, G = 50) and larger samples (N = 30, G = 150). Notable values for bias and RMSE are presented in bold. Bias presented in bold is at least ±5% off the true value for fixed effects, and ±30% off for variance components. For parameters whose true value was zero, a threshold of ±.05 was used. For each simulated condition, the highest RMSE is printed in bold as long as it was significantly larger than that found for the complete datasets (at least twice as large).

As can be seen in Table 2, neither LD nor MI produced strongly biased results, but bias emerged under specific conditions for both MI procedures. The multivariate imputation model underestimated the slope variance by as much as 50% unless it was essentially zero (i.e., .01), but overestimated the intercept variance. Conditional MI (using the RC model) overestimated both the intercept and slope variance (Table 2, top panel). A sufficient sample size reduced bias to acceptable proportions even for the smallest variance components (Table 2, bottom panel). For larger values of the ICC (i.e., .15 and .25) and the slope variance (i.e., .05, .10, and .20), this bias

was reduced to essentially zero (see Supplement D). Using LD, the intercept and slope variance were sometimes biased when samples were not sufficiently large. When data were MNAR, all approaches yielded biased results (see Supplement D).

Listwise deletion has previously been shown to provide essentially unbiased estimates when the outcome is ignorably missing (e.g., Little & Rubin, 2002). Surprisingly, the imputation models overestimated small random effects variances in small samples. We argue that this is a side effect of the least-informative prior which expects variances to be larger, and that bias may be reduced to zero when the prior is set on an appropriate scale (see general discussion). From the data at hand, both LD and conditional MI can be recommended for univariate missing data on Y provided that the sample is sufficiently large or the prior is set on an appropriate scale. Care should be taken when small variance components are to be estimated, as overly non-informative priors may inflate them. The multivariate model is useful if the slope variance is close to zero.

Study 2

The second study examined the performance of MI and LD with missing values on the covariate X. The analyst's model was again the RC model $f(Y|X, \boldsymbol{\theta}_X)$, whereas conditional MI was carried out using the *reversed* RC model $g(X|Y, \boldsymbol{\omega}_Y)$. The two models fit the clustered structure of the data but differed in the way the slope variability was attributed. Multivariate MI and LD were administered as before.

Simulation and Methods

The same procedures as applied in Study 1 were used to simulate data and impose missing values on the covariate X, whereas MAR was now dependent on the outcome Y. Imputations were

created by pan using the least-informative priors as chosen in Study 1. The analyst's model was fit

using lme4, and the bias and RMSE were calculated for each parameter in each setting.

Results

The results of Study 2 are reported in full in Supplement D. Here, we will report the most

important findings. Table 3 provides a brief overview of the results for samples that featured small

variance components. Estimating the fixed effects of the RC model proved to be more accurate

and efficient using MI. Specific difficulties emerged again for small variance components, that is,

when samples featured small ICCs or little slope variation. In contrast to when data were missing

on Y, however, estimates of larger slope variances were not necessarily unbiased.

Fixed effects.   As shown in Table 3, LD led to biased estimates for the fixed effects unless

the data were MCAR (see Supplement D). Bias for the fixed intercept varied between $-.098$ and

$-.161$ with MAR data and between $-.055$ and $-.101$ with MNAR data. The fixed slope was

underestimated by approximately 6-10% when the data were not MCAR. Results from MI were

essentially unbiased, but the reversed model exhibited a small downward bias across conditions.

The RMSE suggested that estimates obtained from MI were at least as efficient as those obtained

by LD across conditions, and more efficient when data were not MCAR.

Interestingly, bias from both LD and conditional MI was dependent on the amount of slope

variation that was present in the dataset. As slope variation increased, bias became weaker with

LD, and stronger with conditional MI. This result is illustrated in Figure 1 for small samples

(N = 10, G = 150), moderate ICCs (i.e., .15), and MAR data. Nonetheless, estimates obtained

from MI were more accurate and efficient across all conditions.

Variance and covariance of random effects.   Conditional and multivariate MI

underestimated the intercept variance when the ICCs were small but provided unbiased estimates

otherwise. Listwise deletion followed the same pattern for MCAR data, but otherwise

underestimated the intercept variance. This bias was strongest in the MAR condition, weaker with

MNAR data, and increased as the ICCs grew larger. Figure 2 (top row) illustrates this finding for

different levels of the ICC.

Results for the slope variance differed from Study 1. Although conditional MI again

overestimated small amounts of slope variation, this bias was much weaker and practically

disappeared in larger samples (see Table 3). Moderate slope variation could be estimated almost

without bias. In contrast to Study 1, however, large and very large slope variances were not

estimated correctly by conditional MI but increasingly suffered from a downward bias. Listwise

deletion provided practically unbiased estimates of the slope variance if the sample size was

sufficiently large. The positive bias for conditional MI was also present with MNAR data, whereas

the negative bias was smaller. Figure 2 (bottom row) illustrates these findings for different levels

of slope variation.

According to the RMSE, the intercept variance could occasionally be estimated more

efficiently using MI, whereas the slope variance could be estimated more accurately using LD.

However, these differences were usually very small. Supplement D even suggests that conditional

MI occasionally estimated the slope variance more efficiently in small samples.

Other parameters.   The covariance between random intercepts and slopes was well

recovered across all conditions. The Level 1 residual variance was overestimated using MI, where

conditional MI was less biased, but it was underestimated by LD when data were not MCAR. For

higher amounts of slope variation, the bias associated with LD became smaller, whereas the bias

grew for MI. These patterns were observed with MAR and MNAR data, but the bias was relatively

small.

Discussion

Regarding most parameters of the analyst's model, better estimates could be obtained using the reversed MI procedure, especially when the covariate X was not MCAR. This was true for the fixed regression coefficients but also applied to the intercept variance and even transferred to MNAR data. However, reversed MI seemed to provide unstable estimates of the slope variance, which could be positively or negatively biased. The positive bias for small slope variances became essentially zero as the samples grew larger. For larger slope variances, the bias did not approach zero (as in Study 1) but turned negative regardless of sample size. The negative bias was, however, rather small and could be viewed as negligible considering that it only occurred for large slope variances, which are rarely found in empirical studies. Furthermore, the overall precision of the estimates, as indicated by the RMSE, was often comparable to LD because the data were handled more efficiently using MI. The reversed model seemed to share many but not all of the desirable properties of the regular RC model.

The multivariate imputation model is applicable if little slope variation is present in the data, but it will suppress even moderate amounts of slope variation and inflate the Level 1 residual variance. Estimates of the fixed slope obtained from multivariate MI were even less biased and more efficient than those from the reversed MI procedure. Listwise deletion offered little benefit as most of its parameter estimates were biased unless the data were MCAR. However, LD provided surprisingly accurate results for the slope variance. Small variance components were again positively biased but less so than in the previous study. We will return to this point in the General Discussion.

Study 3

The final study examined the performance of MI and LD with multivariate missing data. The analyst's model was once again the RC model $f(Y|X, \boldsymbol{\theta}_X)$, but only the multivariate imputation model $g(X, Y|\boldsymbol{\omega}_0)$ could be applied. This imputation model ignores slope variability, but may provide reasonable results for the remaining parameters of the analyst's model.

Simulation and Methods

The same procedures that were used in the previous studies could be used for most tasks. Because the pattern of missing data was no longer univariate, the missing data model had to be adjusted. We excluded unit-nonresponse from our considerations; thus, every participant was expected to have at least one observation on either X or Y. This allowed us to implement the same mechanisms as described before (i.e., MCAR, MAR, MNAR) for both X and Y. For each case, a coin toss decided whether X or Y could be missing (i.e., each was equally likely). The actual missing values were then imposed on either X or Y with the probability that was given in the simulation design. Thus the amount of missing values in each dataset was the same in all three studies.

Results and Discussion

The results of the third study provided little further insight into the performance of LD and multivariate MI because the bias and RMSE were usually halfway between those reported in Studies 1 and 2. Results for small variance components are presented in Table 4. The complete results are available in Supplement D. Multivariate MI provided approximately unbiased estimates of all parameters as long as the slope variance was close to zero and the values were either MCAR

or MAR. The slope variance was underestimated by as much as 40%, especially in larger samples

where more values were imputed under false assumptions. When the data were MNAR,

multivariate MI underestimated the fixed regression coefficient, but the bias was relatively small

compared with the true values. Estimates obtained from LD were approximately unbiased when

data were MCAR. When data were MAR or MNAR, the fixed effects were biased downward and

were estimated less efficiently than with multivariate MI, where higher values for the ICC and

slope variance reduced bias with LD (see Supplement D).

The results of the third study suggest that MI is necessary for proper estimation of the fixed

regression coefficients. Unfortunately, pan's multivariate imputation model could not preserve the

slope variance. If the slope variance was small and the number of missing values was not very

high, then the bias was relatively small in absolute size. Limiting the analysis to complete cases

only distorted the parameter estimates, but provided reasonable estimates of the slope variance.

General Discussion

We investigated the performance of conditional and multivariate MI for univariate and

multivariate patterns of missing data. Both conditional MI and LD provided unbiased estimates if

only the outcome was missing. Care should be taken if covariates are partially unobserved.

Imputing the covariate in a reversed manner accounted for, but also misspecified the slope

variation. Only vague estimates could be obtained for the slope variance, but bias was not

extreme, and the remaining estimates exhibited either no or less bias than what would have been

obtained by deleting cases. The multivariate imputation model rarely induced any bias but

strongly underestimated the slope variance. Thus, it is appropriate only if the true slope variance

is close to zero and not too many values are unobserved. We recommend that LD be avoided

when covariate data are missing unless the data are strictly MCAR.

As is true for all computer simulations, our study was limited in several ways. The missing

data mechanisms were based on linear models and may behave quite differently in nature. Other

implementations are possible, and results may vary especially for MAR and MNAR data (Allison,

2000; Galati & Seaton, 2013). We focused on descriptive measures of approximate performance

but ignored statistical inference. Testing for slope variation (LaHuis & Avis, 2007) as well as

Type-I and Type-II error rates associated with LD and MI should be a subject of future research.

Rather than estimating the slope variance, researchers often wish to explain it using predictor

variables at Level 2 (Aguinis et al., 2013; Mathieu et al., 2012). Cross-level interaction effects

might be relatively easy to recover even if the slope variance is not.

Interestingly, small variance components were positively biased across the three studies. We

argue that this is due to the standard least-informative prior, which induces bias into small

variance components. Ad hoc procedures might combine the specific advantages of LD and MI

and lead to less biased and more stable estimates. For example, choosing $D^{-1} = 2 \cdot \hat{\mathbf{\Psi}}_{LD}$ as the

scale matrix of the inverse Wishart prior for the covariance matrix of random effects, where $\hat{\mathbf{\Psi}}_{LD}$

is an estimate of this covariance matrix obtained from LD, would loosely center the prior

distribution around appropriate values. The computer code for this specification is provided in

Supplement C of the supplemental online materials. We conducted a small simulation to examine

whether the bias for the intercept and slope variance could be reduced by rescaling the prior

distribution in this manner. The simulation featured small samples, univariate MAR data on either

X or Y, small values for the ICCs, as well as small and very large values for the slope variance.

Estimates of small variance components that utilized the adjusted prior did not exhibit any more

bias than LD did and were often more efficient. The positive bias reported in Studies 1 and 2

could therefore be viewed as an artifact of specifying the least-informative prior. The negative

bias for large slope variances in Study 2, however, could not be improved in this manner. Using

least squares or maximum likelihood estimation might further strengthen this approach.

The methodological literature offers alternatives to pan for multilevel MI. It has been

suggested that multilevel data be imputed using dummy variables in random intercept models but

that imputations should be conducted separately for each group if random slopes are involved

(Graham, 2009, 2012). However, Andridge (2011) found that the first approach leads to biased

results, and unreported simulation results indicate that very large samples are needed to treat even

small amounts of missing data with the second approach. Alternative MI procedures include *fully*

conditional specification using chained equations (van Buuren & Groothuis-Oudshoorn, 2011).

These procedures might lead to better results, but may face similar problems with respect to the

slope variance. However, recent developments in the context of substantive model compatible MI

have offered promising results for interaction effects and nonlinear terms among covariates that

have missing values (Bartlett, Seaman, White, & Carpenter, 2014; von Hippel, 2009). Extending

this approach to multilevel MI (Goldstein, Carpenter, Kenward, & Levin, 2009; Goldstein,

Carpenter, & Browne, 2014) and applying it to random slope models should be the subject of

future research. Adaptations of the pan model have been proposed by Shin and Raudenbush

(2010) and Yucel (2011). The latter approach specifies a joint model that allows the within-group

covariance matrix to vary across groups, and has been recently discussed by Carpenter and

Kenward (2013). However, it is currently not available in standard software and has yet to be

evaluated in a systematic manner.

In general, we believe that MI is a flexible and powerful tool that can be used to treat

missing data in multilevel research. More research should be conducted to generalize the current

formulations of MI and to evaluate recent developments as well as sensible ad hoc solutions to

missing data in multilevel models with random slopes.

References

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, *39*, 1490–1528. doi: 10.1177/0149206313478188

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods & Research*, *28*, 301–309. doi: 10.1177/0049124100028003003

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.

Andridge, R. R. (2011). Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biometrical Journal*, *53*, 57–74. doi: 10.1002/bimj.201000140

Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R. (2014). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research*, *0*, 1–26. doi: 10.1177/0962280214521348

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.0-5) [Computer software]. Retrieved from `http://CRAN.R-project.org/package=lme4`

Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. Hoboken, NJ: Wiley.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330–351. doi: 10.1037/1082-989X.6.4.330

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Galati, J. C., & Seaton, K. A. (2013). MCAR is not necessary for the complete cases to constitute

a simple random subsample of the target sample. *Statistical Methods in Medical Research*,

0, 1–8. doi: 10.1177/0962280213490360

Goldstein, H., Carpenter, J., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with

multivariate mixed response types. *Statistical Modelling*, *9*, 173–197. doi:

10.1177/1471082X0800900301

Goldstein, H., Carpenter, J. R., & Browne, W. J. (2014). Fitting multilevel multivariate models

with missing data in responses and covariates that may include interactions and non-linear

terms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *177*,

553–564. doi: 10.1111/rssa.12022

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of

Psychology*, *60*, 549–576. doi: 10.1146/annurev.psych.58.110405.085530

Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really

needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*,

206–213. doi: 10.1007/s11121-007-0070-9

Hochweber, J., Hosenfeld, I., & Klieme, E. (2014). Classroom composition, classroom

management, and the relationship between student attributes and grades. *Journal of

Educational Psychology*, *106*, 289–300. doi: 10.1037/a0033829

Hussong, A. M., Cai, L., Curran, P. J., Flora, D. B., Chassin, L. A., & Zucker, R. A. (2008).

Disaggregating the Distal, Proximal, and Time-Varying Effects of Parent Alcoholism on

Children's Internalizing Symptoms. *Journal of Abnormal Child Psychology*, 335–346. doi:

10.1007/s10802-007-9181-9

LaHuis, D. M., & Avis, J. M. (2007). Using multilevel random coefficient modeling to investigate rater effects in performance ratings. *Organizational Research Methods*, *10*, 97–107. doi: 10.1177/1094428106289394

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*, 951–966. doi: 10.1037/a0028380

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, *9*, 538– 558.

Murray, D. M., & Blitstein, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review*, *27*, 79–103. doi: 10.1177/0193841X02239019

Peters, S. A., Bots, M. L., den Ruijter, H. M., Palmer, M. K., Grobbee, D. E., Crouse, J. R., . . . Koffijberg, H. (2012). Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. *Journal of Clinical Epidemiology*, *65*, 686–695. doi: 10.1016/j.jclinepi.2011.11.012

R Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.2) [Computer software]. Retrieved from http://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.

Schafer, J. L. (1997). *Imputation of missing covariates under a multivariate linear mixed model*

(Technical Report). Retrieved from `http://sites.stat.psu.edu/`

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and

    analysis models differ. *Statistica Neerlandica*, *57*, 19–35. doi: 10.1111/1467-9574.00218

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art.

    Psychological Methods, *7*, 147–177. doi: 10.1037//1082-989X.7.2.147

Schafer, J. L., & Zhao, J. H. (2014). pan: Multiple imputation for multivariate panel or clustered

    data (Version 0.9) [Computer software]. Retrieved from

    `http://CRAN.R-project.org/package=pan`

Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects

    model with missing data. *Journal of Educational and Behavioral Statistics*, *35*, 26–53. doi:

    10.3102/1076998609345252

Snijders, T. A. B., & Bosker, R. J. (2012). Components of variance. In *Multilevel analysis: An*

    introduction to basic and advanced multilevel modeling (pp. 114–117). Thousand Oaks,

    CA: Sage.

Taljaard, M., Donner, A., & Klar, N. (2008). Imputation strategies for missing continuous

    outcomes in cluster randomized trials. *Biometrical Journal*, *50*, 329–345. doi:

    10.1002/bimj.200710423

Twisk, J., de Boer, M., de Vente, W., & Heymans, M. (2013). Multiple imputation of missing

    values was not necessary before performing a longitudinal mixed-model analysis. *Journal*

    of Clinical Epidemiology, *66*, 1022–1028. doi: 10.1016/j.jclinepi.2013.03.017

van Buuren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox (Ed.), *Handbook of*

    advanced multilevel analysis (pp. 173–196). New York, NY: Routledge.

van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC Press.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained

    equations in R. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from

    `http://www.jstatsoft.org/`

von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables.

    Sociological Methodology, *39*, 265–291. doi: 10.1111/j.1467-9531.2009.01215.x

Yucel, R. M. (2011). Random covariances and mixed-effects models for imputing multivariate

    multilevel continuous data. *Statistical Modelling*, *11*, 351–370. doi:

    10.1177/1471082X1001100404

Table 1

Simulation Designs of Study 1, Study 2, and Study 3

Design conditions	Study 1	Study 2	Study 3
Number of groups	50, 150	50, 150	50, 150
Group size	10, 30	10, 30	10, 30
ICC	.05, .15, .25	.05, .15, .25	.05, .15, .25
Fixed slope	.50	.50	.50
Slope variance	.01, .05, .10, .20	.01, .05, .10, .20	.01, .05, .10, .20
MD pattern	univariate Y	univariate X	multivariate X and Y
MD proportion	25%, 50%	25%, 50%	25%, 50%
MD mechanism	MCAR, MAR, MNAR	MCAR, MAR, MNAR	MCAR, MAR, MNAR
Imputation models	RC model $f(Y|X, \boldsymbol{\theta}_X)$, multivariate $f(X, Y|\boldsymbol{\theta}_0)$	reversed RC $f(X|Y, \boldsymbol{\theta}_Y)$, multivariate $f(X, Y|\boldsymbol{\theta}_0)$	multivariate $f(X, Y|\boldsymbol{\theta}_0)$

Note. ICC = intraclass correlation of X and Y; MD = missing data; MCAR = missing completely at random; MAR = missing at random; MNAR = missing not at random; RC = random coefficients.

Table 2

Study 1: Bias and RMSE for Estimates Obtained from LD and MI Given Small Variance Components, Smaller or Larger Samples, and Missing Y

	Bias						RMSE					
	MCAR			MAR			MCAR			MAR		
Est.	LD	MV	RC	LD	MV	RC	LD	MV	RC	LD	MV	RC
A: N=10, G=50, ICC=0.05, SV=0.01												
β_0	.002	.002	.002	.003	.001	.003	.003	.003	.003	.003	.003	.003
β_1	−.001	−.002	−.001	.002	−.001	.002	.002	.002	.002	.002	.002	.002
ψ_{11}^2	−.005	**.014**	.010	−.002	**.013**	.005	.001	.001	.001	.001	.001	.001
ψ_{22}^2	**.006**	.001	**.023**	**.007**	.001	**.032**	.000	.000	**.001**	.000	.000	**.001**
ψ_{12}	−.001	−.002	−.003	.000	.001	.008	.000	.000	.000	.000	.000	**.000**
σ^2	−.001	−.004	−.009	−.007	−.006	−.008	.003	.003	.003	.003	.003	.004
B: N=30, G=150, ICC=0.05, SV=0.01												
β_0	−.000	−.000	−.000	.001	.000	.000	.000	.000	.000	.000	.000	.000
β_1	.000	.000	.000	−.001	−.001	−.001	.000	.000	.000	.000	.000	.000
ψ_{11}^2	−.002	.001	.002	−.000	.002	.001	.000	.000	.000	.000	.000	.000
ψ_{22}^2	.000	−**.004**	**.006**	.000	−**.005**	**.009**	.000	.000	**.000**	.000	.000	**.000**
ψ_{12}	−.000	−.000	−.000	.000	−.000	.003	.000	.000	.000	.000	.000	.000
σ^2	.002	.005	−.002	.000	.004	−.001	.000	.000	.000	.000	.000	.000

Note. LD = listwise deletion; MV = multivariate imputation; RC = conditional imputation (random coefficients); MCAR = missing completely at random; MAR = missing at random; β_0 = fixed intercept; ; β_1 = fixed slope; ψ_{11}^2 = intercept variance; ψ_{22}^2 = slope variance; ψ_{12} = intercept-slope covariance; σ^2 = Level 1 residual variance.

Table 3

Study 2: Bias and RMSE for Estimates Obtained from LD and MI Given Small Variance Components, Smaller or Larger Samples, and Missing X

	Bias						RMSE					
	MCAR			MAR			MCAR			MAR		
Est.	LD	MV	RC	LD	MV	RC	LD	MV	RC	LD	MV	RC
A: N=10, G=50, ICC=0.05, SV=0.01												
β_0	−.001	−.001	−.001	−**.161**	−.000	−.003	.003	.002	.002	**.028**	.002	.002
β_1	.002	−.002	−.010	−**.042**	−.002	−.014	.002	.002	.002	**.004**	.002	.002
ψ_{11}^2	−.005	.000	−.003	−.008	.002	−.000	.001	.001	.001	.001	.001	.001
ψ_{22}^2	**.007**	.002	**.008**	**.005**	.002	**.010**	.000	.000	.000	.000	.000	.000
ψ_{12}	−.001	−.002	−.002	−.002	−.003	−.006	.000	.000	.000	.000	.000	.000
σ^2	−.001	.000	.000	−.063	.000	−.003	.003	.003	.003	**.007**	.003	.003
B: N=30, G=150, ICC=0.05, SV=0.01												
β_0	−.000	−.000	−.000	−**.160**	−.001	−.002	.001	.000	.000	**.026**	.000	.000
β_1	−.000	−.000	−.003	−**.045**	−.001	−.005	.000	.000	.000	**.002**	.000	.000
ψ_{11}^2	−.001	−.001	−.001	−.006	.000	.001	.000	.000	.000	.000	.000	.000
ψ_{22}^2	−.000	−.003	−.000	−.001	−.003	.000	.000	.000	.000	.000	.000	.000
ψ_{12}	−.000	−.000	−.000	−.001	−.000	−.001	.000	.000	.000	.000	.000	.000
σ^2	.002	.005	.003	−.060	.004	.002	.000	.000	.000	**.004**	.000	.000

Note. LD = listwise deletion; MV = multivariate imputation; RC = conditional imputation (random coefficients); MCAR = missing completely at random; MAR = missing at random; β_0 = fixed intercept; ; β_1 = fixed slope; ψ_{11}^2 = intercept variance; ψ_{22}^2 = slope variance; ψ_{12} = intercept-slope covariance; σ^2 = Level 1 residual variance.

Table 4

Study 3: Bias and RMSE for Estimates Obtained from LD and MI Given Small Variance Components, Smaller or Larger Samples, and Missing X and Y

	Bias						RMSE					
	MCAR		MAR		MNAR		MCAR		MAR		MNAR	
Est.	LD	MV	LD	MV	LD	MV	LD	MV	LD	MV	LD	MV
	A: N=10, G=50, ICC=0.05, SV=0.01											
β_0	.002	.001	−**.084**	−.001	−**.097**	−.019	.003	.003	**.010**	.002	**.012**	.003
β_1	.001	−.001	−**.026**	−.001	−**.048**	−.024	.002	.002	.003	.002	**.005**	.003
ψ_{11}^2	−.007	.005	−.006	.006	−.003	.006	.001	.001	.001	.001	.001	.001
ψ_{22}^2	**.006**	.002	**.005**	.001	**.005**	.001	.000	.000	.000	.000	.000	.000
ψ_{12}	−.001	−.002	−.001	−.001	−.001	−.001	.000	.000	.000	.000	.000	.000
σ^2	−.001	−.001	−.026	−.000	−.025	.004	.003	.003	.004	.003	.004	.003
	B: N=30, G=150, ICC=0.05, SV=0.01											
β_0	.000	.000	−**.082**	−.000	−**.099**	−.020	.000	.000	**.007**	.000	**.010**	.001
β_1	.001	.000	−**.027**	.000	−**.046**	−.020	.000	.000	**.001**	.000	**.002**	.001
ψ_{11}^2	−.002	.000	−.003	.001	−.003	.001	.000	.000	.000	.000	.000	.000
ψ_{22}^2	.000	−**.003**	−.001	−**.004**	−.000	−**.004**	.000	.000	.000	.000	.000	.000
ψ_{12}	−.000	−.000	−.001	−.001	−.001	−.000	.000	.000	.000	.000	.000	.000
σ^2	.003	.005	−.024	.004	−.022	.008	.000	.000	**.001**	.000	**.001**	.000

Note. LD = listwise deletion; MV = multivariate imputation; MCAR = missing completely at random; MAR = missing at random; MNAR = missing not at random; β_0 = fixed intercept; ; β_1 = fixed slope; ψ_{11}^2 = intercept variance; ψ_{22}^2 = slope variance; ψ_{12} = intercept-slope covariance; σ^2 = Level 1 residual variance.
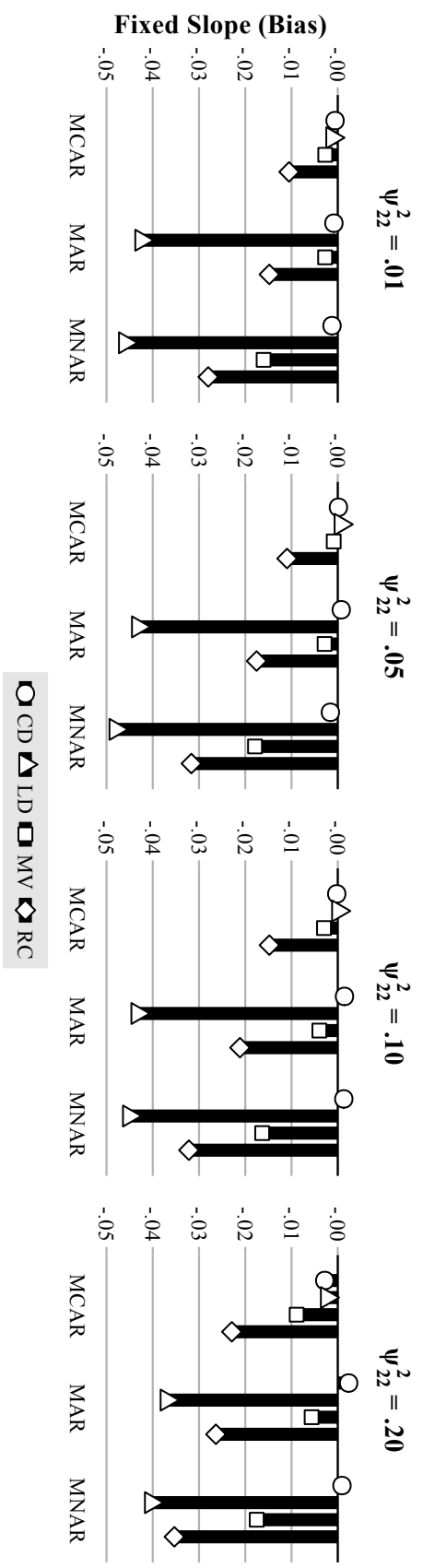
Figure 1. Bias in estimating the fixed slope for univariate missing data on X (Study 2) for different MD mechanisms, MD techniques, and different amounts of slope variance. ψ_{22}^2 = true slope variance; MCAR = missing completely at random; MAR = missing at random; MNAR = missing not at random; CD = complete data; LD = listwise deletion; MV = multivariate imputation; RC = conditional imputation using the reversed RC model.
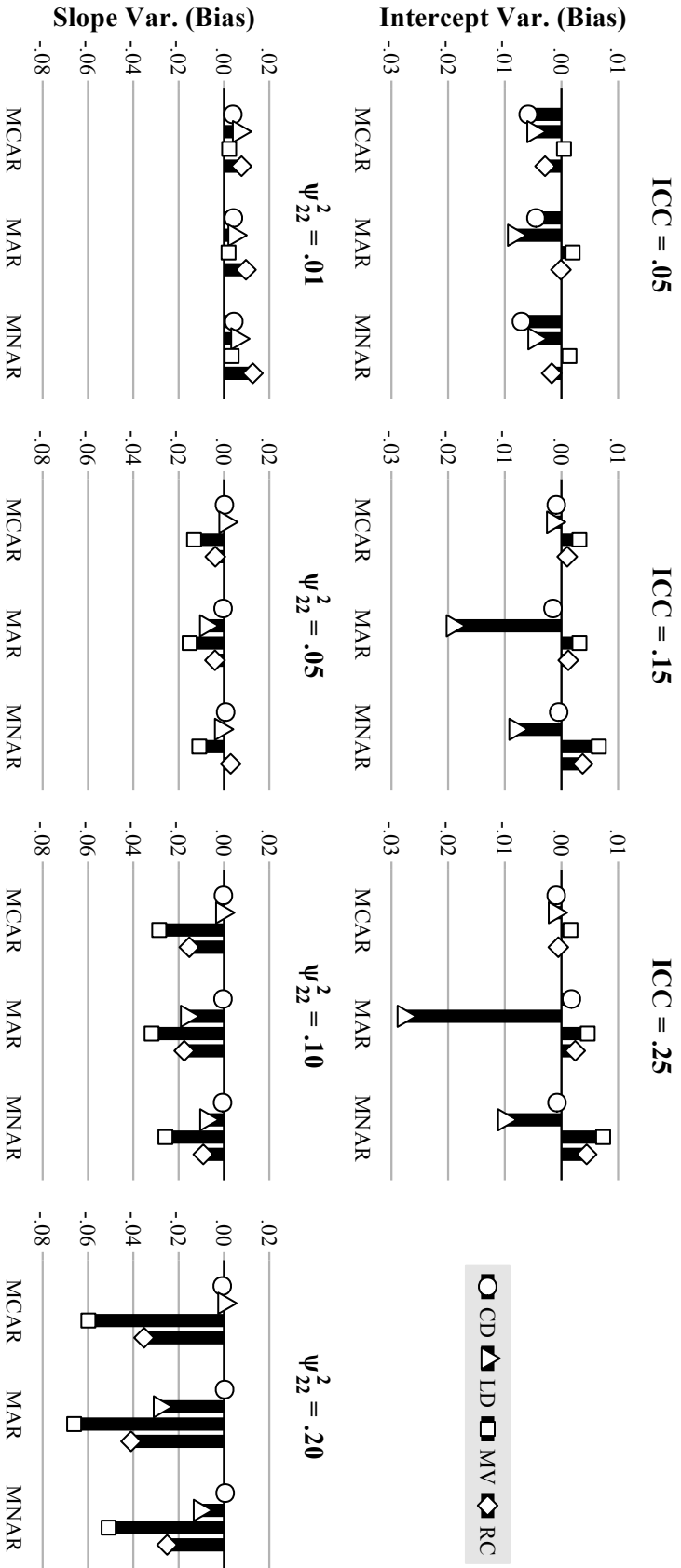
Figure 2. Bias in estimating the intercept (top row) and slope variance (bottom row) for univariate missing data on X (Study 2) for different values of the ICC or slope variance, respectively, and different MD mechanisms and MD techniques. ICC = true intraclass correlation; ψ_{22}^2 = true slope variance; MCAR = missing completely at random; MAR = missing at random; MNAR = missing not at random; CD = complete data; LD = listwise deletion; MV = multivariate imputation; RC = conditional imputation using the reversed RC model.