

# Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations

Organizational Research Methods  
1-39

© The Author(s) 2017

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1094428117703686

journals.sagepub.com/home/orm



Simon Grund<sup>1,2</sup>, Oliver Lüdtke<sup>1,2</sup> and Alexander Robitzsch<sup>1,2</sup>

## Abstract

Multiple imputation (MI) is one of the principled methods for dealing with missing data. In addition, multilevel models have become a standard tool for analyzing the nested data structures that result when lower level units (e.g., employees) are nested within higher level collectives (e.g., work groups). When applying MI to multilevel data, it is important that the imputation model takes the multilevel structure into account. In the present paper, based on theoretical arguments and computer simulations, we provide guidance using MI in the context of several classes of multilevel models, including models with random intercepts, random slopes, cross-level interactions (CLIs), and missing data in categorical and group-level variables. Our findings suggest that, oftentimes, several approaches to MI provide an effective treatment of missing data in multilevel research. Yet we also note that the current implementations of MI still have room for improvement when handling missing data in explanatory variables in models with random slopes and CLIs. We identify areas for future research and provide recommendations for research practice along with a number of step-by-step examples for the statistical software R.

## Keywords

multilevel, missing data, multiple imputation, random intercept model, random coefficients model, random slopes, cross-level interactions

Multilevel models have become one of the standard tools for analyzing clustered empirical data. Such data are often found in organizational psychology, for example, when employees are nested within work groups or enterprises, or in longitudinal studies when measurement occasions are nested within persons. In addition, empirical data are often incomplete, for example, when some

---

<sup>1</sup>Leibniz Institute for Science and Mathematics Education, Kiel, Germany

<sup>2</sup>Centre for International Student Assessment, Germany

## Corresponding Author:

Simon Grund, Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, 24118 Kiel, Germany.

Email: grund@ipn.uni-kiel.de

participants fail to answer all of the items on a questionnaire. Several authors have advocated the use of modern missing data techniques such as multiple imputation (MI) rather than traditional approaches such as listwise deletion (LD; Allison, 2001; Enders, 2010; Little & Rubin, 2002; Newman, 2014; Schafer & Graham, 2002). One central requirement of MI is that the imputation model must be at least as general as the model of interest in order to preserve its key features. In multilevel data, it is important that the imputation model takes the multilevel structure into account (e.g., Andridge, 2011; Drechsler, 2015). However, depending on the research question, the multilevel structure may manifest itself in the analysis model in a number of ways (e.g., random intercepts and slopes, relations between variables within and between groups), leading to a multitude of possible multilevel analysis models, each directed at different research questions (e.g., Aguinis & Culpepper, 2015; Snijders & Bosker, 2012).

The motivation behind the present paper is twofold. First, we offer simulation results regarding the performance of MI when the substantive analysis model belongs to one of several types of multilevel models. Second, we provide an introduction to and recommendations for MI of multilevel data directed toward readers who are not yet familiar with the often technical literature on MI. Our article is divided into four sections. In the first section, we focus on the multilevel random intercept model and discuss imputation procedures that are suitable for application in such models. In the second section, we focus on the random coefficients model and the specific challenges that arise when working with random slopes and cross-level interactions (CLIs). In the third section, we briefly discuss missing data in categorical and group-level variables. In each section, we present results from simulation studies in which we used different MI procedures as well as full-information maximum likelihood (FIML). Finally, in the last section, we provide recommendations for how to handle missing data for different types of multilevel models. We conclude with a discussion of our findings and possible topics for future research.

## Missing Data and Multiple Imputation

The basic idea of MI is to replace missing values by forming an “informed guess” that is based on the observed data and a statistical model (the imputation model). Multiple imputation generates several ( $M$ ) replacements for the missing data by drawing repeatedly from the posterior predictive distribution of the missing data, given the observed data and the parameters of the imputation model. The  $M$  data sets completed in this manner are then analyzed separately, yielding  $M$  sets of parameter estimates. To obtain final estimates and inferences, these results are pooled using the rules described in Rubin (1987; see also Enders, 2010).

The use of MI in most (but not all) implementations is predicated on the assumption that the data are *missing at random* (MAR). The definition of MAR, according to Rubin (1976), assumes that a hypothetical complete data set can be divided into observed and unobserved parts,  $\mathbf{Y} = (\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$ , where an indicator matrix  $\mathbf{R}$  denotes which data are missing or observed. According to Rubin (1976), data are MAR if the probability of observing data,  $P(\mathbf{R})$ , is independent of the unobserved data given the observed data, that is,  $P(\mathbf{R}|\mathbf{Y}) = P(\mathbf{R}|\mathbf{Y}_{obs})$ . In other words, under MAR, there is no link between the chance of observing a value and the value itself, given the data that one has observed. A special case of this occurs when  $P(\mathbf{R})$  is completely independent of the data, that is,  $P(\mathbf{R}|\mathbf{Y}) = P(\mathbf{R})$ . This is referred to as missing *completely* at random (MCAR). If the MAR assumption is violated, that is, data are missing *not* at random (MNAR), the application of MI requires strong assumptions about the missing data mechanism. Such applications are relatively rare and are most often used as sensitivity analyses (see Carpenter & Kenward, 2013). In the present paper, we focus on applications of MI that operate under MAR.

Two aspects make MI a particularly attractive method for dealing with missing data. First, MI recognizes the uncertainty that is due to missing data by generating multiple (as opposed to single)

replacements for each missing value, and by drawing the parameters of the imputation model from Bayesian posterior distributions, given the currently imputed data and a set of prior beliefs. Second, because the imputation phase is separated from the analysis phase, MI is able to make full use of the data by including variables in the imputation model that are either predictive of missingness, thus improving the plausibility that MAR holds, or related to the variables of interest, thus improving the power of its predictions (Collins, Schafer, & Kam, 2001).

## Multiple Imputation for Multilevel Models

A crucial point in the application of MI to multilevel data is that the imputation model not only includes all relevant variables, but also that it “matches” the model of interest (i.e., the substantive analysis model; see Meng, 1994; Schafer, 2003). In other words, the imputation model must capture the relevant aspects of the analysis model, making the imputation model at least as general as (or more general than) the analysis model. If the imputation model is more restrictive than the analysis model, then imputations are generated under a simplified set of assumptions, and the results of subsequent analyses may be misleading. For example, consider the case in which the model of interest is a multilevel random intercept model (Snijders & Bosker, 2012) in which an individual-level outcome  $Y$  is regressed on an individual-level explanatory variable  $X$

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + u_{0j} + e_{ij}, \quad (1)$$

where  $\bar{X}_{\bullet j}$  denotes the group mean of  $X$  in group  $j$ ,  $(X_{ij} - \bar{X}_{\bullet j})$  denotes the individual deviation in  $X$  for a person  $i$  in group  $j$ , and  $\gamma_{10}$  and  $\gamma_{01}$  denote the regression coefficients of  $X$  within and between groups, respectively (see Hofmann & Gavin, 1998; Kreft, de Leeuw, & Aiken, 1995). The intercepts,  $u_{0j}$ , and the residuals,  $e_{ij}$ , are assumed to follow independent normal distributions with mean zero and with variances  $\tau_0^2$  and  $\sigma^2$ , respectively.

Two aspects of the model in Equation 1 are worth noting, and both must be accommodated during MI in order for subsequent analyses to yield proper results. First, the model accounts for the clustered structure of the data by including random effects for each group (Snijders & Bosker, 2012). Therefore, the imputation model must also take the clustered structure into account. Failing to do so, for example, by using single-level MI, might lead to biased parameter estimates and might distort statistical decision making (e.g., Andridge, 2011; Enders, Mistler, & Keller, 2016; Lüdtke, Robitzsch, & Grund, 2017; Taljaard, Donner, & Klar, 2008). Second, the model differentiates between the effects of  $X$  at the individual and the group level (i.e., for  $(X_{ij} - \bar{X}_{\bullet j})$  and  $\bar{X}_{\bullet j}$ ). If the imputation model does not allow these effects to be different, then the parameters will be “conflated” during MI, and estimates obtained in subsequent analyses may be biased (see Enders et al., 2016; Lüdtke et al., 2017; Preacher, Zyphur, & Zhang, 2010). In other words, ignoring the existence of separate effects for  $(X_{ij} - \bar{X}_{\bullet j})$  and  $\bar{X}_{\bullet j}$  in the imputation model will make it more difficult to find them in subsequent analyses. In the following section, we discuss several MI procedures that can be used to accommodate the multilevel random intercept model.

### Joint Modeling and the Fully Conditional Specification of MI

The procedures available for multilevel MI can be roughly divided into two broad paradigms: the joint modeling approach (JM) and the fully conditional specification of MI (FCS). Both approaches offer the necessary tools for dealing with multilevel missing data. Here, we consider the JM approach implemented in the `pan` package (Schafer & Yucel, 2002) and the FCS approach known as “multiple imputation by chained equations” implemented in the `mice` package (van Buuren & Groothuis-Oudshoorn, 2011) in the statistical software R (R Core Team, 2016).

*Joint modeling (JM).* In the JM approach, a single model is specified for all variables with missing data, and imputations are simultaneously generated from this model for all variables with missing data. For individual-level variables, the joint model reads

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{u}_j + \mathbf{e}_{ij}, \quad (2)$$

where  $\mathbf{Y}_{ij}$  contains a number of individual-level target variables with arbitrary patterns of missing data,  $\mathbf{X}_{ij}$  contains fully observed predictor variables with associated fixed effects  $\boldsymbol{\beta}$ ,  $\mathbf{Z}_{ij}$  contains fully observed predictor variables with associated random effects  $\mathbf{u}_j$ , and  $\mathbf{e}_{ij}$  denotes the residuals at the individual level. The random effects,  $\mathbf{u}_j$ , are assumed to follow a multivariate normal distribution with mean zero and covariance matrix  $\boldsymbol{\Psi}$ . The residuals,  $\mathbf{e}_{ij}$ , follow a multivariate normal distribution with mean zero and covariance matrix  $\boldsymbol{\Sigma}$ .

The design matrices,  $\mathbf{X}_{ij}$  and  $\mathbf{Z}_{ij}$ , on the right-hand side of the model equation may contain any number of variables as long as they are fully observed (see Schafer & Yucel, 2002). If the model of interest is a multilevel random intercept model, it is possible to include all variables (both partially and fully observed) as target variables on the left-hand side of the model equation, whereas the right-hand side includes only the intercept (i.e.,  $\mathbf{X}_{ij} = \mathbf{Z}_{ij} = 1$ ). For example, consider the random intercept model in Equation 1 and assume that  $X$  and/or  $Y$  are partially missing. Treating both  $X$  and  $Y$  as target variables, the JM becomes

$$[X_{ij}, Y_{ij}]^T = [\beta_{0(x)}, \beta_{0(y)}]^T + [u_{j(x)}, u_{j(y)}]^T + [e_{ij(x)}, e_{ij(y)}]^T, \quad (3)$$

where the random effects  $[u_{j(x)}, u_{j(y)}]^T$  and the residuals  $[e_{ij(x)}, e_{ij(y)}]^T$  follow independent multivariate normal distributions with mean zero and covariance matrices  $\boldsymbol{\Psi} = \begin{bmatrix} \psi_x^2 & \psi_{xy} \\ \psi_{xy} & \psi_y^2 \end{bmatrix}$  and  $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$ . In this specification, the joint model decomposes the variables into separate within- and between-group components represented by  $[e_{ij(x)}, e_{ij(y)}]^T$  and  $[u_{j(x)}, u_{j(y)}]^T$ , thus allowing for different relations (i.e., covariances) between  $X$  and  $Y$  to be estimated at the individual and the group level (Lüdtke et al., 2017; see also Grund, Lüdtke, & Robitzsch, 2016b).<sup>1</sup> Similar models are also implemented in the statistical software *Mplus* (L. K. Muthén & Muthén, 2012; see also Enders et al., 2016) and in the R package *jomo* (Quartagno & Carpenter, 2016).

*The FCS approach.* In contrast to the JM approach, the FCS approach imputes missing data separately for each variable with missing data, conditioning on some or all of the other variables in the data set. To address multivariate patterns of missing data, the FCS algorithm iterates back and forth between different target variables. Again, consider the analysis model in Equation 1. For missing data in  $X$  and  $Y$ , an appropriate FCS approach may generate imputations on the basis of the following two univariate models

$$\begin{aligned} X_{ij} &= \beta_{0(x)} + \beta_{1(x)}(Y_{ij} - \bar{Y}_{\bullet j}) + \beta_{2(x)}\bar{Y}_{\bullet j} + u_{j(x)} + e_{ij(x)} \\ Y_{ij} &= \beta_{0(y)} + \beta_{1(y)}(X_{ij} - \bar{X}_{\bullet j}) + \beta_{2(y)}\bar{X}_{\bullet j} + u_{j(y)} + e_{ij(y)}. \end{aligned} \quad (4)$$

The FCS approach iterates between these equations, generating imputations for each missing variable in turn. If both variables are affected by missing data, then the group means are updated at each iteration of the sampling algorithm on the basis of the most recent imputations for  $X$  and  $Y$  (passive imputation; see below). Similar to JM, unsystematic differences between groups in  $X$  and  $Y$  are captured by the inclusion of random effects,  $u_{j(x)}$  and  $u_{j(y)}$ . In contrast to JM, however, the FCS approach uses the observed group means,  $\bar{Y}_{\bullet j}$  and  $\bar{X}_{\bullet j}$ , to represent the different relations between  $X$  and  $Y$  at the individual and the group level.<sup>2</sup> For missing  $Y$ , there is no difference between the imputation and the analysis model. For missing  $X$ , the imputation model has similar implications

as in the JM approach, but it relies not only on random effects but also on the observed group means to represent the relation between  $X$  and  $Y$  at the group level. In applications with more than two variables, the general approach remains the same: For each additional variable with missing data, an additional equation must be specified, each conditioning on the other variables and their respective group means.

*Summary.* Summing up, there are two points worth noting. First, both the JM and the FCS approach allow for different relations between variables to be estimated at the individual and the group level. Second, the two approaches differ in the way in which they accomplish this task. In the JM, the group level is represented by random effects, whereas the FCS approach relies on the observed group means. However, even though the general approach is different, it has been argued that the two approaches imply similar covariance structures at the individual and the group level and can be used interchangeably (e.g., Carpenter & Kenward, 2013, p. 220; Lüdtke et al., 2017; Mistler, 2015; however, see also Resche-Rigon & White, in press). Therefore, we expected the two procedures to yield approximately the same, unbiased parameter estimates, making both suitable for MI in quite general applications of the multilevel random intercept model.

### *Model-Based Treatment using FIML*

As an alternative to MI, it is often possible to use model-based procedures such as FIML to treat missing data (for an introduction, see Enders, 2010). FIML is often considered to be very user-friendly because missing data are handled directly during the estimation of the analysis model without requiring any additional steps to be taken by the user (e.g., Allison, 2012; Graham, 2009). Currently, the most popular and versatile implementation of FIML for multilevel models is available in the statistical software *Mplus* (L. K. Muthén & Muthén, 2012). FIML estimates the parameters of the analysis model directly from the incomplete data set by maximizing the observed-data likelihood. As a result, the use of FIML to treat missing data is closely tied to the analysis model (Schafer & Graham, 2002). In the traditional multilevel model (e.g., Equation 1), the observed-data likelihood includes only the dependent variable in the analysis (e.g.,  $Y$ ), and distributional assumptions are imposed only on that variable. For that reason, FIML initially deals with missing data only in the dependent variable, whereas cases with missing data in explanatory variables are often discarded (see also Hox, van Buuren, & Jolani, 2016). To treat missing data in explanatory variables (e.g.,  $X$ ), the model must be extended in such a way that the likelihood function will incorporate all variables with missing data, thus imposing additional distributional assumptions on the data. In *Mplus*, this is typically achieved by specifying a set of latent variables for the explanatory variables with missing data (for an illustration, see Enders, 2010).

Although it may not be immediately obvious, this strategy can have negative side-effects in multilevel modeling because of the way in which *Mplus* estimates multilevel models with latent variables. For example, consider the model in Equation 1 with missing values in  $X$  and  $Y$ . To estimate this model, *Mplus* uses a decomposition approach similar to the JM, in which the two variables are decomposed into (latent) individual- and group-level components, each of which is assumed to follow a multivariate normal distribution (Rabe-Hesketh, Skrondal, & Zheng, 2012). However, in doing so, *Mplus* adopts a different analysis model in which the group-level effects of  $X$  on  $Y$  are represented by latent variables instead of observed group means (i.e.,  $X_{\bullet j}$ ) as they are in the analysis model (for a discussion, see Lüdtke et al., 2008). As a result, parameter estimates may change substantially, both in meaning and in value (Grund, Lüdtke, & Robitzsch, in press). To avoid this shift in the analysis model, the user may calculate the group means beforehand from the observed data and specify a latent variable only for the within-group component of the explanatory variables (i.e.,  $X_{ij} - \bar{X}_{\bullet j}$ ). This strategy tends to reduce bias in group-level effects and will be preferred for the remainder of this article (see also Grund et al., in press). In addition, because the

**Table 1.** Simulation Conditions for the Data-Generating Model and the Generation of Missing Values

	Study 1	Study 2	Study 3a	Study 3b
<b>Data conditions</b>				
No. of individuals	5, 10	5, 10	5, 10	5, 10
No. of groups	50, 100, 200, 500	50, 100, 200, 500	50, 100, 200, 500	50, 100, 200, 500
ICC of $X$ and $Y$	.10, .20, .50	.10, .20, .50	.10, .20, .50	.10, .20, .50
ICC of $D$			.10, .20, .50	
Correlation $XW$		.20		.20
Correlation $XD$			.20	
<b>Model parameters</b>				
Effect of $(X_{ij} - \bar{X}_{\bullet j})$	.20	.50	.50	0
Effect of $\bar{X}_{\bullet j}$	.20, .50	0	.50	.20
Effect of $W_j$		.35		.20
Effect of $D_{ij}$			.20	
CLI $(X_{ij} - \bar{X}_{\bullet j})W_j$		.0, .20		
GLI $\bar{X}_{\bullet j}W_j$		0		
Total slope variance		.10		
Int.-slope covariance		0		
<b>Missing values</b>				
Pattern	$Y \sim X, X \sim Y$	$Y \sim X, X \sim Y$	$D \sim Y$	$W \sim Y$
Mechanism	MCAR, MAR	MCAR, MAR	MCAR, MAR	MCAR, MAR
Proportion (%)	25	25	25	25
No. of conditions	192	192	48	48

Note: The residual intercept and slope variance were determined by the remaining simulation parameters and by setting a target value for the ICC of  $Y$  and the total slope variance ( $\gamma_{11}^2 + \tau_j^2$ ). CLI = cross-level interaction; GLI = group-level interaction.

individual- and group-level components are assumed to follow a multivariate normal distribution, only linear relations are allowed between variables with missing data, and handling missing data in categorical variables may be challenging. The *Mplus* syntax files needed to perform FIML estimation for the models presented here are given in the supplemental online materials.

### Study 1: Random Intercept Models

Next, we present findings from a computer simulation study in which we compared the performance of different MI procedures in the context of multilevel random intercept models. In addition to the JM and the FCS approach, we also investigated single-level MI, which ignores the multilevel structure altogether, LD, and FIML as discussed above. The main question was which procedures would preserve the relevant features of the substantive analysis model. Here, we provide only a brief sketch of the study's design. For interested readers, we provide further details in Appendix A.

The substantive analysis model was the random intercept model in Equation 1, and the data were generated from this model. The parameters of the data-generating model were chosen in such a way that they would imply a given value for the intraclass correlations (ICCs) of  $X$  and  $Y$ . Missing data were generated on  $Y$  or  $X$  in either a random fashion (MCAR) or conditional on the other variable (MAR). A summary of the simulation conditions is provided in Table 1. We varied the number of groups ( $k = 50, 100, 200, 500$ ), the number of individuals within each group ( $n = 5, 10$ ), the ICCs of  $X$  and  $Y$  ( $\rho_{I,X} = \rho_{I,Y} = .10, .20, .50$ ), the effect of  $\bar{X}_{\bullet j}$  ( $\gamma_{01} = .20, .50$ ), and the missing data mechanism. The effect of  $(X_{ij} - \bar{X}_{\bullet j})$  was held constant at .20 ( $\gamma_{10}$ ), thus providing conditions in which the effects at the individual and the group level were equal or different in the population

model. Taken together, these conditions mimic typical applications of multilevel models in cross-sectional and longitudinal organizational research (e.g., smaller and larger ICCs, smaller and larger numbers of observations per unit or group). In addition, they provide information about the small- and large-sample properties of each procedure and about conditions that are interesting from a methodological point of view (e.g., with or without contextual effects). Each condition was replicated 1,000 times. We applied the following procedures to each data set:

1. LD
2. single-level FCS, ignoring the multilevel structure (FCS-SL)
3. multilevel FCS with separate within- and between-group effects (Equation 4; FCS-ML)
4. multilevel JM (Equation 2; JM)
5. FIML

The parameters of interest were the ICC of  $Y$ , estimated from an empty model, and the regression coefficients within and between groups,  $\gamma_{10}$  and  $\gamma_{01}$ , from the substantive analysis model. For each condition, each procedure, and each parameter, we calculated the bias, the RMSE, and the coverage of the 95% confidence interval to evaluate performance. The bias is defined as the difference between an estimator's average value and its true value. The RMSE is the square root of the average squared difference between average estimates and true values, combining information about bias and efficiency of parameter estimates. The coverage of the 95% confidence interval denotes the relative frequency with which the 95% confidence interval covers the true value. The properties of an estimator may be considered suboptimal if the bias exceeds 10%, the RMSE is large in comparison with other procedures, or the coverage rate is below 90% (or very close to 100%).

**Results.** Our findings are summarized in Table 2 and Figure 1. The complete collection of results for the parameters of interest is provided in the supplemental online materials. Consistent with our expectations, single-level MI (FCS-SL) reduced the ICC of  $Y$  when  $Y$  was partially missing. In such a case, the between-group regression coefficient was biased downwards, and the within-group regression coefficient was biased upwards to different extents as determined by the true magnitude of the ICCs of  $X$  and  $Y$ . With missing values in  $X$ , FCS-SL either over- or understated the true size of the regression coefficients, depending on the ICCs. As shown in Figure 1, this bias did not decrease in larger samples. The results from the two appropriate MI procedures (FCS-ML and JM) were similar to one another. Both procedures had a slight tendency to overestimate the ICC of  $Y$  and to underestimate the between-group coefficient in smaller samples (i.e.,  $k = 100$  or lower, with  $n = 5$ ) with low ICCs ( $\rho_{I,X} = \rho_{I,Y} = .10$ ). However, this bias was seldom substantial and decreased as the sample size increased (see Figure 1). FIML produced unbiased estimates of the regression coefficients with missing  $Y$  but biased estimates of the between-group regression coefficient ( $\gamma_{01}$ ) with missing  $X$ . Finally, LD led to substantially biased estimates of all parameters of interest when data were MAR, especially when values were missing in  $X$ . In conditions in which the within- and between-group coefficients were equal ( $\gamma_{01} = .20$ ), the results were essentially the same, and both JM and FCS-ML provided approximately unbiased estimates of the parameters of interest.

The coverage of the 95% confidence interval was acceptable in all conditions for FCS-ML and in all but the most extreme conditions under JM. However, owing to persistent bias, the coverage rates under FCS-SL frequently dropped below 90% in larger samples (i.e., above  $k = 200$ ,  $n = 10$  or  $k = 500$ ,  $n = 5$ ). Coverage rates for FIML were acceptable with missing  $Y$ , but dropped below 90% with missing  $X$ ; those for LD were acceptable under MCAR but unacceptable under MAR. Finally, the RMSE tended to be lowest under FCS-ML and JM as well as under FIML with missing  $Y$ . By contrast, the RMSE for the parameters of interest tended to be larger under FCS-SL as well as under LD and FIML with missing  $X$ , indicating that these procedures were altogether less accurate and

**Table 2.** Bias (in %), RMSE, and Coverage of the 95% Confidence Interval for the ICC of Y and the Within- and Between-Group Regression Coefficients in Study I (Small Groups, n = 5).

	LD			FCS-SL			FCS-ML			JM			FIML		
	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.
Missing Y ~ X (MAR, 25%)															
k = 100	$\rho_{t,x} = \rho_{t,y} = .10$														
$\hat{\gamma}_{t0}$	-1.4	0.06	94.2	9.6	0.06	94.0	-1.3	0.06	93.7	1.6	0.06	95.1	-1.4	0.06	94.5
$\hat{\gamma}_{t1}$	-9.1	0.11	92.6	-10.0	0.10	95.0	1.1	0.10	95.7	-4.5	0.10	97.1	1.1	0.10	95.4
$\hat{\rho}_{t,y}$	-2.0	0.05	—	-30.7	0.04	—	28.6	0.05	—	24.7	0.05	—	12.6	0.04	97.5
k = 500	$\rho_{t,x} = \rho_{t,y} = .10$														
$\hat{\gamma}_{t0}$	0.0	0.03	95.8	10.4	0.03	90.0	-0.1	0.03	96.1	2.6	0.03	95.6	-0.1	0.03	96.0
$\hat{\gamma}_{t1}$	-10.0	0.07	77.4	-10.9	0.07	76.3	-0.1	0.05	95.4	-3.7	0.05	94.4	-0.0	0.05	95.3
$\hat{\rho}_{t,y}$	-4.3	0.02	—	-32.1	0.04	—	13.2	0.02	—	13.7	0.02	—	7.1	0.02	97.9
k = 100	$\rho_{t,x} = \rho_{t,y} = .50$														
$\hat{\gamma}_{t0}$	1.3	0.06	94.5	24.0	0.08	91.1	1.7	0.06	94.2	1.2	0.06	93.8	1.5	0.06	94.2
$\hat{\gamma}_{t1}$	-2.7	0.09	94.6	-6.4	0.09	91.9	-0.5	0.09	95.1	-0.8	0.09	94.9	-0.5	0.09	94.8
$\hat{\rho}_{t,y}$	-1.5	0.05	—	-35.0	0.18	—	-0.8	0.05	—	-0.8	0.05	—	-1.3	0.05	95.8
k = 500	$\rho_{t,x} = \rho_{t,y} = .50$														
$\hat{\gamma}_{t0}$	-0.2	0.03	95.5	22.3	0.05	72.4	-0.2	0.03	95.3	-0.3	0.03	96.0	-0.2	0.03	96.0
$\hat{\gamma}_{t1}$	-2.6	0.04	92.9	-6.2	0.05	83.0	-0.4	0.04	93.5	-0.3	0.04	93.5	-0.3	0.04	93.7
$\hat{\rho}_{t,y}$	-1.3	0.03	—	-34.7	0.17	—	-0.5	0.02	—	-0.4	0.02	—	-0.5	0.02	94.2
Missing X ~ Y (MAR, 25%)															
k = 100	$\rho_{t,x} = \rho_{t,y} = .10$														
$\hat{\gamma}_{t0}$	-6.9	0.06	94.3	5.9	0.06	96.1	1.5	0.06	94.6	4.2	0.06	94.6	1.2	0.06	95.1
$\hat{\gamma}_{t1}$	-17.3	0.12	81.7	-1.7	0.09	98.5	-8.1	0.10	95.7	-11.5	0.10	94.1	-23.2	0.14	68.3
$\hat{\rho}_{t,y}$	-11.9	0.05	—	-0.8	0.04	—	-0.8	0.04	—	-0.8	0.04	—	4.6	0.04	97.3
k = 500	$\rho_{t,x} = \rho_{t,y} = .10$														
$\hat{\gamma}_{t0}$	-6.8	0.03	91.6	6.6	0.03	92.7	1.1	0.03	94.1	3.7	0.03	93.9	1.6	0.03	94.2
$\hat{\gamma}_{t1}$	-18.0	0.10	36.0	-1.8	0.04	97.7	-4.1	0.04	93.9	-7.5	0.05	89.7	-24.1	0.13	8.4
$\hat{\rho}_{t,y}$	-11.8	0.03	—	0.5	0.02	—	0.5	0.02	—	0.5	0.02	—	3.8	0.02	96.8

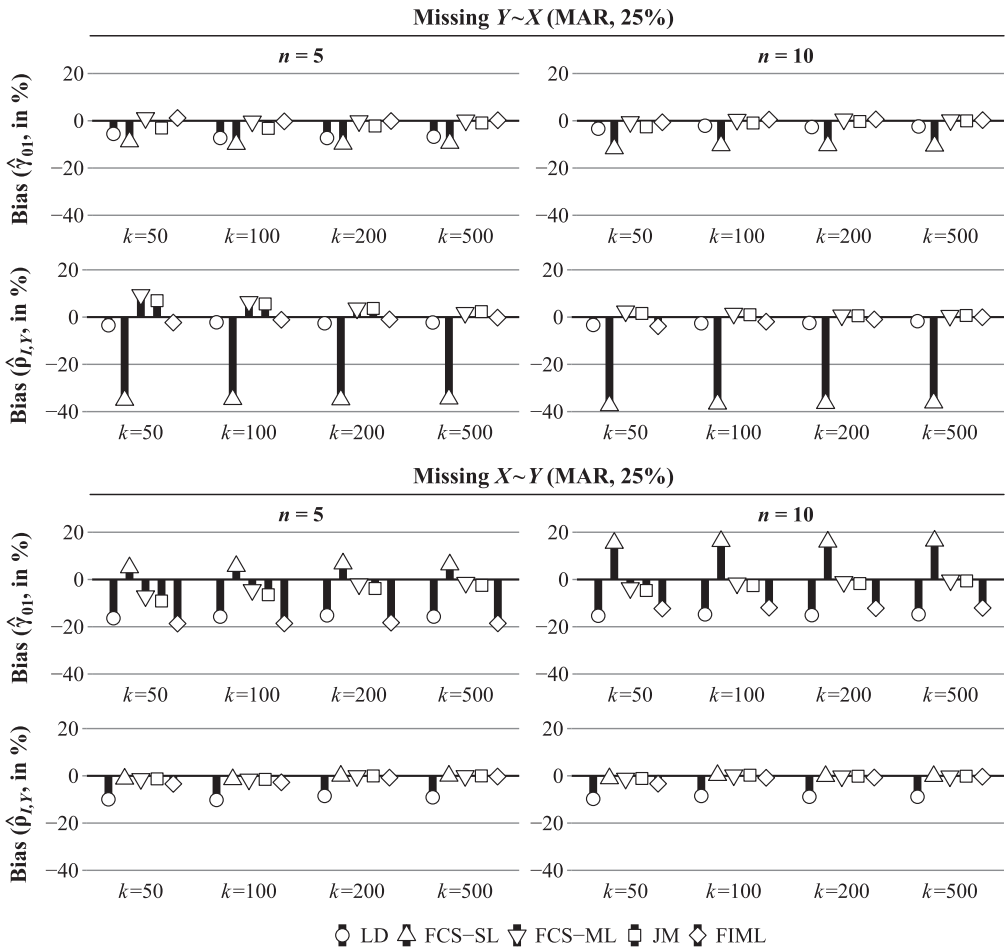
(continued)



**Table 2.** (continued)

	LD			FCS-SL			FCS-ML			JM			FIML		
	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.
$k = 100$															
$\hat{\gamma}_{10}$	-4.7	0.06	94.3	-9.6	0.05	96.7	-1.2	0.06	94.8	-1.1	0.06	94.5	0.1	0.06	94.1
$\hat{\gamma}_{01}$	-10.5	0.10	90.5	20.9	0.14	83.2	-0.9	0.09	95.6	-1.0	0.09	96.0	-9.2	0.09	90.9
$\hat{\rho}_{1,Y}$	-4.8	0.06	—	-0.9	0.05	—	-0.9	0.05	—	-0.9	0.05	—	-1.5	0.05	93.8
$k = 500$															
$\hat{\gamma}_{10}$	-4.0	0.03	93.9	-9.0	0.03	91.8	-0.1	0.03	95.3	-0.2	0.03	95.2	0.5	0.03	94.8
$\hat{\gamma}_{01}$	-9.8	0.06	74.8	21.4	0.12	36.3	-0.4	0.04	94.4	-0.4	0.04	95.2	-8.7	0.06	77.5
$\hat{\rho}_{1,Y}$	-4.1	0.03	—	-0.2	0.02	—	-0.2	0.02	—	-0.2	0.02	—	-0.3	0.02	94.6

Note:  $\hat{\gamma}_{10}$  = within-group regression coefficient;  $\hat{\gamma}_{01}$  = between-group regression coefficient;  $\hat{\rho}_{1,Y}$  = ICC of Y (estimated from an empty model); LD = listwise deletion; FCS-SL = single-level FCS; FCS-ML = multilevel FCS; JM = multilevel JM; FIML = full-information maximum likelihood.



**Figure 1.** Estimated bias for the between-group regression coefficient of  $X$  ( $\gamma_{01}$ ) and the ICC of  $Y$  ( $\rho_{l,Y}$ ) in Study I for different numbers of individuals ( $n$ ) and groups ( $k$ ), moderate ICCs ( $\rho_{l,X} = \rho_{l,Y} = .20$ ), and different missing data mechanisms (MAR;  $Y \sim X$  and  $X \sim Y$ ). LD = listwise deletion; FCS-SL = single-level FCS; FCS-ML = multilevel FCS; JM = multilevel JM; FIML = full-information maximum likelihood.

efficient. For example, the average RMSE for the between-group regression coefficient was 3.3% larger under LD with missing  $Y$  as compared with JM; with missing  $X$ , this difference increased to 9.7%. Note also that the small-sample bias under JM and FCS-ML (e.g., for the ICC of  $Y$ ) did not increase the RMSE, indicating that these procedures remained accurate and efficient overall, even in smaller samples. All in all, our results suggest that the JM and the FCS-ML approach are equally appropriate in the context of the multilevel random intercept model.

### Random Slopes and Cross-Level Interactions

Beyond the scope of random intercept models, those engaged in organizational research often seek to understand how the effects of various quantities differ across higher level organizational units. Multilevel models with random slopes allow (a) individual-level effects to vary across groups and (b) for the inclusion of group-level explanatory variables to explain that variability (i.e., CLIs).

Recently, Aguinis & Culpepper (2015) stated that random slopes and CLIs were “at the heart of [...] any theory that considers outcomes to be a result of combined influences emanating from different levels of analysis,” adding that “the extent to which we understand the presence of cross-level interactions is an indication of theoretical progress” (p. 156).

Consider a multilevel random coefficients model (Snijders & Bosker, 2012) in which an individual-level outcome  $Y$  is regressed on an individual-level variable  $X$  and a group-level variable  $W$ . In addition to the random intercept, we allow the individual-level slope to vary across groups, and we include a CLI to account for some of that variation

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + \gamma_{02}W_j + \gamma_{11}W_j(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{03}W_j\bar{X}_{\bullet j} + u_{0j} + u_{1j}(X_{ij} - \bar{X}_{\bullet j}) + e_{ij}, \quad (5)$$

where  $u_{1j}$  denotes the random slope associated with  $(X_{ij} - \bar{X}_{\bullet j})$  in group  $j$ ,  $\gamma_{11}$  denotes the CLI, and  $\gamma_{03}$  denotes the group-level interaction of  $\bar{X}_{\bullet j}$  and  $W_j$ . The random effects  $(u_{0j}, u_{1j})^T$  are assumed to follow a multivariate normal distribution with mean zero and covariance matrix  $\mathbf{\Gamma}$ .

Two aspects of the model in Equation 5 are worth noting. First, the slope of the regression of  $Y$  on  $X$  is assumed to vary across groups. Incorporating the variability in the slope in the imputation model is particularly important if the slope variance itself is of interest, because ignoring the slope variance may lead one to underestimate it in subsequent analyses. Second, the CLI denotes the degree to which the effect of  $(X_{ij} - \bar{X}_{\bullet j})$  changes as a function of  $W_j$ . Thus, if estimating the CLI is of interest, then the imputation model should allow for the individual-level effect of  $X$  to interact with  $W$  (similarly for the interaction at the group level).

### Accommodating Random Slopes and CLIs

In contrast to applications in the multilevel random intercept model, performing MI is not straightforward when the model of interest includes random slopes and CLIs, particularly when the explanatory variables contain missing data (e.g., Enders et al., 2016; Gottfredson, Sterba, & Jackson, 2016; Grund, Lüdtke, & Robitzsch, 2016a). For example, consider the model of interest in Equation 5. In order for an imputation model to be consistent with this model of interest, it has to acknowledge the fact that the relation between  $X$  and  $Y$  is assumed to vary both systematically as a function of  $W$  (i.e., due to the interaction effects) and unsystematically (i.e., due to random slopes). However, the presence of such terms implies a complex joint distribution for the dependent and explanatory variables which is difficult to emulate in conventional software for multilevel MI (e.g., Kim, Sugar, & Belin, 2015). More advanced methods for accommodating the model of interest when generating imputations are currently being developed, but these are not yet available in standard software for multilevel MI (for further details, see the Discussion section). For this reason, we focus on the procedures that are available in standard software for multilevel MI, which often provide options for accommodating random slopes and CLIs, albeit to different (and arguably imperfect) extents.

*Joint modeling (JM).* As mentioned earlier, modeling the joint distribution of the dependent and explanatory variables in a general manner is not a straightforward endeavor if the model of interest includes random slopes or interaction effects. For this reason, we used `pan` to implement the JM in a manner that is similar to what we presented above. We assume that  $X$  and  $Y$  are treated as target variables (i.e., on the left-hand side), whereas  $W$  is assumed to be completely observed and written on the right-hand side of the model. Thus, the imputation model becomes

$$[X_{ij}, Y_{ij}]^T = [\beta_{0(x)}, \beta_{0(y)}]^T + W_j[\beta_{w(x)}, \beta_{w(y)}]^T + [u_{j(x)}, u_{j(y)}]^T + [e_{ij(x)}, e_{ij(y)}]^T, \quad (6)$$

where  $[\beta_{w(x)}, \beta_{w(y)}]^T$  denotes the vector of regression coefficients from regressing  $X$  and  $Y$  on  $W$ , and the remaining notation is as before. In this specification, the joint model includes possible relations among the three variables at the group level as well as relations between  $X$  and  $Y$  at the individual and the group level. However, the joint model includes only a random intercept for each target variable, whereas the slope variance and the interaction effects in the analysis model are completely ignored. In general, the JM approach may still provide reasonable estimates of the regression coefficients when the substantive analysis model contains random slopes because the inclusion of random slopes does not change the expected value of the estimates for the regression coefficients. However, when the substantive model also includes interaction effects, the integrity of its estimates may be compromised.

*The FCS approach.* To address missing data in multilevel models with random slopes, it has been recommended that researchers specify conditional models that include varying slopes between pairs of variables (Enders et al., 2016). In addition, product terms involving  $W$  can be introduced to accommodate the CLI. If both random slopes and product terms are included, the two conditional models become

$$\begin{aligned} X_{ij} &= \beta_{0(x)} + \beta_{1(x)}(Y_{ij} - \bar{Y}_{\bullet j}) + \beta_{2(x)}\bar{Y}_{\bullet j} + \beta_{w(x)}W_j + \beta_{1wy(x)}W_j(Y_{ij} - \bar{Y}_{\bullet j}) + \beta_{2wy(x)}W_j\bar{Y}_{\bullet j} \\ &\quad + u_{0j(x)} + u_{1j(x)}(Y_{ij} - \bar{Y}_{\bullet j}) + e_{ij(x)} \\ Y_{ij} &= \beta_{0(y)} + \beta_{1(y)}(X_{ij} - \bar{X}_{\bullet j}) + \beta_{2(y)}\bar{X}_{\bullet j} + \beta_{w(y)}W_j + \beta_{1xw(y)}W_j(X_{ij} - \bar{X}_{\bullet j}) + \beta_{2xw(y)}W_j\bar{X}_{\bullet j} \\ &\quad + u_{0j(y)} + u_{1j(y)}(X_{ij} - \bar{X}_{\bullet j}) + e_{ij(y)}, \end{aligned} \quad (7)$$

where  $u_{0j(\cdot)}$  and  $u_{1j(\cdot)}$  denote the random intercepts and slopes in the conditional models, and the coefficients  $\beta_{1wy(x)}$ ,  $\beta_{1xw(y)}$ ,  $\beta_{2wy(x)}$ , and  $\beta_{2xw(y)}$  denote the interaction effects by which the within- and between-group relations of  $X$  and  $Y$  change as a function of  $W$ .

There are two aspects worth noting. The first is related to the way in which random slopes are handled in the conditional models. The imputation model for missing values in  $Y$  is identical to the analysis model. Thus, imputing  $Y$  should be straightforward. However, previous research has shown that missing values in the explanatory variable  $X$  pose a much greater challenge because “reversing” the random slope model may produce biased estimates of the regression coefficients and the slope variance in the analysis model (Gottfredson et al., 2016; Grund et al., 2016a; see also Enders et al., 2016). This is not entirely surprising because the analysis model (Equation 5) and the imputation model for missing  $X$  (Equation 7, first line) make different statements about the varying relation between  $X$  and  $Y$ . In other words, although replacing  $u_{1j}(X_{ij} - \bar{X}_{\bullet j})$  in the analysis model by  $u_{1j(x)}(Y_{ij} - \bar{Y}_{\bullet j})$  in the imputation model may serve as a “proxy” for the relation of interest, the two statements are not equivalent.

The second aspect is related to the presence of nonlinear effects (i.e., interaction effects) in the conditional models. At each iteration of the FCS algorithm, the product terms  $W_j(X_{ij} - \bar{X}_{\bullet j})$  and  $W_j(Y_{ij} - \bar{Y}_{\bullet j})$  must be “updated” to incorporate the most recent imputations of  $X$  and  $Y$ . The simplest strategy for updating the products terms is to recalculate them after  $X$  and  $Y$  have been imputed. This is commonly referred to as “passive imputation” (Royston, 2004; van Buuren, 2012). As an alternative, product terms may be regarded as “just another variable” (von Hippel, 2009). This strategy replaces the passive imputation step with an imputation model for each product term (e.g., a regression model). However, both strategies have been shown to yield biased parameter estimates (e.g., Seaman, Bartlett, & White, 2012; Vink & van Buuren, 2013) because they do not correctly reflect the complex joint distribution of the dependent and explanatory variables in the model when the model of interest includes interaction effects (Kim et al., 2015). In the present study, we used passive imputation because (a) it is easy to use and readily available in standard software and (b) implementing “just another variable” is not straightforward with group-mean-centered data.<sup>3</sup>

*FIML*. Similar to MI, analyzing the incomplete data with FIML can be difficult if the model of interest includes random slopes and CLIs. As before, we focus on FIML estimation in the statistical software *Mplus*. If missing data occur only on  $Y$ , estimating the model of interest in *Mplus* is straightforward because the observed-data likelihood can be evaluated directly on the basis of the incomplete data. However, if missing values occur on  $X$ , it is currently not possible to include  $X$  in the analysis model in *Mplus* without dropping cases with missing  $X$  from the analysis (for a discussion, see also Shin & Raudenbush, 2010).

*Summary*. In contrast to applications in the multilevel random intercept model, missing data pose a greater challenge when the model of interest includes random slopes. Multilevel MI can be expected to provide proper results when only the dependent variable  $Y$  contains missing data. However, if the explanatory variable with a random slope,  $X$ , contains missing data, conducting MI is not straightforward. Specifically, the “reversed” imputation model for missing  $X$  contains only a proxy for the relation of interest, and accommodating product terms (i.e., CLIs and group-level interactions) is still an open area of research (see the Discussion section). As a result, neither JM nor FCS was expected to provide perfect results.

## Study 2: Random Slope Models

In this section, we present findings from a simulation study in which we compared different MI procedures in the context of multilevel models with random slopes and CLIs. The model of interest was the random coefficients model presented in Equation 5, and the data were also generated from this model (see Appendix A). The parameters of the data-generating model were chosen in such a way as to imply a given value for the ICCs of  $X$  and  $Y$  and a given “total” variance for the random slope (i.e.,  $Var(\beta_{1j}) = Var(\gamma_{11}W_j + u_{1j}) = \gamma_{11}^2 + \tau_1^2$ ). Missing data were generated as before (MCAR and MAR on either  $X$  or  $Y$ ). A summary is presented in Table 1. We varied the number of groups ( $k = 50, 100, 200, 500$ ), the number of individuals ( $n = 5, 10$ ), the size of the CLI ( $\gamma_{11} = 0, .20$ ), and the missing data mechanism. The effect of  $(X_{ij} - \bar{X}_{\bullet j})$  was held constant at .50 ( $\gamma_{10}$ ) and the effect of  $W_j$  at .35 ( $\gamma_{02}$ ); the remaining effects were set to zero. Each condition was replicated 1,000 times. We applied the following procedures to each data set:

1. LD
2. single-level FCS, ignoring the multilevel structure and product terms (FCS-SL)
3. multilevel FCS, ignoring random slopes but including passive imputation of product terms (FCS-CLI/no RS)
4. multilevel FCS including random slopes and passive imputation of product terms (Equation 7; FCS-CLI/RS)
5. FIML

The FIML estimation was conducted in *Mplus* as described above. Because FIML could not be used to estimate the model in conditions with missing  $X$ , we included FIML only for conditions with missing  $Y$ . The parameters of interest were the within-group regression coefficient of  $X$  ( $\gamma_{10}$ ), the effect of  $W$  ( $\gamma_{02}$ ), the CLI ( $\gamma_{11}$ ), and the slope variance ( $\tau_1^2$ ). For each condition, each procedure, and each parameter, we calculated the bias, the RMSE, and the coverage rate of the 95% confidence interval as before.

*Results*. Our main findings are summarized in Table 3 and Figure 2. In presenting our results, we focus on the MI procedures because FIML could be applied only in conditions with missing  $Y$ , and the estimates were approximately unbiased in these conditions. For the remaining procedures, the difference between cases with missing data in  $Y$  and  $X$  was substantial, and sample size continued to

**Table 3.** Bias (in %), RMSE, and Coverage of the 95% Confidence Interval for the Within-Group Regression Coefficient of X, the Between-Group Regression Coefficient of W, and the CLI of X with W in Study 2 (Small Groups,  $n = 5$ ).

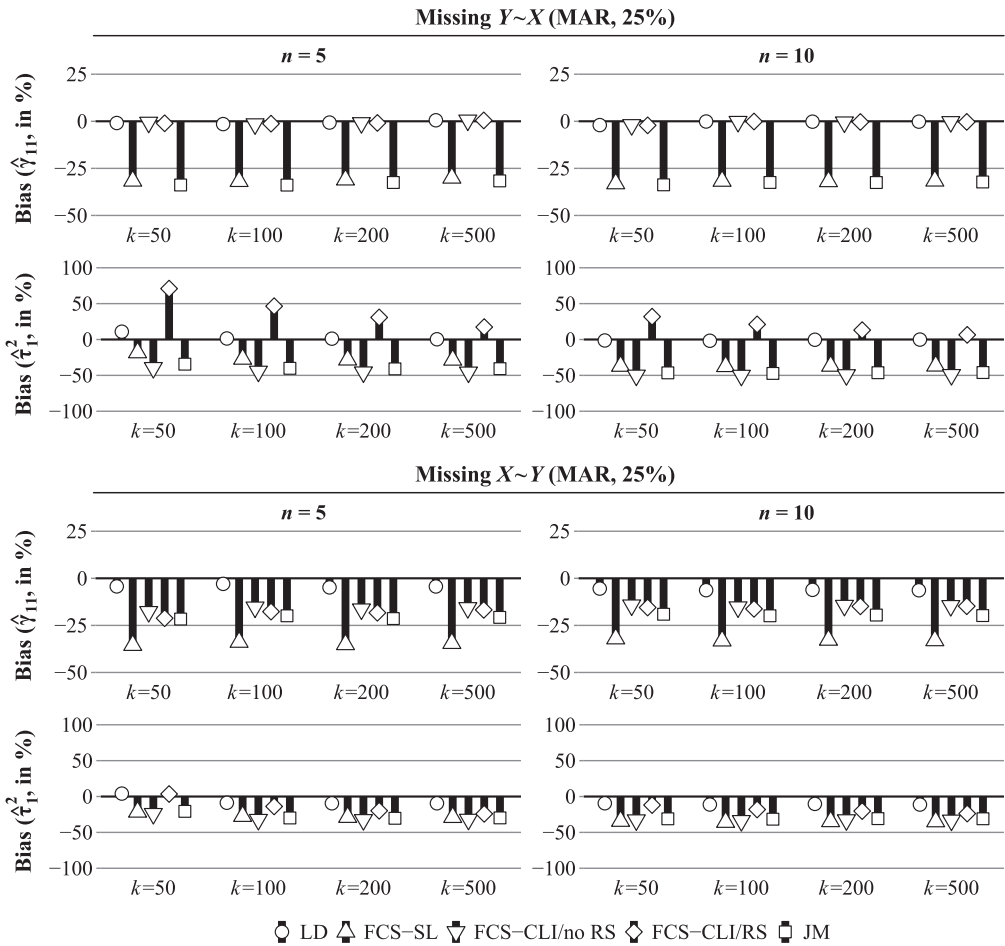
	LD			FCS-SL			FCS-CLI/no RS			FCS-CLI/RS			JM		
	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.
Missing $Y \sim X$ (MAR, 25%)															
$k = 100$															
$\hat{\gamma}_{10}$	0.0	0.06	94.7	-7.2	0.07	91.0	0.2	0.06	92.4	0.0	0.06	95.7	-2.2	0.06	92.8
$\hat{\gamma}_{02}$	-10.3	0.06	89.5	-11.1	0.06	87.8	-0.4	0.05	95.5	-0.2	0.05	94.0	-9.6	0.06	91.2
$\hat{\gamma}_{11}$	0.1	0.06	94.5	-30.7	0.08	83.4	0.2	0.06	92.4	0.3	0.06	96.1	-32.5	0.08	80.1
$k = 500$															
$\hat{\gamma}_{10}$	-0.1	0.03	95.4	-7.6	0.05	68.3	-0.2	0.03	93.1	-0.1	0.02	95.9	-2.4	0.03	90.8
$\hat{\gamma}_{02}$	-9.8	0.04	70.0	-10.6	0.04	57.9	-0.0	0.02	95.5	0.1	0.02	95.6	-9.0	0.04	71.7
$\hat{\gamma}_{11}$	-0.5	0.03	95.2	-30.9	0.06	24.2	-0.5	0.02	94.1	-0.5	0.02	96.0	-32.2	0.07	16.5
Missing $Y \sim X$ (MAR, 25%)															
$k = 100$															
$\hat{\gamma}_{10}$	0.2	0.06	94.0	-15.6	0.10	83.5	0.2	0.06	92.3	0.1	0.06	95.2	-1.3	0.06	92.1
$\hat{\gamma}_{02}$	-6.4	0.08	95.0	-6.7	0.08	90.5	0.3	0.08	95.1	0.4	0.08	94.5	-4.9	0.08	95.1
$\hat{\gamma}_{11}$	1.1	0.06	94.7	-31.0	0.08	92.8	1.0	0.06	93.6	1.1	0.06	96.1	-31.9	0.08	83.9
$k = 500$															
$\hat{\gamma}_{10}$	0.2	0.03	95.4	-15.5	0.08	27.7	0.2	0.03	93.9	0.2	0.03	95.7	-1.3	0.03	92.0
$\hat{\gamma}_{02}$	-6.2	0.04	90.7	-6.2	0.04	83.9	0.4	0.03	94.7	0.4	0.03	94.7	-4.7	0.04	91.5
$\hat{\gamma}_{11}$	-0.8	0.03	94.9	-31.4	0.07	44.2	-0.8	0.03	93.1	-0.7	0.03	95.6	-33.1	0.07	17.1
Missing $X \sim Y$ (MAR, 25%)															
$k = 100$															
$\hat{\gamma}_{10}$	-4.9	0.06	91.4	-12.6	0.08	80.7	-2.9	0.06	92.8	-5.9	0.06	91.0	-2.7	0.06	93.4
$\hat{\gamma}_{02}$	-14.9	0.07	81.6	-1.7	0.04	95.2	0.1	0.04	95.5	0.0	0.04	95.2	-0.2	0.04	95.2
$\hat{\gamma}_{11}$	-4.2	0.06	93.8	-28.4	0.07	87.8	-14.8	0.05	94.3	-17.3	0.06	95.2	-19.9	0.06	93.2
$k = 500$															
$\hat{\gamma}_{10}$	-5.1	0.04	82.4	-12.5	0.07	23.7	-3.2	0.03	87.9	-4.7	0.03	84.3	-2.8	0.03	89.9
$\hat{\gamma}_{02}$	-15.1	0.06	32.6	-2.0	0.02	92.6	-0.1	0.02	96.0	-0.1	0.02	95.8	-0.5	0.02	95.2
$\hat{\gamma}_{11}$	-4.9	0.03	94.0	-28.4	0.06	29.5	-15.3	0.04	77.2	-16.5	0.04	75.0	-20.2	0.04	62.3

(continued)

**Table 3.** (continued)

	LD			FCS-SL			FCS-CLI/no RS			FCS-CLI/RS			JM		
	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.
$k = 100$															
$\hat{\gamma}_{10}$	-2.6	0.06	94.1	-37.9	0.19	4.3	-3.7	0.06	93.3	-6.6	0.06	92.8	-3.1	0.06	93.3
$\hat{\gamma}_{02}$	-10.7	0.08	89.9	-3.1	0.08	95.1	-0.6	0.07	95.3	-0.7	0.07	95.5	-0.4	0.07	95.2
$\hat{\gamma}_{11}$	-2.9	0.06	93.8	-54.5	0.12	46.5	-16.7	0.06	92.2	-19.5	0.06	92.4	-21.7	0.06	90.8
$k = 500$															
$\hat{\gamma}_{10}$	-3.2	0.03	89.5	-38.6	0.19	0.0	-3.4	0.03	87.6	-4.8	0.03	82.0	-2.9	0.03	88.8
$\hat{\gamma}_{02}$	-10.5	0.05	78.2	-2.5	0.03	93.7	-0.1	0.03	95.3	-0.1	0.03	95.5	-0.0	0.03	95.2
$\hat{\gamma}_{11}$	-2.6	0.03	94.4	-54.4	0.11	0.1	-15.6	0.04	76.9	-16.6	0.04	74.6	-20.9	0.05	60.4

Note:  $\hat{\gamma}_{10}$  = within-group regression coefficient of X;  $\hat{\gamma}_{02}$  = between-group regression coefficient of W;  $\hat{\gamma}_{11}$  = CLI; LD = listwise deletion; FCS-CLI/no RS = multilevel FCS including only product terms; FCS-CLI/RS = multilevel FCS including product terms and random slopes; JM = multilevel JM.



**Figure 2.** Estimated bias for the CLI ( $\gamma_{11}$ ) and the slope variance ( $\tau_1^2$ ) in Study 2 for different numbers of individuals ( $n$ ) and groups ( $k$ ), and different missing data mechanisms (MAR;  $Y \sim X$  and  $X \sim Y$ ). LD = listwise deletion; FCS-SL = single-level FCS; FCS-CLI/no RS = multilevel FCS including only product terms; FCS = multilevel FCS including product terms and random slopes; JM = multilevel JM.

play an important role. When only the dependent variable  $Y$  was incomplete, FCS-CLI/RS provided approximately unbiased estimates for the parameters of interest, with bias present for the slope variance ( $\tau_1^2$ ) in smaller samples but tending toward zero as the samples grew larger (Figure 2). The bias in smaller samples was quite large for the slope variance, especially when the samples consisted of smaller groups ( $n = 5$ ). With larger groups ( $n = 10$ ), the bias was reduced by approximately half (Figure 2).<sup>4</sup> When the random slope was ignored (FCS-CLI/no RS), we obtained almost identical estimates for the regression coefficients, but the slope variance was underestimated regardless of sample size. When both the interaction effects and the random slopes were ignored (JM), the estimates of the CLI were biased as well. Moreover, when the imputation model ignored the multi-level structure altogether (FCS-SL), all regression coefficients were biased independent of sample size. In conditions with no CLI ( $\gamma_{11} = 0$ ), the performance of FCS-CLI/RS was the same, but the bias in the slope variance was greatly reduced (see Footnote 4).



In these cases, both FCS-CLI/no RS and JM also provided approximately unbiased estimates of the regression coefficients (see also Enders et al., 2016; Grund et al., 2016a). Finally, LD provided approximately unbiased estimates of the slope variance, but the estimates of the regression coefficient for  $W$  ( $\gamma_{02}$ ) were biased under MAR regardless of sample size. The results for the coverage of the 95% confidence interval and the RMSE were in line with the bias. However, the coverage was slightly too low under FCS-CLI/no RS and JM, illustrating that the confidence intervals were slightly too narrow when the slope variance was omitted from the imputation model.

When missing values occurred in the explanatory variable  $X$ , no procedure provided unbiased estimates of the CLI and the slope variance (see Table 3 and Figure 2). Even when the product terms and random slopes were included in the model (FCS-CLI/RS), multilevel MI provided only biased estimates of the CLI and the slope variance. Ignoring the slope variance (FCS-CLI/no RS) led to slightly better estimates of the regression coefficients but increased the bias in the slope variance. Ignoring both the interaction effects and the random slopes (JM) led to further bias in the CLI but was otherwise comparable to FCS-CLI/no RS. On the other hand, single-level MI (FCS-SL) led to strongly biased estimates of both the main and interaction effects as well as the slope variance. It is interesting that LD provided the least biased estimates of the CLI and the slope variance in conditions with small groups even under MAR. On the other hand, LD introduced bias into the other estimates, particularly the main effect of  $W$  when the data were MAR. In conditions with no CLI ( $\gamma_{11} = 0$ ), FCS-CLI/RS still showed a slight downward bias in the regression coefficient of  $(X_{ij} - \bar{X}_{\bullet j})$  and the slope variance but yielded otherwise unbiased results. Ignoring the slope variance (FCS-CLI/no RS and JM) reduced the bias in the regression coefficients to essentially zero but increased bias in the slope variance. Results for LD were similar to conditions with CLI.

The coverage of the 95% confidence interval and the RMSE were closely related to the bias in the parameter estimates. For FCS-CLI/RS, the coverage was close to the nominal value of 95% for most parameters, but the coverage of the regression coefficient of  $(X_{ij} - \bar{X}_{\bullet j})$  and the CLI dropped below 90% unless the sample was very small ( $k = 50$ ,  $n = 5$ ). As a result of reduced sample size, the coverage under LD was slightly higher but also fell below 90% as the sample size increased. Similar to before, the RMSE indicated a relative loss of efficiency under LD for estimates of the regression coefficients of  $W$  and to a lesser extent  $(X_{ij} - \bar{X}_{\bullet j})$ . For example, the average RMSE for regression coefficients of  $W$  were 13.8% larger under LD with missing  $Y$  as compared with FCS-CLI/RS and 38.7% larger with missing  $X$ . For the CLI, the RMSE was usually lowest under FCS-CLI/no RS and FCS-CLI/RS in smaller samples ( $k \leq 100$ ) and under LD in larger samples ( $k \geq 200$ ). However, these differences were very small: The average RMSE for the CLI under LD as compared with FCS-CLI/RS was approximately equal with missing  $Y$  ( $< 1\%$ ) and only 3.5% larger with missing  $X$ .

Taken together, our results indicate that the FCS approach provides reliable estimates for the parameters of interest when missing values are restricted to  $Y$  and still reasonable (though imperfect) estimates with missing values on  $X$ . Even though some parameter estimates obtained from FCS were biased, they had better statistical properties overall than those of the competing methods. Ignoring the slope variance sometimes reduced bias in the regression coefficients but resulted in confidence intervals for these coefficients that were too narrow. LD provided the least biased estimates of the slope variance and the CLI but introduced bias in other parameters and tended to be slightly less efficient than MI.

## Categorical and Group-Level Missing Data

In the previous two simulation studies, the models of interest were simplified in two ways: (a) the data were always continuous, thus not accounting for missing categorical data, and (b) data were missing only at the individual level, thus not accounting for missing data in group-level variables. Therefore, we conducted two smaller simulation studies that addressed these issues separately.

### Study 3a: Missing Categorical Data

Turning back to the random intercept model, researchers are often interested in estimating the differences between groups of participants by including categorical variables in the model of interest, be it to control for group differences (e.g., due to gender, education, etc.) or to assess the effectiveness of interventions (e.g., treatment vs. control group). Especially in the former case, categorical variables may contain missing data. Here, we briefly discuss two procedures for multi-level MI—one using JM, one using FCS—that address missing data in multilevel categorical variables. We also discuss FIML estimation, and we evaluate their performance in a simulation study.

Here, the model of interest is a multilevel random intercept model with two explanatory variables at the individual level, one continuous and one binary

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{20}D_{ij} + u_{0j} + e_{ij}, \quad (8)$$

where  $D$  is a dummy-coded binary variable that takes on values for each individual  $i$  in group  $j$ . This model was also used to generate the data (see Appendix A). To ensure that  $D$  had a multilevel structure, we simulated a latent background variable  $D^*$  with a given value of its ICC. Binary values were obtained by setting  $D_{ij} = 1$  if  $D_{ij}^* > 0$ , and 0 otherwise, resulting in a 50% prevalence of either category. Missing data were induced in  $D$  as before (MCAR and MAR, based on  $Y$ ). In addition, we varied the number of individuals ( $n = 5, 10$ ) and the number of groups ( $k = 50, 100, 200, 500$ ). The remaining parameters were held constant (see Table 1). For comparison, we included LD and single-level FCS as before.

**Joint modeling (JM).** Quite general procedures that use the JM approach are available for categorical data (e.g., Goldstein, Carpenter, Kenward, & Levin, 2009). These procedures have been implemented recently in the `jomo` package in R (Quartagno & Carpenter, 2016), which allows continuous and categorical variables to be modeled simultaneously, where a categorical variable is represented by  $c - 1$  underlying latent continuous variables (where  $c$  is the number of categories). For our model of interest involving three individual-level variables, the joint model reads

$$[X_{ij}, Y_{ij}, D_{ij}^*]^T = [\beta_{0(x)}, \beta_{0(y)}, \beta_{0(d^*)}]^T + [u_{j(x)}, u_{j(y)}, u_{j(d^*)}]^T + [e_{ij(x)}, e_{ij(y)}, e_{ij(d^*)}]^T, \quad (9)$$

where  $e_{ij(d^*)}$  is constrained to have unit variance to identify the model. For missing data in  $D$ , the model is essentially a generalized linear mixed-effects model conditioning on  $X$  and  $Y$  (see Carpenter & Kenward, 2013; Goldstein et al., 2009). For dichotomous variables, equivalent procedures are available in the statistical software *Mplus*. However, for categorical variables with multiple categories, the procedures in *Mplus* differ from the approach taken in `jomo`.<sup>5</sup>

**The FCS approach.** Imputations for  $D$  may also be generated by directly conditioning on  $X$  and  $Y$  using FCS. Similar to the joint model, imputations may be generated from a generalized linear mixed-effects model (e.g., with a probit or logit link function)

$$D_{ij}^* = \beta_{0(d^*)} + \beta_{1(d^*)}(X - \bar{X}_{\bullet j}) + \beta_{2(d^*)}\bar{X}_{\bullet j} + \beta_{3(d^*)}(Y_{ij} - \bar{Y}_{\bullet j}) + \beta_{4(d^*)}\bar{Y}_{\bullet j} + u_{j(d^*)} + e_{ij(d^*)}, \quad (10)$$

where  $e_{ij(d^*)}$  is constrained in a manner similar to what is done in the JM. Unfortunately, `mice` currently allows for MI of categorical variables only in single-level models. Procedures for multi-level data have been proposed by Snijders & Bosker (2012) and Zinn (2013). The procedure used here is essentially a combination of the two and is implemented in the R package `miceadds` (Robitzsch, Grund, & Henke, 2016).

**FIML.** As an alternative to MI, the model can also be estimated directly by applying FIML. However, because *Mplus* assumes that the variables in the multilevel model are multivariate normal, it was not

straightforward to include  $D$  as a multilevel categorical variable. Instead, we treated  $D$  as a multilevel continuous normal variable to estimate the model with FIML.

**Results.** Our main findings are summarized in Table 4. We restricted our reporting to the overall effects of  $X$  ( $\gamma_{10}$ ) and  $D$  ( $\gamma_{20}$ ) because we felt that they were the most important parameters for judging the performance of each method. Consistent with our expectations, both FCS-ML and JM provided approximately unbiased estimates of the two regression coefficients, whereas the other procedures each yielded biased estimates in some simulated conditions. The coverage was close to the nominal value of 95%, and the RMSE tended to be lowest under FCS-ML and JM. By contrast, LD yielded biased estimates of the two parameters. FCS-SL and FIML introduced bias in the regression coefficient of  $D$  ( $\gamma_{20}$ ) in conditions with large ICCs, although the RMSE and coverage remained acceptable under FIML. We concluded that both multilevel JM and FCS are suitable for MI of multilevel categorical data. Note, however, that we limited our attention to missing binary data. The FCS procedure can be extended to variables with multiple ordered or unordered categories.

### Study 3b: Group-Level Missing Data

The ideal case in which missing data occur only on the lowest level of multilevel data sets (i.e., on the level of individuals) rarely holds in practice. Moreover, data that are missing at the group level can be particularly cumbersome because they can force researchers to discard complete records at lower levels of the data. For example, consider a study in which employees were asked to rate the frequency of benevolent behavior engaged in by supervisors, and supervisors were asked the same question about their employees. If both variables were to be used as explanatory variables in some model of interest, missing data in supervisor ratings would lead one to discard employees' ratings as well, resulting in a severe loss of information. Surprisingly, the methodological literature has focused so far on ad hoc procedures, for example, separate imputation of individual- and group-level variables (Gibson & Olejnik, 2003) or "flat file" imputation using single-level MI (Cheung, 2007; for an overview, see Hox et al., 2016; van Buuren, 2011). However, recent advances in statistical software have greatly improved our ability to treat group-level missing data. Here, we briefly discuss two procedures—one using JM, one using FCS—that can be used to impute missing data at the group level.

Here, the model of interest was a multilevel random intercept model with explanatory variables at the individual (e.g., employee ratings) and the group level (e.g., supervisor ratings)

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + \gamma_{02}W_j + u_{0j} + e_{ij}. \quad (11)$$

Furthermore, we assumed that  $W$  was partially missing. The critical point in this model is that  $W$  is measured at the group level; that is, it does not vary across individuals in the same group. Thus, information located at the group level, including the information provided by individual-level variables, may be used to predict missing scores in  $W$ . For comparison, we also included LD and single-level FCS.

**Joint modeling (JM).** Computationally, the imputation of missing data at the group level is not much different from imputation at the individual level. Specifically, imputations for group-level missing data can be obtained by conditioning on observed group-level variables and on the between-group components of individual-level variables by employing the same general paradigm that is already employed for multilevel MI (for details, see Carpenter & Kenward, 2013; Goldstein et al., 2009). Similar to before, the joint model for the three variables of interest can be written as

$$[X_{ij}, Y_{ij}, W_j]^T = [\beta_{0(x)}, \beta_{0(y)}, \beta_{0(w)}]^T + [u_{j(x)}, u_{j(y)}, u_{j(w)}]^T + [e_{ij(x)}, e_{ij(y)}, 0]^T. \quad (12)$$

where  $u_{j(w)}$  is the residual of  $W$  in group  $j$ . For missing values in  $W$ , imputations are generated in the present case by conditioning on  $[u_{j(x)}, u_{j(y)}]$  at each iteration of the sampling algorithm (see Carpenter

**Table 4.** Bias (in %), RMSE, and Coverage of the 95% Confidence Interval for the Overall Regression Coefficients in Study 3a (Missing  $D \sim Y$ , MAR, 25%).

	LD			FCS-SL			FCS-ML			JM			FIML		
	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.
Missing $D \sim Y$ (MAR, 25%)															
$k = 100$															
$\hat{\gamma}_{10}$	-6.0	0.05	90.2	-0.1	0.04	95.5	-0.1	0.04	95.3	-0.0	0.04	94.8	-0.1	0.04	95.0
$\hat{\gamma}_{20}$	-6.1	0.09	95.4	-1.1	0.09	95.5	-0.3	0.09	95.0	-1.5	0.09	95.0	-0.7	0.09	95.0
$k = 500$															
$\hat{\gamma}_{10}$	-5.9	0.04	69.3	-0.1	0.02	95.3	-0.1	0.02	94.9	-0.1	0.02	95.1	-0.1	0.02	94.8
$\hat{\gamma}_{20}$	-5.9	0.04	94.5	-1.1	0.04	95.2	-0.3	0.04	95.8	-0.2	0.04	96.0	-0.7	0.04	95.5
$\rho_{1,X} = \rho_{1,Y} = .10$															
$k = 100$															
$\hat{\gamma}_{10}$	-4.2	0.05	93.2	0.8	0.04	95.3	0.0	0.04	95.0	-0.1	0.04	95.0	0.4	0.04	95.1
$\hat{\gamma}_{20}$	-2.6	0.09	93.9	-17.2	0.08	93.1	-2.2	0.09	94.5	-0.4	0.09	94.4	-8.1	0.08	92.7
$k = 500$															
$\hat{\gamma}_{10}$	-3.9	0.03	82.6	1.0	0.02	92.8	0.2	0.02	93.9	0.0	0.02	93.9	0.6	0.02	93.0
$\hat{\gamma}_{20}$	-2.7	0.04	95.7	-17.4	0.05	84.4	-2.3	0.04	95.6	0.6	0.04	94.9	-8.6	0.04	91.2

Note:  $\hat{\gamma}_{10}$  = overall regression coefficient of  $X$ ;  $\hat{\gamma}_{20}$  = overall regression coefficient of  $D$ ; LD = listwise deletion; FCS-SL = single-level FCS; FCS-ML = multilevel FCS; JM = multilevel JM; FIML = full-information maximum likelihood.

& Kenward, 2013), thus incorporating the group-level information supplied by  $X$  and  $Y$  in the prediction of missing  $W$ . The joint model can be implemented, for example, with the `jomo` package in R or in the statistical software *Mplus*.

**The FCS approach.** Instead of conditioning on the random effects of individual-level variables as in the JM approach, group-level variables can be imputed by applying an FCS approach based on the observed group means of these variables. Specifically, for missing values in  $W$ , missing data may be imputed by using the linear regression

$$W_j = \beta_{0(w)} + \beta_{1(w)}\bar{X}_{\bullet j} + \beta_{2(w)}\bar{Y}_{\bullet j} + u_{j(w)}, \quad (13)$$

where  $u_{j(w)}$  is the residual of  $W$  given  $\bar{X}_{\bullet j}$  and  $\bar{Y}_{\bullet j}$ . If values are missing at both levels, then the FCS algorithm iterates back and forth between the individual- and group-level equations (Equations 4 and 13; see also Gelman & Hill, 2006; Yucel, 2008). As in the multilevel random intercept model, it can be argued that the FCS and the JM approach imply similar covariance structures that can be used interchangeably (see Study 1; Carpenter & Kenward, 2013).

**FIML.** Missing values in  $W$  can also be addressed using FIML. Because  $W$  is directly measured at the group-level, missing data in  $W$  can be addressed simply by specifying  $W$  as a latent variable in *Mplus*.

**Results.** Our main findings are summarized in Table 5. We restricted our reporting to the group-level effects of  $X$  ( $\gamma_{01}$ ) and  $W$  ( $\gamma_{02}$ ) because we considered them to be the most important in this situation. FCS-ML, JM, and FIML all provided approximately unbiased estimates of the group-level effects in the model of interest. With a smaller number of groups and individuals within each group, FCS-ML and JM exhibited a small negative bias, which tended toward zero in larger samples. By contrast, LD and FCS-SL yielded only biased estimates of these parameters regardless of sample size. The coverage of the 95% confidence interval was close to the nominal value of 95% for FCS-ML, JM, and FIML, and the RMSE was lowest for these procedures. We concluded that multilevel JM and FCS as well as FIML are suitable methods for dealing with group-level missing data.

## Recommendations for Practice

There exist several approaches to the treatment of missing data in multilevel designs. As a result, researchers are faced with a multitude of options, several but not all of which may be suitable for a given task. In order to guide researchers in picking a suitable procedure, we provide a detailed list of recommendations in Table 6. This table covers different applications of multilevel models, including applications with random intercepts, random slopes, different variable types, and interaction effects. For each application, we distinguish between a general case with arbitrary patterns of missing data and a number of cases with missing data on specific variables (e.g., categorical and group-level variables). For each case, we list the recommended and not-recommended procedures as well as the likely consequences of choosing the latter.

Finally, we list statistical software that implements one or more of the recommended procedures, and we provide reference to one of the step-by-step examples in Appendix B, which illustrate the use of multilevel MI for each application in the statistical software R (for a general introduction to multilevel MI, see Enders et al., 2016; Grund et al., 2016b).

For applications in the multilevel random intercept model, multilevel MI—using either JM or FCS—provides an effective and general method for dealing with missing data. Procedures for multilevel JM, for example, are implemented in the software packages `pan` and `jomo` for the statistical software R as well as *Mplus*, MLwiN (Rasbash, Charlton, Browne, Healy, & Cameron,

**Table 5.** Bias (in %), RMSE, and Coverage of the 95% Confidence Interval for the Group-Level Regression Coefficients in Study 3b (Missing  $W \sim Y$ , MAR, 25%).

	LD			FCS-SL			FCS-ML			JM			FIML		
	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.	Bias	RMSE	Covg.
Missing $W \sim Y$ (MAR, 25%)															
$k = 100$															
$\hat{\gamma}_{01}$	-0.8	0.10	94.7	6.3	0.09	94.4	1.3	0.09	94.4	2.9	0.09	95.0	0.7	0.09	94.2
$\hat{\gamma}_{02}$	-3.3	0.05	95.3	-0.6	0.06	94.9	-1.4	0.05	95.0	-5.9	0.05	95.9	-0.3	0.05	95.5
$k = 500$															
$\hat{\gamma}_{01}$	-3.0	0.05	95.4	5.8	0.04	94.1	-0.0	0.04	94.9	0.4	0.04	95.3	-0.1	0.04	94.8
$\hat{\gamma}_{02}$	-3.2	0.03	92.7	-0.8	0.03	94.1	-0.5	0.03	93.5	-2.8	0.03	94.1	-0.3	0.03	93.2
$\rho_{t,x} = \rho_{t,y} = .10$															
$k = 100$															
$\hat{\gamma}_{01}$	-7.1	0.11	95.1	-0.8	0.10	95.3	-0.8	0.10	95.2	-0.7	0.10	95.6	-1.1	0.10	95.1
$\hat{\gamma}_{02}$	-7.9	0.08	95.3	6.8	0.09	92.3	-2.8	0.09	95.0	-2.6	0.09	95.0	-1.1	0.09	94.2
$k = 500$															
$\hat{\gamma}_{01}$	-6.7	0.05	93.6	0.0	0.04	94.8	-0.1	0.04	95.3	-0.1	0.05	94.7	-0.2	0.04	94.9
$\hat{\gamma}_{02}$	-6.9	0.04	93.7	8.3	0.05	89.8	-0.5	0.04	94.6	-0.4	0.04	94.9	-0.1	0.04	94.5

Note:  $\hat{\gamma}_{01}$  = between-group regression coefficient of X;  $\hat{\gamma}_{02}$  = between-group regression coefficient of W; LD = listwise deletion; FCS-SL = single-level FCS; FCS-ML = multilevel FCS; JM = multilevel JM; FIML = full-information maximum likelihood.

2015), REALCOM (Carpenter, Goldstein, & Kenward, 2011), and the SAS macro `MMI_IMPUTE` (Mistler, 2013); multilevel FCS is implemented in the R package `mice` as well as `Mplus` and `Blimp` (Keller & Enders, 2016). When treating missing data in categorical or group-level variables, researchers should choose implementations of multilevel MI that support these types of variables (e.g., `jomo`, `Mplus` and `REALCOM` for multilevel JM; `mice`, `Mplus`, and `Blimp` for multilevel FCS). FIML may be an option if missing data are restricted to the dependent variable in the analysis or if the analysis model includes latent instead of observed (i.e., manifest) group means to estimate group-level effects (Grund et al., in press; Lüdtke et al., 2008). By contrast, single-level MI should be avoided unless only a few cases contain missing data (e.g., less than 5%) and the ICC of the variables is relatively small (e.g., less than .10). Similarly, although LD provided reasonable estimates of model parameters (e.g., the CLI), we do not recommend that it be adopted in practice. This is because LD provides generally unbiased results only under MCAR, whereas its performance under MAR depends on the “strength” of missing data mechanism (i.e., the degree to which the data loss is systematic; see also Newman, 2014). This is problematic, because the missing data mechanism can never be ascertained from the data alone (e.g., Allison, 2001; Enders, 2010). For that reason, LD may provide an alternative if missing data are guaranteed to be MCAR, for example, in “planned missing data designs” (e.g., Graham, Taylor, Olchowski, & Cumsille, 2006). However, under more general conditions, we recommend against using LD. For applications involving random slopes or interaction effects, it is more difficult to provide general recommendations at the present time. Software for multilevel FCS may be used to treat missing data in such models if it supports the specification of random slope imputation models as well as passive imputation steps for the product terms (e.g., `mice`). However, researchers should bear in mind that multilevel FCS with passive imputation is not a definite solution to the problem of missing data in such applications. Instead, model-based procedures may be considered in the future (for a brief exposition, see the Discussion section).

Apart from the procedure selected for the treatment of missing data, the performance of MI also depends on a few general factors. For example, researchers should try to include auxiliary variables in the imputation model, that is, variables that are related to either the occurrence of missing data or the variables with missing data themselves (e.g., Collins et al., 2001; Graham, 2009; Schafer & Graham, 2002). When more information can be included from auxiliary variables, then missing values can be inferred from the observed data with greater accuracy (for a discussion about the use of auxiliary variables under FIML, see Enders, 2008; Graham, 2003). In addition, the quality of estimates and inferences obtained from MI can often be improved by generating a larger number of imputations (Bodner, 2008; Graham, Olchowski, & Gilreath, 2007). In our experience, generating 20 imputations is sufficient for most applications in which the primary goal is to estimate the model parameters, but as many as 100 or more imputations can be useful if the analyses involve testing more elaborate statistical hypotheses (Bodner, 2008; see also Grund et al., 2016b).

## Discussion

In the present article, we outlined several procedures for MI of multilevel missing data, each intended to accommodate typical research questions in organizational psychology and other areas in the social sciences. Through several smaller simulation studies, we tried to provide a broad overview of multilevel MI. We demonstrated that the current implementations of multilevel MI are able to accommodate quite general research questions and multilevel designs. For example, several procedures for multilevel MI, using either the JM or the FCS approach, were suitable in the broad context of random intercept models. In such a context, missing data can be treated fairly accurately and in a very general manner even when missing data occur at different levels of the sample or in categorical and continuous variables simultaneously.

**Table 6.** Recommended Missing Data Treatments and Software for Different Types of Multilevel Analysis Models Recommended Missing Data Treatments and Software for Different Types of Multilevel Analysis Models

Model type (Example)	Missing	Recommended	Not Recommended	Current Software (MI)
<b>Random intercept model</b> $Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + u_{0j} + e_{ij}$	Any	<ul style="list-style-type: none"> <li>• Multilevel FCS</li> <li>▷ Passive imputation of group means</li> <li>• Multilevel JM</li> <li>▷ All variables specified as targets</li> </ul>	<ul style="list-style-type: none"> <li>• Listwise deletion</li> <li>▷ Biased estimates, power loss</li> <li>• Single-level MI</li> <li>▷ Biased estimates and SEs</li> <li>• FIML<sup>a</sup></li> <li>▷ Biased estimates when using group-mean centering</li> </ul>	R (mice, pan, jomo), Mplus, Blimp, SAS (MML IMPUTE), MLwiN, REALCOM → see Example 1.1
... with categorical variables ( $D_{ij}$ ) $Y_{ij} = \gamma_{00} + \gamma_{10}D_{ij} + \gamma_{20}X_{ij} + u_{0j} + e_{ij}$	D	<ul style="list-style-type: none"> <li>• Multilevel FCS</li> <li>▷ Passive imputation of group means</li> <li>▷ Using logistic or probit models</li> <li>• Multilevel JM</li> <li>▷ All variables specified as targets</li> <li>▷ Using models for mixed data types</li> </ul>	<ul style="list-style-type: none"> <li>• Listwise deletion</li> <li>▷ Biased estimates, power loss</li> <li>• Single-level MI</li> <li>▷ Biased estimates and SEs</li> <li>• FIML<sup>a</sup></li> <li>▷ Biased estimates under normality assumption</li> </ul>	R (mice, jomo), Mplus, Blimp, REALCOM → see Example 1.2
... with variables at Level 2 ( $W_j$ ) $Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + \gamma_{02}W_j + u_{0j} + e_{ij}$	W	<ul style="list-style-type: none"> <li>• Multilevel FCS</li> <li>▷ Including group means</li> <li>• Multilevel JM</li> <li>▷ All variables specified as targets</li> <li>▷ Using models for missing data at both levels</li> <li>• FIML</li> </ul>	<ul style="list-style-type: none"> <li>• Listwise deletion</li> <li>▷ Biased estimates, power loss</li> <li>• Single-level MI</li> <li>▷ Biased estimates and SEs</li> </ul>	R (mice, jomo), Mplus, Blimp, REALCOM → see Example 1.3
... with interactions or nonlinear terms $Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{20}Z_{ij} + \gamma_{30}X_{ij}Z_{ij} + u_{0j} + e_{ij}$	X, Z	<ul style="list-style-type: none"> <li>• Multilevel FCS</li> <li>▷ Passive imputation of group means and product terms</li> </ul>	<ul style="list-style-type: none"> <li>• Listwise deletion</li> <li>▷ Biased estimates, power loss</li> <li>• Single-level MI</li> <li>▷ Biased estimates and SEs</li> <li>• Multilevel JM</li> <li>▷ Biased estimates of interaction effects</li> <li>• FIML<sup>a</sup></li> </ul>	R (mice) → see Example 1.4
<b>Random slope model</b> $Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + u_{0j} + u_{1j}(X_{ij} - \bar{X}_{\bullet j}) + e_{ij}$	Any	<ul style="list-style-type: none"> <li>• Multilevel FCS</li> <li>▷ Passive imputation of group means</li> <li>▷ Including random slopes between pairs of variables</li> </ul>	<ul style="list-style-type: none"> <li>• Listwise deletion</li> <li>▷ Biased estimates, power loss</li> <li>• Single-level MI</li> <li>▷ Biased estimates and SEs</li> </ul>	R (mice), Blimp → see Example 2.1

(continued)



**Table 6.** (continued)

Model type (Example)	Missing	Recommended	Not Recommended	Current Software (MI)
<p>... with interactions or nonlinear terms</p> $Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + \gamma_{02}W_j + \gamma_{11}W_j(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{03}W_j\bar{X}_{\bullet j} + u_{0j} + u_{1j}(X_{ij} - \bar{X}_{\bullet j}) + e_{ij}$	X, W	<ul style="list-style-type: none"> <li>• Multilevel FCS</li> <li>▷ Passive imputation of group means and product terms</li> <li>▷ Including random slopes between pairs of variables</li> </ul>	<ul style="list-style-type: none"> <li>• Multilevel JM</li> <li>▷ Biased SEs</li> <li>• FIML<sup>a</sup></li> <li>• Listwise deletion</li> <li>▷ Biased estimates, power loss</li> <li>• Single-level MI</li> <li>▷ Biased estimates and SEs</li> <li>• Multilevel JM</li> <li>▷ Biased estimates of interaction effects and SEs</li> <li>• FIML<sup>a</sup></li> </ul>	R (mice) → see Example 2.2

<sup>a</sup>The present recommendations refer to FIML as it is currently implemented in the statistical software Mplus.

However, we also pointed out applications in which the current implementations of multilevel MI do not correctly accommodate the model of interest. Specifically, it is still challenging to implement multilevel MI for multilevel models with random slopes or interaction effects when the explanatory variables contain missing data (see also Kim et al., 2015). Even though multilevel FCS appears to be slightly more flexible than multilevel JM in accommodating the substantive model, both approaches ultimately contain limitations due to the ways in which they are currently implemented in statistical software. To alleviate this problem, it has been recommended that the substantive analysis model be taken into account when conducting MI, thus ensuring that imputations are generated in a manner consistent with the model of interest (Bartlett, Seaman, White, & Carpenter, 2015). With this procedure, Bartlett et al. (2015) demonstrated that the bias associated with nonlinear and interaction effects in single-level regression models can be greatly reduced (see also Goldstein, Carpenter, & Browne, 2014). Unfortunately, this approach is currently not available in standard software for multilevel MI.

As an alternative to MI, multilevel models can be estimated directly from the incomplete data by applying model-based procedures such as FIML (e.g., in *Mplus*). Even though current implementations of FIML are still quite general and easy to use, it can be challenging to estimate multilevel models with missing values in explanatory variables, for example, when the model of interest uses observed group means to incorporate group-level effects or it includes categorical variables, random slopes, or interaction effects (see also Shin & Raudenbush, 2010). The challenges of FIML are ultimately similar to those of MI, and similar proposals have been made with respect to how one might overcome these challenges. For example, Stubbendick & Ibrahim (2003) proposed a factorization approach to FIML estimation of multilevel models with missing data in explanatory variables (see also Wu, 2010). Unfortunately, this approach is also currently not available in standard software. As an alternative, the model-based treatment of missing data can be implemented in a Bayesian analysis approach (Erler et al., 2016; see also Goldstein et al., 2014; Zhang & Wang, 2016). However, the Bayesian approach requires specialized software for Bayesian analyses such as WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) or JAGS (Plummer, 2016), and such software can be challenging to use in practice (e.g., syntax-based model specification, selection of priors and starting values). Our own experiences indicate that these procedures can provide unbiased estimates with good coverage properties even in multilevel models with random slopes and CLIs. For interested readers, we provide an example of a model-based procedure in the supplemental online materials. This example includes a multilevel model with random slopes and cross- and group-level interactions with missing data in explanatory variables (i.e., the conditions simulated in Study 2). The model syntax for the JAGS software is provided. However, before they can be widely adopted, we recommend that these procedures be subjected to further research and implemented in standard software. Additional software packages that implement FIML for multilevel models are *xxM* (Mehta, 2013) and *Latent GOLD* (Vermunt & Magidson, 2013).

Despite limitations in complex multilevel analyses, multilevel MI provides a more reliable and efficient approach to the treatment of missing data in comparison with simpler methods (e.g., single-level MI). As an alternative, it has been suggested that the multilevel structure might be expressed by including dummy-indicator variables in a single-level imputation model (Drechsler, 2015). Although this strategy substantially increases the complexity of the imputation model when the model of interest includes random slopes or interaction effects, it may be interesting to investigate its performance more thoroughly under such conditions (see also Andridge, 2011; Enders et al., 2016). In the context of multilevel models with random slopes, it has also been recommended that single-level MI be performed separately within each group (Graham, 2009). However, this strategy has been shown to be inefficient (i.e., low power) and should be avoided (Taljaard et al., 2008).

Every simulation study has its limitations, and owing to their smaller frame, the simulation studies presented here are no exception. In each study, we focused on varying the sample sizes

rather than creating a diverse pattern of possible effects and effect sizes. However, this came at the price of choosing constant values for many of the population parameters. Therefore, the results should not be generalized to arbitrary patterns of effects. On the other hand, it is nearly impossible to address the diversity of possible research designs in a single study. Future studies should investigate the performance of multilevel MI in more specialized applications, including settings with very small samples at the individual level (e.g., dyadic data) or the group level (e.g., research in large organizations), a larger variety of patterns of effects and missing data mechanisms (see Newman, 2009), low ICCs, or a large number of continuous and categorical variables (see Vermunt, 2003; Vermunt, van Ginkel, van der Ark, & Sijtsma, 2008). Further topics for future research also include the application of multilevel MI in longitudinal data, which share many but not all of the features of cross-sectional data, and in models with additional levels of hierarchy (see Yucel, 2008). In principle, however, these models can be addressed with existing statistical software.

Summing up, we believe that MI is already a powerful tool for treating missing data in multilevel research. Several procedures that make MI both generally applicable and easy to use have become available. In the present article, we attempted to provide guidance on the application of multilevel MI in research practice by providing both simulation results and recommendations for different applications of multilevel models. Our findings suggest natural directions for future research. For example, even though multilevel MI yielded reliable results in most applications, this was not the case in multilevel models with random slopes or interaction effects when data were missing in explanatory variables. Several procedures that might alleviate these problems have been proposed, but before these procedures can widely be adopted in practice, they must be evaluated more thoroughly in the context of multilevel designs, and they must be implemented in standard software. In this spirit, we hope that the present study and the materials provided with it will stimulate further research in this area and contribute to the regular use of MI in research practice.

## Appendix A

### Simulation design

In all simulation studies, the data were generated on the basis of the following model

$$Y_{ij} = \gamma_{10}(X_{ij} - \bar{X}_{\bullet j}) + \gamma_{01}\bar{X}_{\bullet j} + \gamma_{02}W_j + \gamma_{11}W_j(X_{ij} - \bar{X}_{\bullet j}) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_{\bullet j}) + e_{ij}, \quad (A1)$$

where  $u_{0j}$  and  $u_{1j}$  followed a multivariate normal distribution with mean zero and covariance matrix  $\mathbf{T} = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}$ , and  $e_{ij}$  was normally distributed with mean zero and variance  $\sigma^2$ . All variables were assumed to be standardized with mean zero and unit variance.

The data were generated in several steps. First, we simulated the between- and within-group components of  $X$ , which followed a normal distribution with mean zero and variances  $\tau_x^2$  and  $\sigma_x^2$ , respectively, which were determined by fixing the intraclass correlation (ICC) of  $X$  ( $\rho_{I,X}$ ). Then, we applied group-mean centering in order to obtain  $\bar{X}_{\bullet j}$  and  $(X_{ij} - \bar{X}_{\bullet j})$ . In the next step, we generated  $W_j$  from a linear regression on  $\bar{X}_{\bullet j}$  according to a correlation coefficient  $\rho_{\bar{x}w}$ . The regression coefficient was given by  $\rho_{\bar{x}w}/\sqrt{\tau_x^2 + \sigma_x^2/n}$  and the residual variance by  $(1 - \rho_{\bar{x}w}^2)$ . Finally,  $Y$  was simulated from Equation A1. The residual variance was set to

$$\sigma^2 = 1 - \gamma_{01}^2 \left( \tau_x^2 + \frac{\sigma_x^2}{n} \right) - \gamma_{02}^2 - 2\gamma_{01}\gamma_{02}\rho_{\bar{x}w} \sqrt{\tau_x^2 + \frac{\sigma_x^2}{n}} - (\gamma_{10}^2 + \gamma_{11}^2 + \tau_1^2) \left( \sigma_x^2 - \frac{\sigma_x^2}{n} \right) - \tau_0^2. \quad (A2)$$

The parameters of the simulations were chosen in such a way as to imply certain values for the ICC of  $Y$  (see also Aguinis & Culpepper, 2015, for a very general decomposition of variance). For this purpose, the ICC was calculated as follows

$$\rho_{I,Y} = \gamma_{01}^2 \left( \tau_x^2 + \frac{\sigma_x^2}{n} \right) + \gamma_{02}^2 + 2\gamma_{02} \left( \gamma_{01} - \gamma_{01} \frac{\sigma_x^2/n}{\tau_x^2 + \sigma_x^2/n} \right) \rho_{\bar{x}w} \sqrt{\tau_x^2 + \frac{\sigma_x^2}{n}} - (\gamma_{10}^2 + \gamma_{11}^2 + \tau_1^2) \frac{\sigma_x^2}{n} + \tau_0^2. \quad (\text{A3})$$

Missing data were simulated according to a missing data mechanism  $T \sim P$ , where  $T$  is a target and  $P$  is a predictor variable. For person  $i$  in group  $j$ , we simulated a latent response propensity

$$R_{ij} = \lambda_0 + \lambda_1 P_{ij} + r_{ij} \quad (\text{A4})$$

where  $R$  denotes  $T$ 's response propensity,  $\lambda_0$  is a quantile of the standard normal distribution according to a missing data probability, and  $\lambda_1$  is the weight with which  $P$  drives the missing data mechanism. The residuals  $r_{ij}$  were drawn from a normal distribution with mean zero and variance  $1 - \lambda_1^2$ . A value  $T_{ij}$  was deleted if  $R_{ij} > 0$ . To generate MCAR and MAR data, we set  $\lambda_1 = 0$  and  $.50$ , respectively. In Study 3b, the missing data mechanism was based on the group means.

In Study 3a, which included a dichotomous variable  $D$ , we generated the data as follows: First, we simulated within- and between-group components for a latent background variable  $D^*$  according to the ICC of  $D$  ( $\rho_{I,D}$ ). We set  $D_{ij}$  to one if  $D_{ij}^* > 0$  and zero otherwise. Then,  $X$  was simulated conditional on  $D$ , and  $Y$  was simulated conditional on  $D$  and  $X$ . Because the within- and between-group components of  $D$  could not be partitioned in the same way as for continuous variables, we generated  $X$  and  $Y$  by specifying residual ICCs for both variables ( $\rho_{I,X|D}$  and  $\rho_{I,Y|DX}$ ). These were chosen in such a way that they implied marginal ICCs for both variables that were similar to those in Studies 1, 2, and 3b.

## Appendix B

### Step-by-Step Examples for Multilevel MI

Here, we provide examples for multilevel MI in the statistical software R, using the packages `jomo` (JM), `mice` (FCS), and `miceadds` (FCS). We make use of the `justice` data set included in the `mitml` package, which contains simulated data from 1,400 employees organized in 200 organizations (`id`) including employees' sex (`sex`), their organizational satisfaction (`sat`), their orientation toward procedural justice (`jor`), and scores for justice climate at the level of organizations (`jcl`).

```
library(jomo)      # load "jomo" multilevel JM
library(mice)     # load "mice" for multilevel FCS
library(miceadds) # load "miceadds" for multilevel FCS

library(mitml)    # load "mitml"
data(justice, package="mitml") # load "justice" data set
```

The structure of the data set is as follows.

```
# id sex sat jor jcl
# 1 female 4.783 0.886 -0.336
# 1 male 5.392 0.205 -0.336
# 1 male 6.696 0.328 -0.336
```

All variables contain missing data. For each of the following examples, we focus on one model of interest. For the purpose of illustration, we use a subset of the data in each example in which missing data occur *only* in the variables included in the model of interest. In practice, these examples can (and should) be combined to treat missing data for all variables simultaneously.

### Example 1.1: Random Intercept Model

In this example, we demonstrate multilevel MI in a multilevel random intercept model with only continuous variables. Assume a researcher is interested in estimating the following model:

$$SAT_{ij} = \gamma_{00} + \gamma_{10}(JOR_{ij} - \overline{JOR}_{\bullet j}) + \gamma_{01}\overline{JOR}_{\bullet j} + u_{0j} + e_{ij}. \quad (B1)$$

Either the JM or the FCS approach can be used in this case. In order to set aside the treatment of missing data in categorical and group-level variables (see Examples 1.2 and 1.3), the data set is reduced in such a way that only the variables in the model of interest require imputation.

```
ex1.1 <- subset(justice, !is.na(sex) & !is.na(jcl))
```

**Option 1: Multilevel FCS.** In the FCS approach, it is necessary to specify the imputation model for each variable as well as the predictor variables for each model. Using mice, the imputation model is specified by defining an imputation method and predictor variables for each target variable with missing data. The imputation methods are specified in a character vector (`impMethod`). Here, we assign a two-level model to each variable with missing data.

```
# set up imputation methods
impMethod <- character(ncol(ex1.1)) # create empty vector for
names(impMethod) <- colnames(ex1.1) # imputation methods

# ... define method for each variable
impMethod[ "sat" ] <- "2l.pan" # use '2l.pan' for 'sat' (two-level normal)
impMethod[ "jor" ] <- "2l.pan" # use '2l.pan' for 'jor' (two-level normal)
```

The predictors in each model are then defined in the predictor matrix (`predMatrix`), where each line corresponds to one target variable, and integer values denote the relations between the target and predictor variables.

```
# set up predictor matrix
predMatrix <- matrix(0, ncol(ex1.1), ncol(ex1.1)) # create empty predictor
rownames(predMatrix) <- colnames(predMatrix) <- colnames(ex1.1) # matrix

# ... define predictors for each variable
predMatrix[ "sat" , c("id","sex","jor","jcl") ] <- c(-2,3,3,1) # -2 = cluster variable
predMatrix[ "jor" , c("id","sex","sat","jcl") ] <- c(-2,3,3,1) # 1 = overall effect
# 3 = overall + group-level effect
```

The imputation is then carried out by issuing the following command.

```
imp <- mice::mice(data=ex1.1, maxit=10, m=20, imputationMethod=impMethod,
  predictorMatrix=predMatrix)
```

**Option 2: Multilevel JM.** Because the model of interest is a multilevel random intercept model, all variables (including auxiliary variables) can be treated as target variables in the JM approach. In the `mitml` package, imputation models can be specified as a list of two model formulas, pertaining to individual- and group-level variables, respectively (for an alternative model specification similar to FCS, see the package documentation).

```
fml <- list( sat + jor + sex ~ 1 + (1|id), # level-1 targets = sat + jor + sex (no predictors)
  jcl ~ 1 ) # level-2 targets = jcl (no predictors)
```

The imputation is then carried out by issuing the following command.

```
imp <- mitml::jomoImpute(data=ex1.1, formula=fml, n.burn=5000, n.iter=500, m=20)
```

### Example 1.2: Random Intercept Model With Categorical Variables

In this example, we demonstrate multilevel MI with a mix of continuous and categorical variables. Assume a researcher is interested in the following model

$$SAT_{ij} = \gamma_{00} + \gamma_{10}JOR_{ij} + \gamma_{20}Sex_{ij} + u_{0j} + e_{ij}. \quad (B2)$$

Either the JM or the FCS approach can be used in this case. As before, the data set is reduced in such a way that only these variables require imputation.

```
ex1.2 <- subset(justice, !is.na(jcl))
```

**Option 1: Multilevel FCS.** The specification of the FCS approach is similar to the previous example but now includes a logistic two-level model for the categorical target variable (*sex*).

```
# set up imputation methods
impMethod <- character(ncol(ex1.2)) # create empty vector for
names(impMethod) <- colnames(ex1.2) # imputation methods

# ... define method for each variable
impMethod["sat"] <- "2l.pan" # use '2l.pan' for 'sat' (two-level normal)
impMethod["jor"] <- "2l.pan" # use '2l.pan' for 'jor' (two-level normal)
impMethod["sex"] <- "2l.binary" # use '2l.binary' for 'sex' (two-level logistic)

# set up predictor matrix
predMatrix <- matrix(0, ncol(ex1.2), ncol(ex1.2)) # create empty predictor
rownames(predMatrix) <- colnames(predMatrix) <- colnames(ex1.2) # matrix

# ... define predictors for each variable
predMatrix["sat", c("id", "sex", "jor", "jcl")] <- c(-2, 3, 3, 1) # -2 = cluster variable
predMatrix["jor", c("id", "sex", "sat", "jcl")] <- c(-2, 3, 3, 1) # 1 = overall effect
predMatrix["sex", c("id", "sat", "jor", "jcl")] <- c(-2, 3, 3, 1) # 3 = overall + group-level effect
```

The imputation is then run as follows. Note that categorical variables must be converted into a numeric data type beforehand in order to be imputed using *mice*.

```
# convert "sex" to dummy variable
ex1.2 <- within(ex1.2, sex <- as.integer(sex)-1)

imp <- mice::mice(data=ex1.2, maxit=10, m=20, imputationMethod=impMethod,
  predictorMatrix=predMatrix)
```

**Option 2: Multilevel JM.** The specification of the JM approach is identical to the previous example, treating all variables (including auxiliary variables) as target variables in the imputation model.

```
fml <- list(sat + jor + sex ~ 1 + (1|id), # level-1 targets = sat + jor + sex (no predictors)
  jcl ~ 1) # level-2 targets = jcl (no predictors)
```

The imputation is then carried out by issuing the following command. Note that, in contrast to *mice*, *jomo* requires that categorical variables are formatted as factor variables in R, as is the case in the original data set.

```
imp <- mitml::jomoImpute(data=ex1.2, formula=fml, n.burn=5000, n.iter=500, m=20)
```

### Example 1.3: Random Intercept Model With Group-Level Variables

In this example, we demonstrate multilevel MI with explanatory variables at both the individual and the group level. Assume a researcher is interested in the following model

$$SAT_{ij} = \gamma_{00} + \gamma_{10}(JOR_{ij} - \overline{JOR}_{\bullet j}) + \gamma_{01}\overline{JOR}_{\bullet j} + \gamma_{02}JCL_j + u_{0j} + e_{ij}. \quad (B3)$$

Either the JM or the FCS approach can be used in this case. As before, the data set is reduced in such a way that only these variables require imputation.

```
ex1.3 <- subset(justice, !is.na(sex))
```

**Option 1: Multilevel FCS.** The specification of the FCS approach is similar to the previous example but now includes a group-level regression model for the target variable at the group level (jcl).

```
# set up imputation methods
impMethod <- character(ncol(ex1.3)) # create empty vector for
names(impMethod) <- colnames(ex1.3) # imputation methods

# ... define method for each variable
impMethod[ "sat" ] <- "2l.pan"      # use '2l.pan' for 'sat' (two-level normal)
impMethod[ "jor" ] <- "2l.pan"      # use '2l.pan' for 'jor' (two-level normal)
impMethod[ "jcl" ] <- "2lonly.norm" # use '2lonly.norm' for 'jcl' (group-level regression)

# set up predictor matrix
predMatrix <- matrix(0, ncol(ex1.3), ncol(ex1.3)) # create empty predictor
rownames(predMatrix) <- colnames(predMatrix) <- colnames(ex1.3) # matrix

# ... define predictors for each variable
predMatrix[ "sat" , c("id","sex","jor","jcl") ] <- c(-2,3,3,1) # -2 = cluster variable
predMatrix[ "jor" , c("id","sex","sat","jcl") ] <- c(-2,3,3,1) # 1 = overall effect
predMatrix[ "jcl" , c("id","sex","sat","jor") ] <- c(-2,1,1,1) # 3 = overall + group-level effect
```

The imputation is then carried out by issuing the following command.

```
imp <- mice::mice(data=ex1.3, maxit=10, m=20, imputationMethod=impMethod,
  predictorMatrix=predMatrix)
```

**Option 2: Multilevel JM.** The specification of the JM approach is identical to the previous example, treating all variables (including auxiliary variables) as target variables in the imputation model.

```
fml <- list( sat + jor + sex ~ 1 + (1|id), # level-1 targets = sat + jor + sex (no predictors)
  jcl ~ 1 ) # level-2 targets = jcl (no predictors)
```

The imputation is then carried out by issuing the following command.

```
imp <- mitml::jomoImpute(data=ex1.3, formula=fml, n.burn=5000, n.iter=500, m=20)
```

### Example 1.4: Random Intercept Model With Interaction Effects

In this example, we demonstrate multilevel MI for a multilevel random intercept model with two explanatory variables and an interaction effect. Assume a researcher is interested in the following model:

$$SAT_{ij} = \gamma_{00} + \gamma_{10}(JOR_{ij} - \overline{JOR}_{\bullet j}) + \gamma_{20}Sex_{ij} + \gamma_{30}Sex_{ij}(JOR_{ij} - \overline{JOR}_{\bullet j}) + \gamma_{01}\overline{JOR}_{\bullet j} + u_{0j} + e_{ij}. \quad (B4)$$

In order to include the interaction effect in the imputation model, we make use of passive imputation steps in the FCS approach. As before, the data set is reduced in such a way that only these variables require imputation.

```
ex1.4 <- subset(justice, !is.na(jcl))
```

To define passive steps, new variables must be created for the components of organizational satisfaction (sat) and justice orientation (jor) and for the product terms that allow relations between

individual-level organizational satisfaction (*sat*) and justice orientation (*jor*) to vary by sex (*sex*) and vice versa.

```
ex1.4 <- within(ex1.4,{
  sex <- as.integer(sex)-1 # recode 'sex' as a dummy variable
  sat.GRP <- sat.IND <- NA # passive variables for components of 'sat'
  jor.GRP <- jor.IND <- NA # passive variables for components of 'jor'
  sat.IND_sex <- NA # passive variable for interaction of 'sat' and 'sex'
  jor.IND_sex <- NA # passive variable for interaction of 'jor' and 'sex'
  sat.IND_jor.IND <- NA # passive variable for interaction of 'sat' and 'jor'
})
```

The specification of the imputation methods proceeds in a similar manner as before. In addition, passive steps are defined for the components of organizational satisfaction (*sat*) and justice orientation (*jor*) as well as for the product terms by employing the  $\sim I()$  identity function.

```
# set up imputation methods
impMethod <- character(ncol(ex1.4)) # create empty vector for
names(impMethod) <- colnames(ex1.4) # imputation methods

# ... define method for each variable
impMethod[ "sat" ] <- "2l.pan" # use '2l.pan' for 'sat' (two-level normal)
impMethod[ "jor" ] <- "2l.pan" # use '2l.pan' for 'jor' (two-level normal)
impMethod[ "sex" ] <- "2l.binary" # use '2l.binary' for 'sex' (two-level logistic)

impMethod[ "sat.GRP" ] <- "2l.groupmean" # passive step for updating the group means
impMethod[ "jor.GRP" ] <- "2l.groupmean" # means and within-group deviations of 'sat'
impMethod[ "sat.IND" ] <- "~I(sat-sat.GRP)" # and 'jor'
impMethod[ "jor.IND" ] <- "~I(jor-jor.GRP)"

impMethod[ "sat.IND_sex" ] <- "~I(sat.IND*sex)" # passive steps for updating the
impMethod[ "jor.IND_sex" ] <- "~I(jor.IND*sex)" # product terms involving 'sat',
impMethod[ "sat.IND_jor.IND" ] <- "~I(sat.IND*jor.IND)" # 'jor', and 'sex'
```

The predictors for each model are specified as before. However, two additional entries that allow the group means to be updated are added.

```
# set up predictor matrix
predMatrix <- matrix(0, ncol(ex1.4), ncol(ex1.4)) # create empty predictor
rownames(predMatrix) <- colnames(predMatrix) <- colnames(ex1.4) # matrix

# ... define predictors for each variable
predMatrix[ "sat.GRP", c("id", "sat") ] <- c(-2,1) # -2 = cluster variable
predMatrix[ "jor.GRP", c("id", "jor") ] <- c(-2,1) # 1 = variable to be aggregated

predMatrix[ "sat", c("id", "sex", "jor", "jcl", "jor.IND_sex") ] <- c(-2,1,3,1,1)
predMatrix[ "jor", c("id", "sex", "sat", "jcl", "sat.IND_sex") ] <- c(-2,1,3,1,1)
predMatrix[ "sex", c("id", "sat", "jor", "jcl", "sat.IND_jor.IND") ] <- c(-2,3,3,1,1)
```

Finally, the imputation is carried out by issuing the following command. In addition, we specify a visit sequence (*visitSeq*), which defines the order in which the variables are to be imputed and the passive steps are to be carried out. Passive steps are updated after obtaining a new imputation to the extents to which they are needed in subsequent imputation steps.

```
# set up visit sequence
visitSeq <- c(
  "sat", "sat.GRP", "sat.IND", "sat.IND_sex", # impute 'sat', then update terms for 'jor'
  "jor", "jor.GRP", "jor.IND", "sat.IND_jor.IND", # impute 'jor', then update terms for 'sex'
  "sex", "jor.IND_sex" # impute 'sex', then update terms for 'sat'
)
visitSeq <- match(visitSeq, colnames(ex1.4))

imp <- mice::mice(data=ex1.4, maxit=10, m=20, imputationMethod=impMethod,
  predictorMatrix=predMatrix, allow.na=TRUE)
```



### Example 2.1: Random Slope Model

In this example, we demonstrate multilevel MI for a multilevel model with random slopes. Assume a researcher is interested in the following model:

$$SAT_{ij} = \gamma_{00} + \gamma_{10}(JOR_{ij} - \overline{JOR}_{\bullet j}) + \gamma_{01}\overline{JOR}_{\bullet j} + u_{0j} + u_{1j}(JOR_{ij} - \overline{JOR}_{\bullet j}) + e_{ij}. \quad (B5)$$

To accommodate the fact that the individual-level relation of organizational satisfaction (*sat*) and justice orientation (*jor*) varies across groups, the FCS approach will be used. As before, the data set is reduced in such a way that only the variables in the model of interest require imputation.

```
ex2.1 <- subset(justice, !is.na(sex) & !is.na(jcl))
```

The specification of the FCS approach is very similar to Example 1.1 but now includes a random slope in the imputation model for both organizational satisfaction (*sat*) and justice orientation (*jor*), which is denoted by the number 4 in the predictor matrix.

```
# set up imputation methods
impMethod <- character(ncol(ex2.1)) # create empty vector for
names(impMethod) <- colnames(ex2.1) # imputation methods

# ... define method for each variable
impMethod[ "sat" ] <- "2l.pan" # use '2l.pan' for 'sat' (two-level normal)
impMethod[ "jor" ] <- "2l.pan" # use '2l.pan' for 'jor' (two-level normal)

# set up predictor matrix
predMatrix <- matrix(0, ncol(ex2.1), ncol(ex2.1)) # create empty predictor
rownames(predMatrix) <- colnames(predMatrix) <- colnames(ex2.1) # matrix

# ... define predictors for each variable
predMatrix[ "sat" , c("id","sex","jor","jcl" ) ] <- c(-2,3,4,1) # 4 = individual-level (random)
predMatrix[ "jor" , c("id","sex","sat","jcl" ) ] <- c(-2,3,4,1) # and group-level (fixed) effect
# ... otherwise as above
```

The imputation is then carried out by issuing the following command.

```
imp <- mice::mice(data=ex2.1, maxit=10, m=20, imputationMethod=impMethod,
  predictorMatrix=predMatrix)
```

### Example 2.2: Random Slope Model With Interaction Effects

In this example, we demonstrate multilevel MI with two explanatory variables, one at the individual and one at the group level, including a random slope and interaction effects. Assume a researcher is interested in the following model

$$SAT_{ij} = \gamma_{00} + \gamma_{10}(JOR_{ij} - \overline{JOR}_{\bullet j}) + \gamma_{01}\overline{JOR}_{\bullet j} + \gamma_{02}JCL_j + \gamma_{11}(JOR_{ij} - \overline{JOR}_{\bullet j})JCL_j + \gamma_{03}\overline{JOR}_{\bullet j}JCL_j + u_{0j} + u_{1j}(JOR_{ij} - \overline{JOR}_{\bullet j}) + e_{ij}. \quad (B6)$$

In order to include the random slopes and the interaction effects in the imputation model, we make use of passive imputation steps in the FCS approach. As before, the data set is reduced in such a way that only these variables require imputation.

```
ex2.2 <- subset(justice, !is.na(sex))
```

To define passive steps, new variables must be created for the components of organizational satisfaction (*sat*) and justice orientation (*jor*) as well as for the product terms at the individual and the group level.

```

ex2.2 <- within(ex2.2,{
  sat.GRP <- sat.IND <- NA # passive variables for components of 'sat'
  jor.GRP <- jor.IND <- NA # ... for components of 'jor'
  sat.IND_jcl <- NA # ... for individual-level interaction of 'sat' and 'jcl'
  sat.GRP_jcl <- NA # ... for group-level interaction of 'sat' and 'jcl'
  jor.IND_jcl <- NA # ... for individual-level interaction of 'jor' and 'jcl'
  jor.GRP_jcl <- NA # ... for group-level interaction of 'jor' and 'jcl'
  sat.GRP_jor.GRP <- NA # ... for group-level interaction of 'sat' and 'jor'
})

```

The specification of the imputation methods proceeds as before. Passive steps are defined for the components of organizational satisfaction (*sat*) and justice orientation (*jor*) as well as for the product terms by employing  $\sim I()$ .

```

# set up imputation methods
impMethod <- character(ncol(ex2.2)) # create empty vector for
names(impMethod) <- colnames(ex2.2) # imputation methods

# ... define method for each variable
impMethod[ "sat" ] <- "2l.pan" # use '2l.pan' for 'sat' (two-level normal)
impMethod[ "jor" ] <- "2l.pan" # use '2l.pan' for 'jor' (two-level normal)
impMethod[ "jcl" ] <- "2lonly.norm" # use '2lonly.norm' for 'jcl' (group-level regression)

impMethod[ "sat.GRP" ] <- "2l.groupmean" # passive step for updating the group means
impMethod[ "jor.GRP" ] <- "2l.groupmean" # means and within-group deviations of 'sat'
impMethod[ "sat.IND" ] <- "~I(sat-sat.GRP)" # and 'jor'
impMethod[ "jor.IND" ] <- "~I(jor-jor.GRP)"

impMethod[ "sat.IND_jcl" ] <- "~I(sat.IND*jcl)" # passive steps for updating the
impMethod[ "jor.IND_jcl" ] <- "~I(jor.IND*jcl)" # individual-level product terms

impMethod[ "sat.GRP_jcl" ] <- "~I(sat.GRP*jcl)" # passive steps for updating
impMethod[ "jor.GRP_jcl" ] <- "~I(jor.GRP*jcl)" # the group-level product terms
impMethod[ "sat.GRP_jor.GRP" ] <- "~I(sat.GRP*jor.GRP)"

```

The predictors in each model are specified as before.

```

# set up predictor matrix
predMatrix <- matrix(0, ncol(ex2.2), ncol(ex2.2)) # create empty predictor
rownames(predMatrix) <- colnames(predMatrix) <- colnames(ex2.2) # matrix

# ... define predictors for each variable
predMatrix[ "sat.GRP", c("id","sat") ] <- c(-2,1) # -2 = cluster variable
predMatrix[ "jor.GRP", c("id","jor") ] <- c(-2,1) # 1 = variable to be aggregated
predMatrix[ "sat", c("id","sex","jor","jcl","jor.IND_jcl","jor.GRP_jcl") ] <- c(-2,3,3,1,1,1)
predMatrix[ "jor", c("id","sex","sat","jcl","sat.IND_jcl","sat.GRP_jcl") ] <- c(-2,3,3,1,1,1)
predMatrix[ "jcl", c("id","sex","sat","jor","sat.GRP_jor.GRP") ] <- c(-2,1,1,1,1)

```

The imputation is carried out by issuing the following command. The visit sequence (*visitSeq*) is specified in such a way that passive steps are updated after obtaining a new imputation to the extents to which they are needed in subsequent imputation steps.

```

# set up visit sequence
visitSeq <- c(
  "sat","sat.GRP","sat.IND","sat.IND_jcl","sat.GRP_jcl", # impute 'sat', then update terms for 'jor'
  "jor","jor.GRP","jor.IND","sat.GRP_jor.GRP", # impute 'jor', then update terms for 'jcl'
  "jcl","jor.IND_jcl","jor.GRP_jcl" # impute 'jcl', then update terms for 'sat'
)
visitSeq <- match(visitSeq, colnames(ex2.2))

imp <- mice::mice(data=ex2.2, maxit=10, m=20, imputationMethod=impMethod,
  predictorMatrix=predMatrix, allow.na=TRUE)

```

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Supplemental Material

The online materials are available at <http://journals.sagepub.com/doi/suppl/10.1177/1094428117703686>.

## Notes

1. It is possible to include fully observed variables on the right-hand side of the JM (i.e., in  $X_{ij}$  and  $Z_{ij}$ ). This strategy is conceptually close to the FCS approach described in the next section. Hence, for the JM, we restricted our attention to the multivariate model in which all variables are treated as target variables (see also Enders, Mistler, & Keller, 2016).
2. It is possible to ignore the group means entirely in the FCS approach, for example, by imputing missing data in  $X$  by assuming only an overall effect of  $Y$ . However, this strategy “conflates” the individual- and group-level effects and may introduce bias into the parameter estimates. Here, we consider only the more general model that includes the group means in the set of predictor variables (see also Enders et al., 2016).
3. Under FCS, fitting the imputation models requires that some instances of the product terms are observed. However, with missing data at the individual level, the group means are no longer known, rendering the product terms unobserved as a result. In other words, it is not clear how “just another variable” might differentiate interaction effects at the individual level, the group level, and across levels without knowing the group means.
4. The bias for the slope variance in small samples may be explained by the Bayesian prior distribution employed during MI. In our study, we used the standard “least-informative” inverse-Wishart prior distributions, which imply relatively large values for the variance components as compared with the true size of the slope variance (here,  $Y_2 = .10$  if  $\gamma_{11} = 0$ , and  $\tau_1^2 = .06$  if  $\gamma_{11} = .20$ ). Such problems are well known in the Bayesian literature and can often be mitigated by choosing the prior distribution on the basis of the data (McNeish, 2016).
5. When using the imputation module in *Mplus* to implement multilevel MI, categorical variables are treated as ordinal, and a single latent variable is used for each categorical variable regardless of the number of categories; variables with multiple categories are addressed by estimating  $c - 1$  threshold parameters (Asparouhov & Muthén, 2010).

## References

- Aguinis, H., & Culpepper, S. A. (2015). An expanded decision-making procedure for examining cross-level interaction effects with multilevel modeling. *Organizational Research Methods, 18*(2), 155-176. doi:10.1177/1094428114563618
- Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.
- Allison, P. D. (2012). Handling missing data by maximum likelihood. In *Proceedings of the SAS Global Forum*. Retrieved from <http://support.sas.com/>
- 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
- Asparouhov, T., & Muthén, B. O. (2010). *Multiple imputation with Mplus* (Technical Appendix). Retrieved from <http://statmodel.com/>
- Bartlett, J. W., Seaman, S. R., White, I. R., & Carpenter, J. R. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research, 24*, 462-487. doi:10.1177/0962280214521348.

- Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 651-675. doi:10.1080/10705510802339072
- Carpenter, J. R., Goldstein, H., & Kenward, M. G. (2011). REALCOM-IMPUTE software for multilevel multiple imputation with mixed response types. *Journal of Statistical Software*, *45*(5), 1-14. doi:10.18637/jss.v045.i05
- Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. Hoboken, NJ: Wiley.
- Cheung, M. W.-L. (2007). Comparison of methods of handling missing time-invariant covariates in latent growth models under the assumption of missing completely at random. *Organizational Research Methods*, *10*, 609-634. doi:10.1177/1094428106295499
- 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
- Drechsler, J. (2015). Multiple imputation of multilevel missing data—Rigor versus simplicity. *Journal of Educational and Behavioral Statistics*, *40*, 69-95. doi:10.3102/1076998614563393
- Enders, C. K. (2008). A note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 434-448. doi:10.1080/10705510802154307
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.
- Enders, C. K., Mistler, S. A., & Keller, B. T. (2016). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, *21*, 222-240. doi:10.1037/met0000063
- Erler, N. S., Rizopoulos, D., van Rosmalen, J., Jaddoe, V. W. V., Franco, O. H., & Lesaffre, E. M. E. H. (2016). Dealing with missing covariates in epidemiologic studies: A comparison between multiple imputation and a full Bayesian approach. *Statistics in Medicine*, *35*, 2955-2974. doi:10.1002/sim.6944
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Gibson, N. M., & Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Educational and Psychological Measurement*, *63*, 204-238. doi:10.1177/0013164402250987
- 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
- Goldstein, H., Carpenter, J. R., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling*, *9*, 173-197. doi:10.1177/1471082X0800900301
- Gottfredson, N. C., Sterba, S. K., & Jackson, K. M. (2016). Explicating the conditions under which multilevel multiple imputation mitigates bias resulting from random coefficient-dependent missing longitudinal data. *Prevention Science*. Advance online publication. doi:10.1007/s11121-016-0735-3
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 80-100. doi:10.1207/S15328007SEM1001\_4
- 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., 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
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323-343. doi:10.1037/1082-989X.11.4.323
- Grund, S., Lüdtke, O., & Robitzsch, A. (2016a). Multiple imputation of missing covariate values in multilevel models with random slopes: A cautionary note. *Behavior Research Methods*, *48*, 640-649. doi:10.3758/s13428-015-0590-3

- Grund, S., Lüdtke, O., & Robitzsch, A. (2016b). Multiple imputation of multilevel missing data: An introduction to the R package *pan*. *SAGE Open*, *6*(4), 1-17. doi:10.1177/2158244016668220
- Grund, S., Lüdtke, O., & Robitzsch, A. (in press). Missing data in multilevel research. In S. E. Humphrey & J. M. LeBreton (Eds.), *Handbook for multilevel theory, measurement, and analysis*. Washington, DC: American Psychological Association.
- Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*, 623-641. doi:10.1177/014920639802400504
- Hox, J. J., van Buuren, S., & Jolani, S. (2016). Incomplete multilevel data. In J. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 39-62). Charlotte, NC: Information Age.
- Keller, B. T., & Enders, C. K. (2016). *Blimp Software Manual (Version Beta 6.6)* [Computer software]. Retrieved from <http://www.appliedmissingdata.com>
- Kim, S., Sugar, C. A., & Belin, T. R. (2015). Evaluating model-based imputation methods for missing covariates in regression models with interactions. *Statistics in Medicine*, *34*, 1876-1888. doi:10.1002/sim.6435
- Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1-21. doi:10.1207/s15327906mbr3001\_1
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. O. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203-229. doi:10.1037/a0012869
- Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, *22*, 141-165. doi:10.1037/met0000096
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325-337. doi:10.1023/A:1008929526011
- McNeish, D. M. (2016). Using data-dependent priors to mitigate small sample bias in latent growth models: A discussion and illustration using *Mplus*. *Journal of Educational and Behavioral Statistics*, *41*, 27-56. doi:10.3102/1076998615621299
- Mehta, P. D. (2013). *xxM (Version 0.6.0)* [Computer software]. Retrieved from [xxm.times.uh.edu](http://xxm.times.uh.edu)
- Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, *9*, 538-558. doi:10.1214/ss/1177010269
- Mistler, S. A. (2013). A SAS macro for applying multiple imputation to multilevel data. In *Proceedings of the SAS Global Forum*. Retrieved from <http://support.sas.com/>
- Mistler, S. A. (2015). *Multilevel multiple imputation: An examination of competing methods* (Doctoral dissertation). Retrieved from <http://repository.asu.edu/>
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Newman, D. A. (2009). Missing data techniques and low response rates. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 7-36). New York, NY: Routledge.
- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, *17*, 372-411. doi:10.1177/1094428114548590
- Plummer, M. (2016). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling (Version 4.2.0)* [Computer software]. Retrieved from <http://sourceforge.net/projects/mcmc-jags/>
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209-233. doi:10.1037/a0020141
- Quartagno, M., & Carpenter, J. R. (2016). *Jomo: A package for multilevel joint modelling multiple imputation (Version 2.3-1)* [Computer software]. Retrieved from <http://CRAN.R-project.org/package=jomo>
- R Core Team. (2016). *R: A language and environment for statistical computing (Version 3.3.0)* [Computer software]. Retrieved from <http://www.R-project.org/>

- Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2012). Multilevel structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 512-531). New York, NY: Guilford.
- Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2015). MLwiN (Version 2.34) [Computer software]. Bristol, UK: University of Bristol, Centre for Multilevel Modelling.
- Resche-Rigon, M., & White, I. R. (in press). Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Statistical Methods in Medical Research*. doi:10.1177/0962280216666564
- Robitzsch, A., Grund, S., & Henke, T. (2016). Miceadds: Some additional multiple imputation functions, especially for mice (Version 1.7-8) [Computer software]. Retrieved from <http://CRAN.R-project.org/package=miceadds>
- Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, 4, 227-241.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592. doi:10.1093/biomet/63.3.581
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.
- 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., & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*, 11, 437-457. doi:10.1198/106186002760180608
- Seaman, S. R., Bartlett, J. W., & White, I. R. (2012). Multiple imputation of missing covariates with non-linear effects and interactions: An evaluation of statistical methods. *BMC Medical Research Methodology*, 12(1), 46. Retrieved from <http://www.biomedcentral.com/1471-2288/12/46>
- 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). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Stubbendick, A. L., & Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics*, 59, 1140-1150. doi:10.1111/j.0006-341X.2003.00131.x
- 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
- 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. doi:10.18637/jss.v045.i03
- Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213-239. doi:10.1111/j.0081-1750.2003.t01-1-00131.x
- Vermunt, J. K., & Magidson, J. (2013). Latent GOLD (Version 5.0) [Computer software]. Belmont, MA: Statistical Innovations.
- Vermunt, J. K., van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociological Methodology*, 38, 369-397. doi:10.1111/j.1467-9531.2008.00202.x
- Vink, G., & van Buuren, S. (2013). Multiple imputation of squared terms. *Sociological Methods & Research*, 42, 598-607. doi:10.1177/0049124113502943
- 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

- Wu, L. (2010). *Mixed effects models for complex data*. Boca Raton, FL: CRC Press.
- Yucel, R. M. (2008). Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366, 2389-2403. doi:10.1098/rsta.2008.0038
- Zhang, Q., & Wang, L. (2016). Moderation analysis with missing data in the predictors. *Psychological Methods*. Advance online publication. doi:10.1037/met0000104
- Zinn, S. (2013). *An imputation model for multilevel binary data* (NEPS Working Paper No. 31). Retrieved from <http://www.neps-data.de/>

## Author Biographies

**Simon Grund** graduated in Psychology at the Humboldt University in Berlin and joined the Leibniz Institute for Science and Mathematics Education as a PhD candidate. His research interests include missing data, causality, and multilevel modeling.

**Oliver Lüdtke** is a professor of Educational Measurement at the Leibniz Institute for Science and Mathematics Education and the University of Kiel. His main research interests include the application of multilevel modeling in psychological and educational research, international student achievement studies, and personality development in adolescence.

**Alexander Robitzsch** graduated in Mathematics at the University of Technology in Dresden and earned a PhD in Educational Science at the Humboldt-University in Berlin. His research interests include item response modelling, missing data, and multilevel modeling.