Article

Journal of Educational and Behavioral Statistics
2017, Vol. 42, No. 4, pp. 432–466
DOI: 10.3102/1076998617690869

A Comparison of Joint Model and Fully Conditional Specification Imputation for Multilevel Missing Data

Stephen A. Mistler
SAS Institute

Craig K. Enders
University of California, Los Angeles

Multiple imputation methods can generally be divided into two broad frameworks: joint model (JM) imputation and fully conditional specification (FCS) imputation. JM draws missing values simultaneously for all incomplete variables using a multivariate distribution, whereas FCS imputes variables one at a time from a series of univariate conditional distributions. In single-level multivariate normal data, these two approaches have been shown to be equivalent, but less is known about their similarities and differences with multilevel data. This study examined four multilevel multiple imputation approaches: JM approaches proposed by Schafer and Yucel and Asparouhov and Muthén and FCS methods described by van Buuren and Carpenter and Kenward. Analytic work and computer simulations showed that Asparouhov and Muthén and Carpenter and Kenward methods are most flexible, as they produce imputations that preserve distinct within- and between-cluster covariance structures. As such, these approaches are applicable to random intercept models that posit level-specific relations among variables (e.g., contextual effects analyses, multilevel structural equation models). In contrast, methods from Schafer and Yucel and van Buuren are more restrictive and impose implicit equality constraints on functions of the within- and between-cluster covariance matrices. The analytic work and simulations underscore the conclusion that researchers should not expect to obtain the same results from alternative imputation routines. Rather, it is important to choose an imputation method that partitions variation in a manner that is consistent with the analysis model of interest. A real data analysis example illustrates the various approaches.

Keywords: *achievement; computer applications; dropouts; hierarchical linear modeling*

Multiple imputation is one of the predominant methods for treating missing data in educational and behavioral research. Imputation approaches can be divided into two general frameworks: joint model (JM) imputation (Rubin, 1987; Schafer, 1997) and fully conditional specification (FCS) imputation (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001; van Buuren, 2012; van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006). JM draws missing values for all incomplete variables in a single step from a multivariate distribution, whereas FCS imputes variables one at a time, drawing missing values from a series of univariate distributions. Historically, JM was the predominant method for single-level imputation of multivariate normal data (Rubin, 1987; Schafer, 1997), and FCS was proposed later as a tool for dealing with mixtures of categorical and continuous variables (Raghunathan et al., 2001; van Buuren, 2007). Because single-level JM usually generates imputations from a saturated model, it is able to preserve associations for a wide range of linear models. Single-level FCS possesses the same qualities, and JM and FCS are known to be equivalent with multivariate normal data (Hughes et al., 2014).

The extension of JM and FCS to multilevel data is relatively recent, and little is known about the similarities and differences of the two frameworks in this context. Schafer and Yucel (2002) and Asparouhov and Muthén (2010) outlined JM imputation strategies based on a multivariate linear mixed model. Like their single-level counterparts, these methods draw imputations from a multivariate normal distribution, but they differ in their treatment of complete variables. Consistent with Enders, Mistler, and Keller (2016), we henceforth refer to these methods as JM-SY and JM-AM, respectively. van Buuren (2011) proposed an extension of FCS based on a series of univariate linear mixed models, and Carpenter and Kenward (2013, p. 220) described a modification to FCS developed by Ian White (cited as a personal communication) that incorporates Level-2 cluster means as covariates, much like the classic contextual effects model from the multilevel literature (Longford, 1989; Lüdke, Marsh, Robitzsch, & Trautwein, 2011; Shin & Raudenbush, 2010). We henceforth refer to these variations as FCS-VB and FCS-WCK, respectively. JM-SY is implemented in the PAN, JOMO, and MLMMM packages in R (Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008), the REALCOM-IMPUTE package for MLwiN and Stata (Carpenter, Goldstein, & Kenward, 2011), and SAS (Mistler, 2013); JM-AM is available in Mplus (Muthén & Muthén, 1998–2012). FCS-VB is available in the R package MICE (van Buuren, & Groothuis-Oudshoorn, 2011), and FCS-WCK is implemented in the BLImP application (Enders, Keller, & Levy, 2016).

Although JM and FCS are equivalent to single-level multivariate normal data (Hughes et al., 2014), the same is not necessarily true in the multilevel context, where algorithmic features and variations in the underlying models can lead to different estimates (Enders, Mistler, & Keller, 2016). The differences between JM and FCS approaches are particularly important for multilevel data sets where relations among lower level variables differ at Level 1 and Level 2, a situation that Snijders and Bosker (2012, p. 60) characterize as “the rule rather than the exception.” One such example is the classic contextual effects model that partitions the association between a pair of Level-1 variables into between- and within-cluster components (Longford, 1989; Lüdke et al., 2011; Lüdke et al., 2008; Raudenbush & Bryk, 2002; Shin & Raudenbush, 2010):

$$Y_{2ij} = \beta_0 + \beta_1 Y_{1ij} + \beta_2 X_{ij} + \beta_3 \bar{X}_j + b_j + \varepsilon_{ij}, \quad (1)$$

where β_2 represents the pooled within-cluster regression of Y_2 on X, and β_3 captures the difference between the within-cluster regression and the between-cluster regression of $\bar{Y}_{2j}$ on $\bar{X}_j$ (i.e., the contextual effect). Raudenbush and Bryk (2002) illustrated a contextual effects analysis with student-level socioeconomic status (SES) and school average SES (i.e., X_{ij} and $\bar{X}_j$, respectively) predicting academic achievement, and numerous other examples appear in the applied literature (Harker & Tymms, 2004; Kenny & La Voie, 1985; Lüdke, Köller, Marsh, & Trautwein, 2005; Miller & Murdock, 2007; Simons, Wills, & Neal, 2015). Multilevel structural equation modeling is another common analysis framework for examining within- and between-cluster covariance structures. As an example, Martin, Malmberg, and Liem (2010) used multilevel factor analysis to study the internal structure of individual and school average academic motivation and engagement, and similar applications abound in the applied literature (Dunn, Masyn, Jones, Subramanian, & Koenen, 2015; Huang & Cornell, 2015; Muthén, 1991; Reise, Ventura, Neuchterlein, & Kim, 2005; Toland & De Ayala, 2005).

A chief concern of this article is whether (and which) JM and FCS imputation approaches generate replacement values that are appropriate for modeling unique covariance structures at Level 1 and Level 2 (e.g., the aforementioned contextual effects analysis and multilevel structural equation models). To date, the methodological literature provides little insight into the situations under which multilevel imputation approaches would produce similar (or different) results, and no studies have undertaken a rigorous comparison of the aforementioned strategies. Thus, the purpose of this study is to examine the situations under which JM and FCS reproduce (or preserve) the covariance structure of a population random intercept model with multivariate normal data. To do so, we use analytic methods to examine the model-implied covariance structure of the four imputation approaches, and we then demonstrate the analytic findings with Monte Carlo computer simulations. Because the four methods do not necessarily preserve the population joint distribution, the results from this study have important practical implications for substantive researchers in the behavioral sciences. In particular, the analytic results show that JM-AM and FCS-WCK employ rather unrestrictive imputation models, whereas the JM-SY and FCS-VB models place implicit constraints on the within- and between-cluster covariance matrices. When these constraints are incompatible with the analysis model, imputation can introduce bias, even under a benign missing completely at random (MCAR) mechanism. Thus, our results underscore the need to select an imputation approach that honors the covariance structure of subsequent analysis model.

The organization of this article is as follows: First, we begin with a brief section that establishes some notation. Second, we provide an overview of JM and FCS, highlighting differences among the four approaches. Third, we use analytic methods to examine whether the imputation approaches reproduce the covariance structure from a population model with random intercepts. Fourth, we use computer simulations to verify and illustrate the analytic findings. Fifth, we demonstrate the imputation methods using data from the classic high school and beyond study. Finally, we conclude with a brief discussion, highlighting the practical implications of our findings to behavioral science researchers.

Background and Notation

To establish some notation, let $\mathbf{Y}$ denote a set of Q incomplete Level 1 variables, $\mathbf{Y} = \{Y_{(1)}, \ldots, Y_{(Q)}\}$, and let $\mathbf{X}$ represent a set of S complete covariates, $\mathbf{X} = \{X_{(1)}, \ldots, X_{(S)}\}$. Unless otherwise noted, we use $\mathbf{Y}$ and $\mathbf{X}$ to differentiate incomplete and complete variables, respectively, and this notation is not meant to carry information about a variable's role in an analysis model (e.g., a component of $\mathbf{Y}$ could be a predictor in an analysis model). Consistent with Rubin and colleagues (Little & Rubin, 2002; Rubin, 1976), $\mathbf{M}$ is a set of Q missing data indicators, $\mathbf{M} = \{M_{(1)}, \ldots, M_{(Q)}\}$, where $M_{(q)} = 1$ if $Y_{(q)}$ is missing and $M_{(q)} = 0$ if $Y_{(q)}$ is complete. The missing data indicators partition $\mathbf{Y}$ into observed and missing parts, $\mathbf{Y}_{\text{obs}}$ and $\mathbf{Y}_{\text{mis}}$, respectively, where $\mathbf{Y}_{\text{obs}} = \{Y_{\text{obs}(1)}, \ldots, Y_{\text{obs}(Q)}\}$ and $\mathbf{Y}_{\text{mis}} = \{Y_{\text{mis}(1)}, \ldots, Y_{\text{mis}(Q)}\}$. The goal of multiple imputation is to sample several versions of $\mathbf{Y}_{\text{mis}}$ (e.g., 20 or more; Graham, Olchowski, & Gilreath, 2007) from a distribution that conditions on the observed data in $\mathbf{Y}_{\text{obs}}$ and $\mathbf{X}$. Under the missing at random (MAR) mechanism, $\mathbf{M}$ is conditionally independent of $\mathbf{Y}_{\text{mis}}$, given the values of $\mathbf{Y}_{\text{obs}}$ and $\mathbf{X}$. That is,

$$P(\mathbf{M}|\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}, \mathbf{X}) = P(\mathbf{M}|\mathbf{Y}_{\text{obs}}, \mathbf{X}), \tag{2}$$

where $P(\cdot)$ denotes a probability distribution.

The key difference between JM and FCS is that the former samples values from a Q-dimensional multivariate distribution, whereas the latter draws values from Q univariate conditional distributions, one for each incomplete variable. To accommodate the variable-by-variable imputation scheme of FCS, we must differentiate target variable $Y_{(q)}$ from the remaining variables in $\mathbf{Y}$. Following van Buuren's (2006) notational convention, let $\mathbf{Y}_{(-q)}$ denote all variables in $\mathbf{Y}$ except $Y_{(q)}$. For example, if $Y_{(2)}$ is the target of a particular imputation step, then $\mathbf{Y}_{(-q)} = \mathbf{Y}_{(-2)} = \{Y_{(1)}, Y_{(3)}, \ldots, Y_{(Q)}\}$. Finally, for certain imputation approaches, we must treat the cluster means as distinct variables. When necessary, we use $\bar{\mathbf{Y}}$ and $\bar{\mathbf{X}}$ to represent the cluster means of $\mathbf{Y}$ and $\mathbf{X}$, respectively, $\bar{\mathbf{Y}} = \{\bar{Y}_{(1)}, \ldots, \bar{Y}_{(Q)}\}$ and $\bar{\mathbf{X}} = \{\bar{X}_{(1)}, \ldots, \bar{X}_{(S)}\}$, where $\bar{\mathbf{Y}}$ is computed from the filled-in variables in $\mathbf{Y}$.

JM Imputation

Both JM-SY and JM-AM use a multivariate linear mixed model to define a multivariate normal distribution for the missing values. A multivariate regression model with random intercepts can be written as:

$$\mathbf{Y}_{ij} = \alpha + \mathbf{X}_{ij}\beta + \mathbf{b}_j + \varepsilon_{ij}, \tag{3}$$

where $\mathbf{Y}_{ij}$ is the row vector of k outcome scores for case i in cluster j, $\mathbf{X}_{ij}$ is a corresponding row vector of l covariate scores, α is the row vector containing the k intercepts, β is an $l \times k$ matrix of regression coefficients, $\mathbf{b}_j$ is a row vector containing the k Level-2 residuals for cluster j, and ε_{ij} is the row vector of k Level-1 residuals for case i in cluster j. The Level-1 and Level-2 residuals are assumed to follow a multivariate normal distribution with zero means and covariance matrices $\sum_\varepsilon$ and $\sum_b$, respectively. Note that we use $\sum_\varepsilon$ and $\sum_b$ to denote blocks of the overall covariance matrix, the values of which are assumed constant for all clusters. The model from Equation 3 is the basis for both JM-SY and JM-AM, but the two procedures differ in their treatment of **X** and **Y**.

The imputation phase of JM employs Bayesian estimation machinery that views the missing values, Level-2 residuals, and model parameters as random variables having a joint distribution. Bayesian estimation expresses this joint distribution as a series of full conditional distributions, and it uses a Markov chain Monte Carlo (MCMC) algorithm to iteratively sample variables in three major steps: (a) draw missing values from a multivariate distribution that conditions on the observed data, the current Level-2 residuals, and the current model parameter values, (b) draw Level-2 residuals from a multivariate distribution that conditions on the complete data from the previous step and the current model parameter values, and (c) draw model parameters from distributions that condition on the quantities from the first two steps.

More formally, a single iteration t of the MCMC algorithm can be summarized as:

$$\begin{aligned}
\mathbf{Y}_{\text{mis}}^{(t)} &\sim P\left(\mathbf{Y}_{\text{mis}} | \mathbf{b}^{(t-1)}, \theta^{(t-1)}, \mathbf{Y}_{\text{obs}}, \mathbf{X}\right) \\
\mathbf{b}^{(t)} &\sim P\left(\mathbf{b} | \theta_{(1)}^{(t-1)}, \mathbf{Y}_{\text{mis}}^{(t)}, \mathbf{Y}_{\text{obs}}, \mathbf{X}\right) \\
\theta^{(t)} &\sim P\left(\theta | \mathbf{b}^{(t)}, \mathbf{Y}_{\text{mis}}^{(t)}, \mathbf{Y}_{\text{obs}}, \mathbf{X}\right),
\end{aligned} \tag{4}$$

where P denotes a probability distribution and θ represents the collection of multilevel model parameters, $\theta = \{\alpha, \beta, \Sigma_b, \Sigma_\varepsilon\}$. Because the sampling steps for the model parameters and Level-2 residuals borrow from established complete-data Bayesian estimation procedures (Browne & Draper, 2000; Goldstein, Bonnet, & Rocher, 2007; Goldstein, Carpenter, Kenward, & Levin, 2009; Kasim & Raudenbush, 1998; Schafer & Yucel, 2002; Yucel, 2008), we restrict our attention to the imputation step for $\mathbf{Y}_{\text{mis}}$.

JM-SY

Schafer and Yucel's (2002) seminal work on JM imputation employs the linear mixed model from Equation 3, where the incomplete variables serve as outcomes in **Y** and complete covariates function as predictors in **X**. Using parameter values and Level-2 residual terms from a previous MCMC step, JM-SY imputation draws the missing parts of **Y** from a multivariate normal conditional distribution:

$$\mathbf{Y}_{ij_{\text{mis}}}^{(t)} \sim \text{MVN}\left(\alpha^{(t-1)} + \mathbf{X}_{ij}\beta^{(t-1)} + \mathbf{b}_j^{(t-1)}, \Sigma_\varepsilon^{(t-1)}\right), \tag{5}$$

where the mean vector $\alpha^{(t-1)} + \mathbf{X}_{ij}\beta^{(t-1)} + \mathbf{b}_j^{(t-1)}$ contains predicted values from the model, and the covariance matrix is the within-cluster residual covariance matrix from the multilevel regression of **Y** on **X**. Note that the usual unit vector is not part of **X** because we separate the intercepts from the β matrix (doing so facilitates the analytic work later in the article).

To illustrate JM-SY, consider an imputation model with three variables, Y_1, Y_2, and X, where Y_1 and Y_2 are incomplete and X is complete. Expressed in scalar notation, the conditional distribution that generates imputations is as follows:

$$\begin{bmatrix} Y_{1ij_{\text{mis}}}^{(t)} \\ Y_{2ij_{\text{mis}}}^{(t)} \end{bmatrix} \sim N_2\left(\begin{bmatrix} \alpha_{Y_1}^{(t-1)} + X_{ij}\beta_{Y_1|X}^{(t-1)} + b_{jY_1}^{(t-1)} \\ \alpha_{Y_2}^{(t-1)} + X_{ij}\beta_{Y_2|X}^{(t-1)} + b_{jY_2}^{(t-1)} \end{bmatrix}, \Sigma_\varepsilon^{(t-1)}\right). \tag{6}$$

Note that we subscript the parameters and residual terms with variable names to emphasize that each equation requires unique values. Although not explicit in the previous equation, note that $\sum_\varepsilon$ (an estimate of which is obtained from the MCMC algorithm) defines the variance–covariance matrix of the Level-2 intercept residuals, $\varepsilon_{Y_{1j}}$ and $\varepsilon_{Y_{2j}}$.

Conceptually, the imputations from Equation 6 can be viewed as the sum of predicted values and residual terms (e.g., $Y_{1ij_{\text{mis}}}^{(t)} = \hat{Y}_{1ij} + \varepsilon_{ijY_1}$ and $Y_{2ij_{\text{mis}}}^{(t)} = \hat{Y}_{2ij} + \varepsilon_{ijY_2}$), where the mean vector defines the predicted values (e.g., $\hat{Y}_{1ij} = \alpha_{Y_1} + X_{ij}\beta_{Y_1|X} + b_{jY_1}$) and the within-cluster covariance matrix $\sum_\varepsilon$ defines the variances and covariances of ε_{Y_1} and ε_{Y_2}. A key consideration for this article is whether the above conditional distribution preserves the covariance structure of the population data. To gain some intuition about the imputation model, it is useful to note that setting the Level-2 intercept residuals in the mean vector to zero gives a single-level imputation model where Y_1 and Y_2 have only one source of variability. Thus, the presence of Level-2 residuals infuses the imputations with between-cluster variation, and sampling these residuals from a multivariate normal distribution preserves the between-cluster association between Y_1 and Y_2, the magnitude of which is determined by $\sum_b$. Further, because the b_j and ε_{ij} terms are orthogonal, the imputation model allows the variance–covariance matrix of Y_1 and Y_2 to differ at Level 1 and Level 2.

JM-AM

Asparouhov and Muthén (2010) described a variation of JM that also applies the linear mixed model from Equation 3, but JM-AM treats all variables as outcomes in **Y**, regardless of missing data pattern. That is, the **Y** vector in Equation 3 contains incomplete and complete variables, and **X** is empty. To illustrate JM-AM, reconsider the imputation problem for Y_1, Y_2, and X. Expressed in scalar notation, the conditional distribution that generates imputations is as follows.

$$\begin{bmatrix} Y_{1ij_{\text{mis}}}^{(t)} \\ Y_{2ij_{\text{mis}}}^{(t)} \\ X_{ij} \end{bmatrix} \sim N_3\left(\begin{bmatrix} \alpha_{Y_1}^{(t-1)} + b_{jY_1}^{(t-1)} \\ \alpha_{Y_2}^{(t-1)} + b_{jY_2}^{(t-1)} \\ \alpha_{X}^{(t-1)} + b_{jX}^{(t-1)} \end{bmatrix}, \Sigma_{\varepsilon}^{(t-1)} \right). \tag{7}$$

Consistent with JM-SY, the mean vector contains predicted values from the model (e.g., $\hat{Y}_{ij} = \alpha_{Y_1} + b_{jY_1}$), which in this parameterization are analogous to latent group means (Lüdke et al., 2011). As before, each imputation can be viewed as the sum of a predicted value and a within-cluster residual (e.g., $Y_{1ij_{\text{mis}}} = \bar{Y}_1 + b_{jY_1} + \varepsilon_{ijY_1}$), the variances and covariances of which are given by $\sum_{\varepsilon}$. By treating all variables as outcomes, JM-AM partitions each variable into a within- and between-cluster component, such that $\sum_{\varepsilon}$ and $\sum_{b}$ are unstructured within- and between-cluster covariance matrices, respectively. As explained later, this parameterization requires somewhat stricter distributional assumptions than JM-SY (i.e., multivariate normality vs. conditional normality).

FCS Imputation

The FCS approaches employ a series of Q univariate linear mixed models to define a normal distribution for the missing values, one for each incomplete variable. A univariate regression model with random intercepts can be written as:

$$Y_{ij} = \alpha + \mathbf{X}_{ij}\beta + b_j + \varepsilon_{ij}, \tag{8}$$

where Y_{ij} is the outcome score for case i in cluster j, $\mathbf{X}_{ij}$ is the corresponding row vector of covariate scores, α is the intercept, β is a row vector of regression coefficients, b_j is the Level-2 intercept residual for cluster j, and ε_{ij} is the Level-1 residual for case i in cluster j. The Level-1 and Level-2 residuals are assumed to follow normal distributions with zero means and variances $\sigma^2_{\boldsymbol{\varepsilon}}$ and σ^2_b, respectively. The model from Equation 8 is the basis for both FCS-VB and FCS-WCK, but the contents of **X** differ under the two procedures.

The imputation phase of FCS employs the same Bayesian estimation machinery as JM. The MCMC algorithm iteratively samples values in three major steps (i.e., sample missing values, draw Level-2 residuals, and sample model parameters), but it does so separately for each of the Q incomplete variables in a sequence. More formally, a single iteration t of the MCMC algorithm can be expressed as:

$$\begin{gathered}
Y_{\mathrm{mis}(1)}^{(t)} \sim P\left(Y_{\mathrm{mis}(1)} | b_{(1)}^{(t-1)}, \boldsymbol{\theta}_{(1)}^{(t-1)}, \mathbf{Y}_{(-1)}^{(\mathrm{current})}, \mathbf{X}\right) \\
b_{(1)}^{(t)} \sim P\left(b_{(1)} | \boldsymbol{\theta}_{(1)}^{(t-1)}, Y_{\mathrm{mis}(1)}^{(t)}, Y_{\mathrm{obs}(1)}, \mathbf{Y}_{(-1)}^{(\mathrm{current})}, \mathbf{X}\right) \\
\boldsymbol{\theta}_{(1)}^{(t)} \sim P\left(\boldsymbol{\theta}_{(1)} | b_{(1)}^{(t)}, Y_{\mathrm{mis}(1)}^{(t)}, Y_{\mathrm{obs}(1)}, \mathbf{Y}_{(-1)}^{(\mathrm{current})}, \mathbf{X}\right) \\
\vdots \\
Y_{\mathrm{mis}(Q)}^{(t)} \sim P\left(Y_{\mathrm{mis}(Q)} | b_{(Q)}^{(t-1)}, \boldsymbol{\theta}_{(Q)}^{(t-1)}, \mathbf{Y}_{(-Q)}^{(\mathrm{current})}, \mathbf{X}\right) \\
b_{(Q)}^{(t)} \sim P\left(b_{(Q)} | \boldsymbol{\theta}_{(Q)}^{(t-1)}, Y_{\mathrm{mis}(Q)}^{(t)}, Y_{\mathrm{obs}(Q)}, \mathbf{Y}_{(-Q)}^{(\mathrm{current})}, \mathbf{X}\right) \\
\boldsymbol{\theta}_{(Q)}^{(t)} \sim P\left(\boldsymbol{\theta}_{(Q)} | b_{(Q)}^{(t)}, Y_{\mathrm{mis}(Q)}^{(t)}, Y_{\mathrm{obs}(Q)}, \mathbf{Y}_{(-Q)}^{(\mathrm{current})}, \mathbf{X}\right),
\end{gathered} \tag{9}$$

where $Y_{-q}^{\mathrm{current}}$ represents all incomplete variables except $Y_{(q)}$ at the current iteration and $\boldsymbol{\theta}_{(q)}$ denotes the set of model parameters for variable $Y_{(q)}$, $\boldsymbol{\theta}_{(q)} = \{\alpha_{(q)}, \beta_{1(q)}, \sigma^2_{b(q)}, \sigma^2_{\varepsilon(q)}\}$. We use the (current) superscript in lieu of the iteration index t because the previously imputed variables in $\mathbf{Y}_{(-q)}$ do not necessarily originate from the same MCMC iteration (e.g., when $1 < q < Q$, the filled-in variables $Y_{(1)}$ through $Y_{(q\text{-}1)}$ are obtained from iteration t, while $Y_{(q+1)}$ through $Y_{(Q)}$ are obtained from iteration $t - 1$).

FCS-VB

van Buuren's (2011, 2012) extension of FCS to the multilevel context employs a linear mixed model whereby incomplete variable $Y_{(q)}$ serves as the outcome variable and previously imputed variables and complete covariates function as predictors. That is, the $\mathbf{X}$ vector in Equation 8 contains complete variables and the previously imputed variables in $\mathbf{Y}_{(-q)}$. To illustrate FCS-VB, reconsider the trivariate imputation problem from the previous section. FCS-VB imputes the incomplete variables in a sequence. Expressed in scalar notation, the conditional distribution that generates Y_1 imputations is

$$Y_{1ij_{\mathrm{mis}}}^{(t)} \sim N\left(\alpha_{Y_1}^{(t-1)} + X_{ij}\beta_{Y_1|X}^{(t-1)} + Y_{2ij}^{(t-1)}\beta_{Y_1|Y_2}^{(t-1)} + b_{jY_1}^{(t-1)}, \sigma^2_{\boldsymbol{\varepsilon}(Y_1)}\right), \tag{10}$$

where $Y_{2ij}^{(t-1)}$ is the filled-in variable from the previous iteration. After sampling new residual terms and model parameters for the next round of Y_1 imputation, FCS-VB switches the roles of Y_1 and Y_2, such that Y_1 defines the conditional distribution of $Y_{2(\mathrm{mis})}$, as follows:

$$Y_{2ij_{\mathrm{mis}}}^{(t)} \sim N\left(\alpha_{Y_2}^{(t-1)} + X_{ij}\beta_{Y_2|X}^{(t-1)} + Y_{1ij}^{(t)}\beta_{Y_2|Y_1}^{(t-1)} + b_{jY_2}^{(t-1)}, \sigma^2_{\varepsilon(Y_2)}\right). \tag{11}$$

Consistent with JM imputation, the Level-2 residuals in the mean vector infuse between-cluster variation into the imputations, and $\sigma^2_{\varepsilon(q)}$ provides additional within-cluster variation. Because the MCMC algorithm samples b_{jY_1} and b_{jY_2} from independent univariate normal distributions, FCS-VB effectively imposes a diagonal structure on $\sum_b$, where $\sigma^2_{bY_1}$ and $\sigma^2_{bY_2}$ define the diagonal. The absence of correlation between the Level-2 residuals implies that the fixed effects alone are responsible for preserving between-cluster covariation, and FCS-VB makes no attempt to partition associations into distinct within- and between-cluster components. Finally, it is important to note that Equations 10 and 11 assume a common residual variance for all clusters, but van Buuren (2011) describes another version of FCS-VB that introduces heterogeneous within-cluster variances (Kasim & Raudenbush, 1998). van Buuren (2011) suggested that the heterogeneous model can improve imputations for incomplete predictor variables, but we limit our attention to the standard homogeneous model in order to maintain comparable assumptions for JM and FCS.

FCS-WCK

Carpenter and Kenward (2013, p. 220) describe a modification to FCS proposed by Ian White that introduces the cluster means of Level-1 variables (complete or imputed) as covariates in each imputation model. To illustrate FCS-WCK, reconsider the previous trivariate imputation problem. The conditional distribution that generates Y_1 imputations is:

$$Y^{(t)}_{1ij_{\text{mis}}} \sim N\left(\alpha^{(t-1)}_{Y_1} + X_{ij}\beta^{(t-1)}_{Y_1|X} + Y^{(t-1)}_{2ij}\beta^{(t-1)}_{Y_1|Y_2} + \bar{X}_j\beta^{(t-1)}_{Y_1|\bar{X}} + \bar{Y}^{(t-1)}_{2j}\beta^{(t-1)}_{Y_1|\bar{Y}_2} + b^{(t-1)}_{jY_1}, \sigma^2_{\varepsilon(Y_1)}\right), \quad (12)$$

where $\bar{X}_j$ is the mean of X in cluster j, $Y^{(t-1)}_{2ij}$ is the imputed variable from the previous iteration, and $\bar{Y}^{(t-1)}_{2j}$ is a cluster mean computed from the filled-in data. In a similar vein, the conditional distribution that generates Y_2 imputations is as follows:

$$Y^{(t)}_{2ij_{\text{mis}}} \sim N\left(\alpha^{(t-1)}_{Y_2} + X_{ij}\beta^{(t-1)}_{Y_2|X} + Y^{(t)}_{1ij}\beta^{(t-1)}_{Y_2|Y_1} + \bar{X}_j\beta^{(t-1)}_{Y_2|\bar{X}} + \bar{Y}^{(t)}_{1j}\beta^{(t-1)}_{Y_2|\bar{Y}_1} + b^{(t-1)}_{jY_2}, \sigma^2_{\varepsilon(Y_2)}\right), \quad (13)$$

where $Y^{(t)}_{1ij}$ and $\bar{Y}^{(t)}_{1j}$ are again obtained from the imputed data at iteration t. Note that the logic of FCS-WCK imputation parallels the classic contextual effects model that uses cluster means to partition the relations among Level-1 variables into within- and between-cluster components (Longford, 1989; Lüdke et al., 2011; Shin & Raudenbush, 2010). Unlike FCS-VB, the absence of correlation between b_{jY_1} and b_{jY_2} places no constraints on the imputations because FCS-WCK uses cluster means to model this relation.

Informal Comparison of Imputation Methods

An important issue for our investigation is whether the imputation approaches preserve the multilevel covariance structure of the population data. An inspection of the previous models allows for some informal conclusions, and the next section presents a more rigorous analytic comparison of the four methods. To put the problem in a context, consider the contextual effects analysis model from Equation 1. Further, assume that Y_1 and Y_2 are incomplete and X is complete. Importantly, notice that the analysis model posits a single relation between Y_2 and Y_1, but it uses distinct slopes to capture the within- and between-cluster influence of X on Y_2.

To begin, consider the JM-AM imputation model from Equation 7. As explained previously, JM-AM partitions every variable into within- and between-cluster components, and it uses unrestricted covariance matrices (i.e., $\sum_{\varepsilon}$ and $\sum_{b}$) to generate imputations that allow the magnitude of the variation and covariation to differ at Level 1 and Level 2. As such, JM-AM is more general than this particular analysis model because it includes additional effects that are not present in Equation 1. In particular, JM-AM allows the Y_1–Y_2 relation and the Y_1–X relation to differ at the within- and between-cluster levels, whereas the analysis model posits a common slope for the regression of Y_2 on Y_1. The FCS-WCK imputation models from Equations 12 and 13 also accommodate unique within- and between-cluster covariance structures for all variables, but the models achieve this generality by introducing cluster means as predictor variables. In the single-level context, employing a rich imputation model with additional variables or effects (e.g., auxiliary variables) is usually not detrimental and is often beneficial (Meng, 1994; Schafer, 2003). Thus, we would expect JM-AM and FCS-WCK to produce appropriate imputations for analyses that posit distinct within- and between-cluster effects. The contextual effects model in Equation 1 is one such example, and multilevel structural equation models are another common example (e.g., a confirmatory factor analysis that imposes a different factor structure at Level 1 and Level 2).

Although JM-AM and FCS-WCK are comparable in the sense that they accommodate unique within- and between-cluster covariance matrices, they are subtly different in at least two ways. First, JM-AM uses random effects to model between-cluster associations, whereas FCS-WCK uses cluster means computed from the filled-in data. This distinction is analogous to using a latent versus manifest variable approach to estimating contextual effects models (Lüdke et al., 2011; Lüdke et al., 2008). Second, because JM-AM treats complete variables as outcomes, it assumes multivariate normality for all variables in the imputation model. Although the procedure does not draw replacement values for complete variables, the conditional distributions from which MCMC samples random effects and parameter values derive from the multivariate normal distribution. In line with standard linear mixed models, FCS-WCK instead treats complete variables as fixed predictors, thus requiring the less stringent assumption that only the incomplete variables are multivariate normal. Whether this difference in distributional assumptions has any practical impact on imputation quality is an open question in the literature.

Whereas JM-AM and FCS-WCK are flexible enough to accommodate unique within- and between-cluster associations for all variables, JM-SY and FCS-VB employ simpler imputation models that place restrictions on the covariance structure. As explained previously, JM-SY uses Level-2 residuals with an unstructured covariance matrix (i.e., $\sum_b$) to preserve distinct within- and between-cluster covariance matrices for the incomplete variables. However, returning to Equation 6, notice that the model does not include a distinct between-cluster component for the complete variable X. As such, we would expect JM-SY to yield biased estimates when applied to the contextual effects analysis model from Equation 1. Borrowing from the FCS-WCK method, a simple remedy for this shortcoming is to include the complete variable cluster means in the imputation model's $\mathbf{X}$ matrix, thereby allowing all associations to differ at Level 1 and Level 2. Finally, FCS-VB appears to be the most restrictive of the four imputation approaches because it makes no attempt to partition relations into distinct within- and between-cluster components. Consequently, we would expect FCS-VB to yield biased estimates when applied to analysis models such as that in Equation 1—or more generally, to any analysis that models distinct covariance matrices at Level 1 and Level 2 (e.g., a multilevel factor analysis model with different structures at the two levels).

Analytic Comparison of Multilevel Imputation Models

This section presents an analytic comparison of multilevel imputation methods in the context of a two-level random intercept population model with normally distributed Level-1 variables. We restrict our attention to random intercepts because existing JM approaches have little or no capacity for accommodating random slope variation (Enders, Mistler, & Keller, 2016; Yucel, 2011). Further, we focus on Level-1 variables because there is no established strategy for handling incomplete Level-2 variables with JM-SY and the FCS approaches (Gelman & Hill, 2007; Yucel, 2008). Finally, focusing on a scenario with random intercepts and Level-1 variables is a logical starting point, given the nascent state of the multilevel imputation literature.

To keep the analytic work simple without a loss of generality, we consider a population model with three Level-1 variables. We previously suggested that the four imputation strategies have different capacities for preserving associations between pairs of incomplete variables and between pairs of incomplete and complete variables. To examine this possibility, we consider the situation where two of the three variables are incomplete; in line with our earlier examples, we refer to the incomplete variables as Y_1 and Y_2, and we denote the complete variable as X. Although relatively simple, a trivariate problem with two incomplete variables should yield conclusions that generalize to scenarios with additional variables and general missing data patterns (e.g., a method that fails to partition a particular relation into within- and between-cluster components would necessarily do so when applied to a much larger set of variables).

Population Joint Distribution

The population random intercept model partitions variables into within- and between-cluster components, as follows.

$$\begin{aligned}\begin{bmatrix} Y_{1ij} \\ Y_{2ij} \\ X_{ij} \end{bmatrix} &= \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{X} \end{bmatrix} + \begin{bmatrix} Y_{1ij} - \bar{Y}_{1j} \\ Y_{2ij} - \bar{Y}_{2j} \\ X_{ij} - \bar{X}_j \end{bmatrix} + \begin{bmatrix} \bar{Y}_{1j} - \bar{Y}_1 \\ \bar{Y}_{2j} - \bar{Y}_2 \\ \bar{X}_j - \bar{X} \end{bmatrix} \\ &= \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{X} \end{bmatrix} + \begin{bmatrix} Y_{1(\mathrm{W})} \\ Y_{2(\mathrm{W})} \\ X_{(\mathrm{W})} \end{bmatrix} + \begin{bmatrix} Y_{1(\mathrm{B})} \\ Y_{2(\mathrm{B})} \\ X_{(\mathrm{B})} \end{bmatrix}. \end{aligned} \tag{14}$$

The Level-1 and Level-2 deviation scores are multivariate normal with zero means and unstructured covariance matrices:

$$\begin{bmatrix} Y_{1(\mathrm{W})} \\ Y_{2(\mathrm{W})} \\ X_{(\mathrm{W})} \end{bmatrix} \sim N_3(\mathbf{0}, \Sigma_{\mathrm{w}}) \begin{bmatrix} Y_{1(\mathrm{B})} \\ Y_{2(\mathrm{B})} \\ X_{(\mathrm{B})} \end{bmatrix} \sim N_3(\mathbf{0}, \Sigma_{\mathrm{B}}), \tag{15}$$

where the cluster-specific blocks of the covariance matrices are:

$$\Sigma_{\mathrm{W}} = \begin{bmatrix} \sigma^2_{Y_1(\mathrm{W})} & \sigma_{Y_1Y_2(\mathrm{W})} & \sigma_{Y_1X(\mathrm{W})} \\ \sigma_{Y_2Y_1(\mathrm{W})} & \sigma^2_{Y_2(\mathrm{W})} & \sigma_{Y_2X(\mathrm{W})} \\ \sigma_{XY_1(\mathrm{W})} & \sigma_{XY_2(\mathrm{W})} & \sigma^2_{X(\mathrm{W})} \end{bmatrix} \quad \Sigma_{\mathrm{B}} = \begin{bmatrix} \sigma^2_{Y_1(\mathrm{B})} & \sigma_{Y_1Y_2(\mathrm{B})} & \sigma_{Y_1X(\mathrm{B})} \\ \sigma_{Y_2Y_1(\mathrm{B})} & \sigma^2_{Y_2(\mathrm{B})} & \sigma_{Y_2X(\mathrm{B})} \\ \sigma_{XY_1(\mathrm{B})} & \sigma_{XY_2(\mathrm{B})} & \sigma^2_{X(\mathrm{B})} \end{bmatrix}. \tag{16}$$

Thus, the population joint distribution has a total of 15 unique parameters (i.e., 3 means and 6 unique elements each in the within- and between-cluster covariance matrices), the exact values of which are not important for the ensuing analytic work. Importantly, the population model places no restrictions or equality constraints on the elements of the covariance matrices.

To examine the compatibility (or lack thereof) of each imputation model to the population joint distribution, we employ a strategy similar to that of Shin and Raudenbush (2007) in the context of maximum likelihood estimation. Specifically, we first restate the joint distribution as a conditional distribution having the same form as a given imputation model (e.g., incomplete variables as outcomes, the complete variable as a predictor). Next, we parameterize the conditional distribution in terms of linear mixed model parameters. Finally, we compute the parameters of each imputation model from the parameters of the transformed joint distribution. Methods that fail to preserve the joint distribution use fewer than 15 parameters and thus require constraints on one or more population parameters. This final step highlights incompatibilities between the imputation and population models.

JM-SY

To facilitate comparisons with the JM-SY imputation model in Equation 6, we first restate the population joint distribution as a conditional distribution that expresses Y_1 and Y_2 (the incomplete variables) as a function of X (the complete variable). Because the elements of $\sum_W$ and $\sum_B$ are orthogonal, the covariance matrix of the joint distribution can be written as follows:

$$\begin{bmatrix} Y_{1ij} \\ Y_{2ij} \\ X_{ij} - \bar{X}_j \\ \bar{X}_j \end{bmatrix} \sim N_4 \left(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ 0 \\ \bar{X} \end{bmatrix}, \begin{bmatrix} \sigma^2_{Y_1(W)} + \sigma^2_{Y_1(B)} & \sigma_{Y_1Y_2(W)} + \sigma_{Y_1Y_2(B)} & \sigma_{Y_1X(W)} & \sigma_{Y_1X(B)} \\ \sigma_{Y_2Y_1(W)} + \sigma_{Y_2Y_1(B)} & \sigma^2_{Y_2(W)} + \sigma^2_{Y_2(B)} & \sigma_{Y_2X(W)} & \sigma_{Y_2X(B)} \\ \sigma_{XY_1(W)} & \sigma_{XY_2(W)} & \sigma^2_{X(W)} & 0 \\ \sigma_{XY_1(B)} & \sigma_{XY_2(B)} & 0 & \sigma^2_{X(B)} \end{bmatrix} \right). \tag{17}$$

Next, we use elements from the above covariance matrix to solve the coefficients from the multivariate regression of the incomplete variables on the within- and between-cluster components of X:

$$\begin{gathered} \begin{bmatrix} \gamma_{Y_1|X(W)} & \gamma_{Y_2|X(W)} \\ \gamma_{Y_1|X(B)} & \gamma_{Y_2|X(B)} \end{bmatrix} = \begin{bmatrix} \dfrac{\sigma_{XY_1(W)}}{\sigma^2_{X(W)}} & \dfrac{\sigma_{XY_2(W)}}{\sigma^2_{X(W)}} \\ \dfrac{\sigma_{XY_1(B)}}{\sigma^2_{X(B)}} & \dfrac{\sigma_{XY_2(B)}}{\sigma^2_{X(B)}} \end{bmatrix} \\ \begin{bmatrix} \upsilon_{Y_1|X} \\ \upsilon_{Y_2|X} \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 - \bar{X}\gamma_{Y_1|X(B)} \\ \bar{Y}_2 - \bar{X}\gamma_{Y_2|X(B)} \end{bmatrix}, \end{gathered} \tag{18}$$

where $\upsilon_{Y_1|X}$ and $\upsilon_{Y_2|X}$ denote intercepts, $\gamma_{Y_1|X(W)}$ and $\gamma_{Y_2|X(W)}$ are pure within-cluster regression slopes, and $\gamma_{Y_1|X(B)}$ and $\gamma_{Y_2|X(B)}$ denote between-cluster coefficients from the regression of the incomplete variables on $\bar{X}_j$. Note that we change notation (i.e., υ replaces α for the intercepts and γ replaces β for the slopes) in order to differentiate the population model from the imputation models. Using quantities from Equation 18, the residual covariance matrix of Y_1 and Y_2 conditional on X is:

$$\begin{aligned} \Sigma_{Y_1Y_2|X} = & \begin{bmatrix} \left(\sigma^2_{Y_1(B)} - \sigma^2_{X(B)}\gamma^2_{Y_1|X(B)}\right) + \left(\sigma^2_{Y_1(W)} - \sigma^2_{X(W)}\gamma^2_{Y_1|X(W)}\right) \\ \left(\sigma_{Y_2Y_1(B)} - \sigma^2_{X(B)}\gamma_{Y_2|X(B)}\gamma_{Y_1|X(B)}\right) + \left(\sigma_{Y_2Y_1(W)} - \sigma^2_{X(W)}\gamma_{Y_1|X(W)}\gamma_{Y_2|X(W)}\right) \end{bmatrix} \\ & \| \begin{bmatrix} \left(\sigma_{Y_1Y_2(B)} - \sigma^2_{X(B)}\gamma_{Y_2|X(B)}\gamma_{Y_1|X(B)}\right) + \left(\sigma_{Y_1Y_2(W)} - \sigma^2_{X(W)}\gamma_{Y_1|X(W)}\gamma_{Y_2|X(W)}\right) \\ \left(\sigma^2_{Y_2(B)} - \sigma^2_{X(B)}\gamma^2_{Y_2|X(B)}\right) + \left(\sigma^2_{Y_2(W)} - \sigma^2_{X(W)}\gamma^2_{Y_2|X(W)}\right) \end{bmatrix}, \end{aligned} \tag{19}$$

where || symbol denotes horizontal concatenation of two 2 × 1 vectors.

Finally, to match the form of the JM-SY imputation model, we use the results from Equations 18 and 19 to express the population joint distribution as a conditional distribution that is parameterized in terms of mixed model parameters. The distribution of the incomplete variables given the complete variable is:

$$\begin{bmatrix} Y_{1ij} \\ Y_{2ij} \end{bmatrix} \sim N_2\left(\begin{bmatrix} \upsilon_{Y_1} + X_{ij}\gamma_{Y_1|X(\mathrm{W})} + \bar{X}_j\left(\gamma_{Y_1|X(\mathrm{B})} - \gamma_{Y_1|X(\mathrm{W})}\right) + u_{1j} \\ \upsilon_{Y_2} + X_{ij}\gamma_{Y_2|X(\mathrm{W})} + \bar{X}_j\left(\gamma_{Y_2|X(\mathrm{B})} - \gamma_{Y_2|X(\mathrm{W})}\right) + u_{2j} \end{bmatrix}, \Sigma_e\right), \tag{20}$$

where u_{1j} and u_{2j} denote the population Level-2 residuals, and $\sum_e$ is the population within-cluster residual covariance matrix. Again, we switch notation for the parameters and Level-2 residuals (i.e., Σ_e replaces Σ_{W} for the Level 1 covariance matrix and u replaces b for the intercepts) in order to differentiate the population distribution from the imputation models. Because each element of the residual covariance matrix from Equation 19 contains additive terms involving within- and between-cluster components, we can write the within- and between-cluster covariance matrices as follows:

$$\begin{aligned} \Sigma_e &= \begin{bmatrix} \sigma^2_{Y_1(\mathrm{W})} - \sigma^2_{X(\mathrm{W})}\gamma^2_{Y_1|X(\mathrm{W})} & \sigma_{Y_1Y_2(\mathrm{W})} - \sigma^2_{X(\mathrm{W})}\gamma_{Y_1|X(\mathrm{W})}\gamma_{Y_2|X(\mathrm{W})} \\ \sigma_{Y_2Y_1(\mathrm{W})} - \sigma^2_{X(\mathrm{W})}\gamma_{Y_2|X(\mathrm{W})}\gamma_{Y_1|X(\mathrm{W})} & \sigma^2_{Y_2(\mathrm{W})} - \sigma^2_{X(\mathrm{W})}\gamma^2_{Y_2|X(W)} \end{bmatrix} \\ \Sigma_u &= \begin{bmatrix} \sigma^2_{Y_1(\mathrm{B})} - \sigma^2_{X(\mathrm{B})}\gamma^2_{Y_1|X(\mathrm{B})} & \sigma_{Y_1Y_2(\mathrm{B})} - \sigma^2_{X(\mathrm{B})}\gamma_{Y_2|X(\mathrm{B})}\gamma_{Y_1|X(\mathrm{B})} \\ \sigma_{Y_2Y_1(\mathrm{B})} - \sigma^2_{X(\mathrm{B})}\gamma_{Y_2|X(\mathrm{B})}\gamma_{Y_1|X(\mathrm{B})} & \sigma^2_{Y_2(\mathrm{B})} - \sigma^2_{X(\mathrm{B})}\gamma^2_{Y_2|X(\mathrm{B})} \end{bmatrix}. \end{aligned} \tag{21}$$

Finally, although the traditional mixed model framework treats predictor variables as fixed, our evaluation of JM-SY treats X as a random variable with the following marginal distribution:

$$X \sim N\left(\bar{X}, \sigma^2_{X(\mathrm{W})} + \sigma^2_{X(\mathrm{B})}\right). \tag{22}$$

This marginal distribution is common to both the imputation and population models.

Having expressed the population joint distribution as a multilevel model that matches the form of the JM-SY imputation model from Equation 6, we can now determine whether the imputation model preserve the associations in the population model. Comparing the number of parameters reveals that JM-SY is more restrictive than the population model because it includes a total of 13 parameters (2 intercepts, 2 regression coefficients, 3 unique elements in the Level-2 covariance matrix, 3 unique elements in the Level-1 covariance matrix, and 3 parameters in the marginal distribution of X), whereas the population model includes two additional regression coefficients. A comparison of Equations 6 and 20 reveals that the population distribution contains distinct coefficients for the within- and between-cluster influence of X on the incomplete variables, whereas the JM-SY imputation model uses a single slope to preserve the relation between X and Y_1 and X and Y_2. As such, JM-SY is inappropriate for models that posit unique within- and between-cluster associations between pairs of complete variables and incomplete variables, although it can preserve level-specific relations between pairs of incomplete variables. Returning to the contextual effects analysis model from Equation 1, the analytic results suggest that JM-SY should yield biased estimates of the regression slopes because it generates imputations from a model where the cluster means of X have no additional explanatory power above and beyond the Level-1 scores.

To equate JM-SY to the population model, one of the two situations must occur. First, imposing equality constraints on the within- and between-cluster coefficients from the population model reduces the number of parameters by two, thereby equating the fixed effect portion of the two models, as follows:

$$\begin{aligned} \gamma_{Y_1|X(\mathrm{W})} &= \gamma_{Y_1|X(\mathrm{B})} = \beta_{Y_1|X} \\ \gamma_{Y_2|X(\mathrm{W})} &= \gamma_{Y_2|X(\mathrm{B})} = \beta_{Y_2|X}. \end{aligned} \tag{23}$$

Of course, researchers have no control over the associations in the population model, so the realistic course of action is to adopt a richer imputation model that is capable of preserving a wider range of effects. Although Schafer and Yucel (2002) do not discuss this option, introducing the cluster means of the complete covariates into the fixed effect predictor matrix $\mathbf{X}$ (e.g., as with JM-WCK) would increase the number of parameters by two, thereby equating JM-SY to the population model in Equation 20.

JM-AM

Comparing JM-AM to the population joint distribution does not require the same analytic steps as JM-SY because the imputation model from Equation 7 does not condition on covariates. Rather, the JM-AM imputation model can be compared directly to the joint distribution given by Equations 15 and 16. The equivalence of JM-AM and the joint distribution is immediately apparent when considering that both models require 15 parameters: 3 grand means, 6 unique elements in the Level-2 covariance matrix, and 6 unique elements in the within-cluster covariance matrix.

FCS-VB

Because the FCS imputation models from Equations 10 and 11 have the same form (one incomplete variable predicted from all other variables), we consider only the Y_1 imputation model in this section. Equivalence (or lack thereof) between the FCS-VB imputation model for Y_1 and the population distribution would imply the same for Y_2.

Our evaluation of FCS-VB applies the same analytic steps as the previous JM-SY section. For brevity, we omit intermediate computations and show the population joint distribution expressed in terms of the mixed model parameters. Full computational details are available upon request. The marginal distribution of the predictor variables is

$$\begin{bmatrix} X_{ij} \\ Y_{2ij} \end{bmatrix} \sim N_2\left(\begin{bmatrix} \bar{X} \\ \bar{Y}_2 \end{bmatrix}, \begin{bmatrix} \sigma^2_{X(\mathrm{W})} + \sigma^2_{X(\mathrm{B})} & \sigma_{XY_2(\mathrm{W})} + \sigma_{XY_2(\mathrm{B})} \\ \sigma_{Y_2X(\mathrm{W})} + \sigma_{Y_2X(\mathrm{B})} & \sigma^2_{Y_2(\mathrm{W})} + \sigma^2_{Y_2(\mathrm{B})} \end{bmatrix}\right), \tag{24}$$

and the conditional distribution of Y_1 given X and Y_2 is

$$Y_{1ij} \sim N\left(\begin{matrix} \upsilon_{Y_1} + X_{ij}\gamma_{Y_1|X(\mathrm{W})} + \bar{X}_j\left(\gamma_{Y_1|X(\mathrm{B})} - \gamma_{Y_1|X(\mathrm{W})}\right) + \\ Y_{2ij}\gamma_{Y_1|Y_2(\mathrm{W})} + \bar{Y}_{2j}\left(\gamma_{Y_1|Y_2(\mathrm{B})} - \gamma_{Y_1|Y_2(\mathrm{W})}\right) + u_{1j} \end{matrix}, \Sigma_e\right), \tag{25}$$

where υ_{Y_1} is the intercept, $\gamma_{Y_1|X(\mathrm{W})}$ and $\gamma_{Y_1|Y_2(\mathrm{W})}$ are pure within-cluster regression slopes of Y_1 on X and Y_2, and $\gamma_{Y_1|X(\mathrm{B})}$ and $\gamma_{Y_1|Y_2(\mathrm{B})}$ are the corresponding between-cluster coefficients from the regression of Y_1 on the cluster means. As before, the marginal distribution from Equation 24 applies to the imputation model as well. Finally, population within- and between-cluster residual variances are as follows:

$$\begin{aligned} \sigma^2_e &= \sigma^2_{Y_1(\mathrm{W})} - \sigma^2_{X(\mathrm{W})}\gamma^2_{Y_1|X(\mathrm{W})} - \sigma^2_{Y_2(\mathrm{W})}\gamma^2_{Y_1|Y_2(\mathrm{W})} - 2\sigma_{Y_2X(\mathrm{W})}\gamma_{Y_1|X(\mathrm{W})}\gamma_{Y_1|Y_2(\mathrm{W})} \\ \sigma^2_u &= \sigma^2_{Y_1(\mathrm{B})} - \sigma^2_{X(\mathrm{B})}\gamma^2_{Y_1|X(\mathrm{B})} - \sigma^2_{Y_2(\mathrm{B})}\gamma^2_{Y_1|Y_2(\mathrm{B})} - 2\sigma_{Y_2X(\mathrm{B})}\gamma_{Y_1|X(\mathrm{B})}\gamma_{Y_1|Y_2(\mathrm{B})}. \end{aligned} \tag{26}$$

Again, we switch notation systems here to differentiate the population and imputation models.

Consistent with the procedure for JM-SY, comparing the number of parameters required by the FCS-VB model in Equation 10 reveals that FCS-VB is more restrictive than the population model because it includes a total of 13 parameters (1 intercept, 2 regression coefficients, a single Level-2 variance, a single Level-1 variance, and 8 parameters in the marginal distribution of X and Y_2). A comparison of Equations 10 and 25 shows that the population distribution contains distinct within- and between-cluster regression coefficients for each predictor, whereas the FCS-VB uses a single slope to preserve the relation between X and Y_1 and Y_2 and Y_1. As such, FCS-VB is inappropriate for models that posit unique within- and between-cluster associations between pairs of incomplete variables and/or between pairs of incomplete and complete variables and thus is even more restrictive than JM-SY. Returning to the contextual effects analysis model from Equation 1, the analytic results suggest that FCS-VB should yield biased estimates of the regression slopes because it generates imputations from a model where the cluster means of X have no additional explanatory power above and beyond the Level-1 scores.

To equate FCS-VB to the population model, one of the two situations must occur. First, imposing equality constraints on the within- and between-cluster coefficients from the population model reduces the number of parameters by two, thereby equating the fixed effect portion of the two models, as follows:

$$\begin{aligned}\gamma_{Y_1|X(\mathrm{W})} = \gamma_{Y_1|X(\mathrm{B})} = \beta_{Y_1|X}\\ \gamma_{Y_1|Y_2(\mathrm{W})} = \gamma_{Y_1|Y_2(\mathrm{B})} = \beta_{Y_1|Y_2}.\end{aligned} \tag{27}$$

These constraints are useful for illustrating a population model for which FCS-VB would work well, but a realistic course of action is to adopt a richer imputation model that is capable of preserving a wider range of effects. The FCS-WCK modification to FCS achieves this goal.

FCS-WCK

Because the FCS-WCK imputation models from Equations 12 and 13 have the same form, we again restrict our attention to the Y_1 imputation model from Equation 12. Further, we can reuse the population expressions from the FCS-VB results, given in Equations 24 through 26. The equivalence of JM-WCK and the population joint distribution is immediately apparent when considering that both models require 15 parameters: 1 intercept, 4 regression coefficients, a single Level-2 variance, a single Level-1 variance, and 8 parameters in the marginal distribution of X and Y_2. Returning to the contextual effects analysis model from Equation 1, the analytic results suggest that FCS-WCK should yield accurate estimates of all regression model parameters because it generates replacement values from an imputation model that is more general (i.e., has more parameters) than the analysis model. Finally, we can infer that JM-AM and FCS-WCK are themselves asymptotically equivalent in the case of normally distributed continuous variables because both procedures exactly reproduce the population joint distribution. As such, these procedures are appropriate for a range of analysis models that posit level-specific relations (e.g., multilevel models with contextual effects, multilevel structural equation models).

Computer Simulation

For the simulation study, we generated the data under a population model where correlations among the Level-2 residuals differ from the corresponding within-cluster associations, as this is the situation that differentiates the four imputation approaches. Note that it was not our goal to provide a comprehensive simulation that investigates the performance of multilevel imputation techniques. Rather, the goal was to perform a focused set of simulations that illustrated and tested the propositions derived from the analytic work.

For consistency with the previous sections, we used Equations 14 through 16 as the data-generating model, where Y_1 and Y_2 were incomplete and X was complete. Consistent with Enders, Mistler, and Keller (2016), we did not vary the missing data rate, as this factor has a predictable effect on bias (e.g., as the missing data rate increases, so too does nonresponse bias). Rather, we imposed a constant rate of 20% missingness on Y_1 and Y_2, as this value should be large

TABLE 1.
Within- and Between-Cluster Covariance Matrices for Simulation Study

	ICC = .10			ICC = .50		
Variable	M_1	M_2	C	M_1	M_2	C
			Level 1			
M_1	18.00	**0.50**	**0.00**	18.00	**0.50**	**0.00**
M_2	11.02	27.00	**0.40**	11.02	27.00	**0.40**
C	0.00	6.24	9.00	0.00	6.24	9.00
			Level 2			
M_1	2.00	**−0.50**	**0.00**	18.00	**−0.50**	**0.00**
M_2	−1.22	3.00	**−0.40**	−11.02	27.00	**−0.40**
C	0.00	−0.69	1.00	0.00	−6.24	9.00

Note. Variances and covariances are in the lower diagonal, and correlations in the upper diagonal are in boldface. ICC = intraclass correlation.

enough to confirm the analytic results without harming generalizability. Although an MCAR mechanism may not be a realistic assumption for applied practice, we use it here to demonstrate that certain imputation methods may introduce bias in situations where dropout is completely benign.

The simulation consisted of four manipulated factors: imputation method (JM-SY, JM-AM, FCS-VB, and FCS-WCK), number of clusters (30 and 100), number of observations per cluster (5 and 30), and the intraclass correlation (ICC; .10 and .50). Table 1 gives the population correlation and covariance matrices. Three features of the population model are worth highlighting. First, to introduce level-specific relations, we specified Level-2 correlations that were equal in magnitude but opposite in sign to the Level-1 associations. Second, to manipulate the ICC, we held the Level-1 covariance matrix constant and varied the Level-2 variances. Finally, we set the correlation between Y_1 and X to zero to isolate the influence of imputation on the relation between Y_2 and X (e.g., imputation cannot exploit the relation between Y_1 and X to preserve the association between Y_2 and X). Data were generated in SAS/IML 13.2 by first generating two matrices of random normal variables from a multivariate normal distribution with the desired covariance structure, one representing the between-cluster deviations and the other representing within-cluster deviations. Next, the three variables were computed by summing their respective within- and between-cluster components. An MCAR mechanism was imposed by randomly deleting 20% of the Y_1 and Y_2 values in each data set.

The JM approaches were implemented using Mplus (Muthén & Muthén, 1998–2012), and the FCS methods were implemented with the Blimp application (Enders, Keller, & Levy, 2016; Keller & Enders, 2015). We chose these software packages because they are flexible enough to accommodate the range of methods we are studying, and because JM-AM and FCS-WCK are exclusive to Mplus and Blimp, respectively. As noted previously, JM-SY is also implemented in the PAN and MLMMM packages in R (Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008) as well as in MLwiN and Stata (Carpenter et al., 2011), and FCS-VB is available in MICE (van Buuren et al., 2014). It is important to note that the choice of software package is arbitrary in this case because the MCMC sampling steps for incomplete Level-1 variables are common across platforms. After examining convergence diagnostics, we created 20 imputations with 1,000 burn-in iterations and 500 between-imputation iterations. We performed 1,000 replications in each of the between-group design cells, and we used Mplus to estimate the within-cluster and between-cluster covariance matrices from Equation 16, and we used SAS/IML 13.2 to pool the estimates and manipulate the simulation results. The syntax for the simulations is available upon request.

A brief discussion of our analysis model is warranted. As explained previously, JM-AM uses random effects to model between-cluster associations, whereas FCS-WCK uses cluster means computed from the filled-in data. This distinction is analogous to using a latent versus manifest variable approach to estimating contextual effects models (Lüdke et al., 2011; Lüdke et al., 2008). The analytic work from the previous section shows that these approaches are asymptotically equivalent, but the use of cluster means as covariates is known to produce bias in some situations due to measurement and/or sampling error (Lüdke et al., 2011; Lüdke et al., 2008). We chose Equation 15 as an analysis model because it is identical to the JM-AM imputation model (i.e., it uses random effects to represent the between-cluster associations) and thus provides a "gold standard" against which to compare other methods. In particular, this equality allows us to examine whether the use of cluster means in FCS-WCK model has a detrimental impact on finite samples. Although contextual effects analyses similar to Equation 1 are common in the applied literature, we felt that employing such an analysis model would provide a less precise comparison of JM-AM to FCS-WCK. Nevertheless, there is no reason to believe that insights gleaned from estimating covariance matrices would differ from those of regression models.

It is well known that complete data estimation routines can produce biased estimates of certain parameters in multilevel models (Lüdke et al., 2011; Lüdke et al., 2008). To isolate the impact of imputation on the quality of the estimates, we computed the average complete data estimate (i.e., the estimate based on the full data prior to deletion) in each design cell and used these quantities as the true values. We then calculated raw bias as the difference between the imputation estimates and the average complete data estimates. To facilitate interpretation of the simulation results, we standardized bias by dividing raw bias by the standard deviation of the complete data estimates within each design cell. This standardized metric expresses bias in standard error units (e.g., .40 suggests that the average imputation estimate differs from the average complete data estimate by

TABLE 2.
Raw and Standardized Bias for JM-SY

	ICC = .10				ICC = .50			
	Level 1		Level 2		Level 1		Level 2	
Parameter	Raw	Std.	Raw	Std.	Raw	Std.	Raw	Std.
	$J = 30$, $n_j = 5$				$J = 30$, $n_j = 5$			
Cov. Y_1–Y_2	−.006	−.002	.015	.013	−.036	−.014	−.077	−.015
Cov. Y_1–X	.062	.053	−.067	−.097	−.069	−.053	.057	.020
Cov. Y_2–X	−.083	−.053	.092	.115	−.341	−.203	.290	.082
Mean Y_1	NA	NA	−.004	−.009	NA	NA	−.005	−.006
Mean Y_2	NA	NA	−.008	−.015	NA	NA	.002	.002
	$J = 30$, $n_j = 30$				$J = 30$, $n_j = 30$			
Cov. Y_1–Y_2	.010	.011	.000	.001	−.033	−.034	−.012	−.003
Cov. Y_1–X	−.007	−.015	.010	.030	−.011	−.025	.029	.012
Cov. Y_2–X	−.045	−.074	.031	.075	−.051	−.082	.040	.012
Mean Y_1	NA	NA	.000	.000	NA	NA	.000	.001
Mean Y_2	NA	NA	.000	−.001	NA	NA	.000	.000
	$J = 100$, $n_j = 5$				$J = 100$, $n_j = 5$			
Cov. Y_1–Y_2	.014	.010	−.004	−.006	−.044	−.031	−.006	−.002
Cov. Y_1–X	.047	.070	−.038	−.093	−.082	−.114	.060	.038
Cov. Y_2–X	−.129	−.154	.114	.248	−.353	−.365	.355	.179
Mean Y_1	NA	NA	.001	.005	NA	NA	.000	.000
Mean Y_2	NA	NA	.006	.021	NA	NA	.003	.006
	$J = 100$, $n_j = 30$				$J = 100$, $n_j = 30$			
Cov. Y_1–Y_2	−.005	−.009	.002	.005	−.014	−.026	.014	.005
Cov. Y_1–X	−.007	−.028	.009	.048	−.031	−.117	.023	.017
Cov. Y_2–X	−.036	−.106	.038	.160	−.055	−.166	.052	.030
Mean Y_1	NA	NA	−.002	−.009	NA	NA	.000	.000
Mean Y_2	NA	NA	.000	.001	NA	NA	.000	.001

Note. Values of .000 occur due to rounding; bias values are <.001 in absolute value. ICC = intraclass correlation; Std. = standard; JM-SY = joint model Schafer and Yucel; Cov. = covariance; NA = not applicable.

four tenths of a standard error). We used graphical and tabular displays of standardized bias to identify the salient effects reported below.

Simulation Results

Tables 2 through 5 give the raw and standardized bias values for each imputation approach broken down by ICC and sample size conditions, and Figure 1 displays the standardized bias values from the tables as a trellis plot. The analytic work suggests that the four imputation methods differ in their

TABLE 3.
Raw and Standardized Bias for JM-AM

	ICC = .10				ICC = .50			
	Level 1		Level 2		Level 1		Level 2	
Parameter	Raw	Std.	Raw	Std.	Raw	Std.	Raw	Std.
	$J = 30$, $n_j = 5$				$J = 30$, $n_j = 5$			
Cov. Y_1–Y_2	−.017	−.007	.027	.023	−.021	−.008	−.057	−.011
Cov. Y_1–X	.035	.030	−.037	−.054	.012	.009	−.016	−.006
Cov. Y_2–X	−.043	−.028	.025	.031	−.046	−.028	−.033	−.009
Mean Y_1	NA	NA	.003	.006	NA	NA	.011	.013
Mean Y_2	NA	NA	−.010	−.018	NA	NA	.002	.002
	$J = 30$, $n_j = 30$				$J = 30$, $n_j = 30$			
Cov. Y_1–Y_2	.013	.014	.001	.001	−.028	−.029	−.017	−.004
Cov. Y_1–X	−.001	−.002	.002	.006	.010	.021	.000	.000
Cov. Y_2–X	−.007	−.011	−.003	−.006	.001	.002	−.010	−.003
Mean Y_1	NA	NA	−.002	−.005	NA	NA	.001	.002
Mean Y_2	NA	NA	−.002	−.006	NA	NA	−.001	−.001
	$J = 100$, $n_j = 5$				$J = 100$, $n_j = 5$			
Cov. Y_1–Y_2	−.064	−.047	.042	.063	−.060	−.043	−.009	−.003
Cov. Y_1–X	.004	.006	−.001	−.004	−.017	−.024	−.011	−.007
Cov. Y_2–X	−.011	−.013	.000	.000	.012	.012	−.019	−.009
Mean Y_1	NA	NA	.003	.010	NA	NA	.006	.014
Mean Y_2	NA	NA	−.002	−.006	NA	NA	−.004	−.006
	$J = 100$, $n_j = 30$				$J = 100$, $n_j = 30$			
Cov. Y_1–Y_2	−.002	−.004	−.004	−.012	.005	.009	.000	.000
Cov. Y_1–X	.001	.002	−.002	−.012	−.007	−.028	.000	.000
Cov. Y_2–X	.004	.012	−.002	−.010	−.002	−.006	−.002	−.001
Mean Y_1	NA	NA	−.002	−.010	NA	NA	.001	.002
Mean Y_2	NA	NA	.001	.005	NA	NA	.000	.000

Note. Values of .000 occur due to rounding; bias values are <.001 in absolute value. ICC = intraclass correlation; Std. = standard; JM-AM = joint model Asparouhov and Muthén; Cov. = covariance; NA = not applicable.

ability to preserve unique within- and between-cluster covariances. JM-AM and FCS-WCK should yield the most accurate estimates because these methods place no restrictions on the covariance matrix elements. On the other side of the spectrum, FCS-VB should exhibit the greatest bias because it places restrictions on relations between pairs of complete and incomplete variables as well as pairs of incomplete variables. JM-SY should be somewhat better than FCS-VB because it places restrictions only on relations between pairs of complete and incomplete variables (recall the previous caveat that JM-SY would be

TABLE 4.
Raw and Standardized Bias for FCS-VB

	ICC = .10				ICC = .50			
	Level 1		Level 2		Level 1		Level 2	
Parameter	Raw	Std.	Raw	Std.	Raw	Std.	Raw	Std.
	$J = 30$, $n_j = 5$				$J = 30$, $n_j = 5$			
Cov. Y_1–Y_2	−0.446	−0.183	0.494	0.428	−1.332	−0.504	1.258	0.249
Cov. Y_1–X	0.111	0.094	−0.109	−0.158	0.118	0.090	−0.135	−0.048
Cov. Y_2–X	−0.135	−0.087	0.124	0.154	−0.324	−0.193	0.263	0.075
Mean Y_1	NA	NA	−0.002	−0.004	NA	NA	0.001	0.001
Mean Y_2	NA	NA	−0.014	−0.025	NA	NA	−0.001	−0.001
	$J = 30$, $n_j = 30$				$J = 30$, $n_j = 30$			
Cov. Y_1–Y_2	−0.104	−0.110	0.133	0.214	−0.154	−0.157	0.137	0.030
Cov. Y_1–X	0.017	0.035	−0.017	−0.051	0.030	0.064	−0.013	−0.005
Cov. Y_2–X	−0.037	−0.061	0.025	0.060	−0.032	−0.051	0.026	0.008
Mean Y_1	NA	NA	−0.001	−0.004	NA	NA	0.001	0.001
Mean Y_2	NA	NA	0.000	0.001	NA	NA	0.001	0.001
	$J = 100$, $n_j = 5$				$J = 100$, $n_j = 5$			
Cov. Y_1–Y_2	−0.494	−0.365	0.540	0.810	−1.324	−0.940	1.328	0.468
Cov. Y_1–X	0.095	0.141	−0.089	−0.215	0.102	0.142	−0.122	−0.078
Cov. Y_2–X	−0.116	−0.139	0.098	0.214	−0.285	−0.295	0.289	0.145
Mean Y_1	NA	NA	0.002	0.009	NA	NA	−0.001	−0.002
Mean Y_2	NA	NA	0.001	0.005	NA	NA	0.000	−0.001
	$J = 100$, $n_j = 30$				$J = 100$, $n_j = 30$			
Cov. Y_1–Y_2	−0.121	−0.230	0.138	0.412	−0.127	−0.241	0.160	0.065
Cov. Y_1–X	0.019	0.070	−0.018	−0.099	0.012	0.044	−0.019	−0.014
Cov. Y_2–X	−0.026	−0.075	0.027	0.113	−0.034	−0.101	0.029	0.017
Mean Y_1	NA	NA	−0.001	−0.008	NA	NA	0.001	0.002
Mean Y_2	NA	NA	0.001	0.004	NA	NA	0.001	0.001

Note. Values of .000 occur due to rounding; bias values are <.001 in absolute value. ICC = intraclass correlation; Std. = standard; FCS-VB = fully conditional specification van Buuren; Cov. = covariance; NA = not applicable.

equivalent to JM-AM and FCS-WCK if the complete variable cluster means are used as predictors).

The standardized bias results in Figure 1 confirm the analytic work. Specifically, notice that FCS-VB (denoted by a circle) produced the largest bias values across all conditions, followed by JM-SY (denoted by a plus). As expected, JM-AM and FCS-WCK produced little to no bias. Further, biases were in the expected direction. To illustrate, consider the association between Y_2 and X. Because JM-SY and FCS-VB impose equality constraints on functions of this relation, the Level-1 and Level-2 covariances should be biased toward a common

TABLE 5.
Raw and Standardized Bias for FCS-WCK

	ICC = .10				ICC = .50			
	Level 1		Level 2		Level 1		Level 2	
Parameter	Raw	Std.	Raw	Std.	Raw	Std.	Raw	Std.
	$J = 30$, $n_j = 5$				$J = 30$, $n_j = 5$			
Cov. Y_1–Y_2	−.009	−.004	.074	.064	−.005	−.002	−.164	−.032
Cov. Y_1–X	.023	.019	−.020	−.029	.007	.006	−.005	−.002
Cov. Y_2–X	−.040	−.025	.040	.050	−.037	−.022	−.029	−.008
Mean Y_1	NA	NA	−.005	−.010	NA	NA	.000	.001
Mean Y_2	NA	NA	−.010	−.018	NA	NA	−.001	−.001
	$J = 30$, $n_j = 30$				$J = 30$, $n_j = 30$			
Cov. Y_1–Y_2	.011	.012	−.007	−.011	−.030	−.030	−.034	−.007
Cov. Y_1–X	.000	.000	.002	.007	.008	.017	.003	.001
Cov. Y_2–X	−.010	−.017	−.001	−.003	.004	.006	−.009	−.003
Mean Y_1	NA	NA	−.002	−.006	NA	NA	.000	.000
Mean Y_2	NA	NA	.000	.001	NA	NA	.001	.001
	$J = 100$, $n_j = 5$				$J = 100$, $n_j = 5$			
Cov. Y_1–Y_2	−.033	−.024	.051	.076	−.057	−.041	−.077	−.027
Cov. Y_1–X	−.003	−.004	.007	.016	−.011	−.015	−.013	−.008
Cov. Y_2–X	−.017	−.020	.005	.012	.011	.011	−.017	−.009
Mean Y_1	NA	NA	.000	.001	NA	NA	.000	.000
Mean Y_2	NA	NA	.001	.005	NA	NA	−.003	−.005
	$J = 100$, $n_j = 30$				$J = 100$, $n_j = 30$			
Cov. Y_1–Y_2	.000	.000	−.011	−.033	−.006	−.012	−.014	−.006
Cov. Y_1–X	.003	.013	−.002	−.010	−.005	−.021	.000	.000
Cov. Y_2–X	.003	.010	−.002	−.009	−.002	−.005	−.002	−.001
Mean Y_1	NA	NA	−.001	−.006	NA	NA	.001	.001
Mean Y_2	NA	NA	.001	.007	NA	NA	.000	.000

Note. Values of .000 occur due to rounding; bias values are <.001 in absolute value. ICC = intraclass correlation; Std. = standard; FCS-WCK = fully conditional specification Ian White and Carpenter and Kenward; Cov. = covariance; NA = not applicable.

value. Given the configuration of population values (see Table 1), we expected negative bias for the Level-1 covariance and positive bias for the Level-2 covariance. The bias values in Figure 1 confirm these predictions. To better illustrate this bias, Figure 2 shows the Level-2 covariance between Y_2 and X by imputation method and number of observations per cluster (the salient factors that influenced bias), averaging across other design features. The figure shows that JM-SY and FCS-VB produced nearly identical estimates, with the largest bias values occurring in design cells with five observations per cluster.

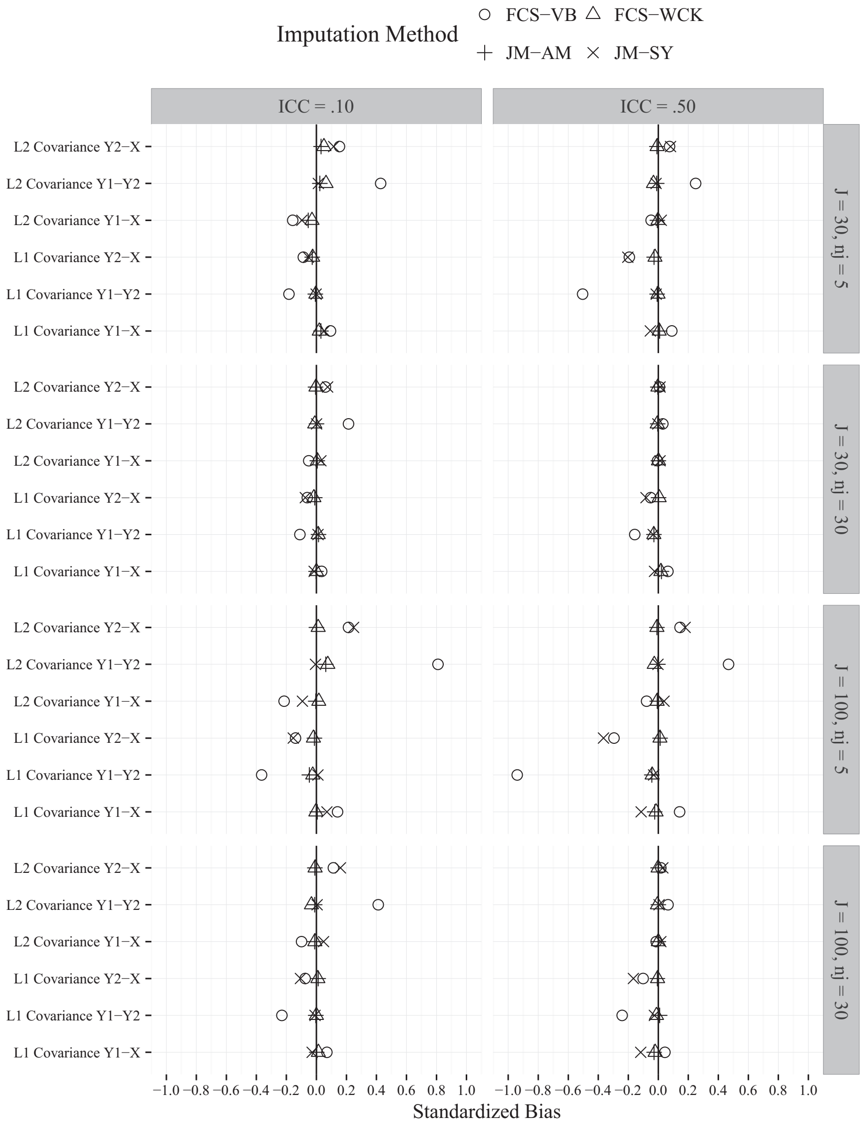

FIGURE 1. *Standardized bias means for the Level-1 and Level-2 covariances by imputation method, intraclass correlation, number of clusters, and number of observations per cluster, averaging across other design cells.*

Next, consider the covariance between Y_1 and Y_2. For this relation, the analytic work suggests that only FCS-VB should produce biased estimates. Given our configuration of population values (see Table 1), the estimates should again be

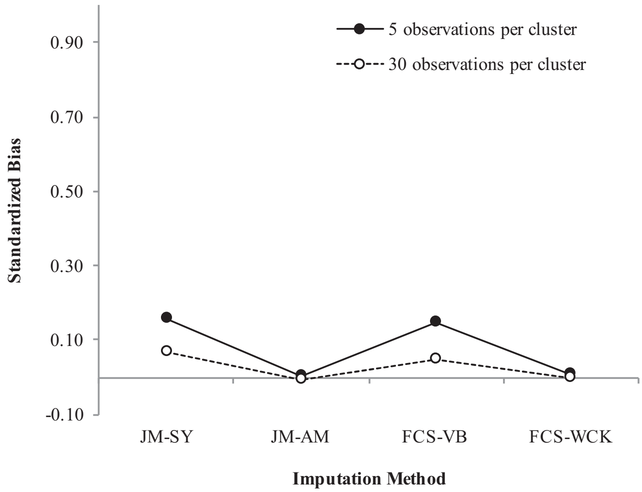

FIGURE 2. *Standardized bias means for the Level-2 covariance between* Y_2 *and* X *by imputation method and number of observations per cluster, averaging across other design cells. It shows that Asparouhov and Muthén's joint modeling approach (FCS-AM) and White, Carpenter, and Kenward's fully conditional specification approach (FCS-WCK) estimates were relatively free of bias, whereas Schafer and Yucel's joint modeling approach (JM-SY) and van Buuren's fully conditional specification approach (FCS-VB) produced positively biased estimates. Additionally, cluster size moderated this bias, such that* $n_j = 5$ *exacerbated bias relative to* $n_j = 30$*. Note that the symbols on the graph are connected by lines to enhance readability, but these lines should not be interpreted as continuous trends because the horizontal axis depicts nominal categories.*

biased toward a common value, with negative bias at Level 1 and positive bias at Level 2. Returning to Figure 1, FCS-VB consistently produced rather large bias values. To better illustrate the bias, Figure 3 shows mean standardized bias for the Level-1 covariance between Y_1 and Y_2 by imputation method, ICC, and number of observations per cluster, averaging across the number of clusters. As seen in the figure, the ICC had little impact on bias when the within-cluster sample size was large ($n_j = 30$), but a large ICC and small cluster size (ICC = .50, $n_j = 5$) combined to produce substantial bias. As seen in Figure 4, the same three-way effect appeared with the Level-2 association between Y_1 and Y_2, although the bias was in the opposite direction and was more pronounced in the low ICC and small cluster size conditions.

Finally, consider the differences between JM-AM and FCS-WCK. JM-AM uses random effects to model between-cluster associations. In contrast, FCS-WCK uses cluster means computed from the filled-in data, a method that is known to produce bias in some situations due to measurement and/or sampling error (Lüdke et al., 2011; Lüdke et al., 2008). Because the analysis model in

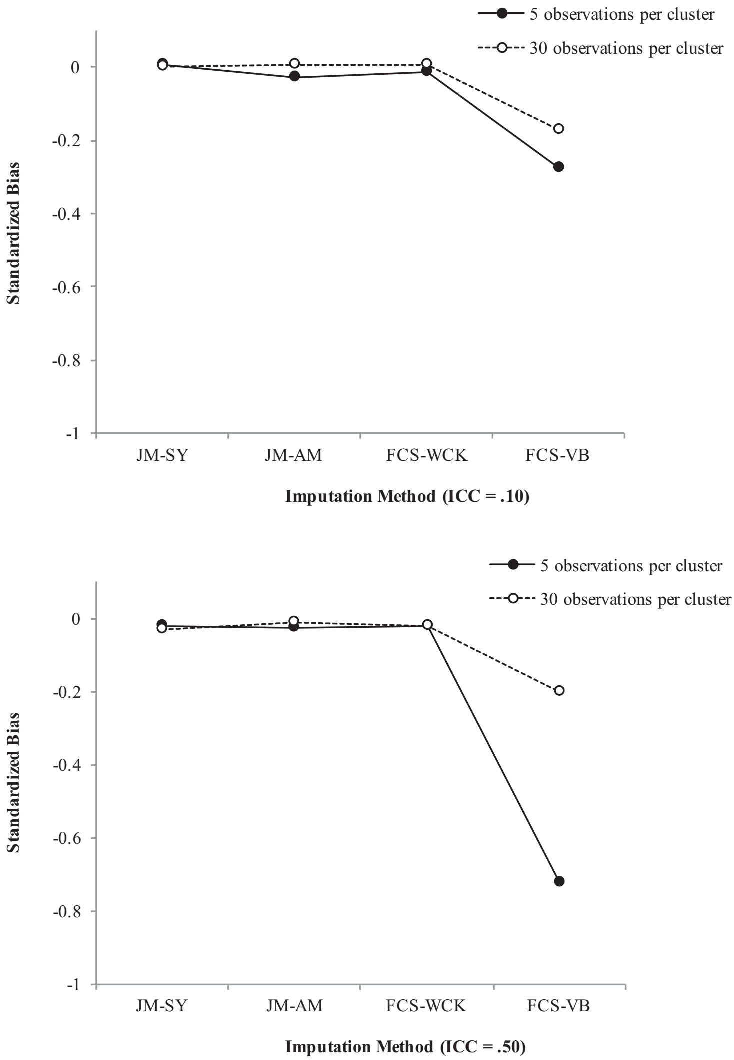

FIGURE 3. *Standardized bias means for the Level-1 covariance between* Y_1 *and* Y_2 *by method, intraclass correlation (ICC), and number of observations per cluster, averaging across the number of clusters. It shows that joint model Schafer and Yucel (JM-SY), joint model Asparouhov and Muthén, and fully conditional specification Ian White and Carpenter and Kenward estimates were relatively free of bias, whereas JM-SY produced negatively biased estimates. Additionally, the graph shows that the negative impact of a small cluster size (*$n_j = 5$*) was exacerbated by a large ICC of .50. Note that the symbols on the graph are connected by lines to enhance readability, but these lines should not be interpreted as continuous trends because the horizontal axis depicts nominal categories.*

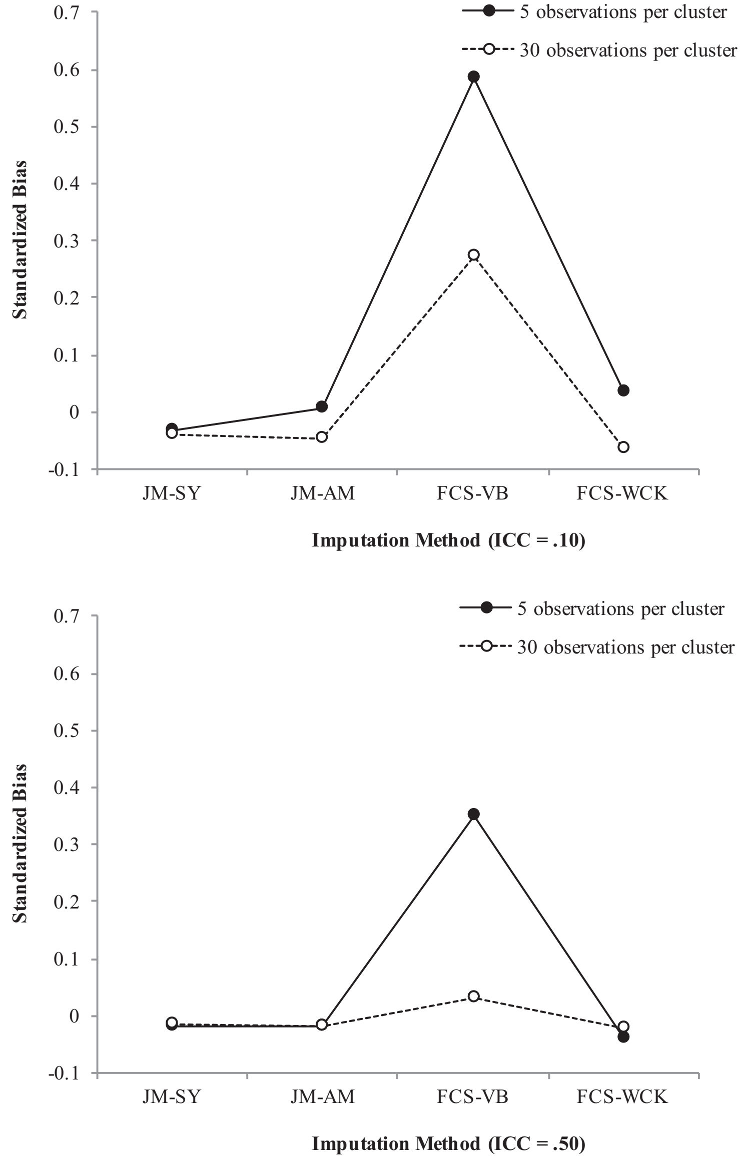

FIGURE 4. *Standardized bias means for the Level-2 covariance between* Y_1 *and* Y_2 *by method, intraclass correlation (ICC), and number of observations per cluster, averaging across the number of clusters. It shows that joint model Schafer and Yucel (JM-SY), joint model Asparouhov and Muthén, and fully conditional specification Ian White and Carpenter and Kenward estimates were relatively free of bias, whereas JM-SY produced negatively biased estimates. Additionally, the graph shows that the negative impact of a small cluster size (*$n_j = 5$*) was exacerbated by a large ICC of .50. Note that the symbols on the graph are connected by lines to enhance readability, but these lines should not be interpreted as continuous trends because the horizontal axis depicts nominal categories.*

Equation 15 is identical to the JM-AM imputation model, we can confidently attribute any differences between JM-AM and FCS-WCK to the latter method's use of (potentially fallible) estimates of the cluster means. Interestingly, the bias values in Tables 3 and 5 are virtually identical, suggesting that the use of cluster means has no detrimental effect on estimates, at least in the conditions that we examined here.

Data Analysis Example

To illustrate the application of the four imputation approaches, we fit a contextual effects analysis model to the high school and beyond data. The analysis example is similar to that of Raudenbush and Bryk (2002), where the influence of SES on math achievement differed at Level 1 and Level 2. The analysis model is as follows:

$$\text{MathAch}_{ij} = \beta_0 + \beta_1(\text{Female}_{ij}) + \beta_2(\text{SES}_{ij}) + \beta_3(\text{MeanSES}_j) + b_j + \varepsilon_{ij}. \tag{28}$$

This analysis is useful because the influence of cluster-average SES is a stronger predictor of achievement than individual SES (i.e., β_3 is a larger positive value than β_2). The previous analytic work and simulation results suggest that the four imputation methods differ in their capacity to partition unique relations at Level 1 and Level 2. To demonstrate this effect, we deleted achievement and SES scores according to an MAR mechanism. Specifically, females had a 25% and 15% probability of missing SES and achievement scores, respectively, whereas males had a 15% and 25% probability of missing these two variables. The gender dummy code was complete.

We used the Mplus software package (Muthén & Muthén, 1998–2012) to implement JM-AM and JM-SY, and we used the Blimp application (Enders, Keller, & Levy, 2016) to implement FCS-VB and FCS-WCK. As noted previously, the version of JM-SY that we present in Equation 6 can be modified to include the means of complete predictor variables such as gender. Although the analysis model in Equation 28 does not partition the gender influence into unique within- and between-cluster components, we nevertheless include this variant of JM-SY in the example, as it should be equivalent to JM-AM and FCS-WCK. Consistent with current recommendations from the literature, we generated 20 imputations (Graham et al., 2007), and we did so by specifying a thinning interval of 1,000 iterations (i.e., we saved a data set for analysis after every 1,000 MCMC cycles). Finally, we used Mplus to estimate the model and pool parameter estimates and standard errors. The Appendix, available in the online version of the journal, gives the imputation and analysis scripts for the analysis example,[1] and all files are available from the second author upon request.

Table 6 gives the pooled parameter estimates and standard errors from each imputation approach. As seen in the table, four of the imputation approaches produced nearly identical estimates: FCS-WCK, JM-AM, and the two variants of

TABLE 6.
*Pooled Parameter Estimates and Standard Errors (*SE*s) From the Real Data Analysis Example*

	FCS-VB		FCS-WCK		JM-AM		JM-SY[a]		JM-SY[b]	
Parameter	Est.	*SE*	Est.	*SE*	Est.	*SE*	Est.	*SE*	Est.	*SE*
Intercept	13.34	0.18	13.35	0.19	13.35	0.19	13.39	0.19	13.37	0.18
Female	−1.21	0.18	−1.19	0.19	−1.18	0.19	−1.23	0.18	−1.21	0.19
SES	2.09	0.14	2.02	0.14	2.04	0.14	2.03	0.14	2.02	0.14
SES Means	3.53	0.37	3.79	0.36	3.75	0.38	3.75	0.37	3.78	0.36
Intercept Var.	2.72	0.46	2.41	0.43	2.55	0.46	2.44	0.44	2.40	0.43
Residual Var.	37.00	0.74	37.30	0.75	37.13	0.76	37.16	0.75	37.15	0.74

Note. JM-SY[a] is the standard approach outlined in the article. JM-SY[b] adds the cluster means of the complete gender variable to the model, following the logic of FCS-WCK. SES = socioeconomic status; FCS-WCK = fully conditional specification Ian White and Carpenter and Kenward; FCS-VB = fully conditional specification van Buuren; JM-SY = joint model Schafer and Yucel; JM-AM = joint model Asparouhov and Muthén.

JM-SY. The analytic and simulation results predict this result, as these methods partition the SES achievement relation into unique within- and between-cluster components. It is interesting to note that JM-AM treats the gender dummy code as a normally distributed variable, whereas FCS-WCK and JM-SY do not apply distributional assumptions to this complete variable. This seemingly egregious violation of normality appeared to have no impact on the analysis results, however. Recall from the analytic work and simulation results that FCS-VB generates imputations from a parsimonious model that does not partition associations into within- and between-cluster components. In the context of the high school and beyond example, FCS-VB assumes that the β_3 coefficient in Equation 28 equals 0, such that the SES cluster means have no predictive power above and beyond the individual scores. The effect of applying a restrictive imputation model is clearly evident in Table 6, where the within-cluster regression is too large and the between-cluster regression is too small (i.e., the two regressions are biased toward a common slope). To put the bias in practical terms, the FCS-VB estimate of β_3 (cluster-level mean SES) differs from the other methods by roughly two thirds of a standard error unit, and the estimate of β_2 (individual SES) differs by about one half of a standard error. The magnitude of these differences is probably large enough to distort inferences and confidence interval coverage in many applied settings.

Discussion

In the single-level context, JM and FCS have been shown to be equivalent with multivariate normal data (Hughes et al., 2014), but less is known about the similarities and differences of these two approaches with multilevel data. Thus, the purpose of this study was to examine the situations under which JM and FCS reproduce (or preserve) the mean and covariance structure of a population random intercept model with multivariate normal data. Our analytic work showed that JM-AM and FCS-WCK were the only methods that reproduced the covariance structure of the population model, and the results also suggested that these methods are asymptotically equivalent. In contrast, JM-SY imposes implicit restrictions on covariance parameters involving pairs of incomplete and complete variables, and FCS-VB imposes implicit restrictions on all covariance parameters involving the incomplete variables. Computer simulation results verified the analytic work, further revealing that differences among the methods were most evident with small within-cluster sample sizes (e.g., $n_j = 5$).

It is important to emphasize that the biases predicted by the analytic work (and subsequently confirmed by the simulation studies) can occur under any mechanism. Although multiple imputation is usually described as an MAR-based approach, it is widely known that multiple imputation estimates are biased when the imputation model is more restrictive than a particular analysis model, in which case the two models are said to be uncongenial (Meng, 1994; Schafer, 2003). Our results show that JM-AM and FCS-WCK employ unrestrictive imputation models where the Level-1 and Level-2 covariance matrices are saturated, whereas the JM-SY and FCS-VB models place implicit constraints on the within- and between-cluster covariance matrices. When these constraints are incompatible with the analysis model, imputation can introduce bias, even under a benign MCAR mechanism. The data analysis example suggests that this bias will also be evident with an MAR mechanism, as would be expected.

The differences between JM and FCS approaches are particularly salient for multilevel data sets where relations among lower level variables differ at Level 1 and Level 2. One such example is the classic contextual effects model that partitions the association between a pair of Level 1 variables into between- and within-cluster components (Longford, 1989; Lüdke et al., 2011; Lüdke et al., 2008; Raudenbush & Bryk, 2002; Shin & Raudenbush, 2010), and multilevel structural equation models that place different restrictions on the within- and between-cluster covariance matrices (e.g., a multilevel factor analysis with different latent variables at Level 1 and Level 2) are a second common example (Dunn et al., 2015; Huang & Cornell, 2015; Muthén, 1991; Reise et al., 2005; Toland & De Ayala, 2005). In applied settings, researchers may not have strong a priori predictions about the structure of the within- and between-cluster covariance matrices. For example, a researcher using multilevel factor analysis may examine a number of different factor structures at Level 1 and Level 2 and choose the solution that best fits the data (e.g., Dedrick & Greenbaum, 2011; Roesch et al., 2010). Taken as a whole, results suggest that researchers should use

JM-AM or FCS-WCK because both employ very general models that are capable of preserving complicated multilevel data structures. Other than personal preference and software access, the current study provides no reason to prefer JM-AM to FCS-WCK (or vice versa) for random intercept analyses, but it is worth noting that JM-AM cannot preserve random slope variation, whereas FCS-WCK can readily accommodate random associations (Enders, Keller, & Levy, 2016; Enders, Mistler, & Keller, 2016). Although not examined in this article, FCS-WCK may provide better estimates of Level 2 variances in some conditions. If this is the case, JM-AM could also provide better results for some models (e.g., multilevel confirmatory factor analysis). We also point out that the original JM formulation was restricted to multivariate normal data, but the approach has since been extended to handle incomplete categorical variables[2] via a probit regression formulation (e.g., Asparouhov & Muthén, 2010; Goldstein et al., 2009; Enders, Keller, & Levy, 2016; Enders, Mistler, & Keller, 2016). The FCS-WCK method in the Blimp application accommodates categorical variables using the same procedure (Enders, Keller, & Levy, 2016).

Finally, our study has a number of limitations that raise possibilities for future research. First, the simulation study examined a relatively limited set of design factors. Our goal for the simulation was to demonstrate and verify the analytic work, but future research could examine a wider set of conditions, including different cluster sizes, ICCs, covariance structures, and missing data mechanisms, to name a few. Note that the size of the contextual effect examined in this study was larger than would commonly be encountered. A smaller contextual effect would better reflect situations likely to be encountered in the real world and would lead to a smaller gap between the two methods that performed poorly in this study (JM-SY and FCS-VB) and the two methods that performed well in this study (JM-AM and FCS-WCK). Second, we focused exclusively on population and imputation models that contained random intercepts but no random slopes. Multilevel imputation routines differ in their ability to include random slopes (Enders, Mistler, & Keller, 2016); JM-AM does not allow for random slopes, JM-SY restricts random slopes to the complete covariates, and both FCS approaches can accommodate random slopes for any pair of variables. Future studies should extend our analytic and simulation work by examining population models where Level-1 associations vary across clusters. Third, we made no attempt to examine data structures with three or more levels. Data sets with three or more levels are fairly common (e.g., longitudinal studies where repeated measures are nested within individuals, and individuals are nested within groups), yet very little work exists on three-level imputation. Yucel (2008) proposed a JM approach to three-level imputation, and extending our analytic work to this context could facilitate the development of a three-level FCS imputation model. Finally, it might be useful to rerun the simulation study using an alternative analysis model (e.g., multilevel SEM), as this might make show differences between the imputation methods that were not anticipated by the current authors.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Institute of Educational Sciences award R305D150056, Craig K. Enders, PI.

Notes

1. The same analysis was fit to each set of imputations, but the Mplus analysis scripts varied slightly to accommodate the fact that variable order differed across methods. For brevity, we include only the joint model Asparouhov and Muthén (JM-AM) analysis script in the Appendix, available in the online version of the journal.
2. Categorical missing data handling via the probit model varies across software platform. For example, Mplus (JM-AM) can accommodate binary and ordinal variables, whereas Blimp (fully conditional specification-WCK) and MLwiN (joint model Schafer and Yucel [JM-SY]) can additionally handle incomplete nominal variables.

References

Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus*. Retrieved from http://www.statmodel.com/download/Imputations7.pdf

Browne, W. J., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, *15*, 391–420.

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.

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

Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental health agencies. *Journal of Emotional and Behavioral Disorders*, *19*, 27–40.

Dunn, E. C., Masyn, K. E., Jones, S. M., Subramanian, S. V., & Koenen, K. C. (2015). Measuring psychosocial environments using individual responses: An application of multilevel factor analysis to examing students in schools. *Prevention Science*, *16*, 718–733.

Enders, C. K., Keller, B. T., & Levy, R. (2016). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Manuscript submitted for publication*.

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.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Goldstein, H., Bonnet, G., & Rocher, T. (2007). Multilevel structural equation models for the analysis of comparative data on educational performance. *Journal of Educational and Behavioral Statistics*, *32*, 252–286.

Goldstein, H., Carpenter, J., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling*, *9*, 173–197.

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.

Harker, H., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement*, *15*, 177–199.

Huang, F. L., & Cornell, D. G. (2015) Using multilevel factor analysis with clustered data: Investigating the factor structure of the positive values scale. *Journal of Psychoeducational Assessment*, *34*, 3–14. doi:10.1177/0734282915570278

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., & Sterne, J. A. C. (2014). Joint modelling rationale for chained equations. *BMC Medical Research Methodology*, *14*, 28.

Kasim, R. M., & Raudenbush, S. W. (1998). Application of Gibbs sampling to nested variance components models with heterogeneous within-group variance. *Journal of Educational and Behavioral Statistics*, *32*, 93–116.

Keller, B. T., & Enders, C. K. (2015). *BLImP software manual* (Version beta 8.62). Los Angeles, CA. Retrieved from http://www.appliedmissingdata.com/multilevel-imputation.html

Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology*, *48*, 339–348.

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

Longford, N. (1989). Contextual effects and group means. *Multilevel Modelling Newsletter*, *1*, 5.

Lüdke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary Educational Psychology*, *30*, 263–285.

Lüdke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods*, 1*6*, 444–467.

Lüdke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 201–229.

Martin, A. J., Malmberg, L.-E., & Liem, G. A. D. (2010). Multilevel motivation and engagement: Assessing construct validity across students and schools. *Educational and Psychological Measurement*, *70*, 973–989.

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

Miller, A. D., & Murdock, T. B. (2007). Modeling latent true scores to determine the utility of aggregate student perceptions as classroom indicators in HLM: The case of classroom goal structures. *Contemporary Educational Psychology*, *32*, 83–104.

Mistler, S. A. (2013). *A SAS macro for applying multiple imputation to multilevel data*. Paper presented at the SAS Global Forum, San Francisco, CA.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*, 338–354.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*, 85–95.

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

Reise, S. P., Ventura, J., Neuchterlein, K. H., & Kim, K. H. (2005). An illustration of multilevel factor analysis. *Journal of Personality Assessment*, *84*, 126–136.

Roesch, S. C., Aldridge, A. A., Stocking, S. N., Villodas, F., Leung, Q., Bartley, C. E., & Black, L. J. (2010). Multilevel factor analysis and structural equation modeling of daily diary coping data: Modeling trait and state variation. *Multivariate Behavioral Research*, *45*, 767–789.

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

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

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall.

Schafer, J. L. (2001). Multiple imputation with PAN. In A. G. Sayer & L. M. Collins (Eds.), *New methods for the analysis of change* (pp. 357–377). Washington, D.C.: American Psychological Association.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, *57*, 19–35.

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

Shin, Y., & Raudenbush, S. W. (2007). Just-identified versus overidentified two-level hierarchical linear models with missing data. *Biometrics*, *63*, 1262–1268.

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.

Simons, J. S., Wills, T. A., & Neal, D. J. (2015). The many faces of affect: A multilevel model of drinking frequency/quantity and alcohol dependence symptoms among young adults. *Journal of Abnormal Psychology*, *123*, 676–694.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Toland, M. D., & De Ayala, R. J. (2005). A multilevel factor analysis of students' evaluations of teaching. *Educational and Psychological Measurement*, *65*, 272–296.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, *16*, 219–242.

van Buuren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 173–196). New York, NY: Routledge.

van Buuren, S. (2012). *Flexible imputation of missing data*. New York, NY: Chapman & Hall.

van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*, 1049–1064.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Yucel, R. M. (2008). Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *366*, 2389–2403.

Yucel, R. M. (2011). Random covariances and mixed-effects models for imputing multivariate multilevel continuous data. *Statistical Modelling*, *11*, 351–370.

Authors

STEPHEN A. MISTLER is an analytical consultant at SAS Institute, Cary, NC 27703; email: stephen.mistler@sas.com. His research interests are missing data handling and generalized linear mixed models.

CRAIG K. ENDERS is a professor in the Department of Psychology at UCLA, Box 951563, Los Angeles, CA 90095-1563; email: cenders@psych.ucla.edu. His research interests include missing data handling and multilevel modeling.

Manuscript received January 28, 2016
Revision received June 23, 2016
Accepted December 19, 2016