Multilevel Multiple Imputation:

An Examination of Competing Methods

by

Stephen Andrew Mistler

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2015 by the
Graduate Supervisory Committee:

Craig Enders, Chair
Leona Aiken
Roy Levy
Stephen G. West

ARIZONA STATE UNIVERSITY

May 2015

ABSTRACT

Missing data are common in psychology research and can lead to bias and reduced power if not properly handled. Multiple imputation is a state-of-the-art missing data method recommended by methodologists. Multiple imputation methods can generally be divided into two broad categories: joint model (JM) imputation and fully conditional specification (FCS) imputation. JM draws missing values simultaneously for all incomplete variables using a multivariate distribution (e.g., multivariate normal). FCS, on the other hand, imputes variables one at a time, drawing missing values from a series of univariate distributions. In the single-level context, these two approaches have been shown to be equivalent with multivariate normal data. However, less is known about the similarities and differences of these two approaches with multilevel data, and the methodological literature provides no insight into the situations under which the approaches would produce identical results. This document examined five multilevel multiple imputation approaches (three JM methods and two FCS methods) that have been proposed in the literature. An analytic section shows that only two of the methods (one JM method and one FCS method) used imputation models equivalent to a two-level joint population model that contained random intercepts and different associations across levels. The other three methods employed imputation models that differed from the population model primarily in their ability to preserve distinct level-1 and level-2 covariances. I verified the analytic work with computer simulations, and the simulation results also showed that imputation models that failed to preserve level-specific covariances produced biased estimates. The studies also highlighted conditions that exacerbated the amount of bias produced (e.g., bias was greater for conditions with small

cluster sizes). The analytic work and simulations lead to a number of practical recommendations for researchers.

For my wife, Amy.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES viii

CHAPTER

1 INTRODUCTION 1

Missing Data Mechanisms 3

Introduction to Multilevel Modeling 6

Single-Level Multiple Imputation 15

Multilevel Multiple Imputation 24

Goals 43

2 LITERATURE REVIEW 46

Multilevel Imputation of Continuous Variables 46

Conclusions from the Multilevel Imputation Literature 50

Multilevel Models and the Population Joint Distribution 52

3 METHODS 58

Imputation Methods 58

Research Questions and Hypotheses 69

Analytic Examination of JM and FCS Expectations 72

Simulation Studies 74

4 RESULTS 90

Analytic Examination Results 92

Simulation Results 130

CHAPTER Page

5 DISCUSSION 143

Summary of Findings 144

Integration with Existing Literature 147

Practical Recommendations 150

Limitations 155

Summary 162

REFERENCES 163

LIST OF TABLES

Table Page

1. Software Implementations of Multilevel Imputaion Methods 168
2. Number of Failed Iterations (out of 1000) for Each Method, by Condition..... 169
3. Standardized Corrected Bias for Level-1 Covariance Parameters for JM-BD, by Study and by Condition 170
4. Standardized Corrected Bias for Level-2 Covariance Parameters for JM-BD, by Study and by Condition 171
5. Standardized Corrected Bias for Mean Parameters for JM-BD, by Study and by Condition..... 172
6. Standardized Corrected Bias for Level-1 Covariance Parameters for JM-UN, by Study and by Condition 173
7. Standardized Corrected Bias for Level-2 Covariance Parameters for JM-UN, by Study and by Condition 174
8. Standardized Corrected Bias for Mean Parameters for JM-UN, by Study and by Condition..... 175
9. Standardized Corrected Bias for Level-1 Covariance Parameters for JM-Mplus, by Study and by Condition 176
10. Standardized Corrected Bias for Level-2 Covariance Parameters for JM-Mplus, by Study and by Condition 177
11. Standardized Corrected Bias for Mean Parameters for JM-Mplus, by Study and by Condition 178

Table Page

12. Standardized Corrected Bias for Level-1 Covariance Parameters for FCS-VB, by Study and by Condition 179

13. Standardized Corrected Bias for level-2 Covariance Parameters for FCS-VB, by Study and by Condition 180

14. Standardized Corrected Bias for Mean Parameters for FCS-VB, by Study and by Condition 181

15. Standardized Corrected Bias for Level-1 Covariance Parameters for FCS-CK, by Study and by Condition 182

16. Standardized Corrected Bias for Level-2 Covariance Parameters for FCS-CK, by Study and by Condition 183

17. Standardized Corrected Bias for Mean Parameters for FCS-CK, by Study and by Condition 184

LIST OF FIGURES

Figure Page

1. Standardized Bias for Level-1 Covariance Terms in Study 1 185

2. Standardized Bias for Level-1 Covariance Terms in Study 2 186

3. Standardized Bias for Mean Terms in Study 1 187

4. Standardized Bias for Mean Terms in Study 2 188

5. Standardized Bias for Level-2 Covariance Terms in Study 1 189

6. Standardized Bias for Level-2 Covariance Terms in Study 2 190

7. Standardized Corrected Bias Least Squared Means for Level-1 Covariance between *W* and *Y* in Study 1, by Method and Number of Observations per Cluster 191

8. Standardized Corrected Bias Least Squared Means for the Level-1 Covariance between *X* and *Y* in Study 1, by Method and Number of Observations per Cluster 192

9. Standardized Corrected Bias Least Squared Means for the Level-2 Covariance between *X* and *Y* in Study 1, by Method, ICC, and Number of Observations per Cluster 193

10. Standardized Corrected Bias Least Squared Means for the Level-2 Covariance between *X* and *Y* in Study 2, by Method and ICC 194

11. Standardized Corrected Bias Least Squared Means for the Level-2 Covariance between *X* and *Y* in Study 2, by Method and Number of Observations per Cluster 195

Chapter 1. Introduction

Missing data are common in psychology research and can lead to reduced power and bias if not properly handled. Multiple imputation is a state-of-the-art missing data method recommended by methodologists (Schafer & Graham, 2002). Although multiple imputation has advanced greatly in recent years, methods for handling missingness in multilevel data have received less attention. For example, implementations of multilevel imputation may assume multivariate normality, may not impute at level two, may not model random slopes between incomplete variables, etc. Each current implementation of multilevel imputation suffers from one or more of these deficiencies, though the exact constellation of deficiencies differs across implementations of multilevel imputation.

Multiple imputation methods can generally be divided into two broad categories: joint model (JM) imputation (Rubin & Schafer, 1990; Schafer, 1997) and fully conditional specification (FCS) imputation (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001; van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006). JM draws missing values simultaneously for all incomplete variables using a multivariate distribution (e.g., multivariate normal). FCS, on the other hand, imputes variables one at a time, drawing missing values from a series of univariate distributions. In the single-level context, these two approaches have been shown to be equivalent with multivariate normal (MVN) data (Hughes et al., 2014). However, less is known about the similarities and differences of these two approaches with multilevel data, and the methodological literature provides no insight into the situations under which the approaches would produce identical results.

Historically, the JM approach became the predominant method for single-level imputation of multivariate normal data beginning with Schafer's (1997) seminal book. FCS was developed later as a tool for dealing with mixtures of categorical and continuous variables (Raghunathan et al., 2001; van Buuren et al., 2006) – a situation that JM could not accommodate at the time (JM can now handle mixtures of categorical and continuous variables through the use of latent variables, e.g., Goldstein, 2011). Because the single-level JM approach usually (but not necessarily) implements a saturated model, it is able to preserve associations for a wide range of additive models. In the context of single-level MVN data, FCS possesses the same qualities, as it yields the same expectations as JM (Hughes et al., 2014). However, the same is not necessarily true of multilevel imputation methods.

Three variations of JM imputation have been proposed for multilevel data (Asparouhov & Muthén, 2010a, 2010f; Schafer, 2001; Schafer & Yucel, 2002). Although these methods share much in common, they possess subtle differences. To date, no methodological research has investigated the differences among the JM models, in particular their ability to produce imputations that preserve characteristics of the population distribution. Thus, one of the overarching goals for this project was to examine the situations under which the three JM methods reproduce (or preserve) the mean and covariance structure of a population random intercept model with multivariate normal data. To my knowledge, two FCS models have been proposed in the literature to date, only one of which is currently implemented in publicly available statistical software (Carpenter & Kenward, 2012, p. 221; van Buuren & Groothuis-Oudshoorn, 2011). Like JM, no methodological research has investigated the differences between the FCS models

or their ability to preserve a multilevel data structure. Thus, the second overarching goal for this project was to examine the situations under which FCS imputation reproduces the mean and covariance structure of a population random intercept model with multivariate normal data. The analytic work for these two goals also provided insight into the situations where the three JM and two FCS methods are equivalent.

The organization of the document is as follows. Chapter 1 provides background on missing data mechanisms, multilevel modeling, single-level imputation, and multilevel multiple imputation. Chapter 2 reviews the literature on multilevel imputation. Chapter 3 discusses the methods used in the dissertation studies. Chapter 4 presents the results from the dissertation studies. Finally, Chapter 5 discusses the findings and presents practical recommendations.

Missing Data Mechanisms

To demonstrate the advantages and limitations of modern methods for handling missing data, it is necessary to first explain the concept of missing data mechanisms. Missing data mechanisms describe the probability of missing data on a variable as a function of the missing (unobserved) values on the variable, the values of other measured variables, and the values of unknown (unmeasured) variables. The concept of missing data mechanisms is important in missing data research, as approaches to handling missing data assume a particular missing data mechanism. Rubin (1976) formalized the concept of missing data mechanisms by treating missing data indicators as variables and assigning distributions to these variables. Given a set of data, *Y*, one can form a new variable, *M*, that equals one whenever a value is missing and zero whenever the value is observed. Because *M* is a variable, it has a distribution that may or may not be related to

other variables. Below I define the missing data mechanisms based on descriptions of the distribution of *M* as a function of the observed data, Y_{obs}, and the unobserved or would-be values, Y_{mis}. The notation and terminology have changed since the seminal article by Rubin (1976). As such, I employ terminology and notation that are commonly used in missing data literature today (e.g., Little & Rubin, 2002; Schafer & Graham, 2002). I focus on the three most commonly used mechanisms: missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR).

Data are missing not at random (MNAR) if the propensity for missingness is related to the unobserved or would-be values, Y_{mis}, and the observed data, Y_{obs}. Mathematically, this conditional probability is expressed by the following equation:

$$\Pr(M \mid Y) = \Pr(M \mid Y_{mis}, Y_{obs}) \tag{1.1}$$

Equation 1.1 indicates that missingness is related to the missing values, and this relationship remains after conditioning on the observed variables included in the model. For example, in survey applications NMAR missingness could occur if questions about income go unanswered by those with very large or very small incomes.

Data are missing at random (MAR) if the propensity for missingness is related to observed data, but is not related to the unobserved data after conditioning on the observed data. Mathematically, this conditional probability is expressed by the following equation:

$$\Pr(M \mid Y) = \Pr(M \mid Y_{obs}) \tag{1.2}$$

Any relationship between missingness and the unobserved values disappears after conditioning on the observed data. Returning to the income example from the previous paragraph, those with very large or very small incomes may not answer questions about their incomes. If this relationship disappears after conditioning on another variable, such as education level, the mechanism is MAR.

Data are missing completely at random (MCAR) if the propensity for missingness is not related to the data. Mathematically, this conditional probability is expressed by the following equation:

$$\Pr(M \mid Y) = \Pr\left(M\right) \tag{1.3}$$

There is no relationship between the data (either observed or unobserved) and missingness. For example, a research assistant accidentally losing a few questionnaires would result in an MCAR mechanism, as would planned missing data designs.

Missing data mechanisms are important because older missing data techniques (e.g., listwise deletion) assume MCAR missingness and are biased for both MAR and MNAR mechanisms. Modern missing data handling techniques (i.e., multiple imputation, maximum likelihood analysis, and Bayesian simulation) are consistent for MCAR and MAR data. Analysis methods have been developed for MNAR data (e.g., adaptations of the aforementioned methods that explicitly model missingness under a set of strict, untestable assumptions), but these approaches are limited in their utility and are not discussed here. The multilevel imputation methods investigated in this document require an MAR (or MCAR) mechanism. The consistency property should hold, provided that

the imputation model adequately preserves the population data structure. As stated previously, the goal of this study is to investigate the situations in which existing imputation approaches achieve this goal.

Introduction to Multilevel Modeling

For the sake of clarity, it is necessary to review single-level univariate regression notation before proceeding to multilevel model notation. Equation 1.4 shows a single-level univariate multiple regression model predicting *y* from two predictors, *x1* and *x2*.

$$y_i = \beta_0 + \beta_1 \left(x1_i \right) + \beta_2 \left(x2_i \right) + \varepsilon_i \qquad (1.4)$$

β_0 is the intercept term, β_1 and β_2 are the slope terms and ε_i is the residual term. Notice that $x1_i$, $x2_i$, ε_i and y_i all have the same *i* subscript, meaning that they are specific to observation *i*. This subscript is normally not included for single-level regression models, but I include it here for comparison against multilevel models presented later. The regression coefficients do not have a subscript to denote the observation number, as they are the same for all observations in the dataset.

Multilevel or hierarchical data consist of one set of units (level one) nested within another set of units (level two). Examples of multilevel data include: students nested within schools; children nested within families; and observations nested within individuals. Level-2 units are also referred to as clusters. In contrast to the univariate single-level model, the univariate multilevel model allows for between-cluster variation in the intercept and slope coefficients. This is important because the cluster-level intercept of the criterion and the regression coefficients may differ between clusters.

As an example of cluster-level variation in coefficients, consider a study of the relationship between daily stressors and daily affect. In this hypothetical study, daily measures of stressors and affect are both nested within people. People may differ greatly in terms of their average levels of affect. As such, one might find that the person-level (cluster-level) intercepts for affect may differ among people. This would correspond to a random intercept term in a multilevel model. At the observation level (daily), one might examine the relationship between daily stressors and daily affect. For example, some people may be relatively immune to daily stressors, and the regression of daily affect on daily stressors would have a coefficient of zero for these people. That is, an increase in daily stressors would not be associated with a change in affect for these people. Other people may be greatly affected by daily stressors, and may exhibit increased levels of negative affect as a result of increased daily stressors. A level-1 slope coefficient that varies across people corresponds to a random slope in a multilevel model. Thus, in such a study, one would expect to see cluster-level variation in both the intercept term (the person-level intercept for affect) and the slope term (the relationship between daily stressors and daily affect).

Multilevel models are often described using a separate model for each level (Raudenbush & Bryk, 2002). A two-level multilevel model would be described as the combination of a level-1 model and a level-2 model. Multilevel models can be applied to data with more than two levels, but such models are beyond the scope of this paper. The level-1 model describes differences between level-1 units. The level-2 model describes differences between clusters. Consider a multilevel model where the criterion, *y*, is

predicted by two level-1 variables, *x1* and *x2*. The level-1 model consists of a single equation:

$$y_{ij} = \beta_{0j} + \beta_{1j}\left(x1_{ij}\right) + \beta_{2j}\left(x2_{ij}\right) + r_{ij} \tag{1.5}$$

The subscripts *i* and *j* for each variable indicate that the value of the variable is for level-1 unit *i* in cluster *j*. β_{0j} is cluster *j*'s intercept. β_{1j} is the cluster's slope on $x1_{ij}$. β_{2j} is the cluster's slope on $x2_{ij}$. r_{ij} is the level-1 residual, and is the difference between the predicted and observed scores on the criterion. The distribution of the level-1 residuals is assumed to be normal, centered at zero, and with spread equal to the pooled within-cluster residual variance, σ_r^2.

$$r_{ij} \sim N\left(0, \sigma_r^2\right) \tag{1.6}$$

The level-2 model consists of a set of equations, one for each of the level-1 coefficients. Because this example has three level-1 parameters (β_{0j}, β_{1j}, and β_{2j}), the level-2 model consists of three equations:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{1.7}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{1.8}$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \tag{1.9}$$

γ_{00} is the average intercept of the dependent variable. u_{0j} is the difference between the intercept of the dependent variable in cluster *j* and the grand mean of the dependent variable. γ_{10} is the weighted average across clusters of the slope of y_{ij} on $x1_{ij}$. u_{1j} is the difference between the average slope of y_{ij} on $x1_{ij}$ and the cluster-specific slope of y_{ij} on $x1_{ij}$ in cluster *j*. γ_{20} is the weighted average across clusters of the slope of y_{ij} on $x2_{ij}$. u_{2j} is the difference between the average slope of y_{ij} on $x2_{ij}$ and the cluster-specific slope of y_{ij} on $x2_{ij}$ in cluster *j*. To summarize, γ_{00}, γ_{10}, and γ_{20} are coefficients that describe the average regression plane across all of the clusters. Collectively, the γ coefficients are referred to as level-2 coefficients, or fixed effects. u_{0j}, u_{1j}, and u_{2j} describe how the cluster-specific regression coefficients differ from the fixed effects. Collectively, the *u* coefficients are referred to as level-2 residuals or random effects. The distribution of the level-2 residuals is assumed to be multivariate normal, centered at zero, and with spread equal to the level-2 covariance matrix, **T**.

$$\mathrm{E}\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{1.10}$$

$$\text{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} = \mathbf{T} \qquad (1.11)^{1}$$

Random effects do not have to be included for all of the level-1 predictors. The random effect for the intercept, u_{0j}, should always be included, but the random slopes, u_{1j} and u_{2j}, are optional terms.

In addition to level-1 variables (the values of which differ between level-1 units), multilevel models allow the inclusion of level-2 (i.e., cluster-level) variables. The values of the level-2 variables differ across clusters but not across level-1 units within a cluster. The following equations describe a model that is identical to the model presented above, but with the addition of a level-2 predictor (w_j).

$$y_{ij} = \beta_{0j} + \beta_{1j}\left(x1_{ij}\right) + \beta_{2j}\left(x2_{ij}\right) + r_{ij} \qquad (1.12)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\left(w_j\right) + u_{0j} \qquad (1.13)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\left(w_j\right) + u_{1j} \qquad (1.14)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \qquad (1.15)$$

Note that w_j has a *j* subscript but not an *i* subscript. This indicates that it is a level-2 predictor and its values are allowed to vary across clusters but not across level-1 units

[1] The use of "Var" in Equation 1.11 refers to the variance of the vector. The variance of a vector is a matrix, which is the variance-covariance matrix of the elements in the vector.

within a cluster. w_j is included as both a main effect and an interaction term. The inclusion of w_j in the level-2 equation for β_{0j} allows w_j to change the cluster intercept, β_{0j}. Thus, γ_{10} quantifies the change in the cluster intercept, β_{0j}, that results from a one-unit change in w_j. This is the main effect of w_j. The inclusion of w_j in the level-2 equation for β_{1j} allows w_j to change β_{1j}, the average slope of y_{ij} on $x1_{ij}$. γ_{11} quantifies the change in β_{1j}, the average slope of y_{ij} on $x1_{ij}$, for a one-unit change in w_j. Although the level-1 and level-2 notational system does not include a product term, the γ_{11} coefficient represents the interaction between $x1_{ij}$ and w_j. The notational system that I describe below clarifies this point.

As the numbers of level-1 and level-2 predictors increase, the level-1/level-2 notation can become tedious to write and to interpret. As such, reduced form notation is often used instead. Level-1/level-2 equations can be written in reduced form notation by substituting the right side of each of the level-2 equations into the level-1 equation wherever a level-1 coefficient appears (e.g., replacing the β_{0j} coefficient in Equation 1.12 with the right side of Equation 1.13). Doing so with Equations 1.13 through 1.15 yields the following result:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{01}\left(w_j\right) + \gamma_{10}\left(x1_{ij}\right) + \gamma_{20}\left(x2_{ij}\right) \\ &+ \gamma_{11}\left(x1_{ij}\right)\left(w_j\right) + u_{0j} + u_{1j}\left(x1_{ij}\right) + u_{2j}\left(x2_{ij}\right) + r_{ij} \end{aligned} \tag{1.16}$$

In reduced form notation, γ_{11} is multiplied by ($x1_{ij}$)(w_j), showing that the coefficient γ_{11} quantifies the interaction between $x1_{ij}$ and w_j.

Multilevel models are a subclass of models called mixed models, which are models that account for both fixed and random effects (Eisenhart, 1947). Mixed model theory was developed primarily in the area of genetics research, and was formalized by Goldberger (1962), Harville (1976a, 1976c), and Henderson (1950, 1963). I employ both mixed model theory and mixed model notation for the remainder of this document. Mixed model notation, as developed by Henderson, is as follows:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j + \mathbf{e}_j \tag{1.17}$$

$\mathbf{y}_j$ is the n_j x 1 dependent variable vector for cluster *j*, where n_j is the number of level-1 units in cluster *j*. $\mathbf{X}_j$ is the n_j by *p* fixed effect covariate matrix for cluster *j*, where *p* is the number of fixed effects. Note that $\mathbf{X}_j$ contains all of the level-1 and level-2 predictor variables, as well as product terms for the interactions and a unit vector for the intercept. **β** is the *p* x 1 vector of fixed effects (level-2 coefficients) that are common to all clusters. Note that **β** contains the γ's from the Raudenbush and Bryk notation. $\mathbf{Z}_j$ is the n_j by *q* random effect covariate matrix for cluster *j*, where *q* is the number of random effects. $\mathbf{Z}_j$ contains the subset of complete level-1 variables that are allowed to have random effects on the dependent variable, and also includes a unit vector for the intercept. So, $\mathbf{Z}_j$ contains a subset of the X's from Raudenbush and Bryk notation, plus

an intercept vector. $\mathbf{u}_j$ is the q x 1 vector of level-2 residuals (i.e., residual intercepts and slopes) for cluster j. $\mathbf{u}_j$ contains the u's from Raudenbush and Bryk notation. $\mathbf{e}_j$ is the n_j x 1 vector of level-1 residuals for cluster j. It contains the r's from Raudenbush and Bryk notation.

To link the mixed model notation to the previous notational system, consider a cluster of n_j = 3 cases. In mixed model notation, the model from Equation 1.16 would be as follows:

$$\begin{bmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \end{bmatrix} = \begin{bmatrix} 1 & w_j & x1_{1j} & x2_{1j} & x1_{1j}w_j \\ 1 & w_j & x1_{2j} & x2_{2j} & x1_{2j}w_j \\ 1 & w_j & x1_{3j} & x2_{3j} & x1_{3j}w_j \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{11} \end{bmatrix} + \begin{bmatrix} 1 & x1_{1j} & x2_{1j} \\ 1 & x1_{2j} & x2_{2j} \\ 1 & x1_{3j} & x2_{3j} \end{bmatrix} \begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} + \begin{bmatrix} r_{1j} \\ r_{2j} \\ r_{3j} \end{bmatrix} \quad (1.18)$$

Notice that the $\mathbf{X}_j$ and $\mathbf{Z}_j$ matrices are made up of column vectors. Each column vector represents the values of one of the variables in cluster j. The number of rows in the $\mathbf{X}_j$ matrix, the $\mathbf{Z}_j$ matrix, and the $\mathbf{e}_j$ vector equals the number of observations in the cluster, which in this case is three.

Equations 1.16 and 1.17 contain the same terms and thus are two equivalent ways to express the same model. The mixed model notational system similarly offers an alternate way to express the covariance structure. Specifically, the **T** matrix from Equation 1.11 is equivalent to the **Ψ** matrix in the mixed model notation, as follows.

$$\mathbf{u}_j \sim \mathrm{N}\left(0, \boldsymbol{\Psi}\right) \tag{1.19}$$

That is, the distribution of the level-2 residuals is multivariate normal, is centered at zero, and has spread equal to the level-2 residual covariance matrix, $\boldsymbol{\Psi}$ (this matrix was referred to as **T** in the Raudenbush and Bryk notation). The use of $\boldsymbol{\Psi}$ for the level-2 residual covariance matrix is common in multilevel imputation literature (e.g., Schafer, 2001; Schafer & Yucel, 2002), but traditional mixed model notation would refer to this matrix as **G**. Equation 1.19 is equivalent to Equations 1.10 and 1.11. The level-1 residuals for the linear mixed model are distributed as

$$\mathbf{e}_j \sim \mathrm{N}\left(0, \boldsymbol{\Sigma}\right) \tag{1.20}$$

That is, the distribution of the level-1 residuals is multivariate normal, is centered at zero, and has spread equal to the level-1 residual covariance matrix, $\boldsymbol{\Sigma}$. The use of $\boldsymbol{\Sigma}$ for the level-2 residual covariance matrix is common in multilevel imputation literature (e.g., Schafer, 2001; Schafer & Yucel, 2002), but traditional mixed model notation would refer to this matrix as **R**. For this particular application of the mixed linear model, the level-1 residual covariance matrix reduces to a scalar, σ_r^2, which is the level-1 residual variance term. So, the previous equation becomes

$$\mathbf{e}_j \sim \mathrm{N}\left(0, \sigma_r^2\right) \tag{1.21}$$

Equation 1.21 is equivalent to Equation 1.6.

Single-Level Multiple Imputation

Multiple imputation (MI) analysis consists of two distinct phases: the imputation phase and the analysis/pooling phase. In the imputation phase, the algorithm draws parameters for the imputation model from their respective distributions and then uses these parameter estimates to fill in the missing data. This process is repeated for a very large number of iterations. Whenever a pre-determined number of iterations has passed, the algorithm outputs an imputed data set. The algorithm continues in this vein until *m* distinct data sets have been generated, where *m* is a number chosen by the analyst (e.g., a minimum of 20 is a rule of thumb; Graham, Olchowski, & Gilreath, 2007).The multiple imputation algorithm amounts to sampling missing values from a posterior predictive distribution. In the analysis/pooling phase, the data analyst runs an identical analysis on each of the *m* data sets. The point estimates and standard errors from these analyses are then pooled using methods described later in this section. The analysis/pooling phase can be repeated for multiple analyses without repeating the imputation phase, provided that all of the necessary variables/effects are included in the imputation model.

Joint Modeling vs. Fully Conditional Specification. The methods used in the imputation phase of multiple imputation can be divided into two categories: joint modeling (JM) and fully conditional specification (FCS) (van Buuren, 2007). In the JM approach, the observations are partitioned into groups of identical missing data patterns. Within each group, the missing values are imputed using a joint model (e.g., a multivariate normal model). In contrast, the FCS method (also referred to as variable-by-variable imputation, imputation by chained equations, etc.) imputes one variable at a

time. A separate imputation model is specified for each variable. After a variable is imputed in FCS, it is then treated as a complete-data predictor in the next imputation model. Within each iteration, the algorithm cycles through all of the incomplete variables.

JM tends to be computationally more efficient due to shortcuts such as the sweep operator (van Buuren, 2007). However, FCS tends to be more flexible. FCS allows the data analyst to incorporate unique features in the data (e.g., bounds and skip patterns; van Buuren, 2007). The flexibility of FCS can also be problematic, as it is easy for an analyst to specify a set of distributions that do not correspond to a multivariate density (van Buuren, 2007). This is referred to as incompatibility of conditionals. Incompatibility of conditionals can lead to convergence problems. The effect of incompatibility of conditionals on the quality of imputations is not well known, and requires further study. Both the JM and FCS approaches are described in greater detail below.

Single-Level Multiple Imputation with Joint Modeling. This section outlines the computational steps for single-level multiple imputation using joint modeling (JM) as outlined by Schafer (1997). The Markov Chain Monte Carlo (MCMC) algorithm begins by generating initial parameter estimates (e.g., the mean vector and covariance matrix) and placeholder values for the missing data. In subsequent iterations, draws from a posterior distribution provide the necessary parameter values and imputations replace the placeholder values. The following steps summarize the algorithmic details for single-level JM imputation.

1. Draw a new covariance matrix from an inverse Wishart distribution, based on assumed values for the missing data:

$$\boldsymbol{\Sigma}^{(t)}|\boldsymbol{\mu}^{(t\text{-}1)}, \mathbf{Y} \sim \mathrm{W}^{-1}\left(n-1, (n-1)\hat{\boldsymbol{\Sigma}}\right) \tag{1.22}$$

where *n* is the number of observations in the data set, and $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix from the filled-in data. It is worth noting that in cases where software cannot produce samples from an inverse Wishart distribution, the inverse of the covariance matrix can be drawn from a Wishart distribution:

$$\boldsymbol{\Sigma}^{-1(t)}|\boldsymbol{\mu}^{(t-1)}, \mathbf{Y} \sim \mathrm{W}\left(n-1, \left((n-1)\hat{\boldsymbol{\Sigma}}\right)^{-1}\right) \tag{1.23}$$

2. Draw a new mean vector based on assumed values for the missing data and the covariance matrix drawn in step 1.

$$\boldsymbol{\mu}^{(t)}|\boldsymbol{\Sigma}^{(t)}, \mathbf{Y} \sim \mathrm{N}\left(\hat{\boldsymbol{\mu}}, n^{-1}\boldsymbol{\Sigma}^{(t)}\right) \tag{1.24}$$

where $\hat{\mu}$ is the mean vector estimated from the filled-in data and $\boldsymbol{\Sigma}^{(t)}$ is the covariance matrix drawn in the current iteration.

3. Draw new values for the missing data based on the parameters drawn in steps 1 and 2. This is done separately for each missing data pattern. For each missing data pattern, we use the covariance matrix and mean vector drawn in the current step to calculate the regression parameters for the incomplete variables regressed on

the complete variables. This transformation is often carried out using an algorithm called the sweep operator (Goodnight, 1979; Little & Rubin, 2002). After we have obtained the matrix of regression coefficients and residual covariance matrix for a pattern, *k*, we draw new values for the missing variables for each observation *i* in each pattern *k* from the following distribution:

$$\mathbf{y}_{miss,ik}|\boldsymbol{\mu}^{(t)},\boldsymbol{\Sigma}^{(t)},\mathbf{y}_{obs,ik} \sim \text{MV N}\left(\left(\mathbf{y}_{obs,ik}^{\text{T}}\hat{\boldsymbol{\beta}}_k\right)^{\text{T}},\boldsymbol{\Sigma}_{res,k}^{(t)}\right) \tag{1.25}$$

where $\mathbf{y}_{miss,ik}$ is the vector of missing variables for observation *i* in pattern *k*, $\mathbf{y}_{obs,ik}$ is the vector of observed variables for observation *i* in pattern *k*, $\hat{\boldsymbol{\beta}}_k$ is the estimated matrix of regression coefficients for pattern *k*, and $\boldsymbol{\Sigma}_{res,k}^{(t)}$ is the residual covariance matrix for pattern *k*.

After completing an iteration, the MCMC algorithm returns to step 1 and repeats the computational steps at the next iteration. All of the steps above assume a joint model (i.e., a multivariate normal model). The JM approach can be adapted to handle categorical variables by treating the categorical variables as having underlying latent normal variables. The JM approach assumes that these latent variables have a joint multivariate normal distribution. Multiple imputation with latent variables is not discussed further in this document.

Single-Level Multiple Imputation with Fully Conditional Specification. This section outlines the computational steps for single-level multiple imputation using fully

conditional specification (FCS) as described by van Buuren (2007). To be consistent with the previous section, I examine FCS imputation in the context of multivariate normal data. Before beginning the MCMC algorithm, the analyst must specify an imputation model for each of the incomplete variables in the data set. For example, suppose that we have two incomplete variables, *y1* and *y2*, as well as a complete variable, *x*. We need to specify distributions for *y1* and *y2*. For simplicity, we might specify the following distribution for *y1*:

$$y1_i \mid y2_i, x_i \sim \mathrm{N}\left(\beta_0^{(y1)} + \beta_1^{(y1)} x_i + \beta_2^{(y1)} y2_i, \sigma^2_{y1|y2,x}\right) \tag{1.26}$$

where $\beta_0^{(y1)}$ is the intercept for the regression predicting the mean of the conditional distribution of *y1*, $\beta_1^{(y1)}$ and $\beta_2^{(y1)}$ are coefficients for the regression predicting the mean of the conditional distribution of *y1*, and $\sigma^2_{y1|y2,x}$ is the residual variance for *y1*. We might specify the following distribution for *y2*:

$$y2_i \mid y1_i, x_i \sim \mathrm{N}\left(\beta_0^{(y2)} + \beta_1^{(y2)} x_i + \beta_2^{(y2)} y1_i, \sigma^2_{y2|y1,x}\right) \tag{1.27}$$

where $\beta_0^{(y2)}$ is the intercept for the regression predicting the mean of the conditional distribution of *y2*, $\beta_1^{(y2)}$ and $\beta_2^{(y2)}$ are coefficients for the regression predicting the mean of the conditional distribution of *y2*, and $\sigma^2_{y2|y1,x}$ is the residual variance for *y2*.

Before beginning the MCMC algorithm, we obtain initial estimates of each of the regression parameters in the above equations, as well as placeholder values for the missing data.

The MCMC algorithm for single-level FCS imputation consists of two nested loops: an incomplete variable loop within an iteration loop. That is, for a given iteration the algorithm cycles through the incomplete variables, performing the same steps for each of the incomplete variables. The algorithm then begins the next iteration, and once again cycles through the incomplete variables. This continues until a pre-specified number of iterations is complete. For the three-variable example outlined in the previous paragraph, a single iteration would consist of the following steps:

1a. First, we draw a new residual variance for *y1* from its posterior distribution, conditional on the other variables and the regression coefficients drawn in the previous iteration of the algorithm:

$$\sigma^{2(t)}_{y1|y2,x} \,|\, \boldsymbol{\beta}^{(y1,t-1)}, \mathbf{y1}^{(t-1)}, \mathbf{y2}^{(t-1)} \sim \text{Inv}\chi^2\left(n-p-1, \hat{\sigma}^2_{y1|y2,x}\right) \tag{1.28}$$

where $\boldsymbol{\beta}^{(y1,t-1)}$ contains the regression parameters for *y1* drawn in the previous iteration, *n* is the number of observations, *p* is the number of predictors, and $\hat{\sigma}^2_{y1|y2,x}$ is the residual variance for *y1* estimated from the filled-in data at the previous step. As an alternative, the inverse of the residual variance can be drawn from a chi-square distribution rather than an inverse chi-square distribution:

$$\sigma_{y1|y2,x}^{-2(t)} \,|\, \boldsymbol{\beta}^{(y1,t-1)}, \mathbf{y1}^{(t-1)}, \mathbf{y2}^{(t-1)} \sim \chi^2\left(n - p - 1, \hat{\sigma}_{y1|y2,x}^{-2}\right) \tag{1.29}$$

1b. Next, we draw a new vector of regression coefficients for *y1* from its posterior distribution, conditional on the other variables and the residual variance for *y1* drawn in the current iteration. The vector of regression coefficients for predicting *y1*, $\boldsymbol{\beta}^{(y1,t)}$, is drawn from a multivariate normal distribution:

$$\boldsymbol{\beta}^{(y1,t)} \,|\, \sigma_{y1|y2,x}^{2(t)}, \mathbf{y1}^{(t-1)}, \mathbf{y2}^{(t-1)} \sim \text{MVN}\left(\hat{\boldsymbol{\beta}}^{(y1)}, \sigma_{y1|y2,x}^{2(t)} * \widehat{\mathbf{SS}}_{x,y2}^{-1}\right) \tag{1.30}$$

where $\hat{\boldsymbol{\beta}}^{(y1)}$ is the vector of regression coefficients estimated from the filled in data at the previous step and $\widehat{\mathbf{SS}}_{x,y2}^{-1}$ is the inverse sum of squares matrix for the predictors (*x* and *y2*) estimated from the filled-in data at the previous step.

1c. Next, we draw new values for the missing values of *y1*. So, as a third step in the MCMC algorithm we draw new values of the criterion variable conditioned on the current draws of the regression parameters and the other variables:

$$y1_i^{(t)} \,|\, y2_i^{(t-1)}, \boldsymbol{\beta}^{(y1,t)}, \sigma_{y1|y2,x}^{2(t)} \sim \text{N}\left(\beta_0^{(y1,t)} + \beta_1^{(y1,t)} x_i + \beta_2^{(y1,t)} y2_i^{(t-1)}, \sigma_{y1|y2,x}^{2(t)}\right) \tag{1.31}$$

2a. Having completed the imputation steps for *y1*, the algorithm repeats the steps for *y2*. As before, we draw a new residual covariance for *y2* from its posterior

distribution, conditional on the other variables and the regression coefficients drawn in the previous iteration of the algorithm:

$$\sigma^{2(t)}_{y2|y1,x} \,|\, \boldsymbol{\beta}^{(y2,t-1)}, \mathbf{y1}^{(t)}, \mathbf{y2}^{(t-1)} \sim \text{Inv}\chi^2\left(n-p-1, \hat{\sigma}^2_{y2|y1,x}\right) \tag{1.32}$$

where $\boldsymbol{\beta}^{(y2,t-1)}$ contains the regression parameters for *y2* drawn in the previous iteration, *n* is the number of observations, *p* is the number of predictors, and $\hat{\sigma}^2_{y2|y1,x}$ is the residual variance for *y2* estimated from the filled-in data at the end of Step 1c.

2b. Next, we draw a new vector of regression coefficients for *y2* from its posterior distribution, conditional on the other variables and the residual variance for *y2* drawn in the current iteration:

$$\boldsymbol{\beta}^{(y2,t)} \,|\, \sigma^{2(t)}_{y2|y1,x}, \mathbf{y1}^{(t)}, \mathbf{y2}^{(t-1)} \sim \text{MVN}\left(\hat{\boldsymbol{\beta}}^{(y2)}, \sigma^{2(t)}_{y2|y1,x} * \widehat{\mathbf{SS}}^{-1}_{x,y1}\right) \tag{1.33}$$

$\hat{\boldsymbol{\beta}}^{(y2)}$ is the vector of regression coefficients estimated from the filled in data at the previous step and $\widehat{\mathbf{SS}}^{-1}_{x,y1}$ is the inverse sum of squares matrix for the predictors (*x* and *y1*) estimated from the filled-in data at the previous step.

2c. Next, we draw new values for the missing values of *y2*, conditional on the current draws of the regression parameters and the other variables:

$$y2_i^{(t)} \mid y1_i^{(t)}, \boldsymbol{\beta}^{(y2,t)}, \sigma^{2(t)}_{y2|y1,x} \sim \mathrm{N}\left(\beta_0^{(y2,t)} + \beta_1^{(y2,t)} x_i + \beta_2^{(y2,t)} y1_i^{(t)}, \sigma^{2(t)}_{y2|y1,x}\right) \quad (1.34)$$

The above six steps form a single iteration of the MCMC algorithm. The process repeats until a pre-specified number of iterations is reached.

Multiple Imputation Analysis and Pooling. The analysis and pooling phase of multiple imputation is identical for both the joint modeling and fully conditional specification approaches for both single-level and multilevel imputation. The analysis phase involves performing the same analysis once for each of the *m* datasets generated by the imputation procedure. The pooling of estimates is the same for both multilevel data and single-level data. A more thorough discussion of the topic is available in Rubin (1987). Because the analysis yields *m* estimates of each parameter, it is necessary to pool them into a single point estimate. This is done by averaging the parameter estimates across the *m* datasets

$$\bar{\theta} = \frac{1}{m}\sum_{l=1}^{m}\hat{\theta}_l \quad (1.35)$$

where $\hat{\theta}_l$ is a parameter estimate (e.g., regression coefficient, variance estimate) from data set *l*.

Standard errors are slightly more difficult to combine, as the pooled standard error is the combination of the sampling error that would be present if the data were complete (within-imputation variance) and the sampling error resulting from missing data

(between-imputation variance). The within-imputation variance is merely the average of the sampling variances, as follows

$$V_W = \frac{1}{m}\sum_{l=1}^{m} SE_l^2 \tag{1.36}$$

where SE_l^2 is the squared standard error from data set *l*. The between-imputation variance is calculated as the variance of a parameter across the imputations.

$$V_B = \frac{1}{m-1}\sum_{l=1}^{m}\left(\hat{\theta}_l - \bar{\theta}\right)^2 \tag{1.37}$$

The total variance is calculated as the combination of the within-imputation and between-imputation variances.

$$V_T = V_W + V_B + \frac{V_B}{m} \tag{1.38}$$

Finally, the standard error of the parameter is the square root of the parameter variance.

$$SE = \sqrt{V_T} \tag{1.39}$$

Multilevel Multiple Imputation

Single-level imputation methods do not incorporate the effects of clustering into the imputation process. To correctly handle clustering in the data, it is necessary to use an imputation model that accounts for random effects. Such models are referred to as multilevel imputation models. Note that the modeling of categorical data is not discussed in this document, as the focus of this dissertation is on multilevel imputation of normal variables. Multilevel multiple imputation procedures use imputation models based on the linear mixed-effects model for clustered data. Recall from earlier that the linear mixed model is:

$$\boldsymbol{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{B}_j + \mathbf{e}_j \tag{1.40}$$

$\boldsymbol{y}_j$ is the n_j x 1 dependent variable vector for cluster *j*, where n_j is the number of level-1 units in cluster *j*. $\mathbf{X}_j$ is the n_j by *p* fixed effect covariate matrix for cluster *j*, where *p* is the number of fixed effects. Note that $\mathbf{X}_j$ contains all of the level-1 and level-2 predictor variables, as well as product terms for any interactions and a unit vector for the intercept. $\boldsymbol{\beta}$ is the *p* x 1 vector of fixed effects (level-2 coefficients) that are common to all clusters. Note that $\boldsymbol{\beta}$ contains the γ's from the Raudenbush and Bryk notation. $\mathbf{Z}_j$ is the n_j by *q* random effect covariate matrix for cluster *j*, where *q* is the number of random effects. $\mathbf{Z}_j$ contains the subset of complete and variables that are allowed to have random effects on the dependent variable and includes a unit vector for the intercept. $\mathbf{B}_j$ is the *q* x 1 vector of level-2 residuals (i.e., residual intercepts and slopes) for cluster *j*. It

contains the u's from Raudenbush and Bryk notation. $\mathbf{e}_j$ is the n_j x 1 vector of level-1 residuals for cluster *j*. It contains the r's from Raudenbush and Bryk notation. Equation 1.40 corresponds to an imputation model for a single incomplete variable (with the notable exception of the Mplus JM imputation model, as will be explained later). This imputation model can be adapted to handle multivariate missingness using the JM or FCS approaches, as described below.

An imputation model can be very different from the analysis model. The imputation model predicts one or more incomplete variables from the complete variables, regardless of the role of the variables in the analysis model. As with single-level imputation, there are both joint model (JM) and fully conditional specification (FCS) approaches to multilevel imputation, and both employ a variant of Equation 1.40 to generate imputations. The form of Equation 1.40 and the contents of the matrices differ between the two approaches, as discussed below. The remainder of this chapter describes both the JM and FCS multilevel imputation approaches, followed by a discussion of the relative merits of each.

JM for Multilevel Imputation. This document previously described the use of joint imputation models for single-level data. Single-level JM imputation models draw imputations from a multivariate normal distribution. Distributions other than normal can be used, but these distributions are beyond the scope of this document. Researchers extended the logic of single-level joint imputation to multilevel models. One of the most widely used JM multilevel imputation methods, PAN (Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008), begins with the specification of a multivariate mixed model

predicting the incomplete level-1 variables from the complete level-1 and level-2 variables:

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{B}_j + \mathbf{E}_j \tag{1.41}$$

where $\mathbf{Y}_j$ is the n_j by *r* matrix of incomplete level-1 variables for cluster *j* and *r* is the number of incomplete level-1 variables. Note that $\mathbf{Y}_j$ contains all of the incomplete level-1 variables, regardless of their role in the analysis model. $\mathbf{X}_j$ is the n_j by *f* fixed effect covariate matrix for cluster *j*, where *f* is the number of fixed effect covariates. $\mathbf{X}_j$ contains all of the complete level-1 and level-2 variables (including imputed level-2 variables from the previous iteration), as well as a unit vector for the intercept. $\boldsymbol{\beta}$ is the *f* by *r* matrix of fixed effects that are common for all clusters. $\mathbf{Z}_j$ is the n_j by *q* random effect covariate matrix for cluster *j*, where *q* is the number of random effects. $\mathbf{Z}_j$ contains the subset of complete level-1 variables that are allowed to have random effects with the variables in $\mathbf{Y}_j$, as well as a vector of ones for the random intercepts. $\mathbf{B}_j$ is the *q* by *r* matrix of level-2 residuals (i.e., residual intercepts and slopes) for cluster *j*. $\mathbf{E}_j$ is the n_j by *r* matrix of level-1 residuals for cluster *j*.

Given the imputation model in Equation 1.41, multilevel joint imputation draws new values from a conditional multivariate normal distribution:

$$\mathbf{Y}_j \,|\, \mathbf{X}_j \sim \text{MVN}\left(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{B}_j, \boldsymbol{\Sigma}\right) \tag{1.42}$$

That is, the missing values are drawn from a conditional multivariate normal distribution, centered at the predicted value from the imputation model ($\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{B}_j$), with spread equal to the level-1 residual covariance matrix ($\boldsymbol{\Sigma}$). The parameter values and the level-2 residuals are obtained from a previous MCMC step. The parameters in Equation 1.42 are generated from a Markov Chain Monte Carlo algorithm similar to that described in the single-level imputation section. Because these sampling steps are described throughout the literature (Browne & Draper, 2000; Goldstein, 2011; Goldstein, Carpenter, Kenward, & Levin, 2009; Kasim & Raudenbush, 1998; Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008), I do not detail them here.

At this point, I switch to scalar notation for the sake of clarity. To illustrate the points presented earlier in this section, consider the following multilevel analysis model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 w_{ij} + b_{0j} + b_{1j} x_{ij} + b_{2j} w_{ij} + e_{ij} \tag{1.43}$$

y_{ij} is the value of the dependent variable for observation *i* in cluster *j*. x_{ij} and w_{ij} are the values of the two level-1 predictor variables for observation *i* in cluster *j*. β_0 is the intercept and β_1 and β_2 are the regression coefficients. b_{0j} is the random intercept term. b_{1j} and b_{2j} are the random slope terms. e_{ij} is the level-1 residual term. In this model, *y* is modeled on two predictors, *x* and *w*. The intercept and both slopes are allowed to vary across clusters.

Suppose that *y* and *w* are both incomplete, whereas *x* is complete. The PAN method (the first of the multilevel JM imputation methods; Schafer, 2001; Schafer & Yucel, 2002) would generate imputations based on the following conditional multivariate normal distribution:

$$\begin{matrix} y_{ij} \mid x_{ij} \\ w_{ij} \mid x_{ij} \end{matrix} \sim \text{MVN}\left(\begin{matrix} \beta_0^{(y)} + \beta_1^{(y)} x_{ij} + b_{0j}^{(y)} + b_{1j}^{(y)} x_{ij} \\ \beta_0^{(w)} + \beta_1^{(w)} x_{ij} + b_{0j}^{(w)} + b_{1j}^{(w)} x_{ij} \end{matrix}, \boldsymbol{\Sigma} \right) \tag{1.44}$$

where $\beta_0^{(y)}$ and $\beta_1^{(y)}$ are the fixed effects for the imputation model predicting *y* from *x*, $b_{0j}^{(y)}$ and $b_{1j}^{(y)}$ are the random effects in the imputation model predicting *y* from *x*, $\beta_0^{(w)}$ and $\beta_1^{(w)}$ are the fixed effects in the imputation model predicting *w* from *x*, and $b_{0j}^{(w)}$ and $b_{1j}^{(w)}$ are the random effects in the imputation model predicting *w* from *x*. Equation 1.44 illustrates two important points about the PAN method. First, the imputation model allows the regression of *w* on *x* to vary across clusters (capturing a random effect of *x* on *w*, or vice versa). This effect was not specified in the analysis model in Equation 1.43, but it may lead to better imputations if the regression of *w* on *x* varies across clusters in the population. If the regression of *w* on *x* does not vary across clusters, the inclusion of this random effect may result in computational issues for the MCMC algorithm (inverting matrices that may not be positive definite, trying to draw matrices where the sum of squares and cross products matrix is not positive definite, etc.). Second, though the analysis model contains the random effect of *w* on *y*, this effect is not included in the imputation model in Equation 1.44. Rather, the level-1 regression of *y* on *w* (conditional

on *x*) is captured entirely by the level-1 residual covariance matrix, $\boldsymbol{\Sigma}$, the elements of which do not vary (i.e., the association is assumed to be fixed across clusters). Due to this mismatch, the imputation model is said to be uncongenial with the analysis model (an analysis model and an imputation model are said to be uncongenial if the analysis model cannot be derived from the imputation model, or vice versa; Meng, 1994; Schafer, 1997, 2003). The mismatch between the random effects in the analysis and imputation models would negatively bias the estimate of the slope variance of *w*, and could potentially affect other estimates as well (Enders, Mistler, & Keller, 2014).

To reiterate an earlier point, the level-1 covariances between pairs of incomplete variables, conditioned on the complete variables, are captured solely by the level-1 residual covariance matrix, $\boldsymbol{\Sigma}$. This is important because the residual covariance between a pair of incomplete variables is not allowed to vary between clusters. As a remedy for this problem, Yucel (2011) proposed a modification to the PAN method in which the level-1 residual covariance matrix is allowed to vary across clusters. This modified PAN method could be used to model random effects between the incomplete level-1 variables. Unfortunately, allowing the level-1 residual covariance matrix to vary across clusters may result in greatly increased computational complexity. As such, the method may not be practical in realistic data sets. The article in which the method was published included an application of the method to an example data set, rather than a simulation study. Further research is needed to determine whether the method is practical for general use (e.g., when cluster sizes are very small relative to the number of incomplete variables).

So far I have provided a general description of the PAN method. It should be noted, however, that Schafer suggested two different implementations of PAN (Schafer,

2001; Schafer & Yucel, 2002). Though not shown previously, one must draw new random effects for each cluster at each iteration:

$$\text{vec}(\mathbf{b}_j) \sim N(0, \boldsymbol{\Psi}) \tag{1.45}$$

where $\text{vec}(\mathbf{b}_j)$ refers to stacking all of the random effects for cluster *j* in a column vector and $\boldsymbol{\Psi}$ is the covariance matrix of the random effects for all of the incomplete variables. Schafer proposed two options for $\boldsymbol{\Psi}$: (1) an unstructured covariance matrix or (2) a block-diagonal covariance matrix. I refer to the first method as JM-UN and the second method as JM-BD. The unstructured level-2 covariance matrix used in the JM-UN method allows the random effects for any two incomplete variables to be correlated. The practical implication of an unstructured level-2 covariance matrix is that the level-2 residual covariances among the incomplete variables need not be of the same sign or magnitude as the level-1 residual covariances among those variables, which are captured in $\boldsymbol{\Sigma}$. This means that if the level-1 and level-2 residual covariances between pairs of incomplete variables differ in the data (referred to as a contextual effect in some disciplines; Firebaugh, 1978; Kreft, De Leeuw, & Aiken, 1995; Raudenbush & Bryk, 2002), these covariances will be allowed to differ in the imputation model. However, the number of parameters to be estimated in the unstructured level-2 covariance matrix may be large with many incomplete variables.

If the number of clusters in a data set is small, it may not be possible to estimate covariances among all of the random effects for all of the incomplete variables (Schafer & Yucel, 2002). In such cases, it may be advantageous to model $\boldsymbol{\Psi}$ as a block-diagonal

matrix in which the covariances between random effects for different incomplete variables are constrained to zero (JM-BD):

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Psi}_r \end{bmatrix} \tag{1.46}$$

where *r* is the number of incomplete variables.

To clarify the differences between the two PAN methods, consider the analysis model in Equation 1.43 and the JM model in Equation 1.44. The JM-UN method would estimate covariances between all of the random effects in the imputation model ($b_{0j}^{(y)}$, $b_{1j}^{(y)}$, $b_{0j}^{(w)}$, and $b_{1j}^{(w)}$). The covariance between the two random intercepts $\left(b_{0j}^{(y)}\right.$ and $\left.b_{0j}^{(w)}\right)$ would preserve the level-2 covariance between *y* and *w*, conditional on *x*. This does not correspond to any term in the analysis model, as Equation 1.44 did not include the cluster means for *w*. The JM-BD method would estimate covariances between random effects for the same incomplete variable, but would constrain covariances between random effects for different variables to zero. So, the covariance between the random intercept and random slope for *y* ($b_{0j}^{(y)}$ and $b_{1j}^{(y)}$) would be estimated, as would the covariance between the random intercept and random slope for *w* ($b_{0j}^{(w)}$, and $b_{1j}^{(w)}$). However, the covariances between the random effects for *y* ($b_{0j}^{(y)}$ and $b_{1j}^{(y)}$) and the

random effects for *w* ($b_{0j}^{(w)}$, and $b_{1j}^{(w)}$) would be constrained to zero. Because the covariance between the two random intercepts ($b_{0j}^{(y)}$ and $b_{0j}^{(w)}$) would be constrained to zero, JM-BD would not preserve the level-2 covariance between *y* and *w*, conditional on *x*. JM-BD and JM-UN are available for R and S-Plus in packages called PAN (Schafer, 2001; Schafer & Yucel, 2002). JM-BD is available in SAS as a macro called MMI_IMPUTE (Mistler, 2013). JM-BD and JM-UN imputation can also be performed in a standalone software package called REALCOM-IMPUTE (Carpenter, Goldstein, & Kenward, 2011) or the latent variable modeling package Mplus (Asparouhov & Muthén, 2010f).

The two JM methods just described (which differ only in the structure of $\boldsymbol{\Psi}$) both modeled the incomplete variables as a function of the complete variables. In contrast, the JM imputation method implemented in Mplus (referred to as H1 imputation in the Mplus documentation, but which I call JM-Mplus) treats all variables, both incomplete and complete, as response variables and includes no predictors in the model (Asparouhov & Muthén, 2010a, 2010f):

$$\mathbf{y}_{ij} = \boldsymbol{\beta} + \mathbf{b}_j + \mathbf{e}_{ij} \tag{1.47}$$

where $\mathbf{y}_{ij}$ is a vector of level-1 variables for level-1 unit *i* in cluster *j*. $\mathbf{y}_{ij}$ contains all of the level-1 variables, both complete and incomplete. $\boldsymbol{\beta}$ is a vector containing the grand mean for each of the variables. $\mathbf{b}_j$ contains the random intercept for each variable for

cluster *j*. $\mathbf{e}_{ij}$ contains the discrepancies between the cluster intercepts for cluster j and the values (observed or imputed) for observation *i* in cluster *j*. Note that all of the objects in Equation 3.5 are vectors, as the equation describes only a single observation, *i*. Matrices $\mathbf{X}_j$ and $\mathbf{Z}_j$ from Equation 3.1 are not shown in Equation 3.5, as the two matrices reduce to scalars equal to 1. Vectors $\mathbf{e}_{ij}$ and $\mathbf{b}_j$ are distributed as:

$$\mathbf{e}_{ij} \sim \mathrm{N}(0, \boldsymbol{\Sigma}) \tag{1.48}$$

$$\mathbf{b}_j \sim \mathrm{N}(0, \boldsymbol{\Psi}) \tag{1.49}$$

where $\boldsymbol{\Sigma}$ is the unstructured level-1 covariance matrix of the variables and $\boldsymbol{\Psi}$ is the unstructured level-2 covariance matrix of the residuals. JM-Mplus does not allow for the inclusion of random slopes among ANY variables. This is in contrast to the PAN methods (JM-BD and JM-UN), which allow complete variables to exert random influences on the incomplete variables.

In the context of the analysis model in Equation 1.43, JM-Mplus would treat all three variables, *w*, *x*, and *y*, as having the following multivariate normal distribution:

$$\begin{matrix} w_{ij} \\ x_{ij} \\ y_{ij} \end{matrix} \sim \mathrm{MVN}\left(\begin{matrix} \beta_0^{(w)} + b_{0j}^{(w)} \\ \beta_0^{(x)} + b_{0j}^{(x)} \\ \beta_0^{(y)} + b_{0j}^{(y)} \end{matrix}, \boldsymbol{\Sigma} \right) \tag{1.50}$$

where $\beta_0^{(w)}$, $\beta_0^{(x)}$, and $\beta_0^{(y)}$ are the grand means of the variables, $b_{0j}^{(w)}$, $b_{0j}^{(x)}$, and $b_{0j}^{(y)}$ are the differences between the grand means and the means of cluster *j*, and $\boldsymbol{\Sigma}$ is the unstructured level-1 covariance matrix for the three variables. The random intercepts for the three variables are distributed as:

$$\mathbf{b}_j \sim \mathrm{N}(0, \boldsymbol{\Psi}) \tag{1.51}$$

where $\boldsymbol{\Psi}$ is the unstructured level-2 covariance matrix for the three variables. Because covariances between all three variables are captured at level one by $\boldsymbol{\Sigma}$ and at level two by $\boldsymbol{\Psi}$, the covariances between the variables are allowed to differ across levels. However, the random effects of *x* on *w* and *y* in the analysis model are not captured by this imputation method. JM-Mplus can be implemented in the latent variable modeling package Mplus (Asparouhov & Muthén, 2010f).

FCS for Multilevel Imputation. In contrast to the JM approach, which uses a multivariate mixed model, the FCS approach employs a series of univariate mixed models, one for each incomplete variable. The FCS approach requires that a separate univariate imputation model be specified for each of the incomplete level-1 variables. FCS imputation predicts each incomplete variable as a function of all the other (filled-in) incomplete variables and all of the complete variables in a univariate mixed model:

$$\mathbf{y}_{jk} = \mathbf{X}_{jk}\boldsymbol{\beta}_k + \mathbf{Z}_{jk}\mathbf{b}_{jk} + \mathbf{e}_{jk} \tag{1.52}$$

In the above equation, $\mathbf{y}_{jk}$ is the n_j by 1 vector of values for incomplete level-1 variable *k* in cluster *j*. $\mathbf{y}_{jk}$ is used to represent each incomplete variable once. $\mathbf{X}_{jk}$ is the fixed effect covariate matrix predicting incomplete variable *k* for cluster *j*. $\mathbf{X}_{jk}$ contains all of the complete and imputed level-1 and level-2 variables specified as predictors for incomplete variable *k*, as well as a unit vector for the intercept. Note that in each imputation, $\mathbf{X}_{jk}$ is normally specified to include all of the (filled-in) incomplete variables except for the incomplete variable currently being imputed. X_{jk} makes no distinction between the independent variables and the dependent variable from the analysis model. $\boldsymbol{\beta}_k$ is the vector of fixed effects for variable *k* that are common for all clusters. $\mathbf{Z}_{jk}$ is the n_j by q_k random effect covariate matrix for cluster *j* for variable *k*. $\mathbf{Z}_{jk}$ contains the subset of level-1 variables (complete and imputed) that are allowed to have random effects on variable *k*. $\mathbf{b}_{jk}$ is the vector of level-2 residuals (i.e., residual intercepts and slopes) for cluster *j* for variable *k*. $\mathbf{e}_{jk}$ is the vector of level-1 residuals for cluster *j* for variable *k*.

Multilevel FCS imputation draws new values from a separate conditional normal distribution for each incomplete variable:

$$\mathbf{y}_{jk} \mid \mathbf{X}_{jk} \sim \mathrm{N}\left(\mathbf{X}_{jk}\boldsymbol{\beta}_k + \mathbf{Z}_{jk}\mathbf{b}_{jk}, \sigma_k^2\right) \tag{1.53}$$

That is, the missing values for incomplete variable *k* are drawn from a conditional univariate normal distribution, centered at the predicted value from the model for variable

k (i.e., $\mathbf{X}_{jk}\boldsymbol{\beta}_k + \mathbf{Z}_{jk}\mathbf{b}_{jk}$), with spread equal to the level-1 residual variance of variable *k* (i.e., σ_k^2). Notice that, in contrast to the JM imputation model, the random coefficient matrix for the FCS imputation model, $\mathbf{Z}_{jk}$, has an additional subscript, *k*. This means that a different set of random effects may be specified for each incomplete variable. This allows random effects to be included where they are needed and to be excluded when they would result in computational problems. Also worth noting is that, because each incomplete variable is predicted by all of the other incomplete variables, the regression of each incomplete variable on all other variables can be allowed to vary across clusters. If the analysis model uses the random effect of one incomplete variable to predict a second incomplete variable, this effect can be included in the imputation model, making it congenial. Remember, non-congenial imputation models lead to biased results. The issues described here are demonstrated in subsequent paragraphs. As an aside, van Buuren (2011, 2012) recommends the use of cluster-specific level-one residual variances in FCS imputation. This would entail replacing σ_k^2 with σ_{jk}^2. The use of cluster-specific level-1 residual variances is not a requirement of FCS imputation, but may increase the flexibility of the method with random slopes. The parameters in Equation 1.53 are generated from a Markov Chain Monte Carlo algorithm similar to that described in the single-level imputation section. Because these sampling steps are described throughout the literature (Browne & Draper, 2000; Goldstein, 2011; Goldstein et al., 2009; Kasim & Raudenbush, 1998; Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008), I do not detail them here.

Returning to the example used for JM multilevel imputation, recall that we wish to analyze the following model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 w_{ij} + b_{0j} + b_{1j} x_{ij} + b_{2j} w_{ij} + e_{ij} \quad (1.54)$$

Variables *y* and *w* are incomplete, whereas variable *x* is complete. The multilevel FCS algorithm would draw new values for one incomplete variable at a time. First, we start with *y*. A series of MCMC steps provides the parameter values and residuals. Then the algorithm draws *y* values from the following distribution:

$$y_{ij}^{(t)} \mid x_{ij}, w_{ij}^{(t-1)} \sim \mathrm{N}\left(\beta_0^{(y)} + \beta_1^{(y)} x_{ij} + \beta_2^{(y)} w_{ij}^{(t-1)} + b_{0j}^{(y)} + b_{1j}^{(y)} x_{ij} + b_{2j}^{(y)} w_{ij}^{(t-1)}, \sigma^2_{y|x,w}\right) \quad (1.55)$$

$\beta_0^{(y)}$ is the intercept for the imputation model predicting *y*. $\beta_1^{(y)}$ and $\beta_2^{(y)}$ are the fixed effects for predicting *y* from *x* and *w*. $b_{0j}^{(y)}$ is the random intercept for the imputation model predicting *y*. $b_{1j}^{(y)}$ and $b_{2j}^{(y)}$ are the random slopes for predicting *y* from *x* and *w*. $\sigma^2_{y|x,w}$ is the residual variance for *y*. Because FCS imputation uses multiple univariate imputation models, it is necessary to specify which model is referred to when describing model terms. The parameters in the imputation model each include a superscript to indicate which incomplete variable is being predicted. So, (*y*) indicates that $\beta_0^{(y)}$ is the intercept for the imputation model predicting *y*. Importantly, the *w* values on the right side of the equation are imputed from a previous step.

Once imputed, *y* becomes a predictor in the *w* imputation model. Again, a series of MCMC steps provide parameter values, and *w* values are drawn from the following distribution:

$$w_{ij}^{(t)} \mid x_{ij}, y_{ij}^{(t)} \sim \mathrm{N}\left(\beta_0^{(w)} + \beta_1^{(w)} x_{ij} + \beta_2^{(w)} y_{ij}^{(t)} + b_{0j}^{(w)} + b_{1j}^{(w)} y_{ij}^{(t)}, \sigma^2_{\mathrm{w}|x,\mathrm{y}}\right) \qquad (1.56)$$

$\beta_0^{(w)}$ is the intercept for the imputation model predicting *w*. $\beta_1^{(w)}$ and $\beta_2^{(w)}$ are the fixed effects for predicting *w* from *x* and *y*. $b_{0j}^{(w)}$ is the random intercept for the imputation model predicting *w*. $b_{1j}^{(w)}$ is the random slope for predicting *w* from *y*. $\sigma^2_{w|x,\mathrm{y}}$ is the residual variance for *w*. As mentioned previously, the *y* values on the right side of the equation are imputed from a previous step. The FCS method described above is implemented in the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011) and also in a standalone software package called BLImP (Keller & Enders, 2014). As of the writing of this dissertation, BLImP is not yet available to the public. FCS-VB is currently the only FCS multilevel imputation method implemented in publicly available software. I refer to this implementation of FCS as FCS-VB.

Carpenter and Kenward (2012, p. 221) state that the cluster means of the level-1 variables (complete or imputed incomplete) should be included as predictors in each of the univariate FCS imputation models. This approach, which I refer to as FCS-CK, is straightforward and very similar to FCS-VB imputation. However, users cannot add incomplete variable cluster means as predictors in MICE, the software implementation of

the FCS-VB method, because incomplete variables need to be recalculated at each MCMC iteration. The recalculation of incomplete variables cluster means at each MCMC iteration is not supported in MICE, so FCS-CK imputation cannot be performed in the MICE software package. Adding cluster means to Equations 1.55 and 1.56 would result in the following univariate imputation distributions:

$$y_{ij}^{(t)} \mid x_{ij}, w_{ij}^{(t-1)} \sim \mathrm{N}\begin{pmatrix} \beta_0^{(y)} + \beta_1^{(y)} x_{ij} + \beta_2^{(y)} w_{ij}^{(t-1)} + \beta_3^{(y)} \bar{x}_j + \beta_4^{(y)} \bar{w}_j^{(t-1)} \\ + b_{0j}^{(y)} + b_{1j}^{(y)} x_{ij} + b_{2j}^{(y)} w_{ij}^{(t-1)}, \sigma_{y|x,w}^2 \end{pmatrix} \quad (1.57)$$

$$w_{ij}^{(t)} \mid x_{ij}, y_{ij}^{(t)} \sim \mathrm{N}\begin{pmatrix} \beta_0^{(w)} + \beta_1^{(w)} x_{ij} + \beta_2^{(w)} y_{ij}^{(t)} + \beta_3^{(w)} \bar{x}_j \\ + \beta_4^{(w)} \bar{y}_j^{(t)} + b_{0j}^{(w)} + b_{1j}^{(w)} y_{ij}^{(t)}, \sigma_{\mathrm{w}|x,\mathrm{y}}^2 \end{pmatrix} \quad (1.58)$$

where $\bar{x}_j$ is the mean of x in cluster j, $\bar{w}_j$ is the mean of w in cluster j, and $\bar{y}_j$ is the mean of y in cluster j. After the addition of the cluster means, each univariate distribution includes two parameters to represent the regression of an incomplete variable on a predictor. For example, the regression of y on x in Equation 1.57 is captured by $\beta_1^{(y)}$ and $\beta_3^{(y)}$. $\beta_1^{(y)}$ captures the within-cluster regression of y on x and $\beta_3^{(y)}$ captures the difference between the between-cluster and within-cluster regressions of y on x. The inclusion of these two parameters allows the level-1 and level-2 coefficients for the regression of y on x to both freely vary. That is, the level-1 and level-2 regression coefficients can both be non-zero and can differ from one another. Similarly, the regressions of y on w and x on w can be different between levels one and two. FCS-CK is

implemented in a standalone software package called BLImP (Enders et al., 2014). As of the writing of this dissertation, BLImP is not yet available to the public.

The previous equations illustrate an important point about FCS imputation. In contrast to JM imputation, FCS imputation allows for the inclusion of random slopes between incomplete variables. For example, notice that *w* serves as a random slope predictor of *y* in Equations 1.55 and 1.57. The filled-in *y* values, in turn, serve as a random slope predictor of *w* in Equations 1.56 and 1.58. Additionally, imputing a single variable at a time allows the analyst to add additional features to the imputation process, such as skip logic. For example, it is desirable to avoid impossible combinations of responses such as pregnant and male. Due to these differences in the imputation models of JM and FCS imputation methods, FCS is more flexible than is the JM method.

Fully Conditional Specification vs. Joint Modeling. The FCS and JM approaches to multilevel imputation each have advantages and disadvantages. The bullets below highlight the advantages and disadvantages of the two approaches to multilevel imputation.

Benefits of JM:

- The JM approach is computationally more efficient for multiple incomplete variables because it requires fewer parameter draws (e.g., the level-2 covariance matrix is drawn once per iteration rather than once per incomplete variable per iteration).
- JM is usually easier to use, as the analyst need only specify a single imputation model (though some software packages allows the user to specify different random effects for each missing variable, at the cost of greater effort).

Disadvantages of JM:

- JM (as currently implemented in publicly available software) cannot be used to model the random effects of incomplete variables on other incomplete variables. This may be very problematic if such random effects are of interest in the analysis model. Though Yucel (2011) suggested a modified multilevel JM imputation method that may solve this problem, it has not yet been thoroughly tested and is not included in any publicly available software.
- Complex features of survey data, such as skip logic, are difficult to include in the imputation process.

Advantages of FCS:

- FCS can be used to model random effects of incomplete variables on other incomplete variables.
- FCS allows complex features of survey data, such as skip logic, to be incorporated in the imputation process.

Disadvantages of FCS

- FCS is harder to use, as the user must specify a distinct imputation model for each incomplete variable.
- As with single-level FCS, multilevel FCS can have convergence problems if the user specifies a set of distributions that do not correspond to a multivariate density.
- FCS is much less computationally efficient, and can take a long time to run.

Table 1 lists the software packages able to implement each of the five imputation methods examined in this dissertation.

Goals

A common approach to handling missing data is to perform multiple imputation once and later perform many analyses on the imputed data sets. If a multiply imputed data set is to be used in multiple analyses, the imputation model must be congenial with all of the analysis models to be used. This has led researchers to recommend a "kitchen-sink" approach to multiple imputation in which all variables and effects that may later be of interest are included in the imputation process. For example, Schafer and Olsen (1998) stated that "a rich imputation model that preserves a large number of associations is desirable because it may be used for a variety of post-imputation analyses." Rubin (1996) stated that "the press to include all possibly relevant predictors is demanding in practice, but it is generally a worthy goal." This issue has received little attention in the multilevel imputation literature, and researchers currently have few (if any) recommendations for choosing among competing multilevel imputation methods. The results from this study have practical implications for substantive research.

In the context of traditional multilevel models, there is a literature on contextual effects, whereby a predictor's influence differs between level-1 and level-2 (e.g., the influence of school-average SES on achievement differs from the influence of individual SES on achievement; the influence of daily pain fluctuations on positive affect differs from the influence of average or chronic pain on positive affect). "For theoretically important variables in multilevel studies, it is the rule rather than the exception that within-group regression coefficients differ from between-group regression coefficients" (Snijders & Bosker, 2012). Another situation where associations can differ across levels is multilevel structural equation modeling. For example, in a multilevel confirmatory

factor analysis model, the loadings and other estimates (even the structure itself) can differ between level-1 and level-2.

Given the fact that level-1 and level-2 correlations between variables can, and often do, differ from one another, attempting to include all possible relationships between variables in a multilevel data set in an imputation model requires using an imputation method that allows level-1 and level-2 correlations to differ. Researchers have several choices of imputation method (JM-BD, JM-UN, JM-Mplus, FCS-VB, and FCS-CK), and the ability of these methods to separately model level-1 and level-2 correlations between variables has not yet been examined.

In this chapter, I described the three JM imputation methods and the two FCS imputation methods that have been proposed for multilevel data. Although the previous illustrations demonstrated JM and FCS in the context of a random slope analysis model, my study examines JM and FCS for a population model consisting of a multivariate normal joint distribution containing only random intercepts. Because there is no literature examining the ability of either JM or FCS multilevel imputation models to preserve both level-1 and level-2 covariances between variables, as well as differences between these covariances across levels, this is a logical starting point.

Three variations of JM imputation have been proposed for multilevel data (Asparouhov & Muthén, 2010a, 2010f; Schafer, 2001; Schafer & Yucel, 2002). Although these methods share much in common, they possess subtle differences. To date, no methodological research has investigated the differences among the JM models, in particular their ability to produce imputations that preserve characteristics of the population distribution. Thus, one of the overarching goals for this project was to

examine the situations under which the three JM methods reproduce (or preserve) the mean and covariance structure of a population random intercept model with multivariate normal data. To date, two multilevel FCS methods have been proposed in the literature, only one of which is currently implemented in statistical software (Carpenter & Kenward, 2012, p. 221; van Buuren & Groothuis-Oudshoorn, 2011). Like JM, no methodological research has investigated the differences between the FCS methods or their ability to preserve a multilevel data structure. Thus, the second overarching goal for this project was to examine the situations under which FCS imputation reproduces the mean and covariance structure of a population random intercept model with multivariate normal data. The analytic work for these two goals also provided insight into the situations where JM and FCS methods are equivalent.

Chapter 2. Literature Review

Multilevel Imputation of Continuous Variables – Literature Review

This chapter provides an overview of the literature on multilevel multiple imputation. Specifically, this chapter describes published simulation studies that examine multilevel multiple imputation for continuous variables. Any article that does not include a simulation study is not included in this chapter. In describing multilevel multiple imputation in Chapter 1, I referenced multiple published works that proposed the methods examined in this document or which proposed modifications to the methods examined in this document (e.g., Asparouhov & Muthén, 2010f; Carpenter & Kenward, 2012; Schafer, 2001; Schafer & Yucel, 2002; van Buuren & Groothuis-Oudshoorn, 2011; Yucel, 2008, 2011; Yucel, Schenker, & Raghunathan, 2006). Though some of these publications included an illustrative analysis to demonstrate a proposed method, none of the publications included a simulation study and as such they are not reviewed in this chapter. The brevity of this chapter shows the dearth of simulation studies examining multilevel imputation methods in the literature and in doing so emphasizes the need for such simulation studies.

Taljaard, Donner, and Klar (2008) performed a simulation study to examine the usefulness of JM imputation for analyzing data from cluster randomized trials (CRTs). To mimic CRT data, the authors specified a population model containing a dichotomous level-2 predictor and a random intercept. No other predictors were included. The effect size was set to zero to measure type-1 error rate. The authors varied cluster size between 30 and 500. The authors varied the number of clusters between 6 and 30. They also varied the intraclass correlation between .001 and .10. The dependent variable was

specified to have a 30% chance of missingness under an MCAR mechanism. The authors also varied cluster imbalance in missingness, ranging from low imbalance where all clusters had equal rates of missingness and high imbalance where entire clusters were missing. The authors generated 10 imputed data sets using the PAN library in SPlus. The description of the imputation method did not mention which implementation of PAN was used (JM-BD or JM-UN). The authors specified 2000 burn-in iterations and 199 between-imputation iterations. The analysis model consisted of a two-sample t-test to compare the treatment groups. Note that this analysis model is not consistent with the data. If the clusters had equal sizes, then MLM would give identical results to a t-test. However, when the clusters have unequal sizes the two methods are not identical. As such, the results of this study should be interpreted with caution.

For conditions with a small intraclass correlation (i.e., $\rho = .001$), multilevel imputation resulted in type-1 errors that tended to be below nominal (ranging from .004 to .062). The range of type-1 error rates comes from the authors varying the sample size at both level 1 and level 2. Because the authors varied the sample size at both levels simultaneously, it was not possible to detect whether the type-1 error rate varied systematically cross sample size conditions. For conditions with a higher intraclass correlation (i.e., $\rho = .10$), the type-1 error rate stayed closer to the nominal value (ranging from .035 to .052). The authors did not report other outcome measures.

A simulation study by Andridge (2011) examined JM imputation for two-level data. The number of level-2 units was held constant at 50, and the number of level-1 units per cluster was held constant at 20. The population model consisted of a single level-1 predictor and a random intercept. ICC values varied between 0.001 and 0.5. The

correlation between the outcome and the level-1 predictor varied between 0 and 0.9. The dependent variable was made missing at a rate of .3 for MCAR. For the MAR condition, the missingness rate depended on the level-1 predictor. Multiple imputation was performed using the PAN package in R. The description of the imputation method did not mention which implementation of PAN was used (JM-BD or JM-UN). Note that the imputation model contained the level-1 predictor from the population model, but this predictor was not included in the analysis model. The authors generated 10 imputed data sets using 1000 burn-in iterations and 99 between-imputation iterations.

The analysis goal was to examine the variance of the unconditional mean of the dependent variable. To rephrase, the analysis model contained only the intercept as a predictor, and the intercept was allowed to vary across clusters. The variance of the intercept across clusters was the parameter of interest in the simulation. The authors assessed the performance of the method by examining the coverage rate of the bootstrapped 95% confidence interval. Across all simulation conditions, the 95% confidence interval coverage rate was close to nominal (ranging between 94.3 and 97.7).

A simulation study by van Buuren (2011) examined the performance of FCS-VB imputation for multilevel data under the assumption of MAR missingness. The author simulated data based on a population model containing a level-1 predictor and a random intercept, but without a random slope. The study contained three missingness conditions: 1) incomplete dependent variable; 2) incomplete level-1 predictor; and 3) incomplete dependent variable and incomplete predictor. Intraclass correlations ranged from 0.0 to 0.67. The total number of observations was set to 1,200. The number of level-2 units and the number of level-1 units per cluster were varied simultaneously. That is, the following

conditions were included: 12 clusters with 100 units each; 24 clusters with 50 units each, and 60 clusters with 20 units each. Probability of nonresponse ranged from 10% to 90%. The authors carried out the imputation procedure using the R package, MICE. The authors generated 5 imputed datasets. Each imputation was generated using a separate imputation chain with 20 burn-in iterations. The author repeatedly found 95% confidence interval coverage rates below nominal (discussed in greater depth below). It is possible that the use of only 20 burn-in iterations resulted in these poor coverage rates, as the MCMC algorithm can take a long time to become stable. Though the author made claims regarding whether estimates were biased or unbiased, the statistics necessary for computing bias were not reported in the publication. As such, this article summary merely copies the author's claims regarding bias.

When only the dependent variable was missing, FCS-VB imputation produced unbiased estimates of both the fixed and random effects. The 95% confidence interval coverage rate for the fixed intercept term was below nominal, ranging from .84 to .97. The 95% confidence interval coverage rate for the fixed slope term was slightly below nominal, ranging from .86 to .94. This coverage did not appear to change across ICC conditions or cluster size conditions. That is, though the coverage rates differed across experimental conditions they did not do so in any systematic way. The author did not explain this phenomenon.

When only a level-1 predictor was incomplete, FCS-VB imputation resulted in unbiased estimates for both the fixed and random effects. The 95% confidence interval coverage rate for the fixed intercept term was very poor, ranging from .87 to .95. This coverage did not appear to change across ICC conditions or cluster size conditions. That

is, though the coverage rates differed across sampling conditions they did not do so in any systematic way. The author did not explain this phenomenon. The 95% confidence interval coverage rate for the fixed slope term was close to nominal for the large sample size condition (100 level-1 units per cluster), ranging from .93 to .95, but was below nominal for the small sample size condition (20 level-1 units per cluster), ranging from .79 to .85.

When both the dependent variable and the level-1 predictor were incomplete, FCS-VB imputation produced positively biased estimates of the fixed intercept term (estimates around .08 compared to the complete data estimate of 0). The procedure also produced negatively biased estimates of the fixed slope term (estimates around .44 compared to the complete data estimate of .5). The confidence interval coverage rate for the fixed slope parameter was atrocious when the ICC was equal to zero (ranging from .34 to .37). For the large ICC condition, however, the coverage rates were much better (.90 for the large sample size condition and .82 for the small sample size condition). The confidence interval coverage rates for the fixed slope parameter appeared to be affected by the sample size but not by the ICC. For the small sample size condition the coverage rate ranged from .42 to .57. For the large sample size condition the coverage rate ranged from .76 to .85. Though the coverage rates varied across ICC conditions within a particular sample size, they did not do so in any systematic fashion.

Conclusions from the Multilevel Imputation Literature

The simulation studies by Taljaard et al. (2008) and Andridge (2011) both examine JM multilevel imputation using the PAN method, though neither paper specified which version of PAN (JM-BD or JM-UN) was used. Both studies used a random

intercept model as their population model. One of the two studies included a level-2 predictor, but neither study included a level-1 predictor. As such, neither study provides information about whether JM imputation preserves the covariances between pairs of level-1 variables. Related to my proposed work, the studies do provide information about whether JM imputation preserves other aspects of the joint distribution. The study by Taljaard et al. (2008) showed that JM imputation can preserve the regression of a level-1 variable on a level-2 predictor. In terms of the current study, this indicates that the covariances between the cluster means (which are level-2 variables) and the level-1 variables should be correctly preserved. The study by Andridge (2011) also indicated that JM imputation can preserve the variance of cluster means.

The simulation study by van Buuren (2011) generated the data under a random intercept model with a single level-1 predictor. The regression of the dependent variable on the level-1 predictor was modeled by a single variable, constraining the level-1 and level-2 regression coefficients to be equal. Recall from Chapter 1 that the FCS-VB method (which was used for the imputation in this study) assumes that the coefficients for the regression of each incomplete variable on all other variables are constant across levels. Because the population model was constrained to match the assumptions of the imputation model, the simulation study does not address potential problems that might arise from applying FCS-VB imputation to population data where regression coefficients differ between level 1 and level 2. Though the study by van Buuren (2011) did not address the ability of FCS-VB imputation to preserve contextual effects (when regression coefficients differ between level-1 and level-2) the study did show that FCS-VB imputation can preserve the covariance between pairs of level-1 variables when no

contextual effect is present. The study also indicated that the level-1 residual variances are correctly preserved.

In summary, the current literature does not examine whether any of the three JM imputation methods (JM-BD, JM-UN, or JM-Mplus) preserve the covariances among pairs of level-1 variables. The simulation study by van Buuren (2011) examined FCS-VB imputation under the strict assumption that the level-1 and level-2 coefficients for regressions of level-1 variables on other level-1 variables are equal in the population. JM and FCS imputation methods both need to be evaluated for their ability to preserve relationships between variables when level-1 and level-2 regression coefficients differ from one another and from zero.

Multilevel Models and the Population Joint Distribution

As discussed in Chapter 1, no publications have shown the theoretical equivalence of the multilevel imputation methods examined in this document and the population random intercept distribution. However, Shin and Raudenbush (2007) showed the equivalence between the model-implied moments of a univariate multilevel model and the moments of a population random intercept distribution. Specifically, the authors showed that the parameters of a univariate multilevel model can be directly calculated from the parameters of the joint distribution. Their approach provides the basis for the analytic examination of the multilevel imputation methods proposed in this document. I review Shin and Raudenbush (2007) here to provide background on the method I used to evaluate the multilevel imputation methods. For pedagogical reasons, I limit the review to the case of two level-1 variables with random intercepts. For consistency, I replace the notation used in Shin and Raudenbush (2007) with the notation that used in Chapter 3.

Consider the following random intercept model predicting level-1 variable *Y* from level-1 variable *X*:

$$\begin{aligned} Y_{ij} &= \gamma_0 + \gamma_1 X_{ij} + u_j + e_{ij} \\ u_j &\sim N(0, \tau^2) \\ e_{ij} &\sim N(0, \sigma^2) \end{aligned} \tag{2.1}$$

γ_0 is the intercept, γ_1 is the fixed effect of X_{ij} on Y_{ij}, u_j is the random intercept, and e_{ij} is the level-1 residual. τ^2 is the level-2 variance of the random intercept, and σ^2 is the level-1 residual variance. For the sake of completeness, we should also describe the distribution of the predictor variable, *X*:

$$X_{ij} \sim \mathrm{N}(\beta_x, \psi_{xx} + \sigma_{xx}) \tag{2.2}$$

where β_x is the mean of *X*, ψ_{xx} is the level-2 variance of *X*, and σ_{xx} is the level-1 variance of *X*. Though a multilevel modeling analysis typically does not report the distribution of the predictors in the results, the information contained in the distribution of the predictors is used to calculate the parameters in the analysis model. As such, the parameters describing the distribution of the predictors can be thought of as implicit parameters in the analysis model.

The random intercept population model for the two variables would be:

$$\begin{bmatrix} X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} \beta_x \\ \beta_y \end{bmatrix} + \begin{bmatrix} b_{xj} \\ b_{yj} \end{bmatrix} + \begin{bmatrix} \varepsilon_{xij} \\ \varepsilon_{yij} \end{bmatrix}$$
$$\begin{bmatrix} b_{xj} \\ b_{yj} \end{bmatrix} \sim \mathrm{N}\left(0, \begin{bmatrix} \psi_{xx} & \psi_{xy} \\ \psi_{xy} & \psi_{yy} \end{bmatrix}\right) \qquad (2.3)$$
$$\begin{bmatrix} \varepsilon_{xij} \\ \varepsilon_{yij} \end{bmatrix} \sim \mathrm{N}\left(0, \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix}\right)$$

where β_x and β_y are the means of *X* and *Y*, b_{xj} and b_{yj} are the level-2 error terms for *X* and *Y*, and ε_{xij} and ε_{yij} are the level-1 residuals for *X* and *Y*. The **Ψ** and **Σ** matrices are the level-2 and level-1 covariance matrices for *X* and *Y*.

Comparing the analysis model and the population model, we see that they differ in the number of parameters. Specifically, the analysis model contains seven parameters (γ_0, γ_1, τ^2, σ^2, β_x, ψ_{xx}, and σ_{xx}) including the parameters describing the distribution of the predictors, and the population model contains eight parameters (β_x, β_y, ψ_{xx}, ψ_{xy}, ψ_{yy}, σ_{xx}, σ_{xy}, and σ_{yy}). As will be shown below, the analysis model describes the regression of *Y* on *X* at both level-1 and level-2 using a single parameter, γ_1, whereas the population model uses two parameters, ψ_{xy} and σ_{xy}, to capture the covariances between the two varables. As such, the analysis model is underparameterized in representing the population model. This also means that the parameters of the population model cannot be obtained from the parameters of the analysis model. The parameters in the analysis model, Equations 2.1 and 2.2, can be calculated directly from the multilevel joint distribution, as shown below. The parameters from the analysis model are listed on

the left side of each equation and the parameters from the population model are listed on the right side of each equation.

$$\beta_x = \beta_x \tag{2.4}$$

$$\sigma_{xx} = \sigma_{xx} \tag{2.5}$$

$$\psi_{xx} = \psi_{xx} \tag{2.6}$$

$$\sigma^2 = \sigma_{yy} - \frac{\left(\psi_{yx} + \sigma_{yx}\right)^2}{\psi_{xx} + \sigma_{xx}} \tag{2.7}$$

$$\tau^2 = \psi_{yy} - \frac{\left(\psi_{yx} + \sigma_{yx}\right)^2}{\psi_{xx} + \sigma_{xx}} \tag{2.8}$$

$$\gamma_0 = \beta_y - \gamma_1 \beta_x \tag{2.9}$$

$$\gamma_1 = \frac{\psi_{yx} + \sigma_{yx}}{\psi_{xx} + \sigma_{xx}} \tag{2.10}$$

Because the cluster means for *Y* were not included in the analysis model, a single parameter, γ_1, was used to represent the regression of *Y* on *X* at both level one and level two. Because of this, the slope coefficient for the regression of *Y* on *X* is a combination of the level-1 and level-2 covariances between the variables (equivalently, a combination of the level-1 and level-2 regression coefficients) as shown in Equation 2.10. This estimation of level-1 and level-2 regression coefficients as a single parameter is also required by some of the multilevel imputation methods (i.e., FCS-VB, as well as JM-BD and JM-UN when cluster means are not included). Because this method of equating a

joint distribution with an analysis (or imputation) model makes the links between joint distribution parameters and analysis/imputation model parameters explicit, it highlights areas of the joint distribution that may not be preserved by the analysis/imputation model (e.g., the aliasing of level-1 and level-2 covariances in Equation 2.10 indicates that the level-1 and level-2 covariances may not be fully preserved by the analysis model). As such, the method is useful for evaluating each of the multilevel imputation methods.

If an analysis model is underparameterized in representing the population joint distribution, the parameters of the analysis model can be obtained from the parameters of the population model but it is not possible to obtain the parameters of the population model from the parameters of the analysis model when the analysis model is underparameterized in representing the population model. This concept is important for the analytic work in this study, described in Chapter 3. That is, the analytic piece of the dissertation study shows the calculation of the parameters in each imputation model from the population joint model. However, because some of the imputation models are underparameterized in representing the population model, the population model parameters cannot be calculated from the imputation model parameters, indicating that those imputation models are incapable of fully preserving all of the information contained in the population model. For imputation models that are not underparameterized in representing the population joint distribution, I show the calculation of the population parameters from the imputation model parameters to demonstrate the ability of these models to preserve the information in the population model.

Although JM and FCS imputation methods for single-level data have been studied with analytic methods (Hughes et al., 2014), no studies have examined JM and FCS for

multilevel data. Also, the simulation studies of multilevel imputation (described in the previous section) provide limited evidence of the capability of multilevel imputation methods to preserve population covariances between level-1 variables. Thus, the goal of this dissertation was to examine the situations under which the three JM methods and the two FCS imputation methods reproduce (or preserve) the mean and covariance structure of a population random intercept model with multivariate normal data. The analytic work was based on the article by Shin and Raudenbush (2007). The next chapter describes the methods that I used to investigate this issue.

Chapter 3. Methods

Imputation Methods

Chapter 1 categorized multiple imputation methods into two categories: JM and FCS. Recall that Chapter 1 described three JM multilevel imputation methods and two FCS multilevel imputation methods: JM-BD (the PAN method with a block-diagonal level-2 residual covariance matrix), JM-UN (the PAN method with an unstructured level-2 residual covariance matrix), JM-Mplus (JM imputation treating all variables as responses and using an unstructured level-2 covariance matrix), FCS-VB (van Buuren's FCS method implemented in MICE), and FCS-CK (FCS imputation with added cluster means for all level-1 variables). I briefly recap the five multilevel imputation methods here with an emphasis on their differences, as these methods were be the focus of this study. Two of the JM methods are variants of the PAN method (Schafer, 2001; Schafer & Yucel, 2002). In the PAN method, the incomplete variables are modeled conditional upon the complete variables:

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{B}_j + \mathbf{E}_j \tag{3.1}$$

where $\mathbf{Y}_j$ is the n_j by r matrix of incomplete level-1 variables for cluster j and r is the number of incomplete level-1 variables. Note that $\mathbf{Y}_j$ contains all of the incomplete level-1 variables, regardless of their role in the analysis model. $\mathbf{X}_j$ is the n_j by f fixed effect covariate matrix for cluster j, where f is the number of fixed effect covariates. $\mathbf{X}_j$ contains all of the complete level-1 and level-2 variables, as well as a unit vector for the

intercept. $\boldsymbol{\beta}$ is the f by r matrix of fixed effects that are common for all clusters. $\mathbf{Z}_j$ is the n_j by q random effect covariate matrix for cluster j, where q is the number of random effects. $\mathbf{Z}_j$ contains the subset of complete level-1 variables that are allowed to have random effects with the variables in $\mathbf{Y}_j$, as well as a vector of ones for the random intercepts. $\mathbf{B}_j$ is the q by r matrix of level-2 residuals for cluster j. Although $\mathbf{B}_j$ can contain both random intercepts and random slopes, this study focuses solely on random intercepts. $\mathbf{E}_j$ is the n_j by r matrix of level-1 residuals for cluster j.

The parameter values and residual terms in Equation 3.1 are obtained from an iterative MCMC algorithm that draws the necessary terms from their theoretical probability distributions. The parameters in Equations 3.1, 3.2, and 3.3 are generated from a Markov Chain Monte Carlo algorithm similar to that described in the single-level imputation section. Because these sampling steps are described throughout the literature (Browne & Draper, 2000; Goldstein, 2011; Goldstein et al., 2009; Kasim & Raudenbush, 1998; Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008), I do not detail them here. Rather, I focus on the structure of the residual covariance matrices, as these play an important role in imputation. For each MCMC iteration, one must draw new residuals for each observation and new random effects for each cluster:

$$\mathbf{e}_{ij} \sim \mathrm{N}(0, \boldsymbol{\Sigma}) \tag{3.2}$$

$$\mathrm{vec}(\mathbf{b}_j) \sim \mathrm{N}(0, \boldsymbol{\Psi}) \tag{3.3}$$

where $\mathbf{e}_{ij}$ is the vector of residuals for observation *i* in cluster *j*, $\boldsymbol{\Sigma}$ is the level-1 residual covariance matrix, $\text{vec}(\mathbf{b}_j)$ refers to stacking all of the random effects for cluster *j* in a column vector, and $\boldsymbol{\Psi}$ is the level-2 random effects covariance matrix. $\boldsymbol{\Sigma}$ preserves the residual level-1 associations among the incomplete variables and $\boldsymbol{\Psi}$ preserves the residual level-2 associations among the incomplete variables. Schafer proposed two options for $\boldsymbol{\Psi}$: an unstructured covariance matrix or a block-diagonal covariance matrix. The unstructured level-2 covariance matrix allows the random effects for any two incomplete variables to be correlated. The unstructured level-2 matrix allows the residual cluster-level associations to freely vary in ways that may or may not be the same as the residual level-1 associations between the variables, which are captured in the level-1 residual covariance matrix, $\boldsymbol{\Sigma}$. To illustrate, consider a scenario with two incomplete variables, both of which are imputed under a random intercepts model. The unstructured covariance matrix allows the intercepts to covary, such that the association between the cluster means at level-2 can differ from the level-1 association, which is captured by the covariance between the level-1 error terms. As a reminder, I refer to the PAN method with an unstructured random-effect covariance matrix as the JM-UN method.

In some cases, the number of parameters to be estimated in the unstructured level-2 covariance matrix may be large with many incomplete variables. If the number of clusters in a data set is small, it may not be possible to estimate covariances among all of the random effects for all of the incomplete variables (Schafer & Yucel, 2002). In such cases, Schafer suggests that it may be advantageous to model $\boldsymbol{\Psi}$ as a block-diagonal matrix in which the covariances between random effects for different incomplete variables are constrained to zero:

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Psi}_r \end{bmatrix} \tag{3.4}$$

where *r* is the number of incomplete variables. In contrast to the unstructured covariance matrix, the block diagonal structure presumes that the intercepts are uncorrelated. Under this specification the residual level-1 association is still captured by the residual covariance between the level-1 error terms. However, the residual covariances between the cluster means at level-2 are assumed to be nonexistent. If the data contains any residual level-2 covariances between pairs of incomplete variables, the imputation model constrains the sizes of these covariances to zero. To clarify, this is not the same as constraining the level-1 and level-2 residual covariances to be equal. In a situation where all of the variables are incomplete, the block-diagonal structure assumes that the level-2 cluster means are uncorrelated. If some of the variables are complete (and thus serve as predictors of the incomplete variables), the block-diagonal structure assumes conditional independence of the incomplete variable cluster means. That is, the cluster means are uncorrelated apart from their mutual dependence on the X variables. Constraining the level-1 and level-2 residual covariances to be equal would allow both the level-1 and level-2 residual covariances to be non-zero as long as they are equal. The block diagonal psi matrix, on the other hand, places no constraints on the residual level-1 covariances between the variables but constrains the covariances between the random intercepts (and thus the residual level-2 covariances between the variables) to be zero. As a reminder, I

refer to the PAN method with a block-diagonal random-effect covariance matrix as the JM-BD method.

We turn now to the associations between the complete and incomplete variables (i.e., between the variables in matrix **X** and the variables in matrix **Y** in Equation 3.1). For the two PAN imputation methods (JM-UN and JM-BD), a single parameter is used to estimate the regression of an incomplete variable on a complete variable in the imputation model by default. This means that the level-2 regression coefficients and the level-1 regression coefficients predicting the incomplete variables from the complete variables are not allowed to differ from one another. As such, level-1 and level-2 regression coefficients predicting incomplete variables from complete variables will be constrained to equality in the imputation model. Note that the analyst can attempt to remedy this by adding the cluster-means of the complete variables to the data set, and including these calculated variables as predictors. This would result in the estimation of two parameters for each incomplete-complete variable pair, one measuring the within-cluster regression and the other measuring the difference between the within-cluster and between-cluster regression coefficients. However, it is likely that the majority of users of multilevel imputation software are unaware of the need to include complete variable cluster means as predictors in their imputation models. As such, this document focuses on the default models used for each of the methods, and does not examine modifications to the JM methods.

The two PAN methods just described (which differ only in the structure of $\boldsymbol{\Psi}$) both modeled the incomplete variables as a function of the complete variables. In contrast, the third JM imputation method, the so-called H1 imputation approach in Mplus

(which I refer to as JM-Mplus), treats all variables, both incomplete and complete, as dependent variables and includes no predictors in the model (Asparouhov & Muthén, 2010a, 2010f):

$$\mathbf{y}_{ij} = \boldsymbol{\beta} + \mathbf{b}_j + \mathbf{e}_{ij} \tag{3.5}$$

where $\mathbf{y}_{ij}$ is a vector of level-1 variables for level-1 unit *i* in cluster *j*. $\mathbf{y}_{ij}$ contains all of the level-1 variables, both complete and incomplete. $\boldsymbol{\beta}$ is a vector containing the grand mean for each of the variables. $\mathbf{b}_j$ contains the random intercepts for cluster *j*. $\mathbf{e}_{ij}$ contains the discrepancies between the cluster intercepts for cluster j and the values (observed or imputed) for observation *i* in cluster *j*. Note that all of the objects in Equation 3.5 are vectors, as the equation describes a single observation, *i*, and multiple variables. Matrices $\mathbf{X}_j$ and $\mathbf{Z}_j$ from Equation 3.1 are not shown in Equation 3.5, as the two matrices reduce to scalars equal to 1. Vectors $\mathbf{e}_{ij}$ and $\mathbf{b}_j$ are distributed as:

$$\mathbf{e}_{ij} \sim \mathrm{N}(0, \boldsymbol{\Sigma}) \tag{3.6}$$

$$\mathbf{b}_j \sim \mathrm{N}(0, \boldsymbol{\Psi}) \tag{3.7}$$

where $\boldsymbol{\Sigma}$ is the unstructured level-1 covariance matrix and $\boldsymbol{\Psi}$ is the unstructured level-2 covariance matrix. Although it is not relevant to this study, notice that the method does not allow for the inclusion of random slopes among ANY variables. This is in contrast to the PAN JM methods, which allow complete variables to exert random influences on the

incomplete variables. Because JM-Mplus models the entire level-1 covariance matrix and the entire level-2 covariance matrix, the within-cluster associations and the between-cluster associations are freely estimated for all variables in the model. This means that the within-cluster and between-cluster associations between a pair of variables can both be non-zero and can differ in their magnitude and sign. JM-Mplus allows level-1 covariances and level-2 covariances between pairs of complete and incomplete variables to differ, which PAN methods do not allow (though PAN can do this for pairs of complete/incomplete variables if the complete variable cluster means are added as predictors).

To recap, this section described three JM multilevel imputation methods: the JM-BD method, the JM-UN method, and JM-Mplus. The two PAN methods both predict the incomplete variables from the complete variables. Unless the analyst calculates the complete variable cluster means and includes these cluster means as predictors in the imputation model as level-2 predictors, both PAN methods constrain the level-1 and level-2 coefficients for the regression of the incomplete variables on the complete variables to be equal,. The JM-BD method constrains the residual covariances between the incomplete variable intercepts to be zero, whereas the JM-UN method does not. This implies that the level-2 covariances between pairs of incomplete variables above and beyond what is accounted for by regression on the complete variables is constrained to be zero. Unlike the PAN methods, JM-Mplus treats all level-1 variables (complete and incomplete) as response variables in the imputation model. JM-Mplus freely estimates the level-1 and level-2 covariances between all of the parameters without constraining any parameters to equality or to zero.

In Chapter 1, I presented two FCS multilevel imputation methods. Both FCS methods impute variables one at a time, drawing missing values from a series of univariate distributions. Each univariate distribution is used to draw a single incomplete variable conditional on the complete variables and all other incomplete, imputed variables.

$$\mathbf{y}_{jk} = \mathbf{X}_{jk}\boldsymbol{\beta}_k + \mathbf{Z}_{jk}\mathbf{b}_{jk} + \mathbf{e}_{jk} \tag{3.8}$$

In the above equation, $\mathbf{y}_{jk}$ is the n_j by 1 vector of values for incomplete level-1 variable *k* in cluster *j*. $\mathbf{y}_{jk}$ is used to represent each incomplete variable once. $\mathbf{X}_{jk}$ is the fixed effect covariate matrix predicting incomplete variable *k* for cluster *j*. $\mathbf{X}_{jk}$ contains all of the complete and imputed level-1 and level-2 variables specified as predictors for incomplete variable *k*, as well as a unit vector for the intercept. Note that in each imputation $\mathbf{X}_{jk}$ is normally specified to include all of the (filled-in) incomplete variables except for the incomplete variable currently being imputed. $\mathbf{X}_{jk}$ makes no distinction between the independent variables and the dependent variable from the analysis model. $\boldsymbol{\beta}_k$ is the vector of fixed effects for variable *k* that are common for all clusters. $\mathbf{Z}_{jk}$ is the n_j by q_k random effect covariate matrix for cluster *j* for variable *k*. $\mathbf{Z}_{jk}$ contains the subset of level-1 variables (complete and imputed) that are allowed to have random effects on variable *k*. $\mathbf{b}_{jk}$ is the vector of level-2 residuals (i.e., residual intercepts and slopes) for cluster *j* for variable *k*. $\mathbf{e}_{jk}$ is the vector of level-1 residuals for cluster *j* for

variable *k*. The parameters in Equation 3.8 are generated from a Markov Chain Monte Carlo algorithm similar to that described in the single-level imputation section. Because these sampling steps are described throughout the literature (Browne & Draper, 2000; Goldstein, 2011; Goldstein et al., 2009; Kasim & Raudenbush, 1998; Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008), I do not detail them here.

Of the two FCS methods presented in Chapter 1, the FCS-VB method is the only FCS multilevel imputation method implemented in publicly available software (van Buuren & Groothuis-Oudshoorn, 2011). Chapter 1 contained an example of the FCS-VB imputation method in which two variables, *y* and *w*, were imputed using two conditional imputation models. For the sake of clarity, I will now use the variable names *y1* and *y2* for the incomplete variables, and x for the complete variable. Because this document focuses on random intercept imputation models for simplicity, we can exclude random slopes. In the FCS-VB method, the univariate distributions for *y1* and *y2* are:

$$y1_{ij}^{(t)} \mid x_{ij}, \mathrm{y}\,2_{ij}^{(t-1)} \sim \mathrm{N}\left(\beta_0^{(y1)} + \beta_1^{(y1)} x_{ij} + \beta_2^{(y1)}\,\mathrm{y}\,2_{ij}^{(t-1)} + b_{0j}^{(y1)}, \sigma^2_{y1|y2,x}\right) \qquad (3.9)$$

$$y2_{ij}^{(t)} \mid x_{ij}, y1_{ij}^{(t)} \sim \mathrm{N}\left(\beta_0^{(y2)} + \beta_1^{(y2)} x_{ij} + \beta_2^{(y2)} y1_{ij}^{(t)} + b_{0j}^{(y2)}, \sigma^2_{y2|y1,x}\right) \qquad (3.10)$$

$\beta_0^{(y1)}$ is the intercept for the imputation model predicting *y1*. $\beta_1^{(y1)}$ and $\beta_2^{(y1)}$ are the fixed effects for predicting *y1* from *x* and *y2*. $b_{0j}^{(y1)}$ is the random intercept for *y1*. $\sigma^2_{y1|y2,x}$ is the residual variance for *y1*. $\beta_0^{(y2)}$ is the intercept for the imputation model

predicting *y2*. $\beta_1^{(y2)}$ and $\beta_2^{(y2)}$ are the fixed effects for predicting *y2* from *x* and *y1*. $b_{0j}^{(y2)}$ is the random intercept for *y2*. $\sigma^2_{y2|y1,x}$ is the residual variance for *y2*.

The FCS-VB method uses a single parameter to capture the regression of each incomplete variable on each other variable. That is, the level-1 and level-2 coefficients for the regression of an incomplete variable on a complete variable are represented a single regression coefficient, constraining the level-1 and level-2 regression coefficients to be equal. In Equation 3.9 the entire regression of *y1* on *x* is captured by $\beta_1^{(y1)}$. In Equation 3.10 the entire regression of *y2* on *x* is captured by $\beta_1^{(y2)}$. Similarly, the level-1 and level-2 coefficients for the regression of one incomplete variable on another incomplete variable are represented with a single parameter, constraining the two level-1 and level-2 coefficients to be equal. In Equation 3.9 the entire regression of *y1* on *y2* is captured by $\beta_2^{(y1)}$. In Equation 3.10 the entire regression of *y2* on *y1* is captured by $\beta_2^{(y2)}$. Because of this, the level-1 and level-2 coefficients for the regression of each dependent variable on each predictor variable are constrained to be equal. The FCS-VB method makes the same assumption about the regression of each variable on each other variable that the PAN methods make about the regression of each incomplete variable on each incomplete variable: the level-1 and level-2 regression coefficients are equal.

Although the approach has not been implemented in publicly available software, Carpenter and Kenward (2012, p. 221) state that the cluster means of each level-1 variable (complete or imputed incomplete) should be included as predictors in each of the univariate imputation models used for FCS. This method, which I refer to as FCS-CK, is

straightforward and very similar to the FCS-VB imputation method. However, users cannot add incomplete variable cluster means as predictors in MICE because incomplete variables need to be recalculated at each MCMC iteration. The recalculation of incomplete variables cluster means at each MCMC iteration is not supported in MICE, so FCS-CK imputation cannot be performed in the MICE software package. Adding cluster means to Equations 3.9 and 3.10 would result in the following univariate imputation distributions:

$$y1_{ij}^{(t)} \mid x_{ij}, \mathrm{y}\,2_{ij}^{(t-1)} \sim \mathrm{N}\begin{pmatrix} \beta_0^{(y1)} + \beta_1^{(y1)} x_{ij} + \beta_2^{(y1)} y2_{ij}^{(t-1)} + \beta_3^{(y1)} \bar{x}_j \\ +\beta_4^{(y1)} \overline{y2}_j^{(t-1)} + b_{0j}^{(y1)}, \sigma^2_{y1|y2,x} \end{pmatrix} \tag{3.11}$$

$$y2_{ij}^{(t)} \mid x_{ij}, y1_{ij}^{(t)} \sim \mathrm{N}\begin{pmatrix} \beta_0^{(y2)} + \beta_1^{(y2)} x_{ij} + \beta_2^{(y2)} y1_{ij}^{(t)} + \beta_3^{(y2)} \bar{x}_j \\ +\beta_4^{(y2)} \overline{y1}_j^{(t)} + b_{0j}^{(y2)}, \sigma^2_{y2|y1,x} \end{pmatrix} \tag{3.12}$$

where $\bar{x}_j$ is the mean of *x* in cluster *j*, $\overline{y2}_j$ is the mean of *y2* in cluster *j*, and $\overline{y1}_j$ is the mean of *y1* in cluster *j*. With the addition of the cluster means, each univariate distribution now includes two parameters to represent the regression of an incomplete variable on a predictor. For example, the regression of *y1* on *x* in Equation 3.11 is captured by $\beta_1^{(y1)}$ and $\beta_3^{(y1)}$. $\beta_1^{(y1)}$ captures the within-cluster regression of *y1* on *x* and $\beta_3^{(y1)}$ captures the difference between the between-cluster and within-cluster coefficients for the regression of *y1* on *x*. The inclusion of these two parameters allow the level-1 and level-2 coefficients for the regression of *y1* on *x* to both freely vary. That is, the level-1

and level-2 regression coefficients can both be non-zero and not constrained to equality. Similarly, the coefficients for the regressions of *y2* on *y1*, *y2* on *x*, and *y1* on *y2* can be different between levels 1 and 2.

In summary, this document examines the following five multilevel imputation methods:

1) JM imputation predicting the incomplete variables from the complete variables using a block-diagonal level-2 covariance matrix (JM-BD)
2) JM imputation predicting the incomplete variables from the complete variables using an unstructured level-2 covariance matrix (JM-UN)
3) JM imputation treating all variables as response variables and using an unstructured level-2 covariance matrix (JM-Mplus)
4) FCS imputation without cluster means (FCS-VB)
5) FCS imputation including cluster means for both incomplete and complete variables (FCS-CK imputation)

Research Questions and Hypotheses

As stated in Chapter 1, the first overarching goal for this project was to examine the situations under which the three JM imputation methods reproduced (or preserved) the mean and e structure of a population random intercept model with multivariate normal data. The second overarching goal for this project was to examine the situations under which FCS imputation reproduced the mean and covariance structure of a population random intercept model with multivariate normal data. These overarching goals corresponded to the five following research questions:

1) Does JM-BD produce expectations that are equivalent to the joint distribution of the data?
2) Does JM-UN produce expectations that are equivalent to the joint distribution of the data?
3) Does JM-Mplus produce expectations that are equivalent to the joint distribution of the data?
4) Does FCS-VB produce expectations that are equivalent to the joint distribution of the data?
5) Does FCS-CK produce expectations that are equivalent to the joint distribution of the data?

Answering the above questions also provided insight into the following question:

6) Do any of the five imputation methods examined in this document produce equivalent expectations? If so, which?

I expected that two of the three JM methods would produce expectations that were not equivalent to the joint distribution containing only random intercepts. I hypothesized that method one, JM-BD imputation, would not produce expectations equivalent to the joint distribution containing only random intercepts. This is because JM-BD imputation uses a single parameter to represent the regression of an incomplete variable on a complete variable by default, and because JM-BD imputation assumes no covariance between the random intercepts of the incomplete variables. I hypothesized that method two, JM-UN imputation, would not produce expectations equivalent to the joint distribution containing only random intercepts. This is because JM-UN imputation uses a single parameter to represent the regression of an incomplete variable on a

complete variable by default. I hypothesized that method three, JM-Mplus, would produce expectations that were equivalent to the joint distribution containing only random intercepts. This is because JM-Mplus imputation uses two parameters to represent the covariances between each pair of variables in the model and makes no assumptions about level-2 covariances being equal to level-1 covariances or to zero.

I expected that, of the two FCS methods, only FCS-CK imputation would produce expectations that were equivalent to the joint distribution containing only random intercepts. I hypothesized that method 4, FCS-VB imputation, would not produce expectations equivalent to the joint distribution containing only random intercepts. This is because FCS-VB imputation uses a single parameter to represent level-1 and level-2 regressions of each incomplete variable on each complete variable by default, and always uses a single parameter to represent level-1 and level-2 regressions of each incomplete variable on every other incomplete variable. I hypothesized that method five, FCS-CK imputation, would produce expectations equivalent to the joint distribution containing only random intercepts. Because FCS-CK imputation uses two parameters to represent the regression of each incomplete level-1 variable on every other level-1 variable, the level-1 and level-2 regression coefficients should not be constrained to be equal or to be zero.

The primary hypotheses just listed also implied a set of secondary hypotheses. JM-Mplus and FCS-CK were predicted to both produce expectations equivalent to those produced by the population joint model. As such, JM-Mplus and FCS-CK were predicted to be equivalent to one another, but not to any of the other methods. JM-BD is the only imputation method that assumes no covariance between the random intercepts of the

incomplete variables. As such, JM-BD was predicted to not be equivalent to any of the other methods. FCS-VB imputation uses a single parameter to represent the regression of an incomplete variable on any other variable. JM-UN imputation uses a single parameter to represent the regression of an incomplete variable on any complete variable, but uses two parameters to represent the covariance between any pair of incomplete variables. As such, the expectations produced by JM-UN were predicted to differ from the expectations produced by FCS-VB.

In summary, I hypothesized that:

1) JM-BD, JM-UN, and FCS-VB produce expectations that are not equivalent to the joint distribution of the data.
2) JM-Mplus and FCS-CK imputation produce expectations that are equivalent to the joint distribution of the data.
3) JM-BD, JM-UN, and FCS-VB do not produce expectations that are equivalent to one another or to JM-Mplus and FCS-CK imputation.
4) JM-Mplus and FCS-CK imputation produce equivalent expectations.

Analytic Examination of JM and FCS Expectations

In order to describe the exact differences between JM and FCS imputation methods, it is helpful to first specify a population model. The analytic work assumed the following population model:

$$
\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix} + \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}
$$

$$
\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \psi_{ww} & \psi_{wx} & \psi_{wy} \\ \psi_{wx} & \psi_{xx} & \psi_{xy} \\ \psi_{wy} & \psi_{wy} & \psi_{yy} \end{bmatrix} \right) \tag{3.13}
$$

$$
\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sigma_{ww} & \sigma_{wx} & \sigma_{wy} \\ \sigma_{wx} & \sigma_{xx} & \sigma_{xy} \\ \sigma_{wy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \right)
$$

The model in Equation 3.13 is simply the joint multilevel distribution of variables *W*, *X*, and *Y* for two-level multivariate normal data. The exact values of the parameters are not of importance for the theoretical examination of the imputation models and are not stated here. It is important that any imputation model (FCS or JM) be able to correctly represent the joint distribution in Equation 3.13.

I began by examining the joint multilevel distribution for *W*, *X*, and *Y*, and compared it to the model-implied moments for the three multilevel JM methods and the two FCS methods for incomplete *X* and incomplete *Y*. Specifying a population model with three variables and treating two of the three variables as incomplete allows for the examination of the covariances between a complete variable and an incomplete variable, as well as the covariances between a pair of incomplete variables. I showed that each parameter in the imputation models can be calculated from the parameters in the joint multilevel distribution. This was done in a method similar to the one described for Shin and Raudenbush (2007) in Chapter 2. Calculating the parameters of each imputation

model from the parameters of the joint population model highlighted areas of the population model that are not fully preserved by the imputation model.

For the methods which I hypothesized to produce expectations equivalent to those produced by the joint model, I proved the hypothesis by showing that the parameters of the population joint model can be calculated from the parameters of the imputation model. Note that it would not possible to calculate the parameters in the joint multilevel distribution from the parameters in the imputation model for the imputation models that do not produce expectations equivalent to those of the joint model. That is, the parameters of the joint population model could not be calculated from the parameters of the imputation model if the imputation model has fewer parameters than does the joint model. So, calculating the parameters of the joint distribution from the parameters of the imputation model demonstrate that they should produce equivalent expectations.

Simulation Studies

Overview of simulations. Following the theoretical examination of the FCS and JM imputation methods, I conducted two simulation studies. The first study generated the data under a population model where the level-2 correlations among variables differed from the level-1 correlations among the variables. The results of study 1 reflect how well the imputation methods preserved the level-1 and level-2 covariances from the general distribution in Equation 3.13. The second study generated the data under a population model where the correlations between variables are identical across levels. The results of the second study enabled me to determine whether the more general multilevel imputation methods (JM-Mplus and FCS-CK imputation) perform well even when the models are overparameterized. Note that it was not my 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.

The following four factors were manipulated in the two studies: the imputation method, the number of clusters, the number of observations per cluster, and the intraclass correlation (ICC). The imputation method factor incorporated the five approaches described at the beginning of this chapter (JM-BD, JM-UN, JM-Mplus, FCS-VB, and FCS-CK). The number of clusters took on values of 30 and 100. The number of observations per cluster took on values of 5 and 30. The ICC took on values of 0.1 and 0.5. The rationale for these choices is provided later in this section.

Population models. The multiple imputation methods examined in this document differ in how well they preserve covariances between complete and incomplete variables, as well as how well they preserve covariances between pairs of incomplete variables. So that the methods may be compared in terms of how they handle both types of covariances, all of the population models included two incomplete variables and one complete.

As described in the above paragraph, the data generation models for both simulation studies included three variables, which I label *W*, *X*, and *Y*. The joint distribution for the three variables with random intercepts but no random slopes can be written as:

$$
\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix} + \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}
$$

$$
\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \psi_{ww} & \psi_{wx} & \psi_{wy} \\ \psi_{wx} & \psi_{xx} & \psi_{xy} \\ \psi_{wy} & \psi_{xy} & \psi_{yy} \end{bmatrix} \right) \tag{3.14}
$$

$$
\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sigma_{ww} & \sigma_{wx} & \sigma_{wy} \\ \sigma_{wx} & \sigma_{xx} & \sigma_{xy} \\ \sigma_{wy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \right)
$$

The values of the fixed intercept terms ($\beta^{(w)}$, $\beta^{(\mathrm{x})}$, and $\beta^{(\mathrm{y})}$) are arbitrary, so, for convenience, I set them to zero across all simulation conditions. Because the sizes of the correlations between the variables at level-2 differed between study one and study two (i.e., the magnitude of the contextual effect differed), and because I also manipulated the variances of the variables at level-2 (to manipulate the ICCs), it is useful to rewrite the level-1 and level-2 covariance matrices as functions of the standard deviation vectors and correlation matrices at each level. Making these changes to the joint distribution yields the following population model:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}$$

$$\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{MVN}\left(\begin{array}{l} 0, \begin{bmatrix} \sqrt{\psi_{ww}} & 0 & 0 \\ 0 & \sqrt{\psi_{xx}} & 0 \\ 0 & 0 & \sqrt{\psi_{yy}} \end{bmatrix} * \begin{bmatrix} 1 & 0 & r_{wy}^{L2} \\ 0 & 1 & r_{xy}^{L2} \\ r_{wy}^{L2} & r_{xy}^{L2} & 1 \end{bmatrix} \\ * \begin{bmatrix} \sqrt{\psi_{ww}} & 0 & 0 \\ 0 & \sqrt{\psi_{xx}} & 0 \\ 0 & 0 & \sqrt{\psi_{yy}} \end{bmatrix} \end{array} \right)$$

$$\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \begin{bmatrix} 1 & 0 & .4 \\ 0 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \right) \quad (3.15)$$

where the two outer matrices for each of the covariance matrix calculations are diagonal matrices containing the standard deviation of each variable on the main diagonal and the inner matrix is the correlation matrix. Note that the correlations between the *W* and *X* at each level were fixed at zero. This was done to isolate potential sources of bias. The non-zero level-1 correlations were not manipulated. Rather, the level-1 correlation between *W* and *Y* is fixed at .4 and the level-1 correlation between *X* and *Y* were fixed at .5 across all study conditions. A correlation of .5 corresponds to Cohen's benchmark for a large effect size. The level-1 variance of *X* was set equal to 9, the level-1 variance of *W* was set equal to 18, and the level-1 variance of *Y* was set equal to 27. The size of the variances was arbitrary. The level-2 variance changed across conditions to achieve the two ICC

conditions under investigation (ICC = .1 and ICC = .5). The expressions in Equation 3.15 produced the following level-1 and level-2 covariance matrices:

$$\Psi = \begin{bmatrix} \psi_{ww} & 0 & r_{wy}^{L2}\sqrt{\psi_{ww}}\sqrt{\psi_{yy}} \\ 0 & \psi_{xx} & r_{xy}^{L2}\sqrt{\psi_{xx}}\sqrt{\psi_{yy}} \\ r_{wy}^{L2}\sqrt{\psi_{ww}}\sqrt{\psi_{yy}} & r_{xy}^{L2}\sqrt{\psi_{xx}}\sqrt{\psi_{yy}} & \psi_{yy} \end{bmatrix} \qquad (3.16)$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 9 & 0 & 6.24 \\ 0 & 18 & 11.02 \\ 6.24 & 11.02 & 27 \end{bmatrix}$$

The goal of study one was to investigate the case where the level-1 and level-2 associations differ. In order to produce a non-zero difference between the level-1 and level-2 correlations as well as a non-zero level-2 correlations between *Y* and the other two variables, the level-2 correlation between *W* and *Y* (r_{wy}^{L2} in Equations 3.15 and 3.16) was set to -.4 and the level-2 correlation between *X* and *Y* (r_{xy}^{L2} in Equations 3.15 and 3.16) was set to -.5. These two correlations are equal in magnitude to the level-1 correlations, but opposite in sign. I chose rather large correlations of -.4 and -.5 because the large effect size would allow me to clearly demonstrate the implications of the analytic work. To achieve an ICC of 0.5 for all of the variables, the level-2 variance terms were set equal to the level-1 variance terms. That is, the level-2 variance of *X* was set equal to 9, the level-2 variance of *W* was set equal to 18, and the level-2 variance of *Y* was set equal to 27. So, the population model for study one in the ICC = .5 condition was:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} b_j^{(w)} \\ b_j^{(x)} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(y)} \end{bmatrix}$$

$$\begin{bmatrix} b_j^{(w)} \\ b_j^{(x)} \\ b_j^{(y)} \end{bmatrix} \sim \text{MVN}\left(0, \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \begin{bmatrix} 1 & 0 & -.4 \\ 0 & 1 & -.5 \\ -.4 & -.5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix}\right) \quad (3.17)$$

$$\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(y)} \end{bmatrix} \sim \text{MVN}\left(0, \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \begin{bmatrix} 1 & 0 & .4 \\ 0 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix}\right)$$

The expressions in Equation 3.17 produce the following level-1 and level-2 covariance matrices:

$$\Psi = \begin{bmatrix} 9 & 0 & -6.24 \\ 0 & 18 & -11.02 \\ -6.24 & -11.02 & 27 \end{bmatrix}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} 9 & 0 & 6.24 \\ 0 & 18 & 11.02 \\ 6.24 & 11.02 & 27 \end{bmatrix} \quad (3.18)$$

For the ICC=.1 condition of study one, the level-2 variance of *X* was set equal to 1, the level-2 variance of *W* was set equal to 2, and the level-2 variance of *Y* was set equal to 3:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}$$

$$\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sqrt{1} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & -.4 \\ 0 & 1 & -.5 \\ -.4 & -.5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{1} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}\right) \tag{3.19}$$

$$\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \begin{bmatrix} 1 & 0 & .4 \\ 0 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix}\right)$$

Other than using different level-2 variances, Equation 3.19 is identical to Equation 3.17. The expressions in Equation 3.19 produce the following level-1 and level-2 covariance matrices:

$$\Psi = \begin{bmatrix} 1 & 0 & -.69 \\ 0 & 2 & -1.22 \\ -.69 & -1.22 & 3 \end{bmatrix}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} 9 & 0 & 6.24 \\ 0 & 18 & 11.02 \\ 6.24 & 11.02 & 27 \end{bmatrix} \tag{3.20}$$

The goal of study two was to investigate the case where the level-1 and level-2 associations were equal. To achieve this, the level-2 correlation between *W* and *Y* (r_{wy}^{L2} in Equations 3.15 and 3.16) was set to .4 and the level-2 correlation between *X* and *Y* (r_{xy}^{L2} in Equations 3.15 and 3.16) was set to .5. The other parameters in the two population

models (one for ICC=.5 and one for ICC=.1) used for study two were identical to the parameters in the two population models used for study one. So, the population model for study two in the ICC=.5 condition was:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}$$

$$\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \begin{bmatrix} 1 & 0 & .4 \\ 0 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix}\right) \qquad (3.21)$$

$$\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \begin{bmatrix} 1 & 0 & .4 \\ 0 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix}\right)$$

The expressions in Equation 3.21 produce the following level-1 and level-2 covariance matrices:

$$\Psi = \begin{bmatrix} 9 & 0 & 6.24 \\ 0 & 18 & 11.02 \\ 6.24 & 11.02 & 27 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 9 & 0 & 6.24 \\ 0 & 18 & 11.02 \\ 6.24 & 11.02 & 27 \end{bmatrix} \qquad (3.22)$$

The population model for study two in the ICC=.1 condition was:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}$$

$$\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sqrt{1} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & .4 \\ 0 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{1} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}\right) \tag{3.23}$$

$$\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix} \begin{bmatrix} 1 & 0 & .4 \\ 0 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{9} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & \sqrt{27} \end{bmatrix}\right)$$

The expressions in Equation 3.23 produce the following level-1 and level-2 covariance matrices:

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0 & .69 \\ 0 & 2 & 1.22 \\ .69 & 1.22 & 3 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 9 & 0 & 6.24 \\ 0 & 18 & 11.02 \\ 6.24 & 11.02 & 27 \end{bmatrix} \tag{3.24}$$

Though the two covariance matrices in Equation 3.24 are different, the multiple regression of *Y* on *X* and *W* would yield the same estimates for level-1 and level-2. That is, the partial regression coefficient predicting *Y* from *W* would be 0.69 at both levels and the partial regression coefficient predicting *Y* from *X* would be 0.61 at both levels.

Data generation. The population models just described were used to generate the data using the SAS Interactive Matrix Language, Version 13.2 (SAS Institute Inc., 2014). Recall from previous equations that, regardless of the population model, the vector of values for the three variables for person *i* in cluster *j* can be written as the sum of the vector of random intercepts for cluster *j* and the vector of residuals for person *i* in cluster *j*:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \tag{3.25}$$

Note that a vector of means is not included because the mean of each variable is fixed to zero in all of the population models. To generate the vector of values for simulated observation 1 in simulated cluster 1, one would simply draw the level-1 and level-2 vectors and calculate their sum. Note that one would use the same level-2 vector for all observations in the same cluster. SAS/IML can be used to directly sample a vector from a multivariate normal distribution specified using a mean vector and covariance matrix (SAS Institute Inc., 2014; Wicklin, 2013). So, for the ICC=0.5 condition of study one, one could draw the vector of random intercepts for cluster one from the following distribution:

$$\begin{bmatrix} b_1^{(w)} \\ b_1^{(x)} \\ b_1^{(y)} \end{bmatrix} \sim \text{MVN}\left(0, \begin{bmatrix} 9 & 0 & -6.24 \\ 0 & 18 & -11.02 \\ -6.24 & -11.02 & 27 \end{bmatrix}\right) \tag{3.26}$$

For person one in cluster one, one would draw the level-1 residuals from the following distribution:

$$\begin{bmatrix} \varepsilon_{1,1}^{(w)} \\ \varepsilon_{1,1}^{(x)} \\ \varepsilon_{1,1}^{(y)} \end{bmatrix} \sim \text{MVN}\left(0, \boldsymbol{\Sigma} = \begin{bmatrix} 9 & 0 & 6.24 \\ 0 & 18 & 11.02 \\ 6.24 & 11.02 & 27 \end{bmatrix}\right) \tag{3.27}$$

The covariance matrices in the previous two equations came from Equation 3.17. The values for observation 1 in cluster 1 are the result of summing the two vectors that were just drawn:

$$\begin{bmatrix} W_{1,1} \\ X_{1,1} \\ Y_{1,1} \end{bmatrix} = \begin{bmatrix} b_1^{(w)} \\ b_1^{(x)} \\ b_1^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,1}^{(w)} \\ \varepsilon_{1,1}^{(x)} \\ \varepsilon_{1,1}^{(y)} \end{bmatrix} \tag{3.28}$$

As previously mentioned, the vector of random intercepts was drawn once for each cluster, and the vector of level-1 residuals was drawn once for each observation in the cluster.

Consistent with Enders et al. (2014), I did not vary the missing data rate, as this factor has a predictable and uninteresting effect on the outcomes (e.g., as the missing data rate increases, nonresponse bias increases and efficiency decreases). Rather than varying the missing data rate, I fixed probability of missingness at .2 for *X* and *Y*. This means that each value of *X* and *Y* had a probability of being deleted equal to .2. However, the exact percentage of missing observations was allowed to vary across sample data sets. This missingness probability was sufficient to highlight differences between the imputation methods. The missing data were generated under an MCAR mechanism. Though MCAR may not be a realistic assumption, it is useful to first examine whether the imputation methods under a benign mechanism. If a missing data handling method produces bias for MCAR, one can conclude that using such a method is worse than simply ignoring the missing data and using listwise deletion (which is unbiased for MCAR). As such, it is standard practice to test missing data handling methods under an MCAR mechanism prior to testing these methods under an MAR.

The deletion of values to simulate MCAR missingness was performed in SAS/IML. After a complete data set was generated and saved, a copy of the data set needed to be created but with values of the variables deleted at random. An indicator variable was created for each of the two variables to be made incomplete (*X* and *Y*). Values for the indicator variables were drawn from a Bernoulli(.2) distribution. That is, for each observation, each indicator took a value of 1 with probability of .2, and a value of 0 with probability of .8. If the indicator for *X* was equal to one for an observation, *X* was deleted for that observation. The same was be done for *Y*. This resulted in roughly 20% of values being deleted from the data set for both *X* and *Y*. Because the distribution

of each indicator is the same for all values in the data set, the missing data mechanism was MCAR.

Manipulated factors. This subsection describes the rationale for the manipulated factors and their values. As previously mentioned, the studies manipulated the following factors: imputation method, number of clusters, number of observations per cluster, and ICC. The imputation method factor incorporated the five approaches described at the beginning of this chapter (JM-BD, JM-UN, JM-Mplus, FCS-VB, and FCS-CK) for the reasons described at the beginning of the chapter. JM-BD imputation, JM-UN imputation, and JM-Mplus imputation were performed in Mplus. FCS-VB imputation and FCS-CK imputation were performed using custom software, BLImP, written by Keller and Enders (2014). Each imputation method was applied to the same data set, making imputation method a within-subjects factor.

Two conditions were used for both the number of clusters (30 and 100) and the number of observations per clusters (5 and 30). I chose these conditions to represent two extremes in psychological research: cross-sectional studies where participants are nested in groups (e.g., children nested within schools) and longitudinal research where repeated observations are nested within participants. Kreft and de Leeuw (1998) and Hox (2010) suggest that 30 clusters is the minimum acceptable number for a multilevel model. Consistent with Maas and Hox (2005), 100 clusters were used for the upper extreme. Five level-1 observations per cluster is typical for longitudinal studies and for family research, whereas 30 observations per cluster is typical in education studies (Maas & Hox, 2005).

Differences among the imputation methods may be highlighted by varying the intraclass correlations (ICCs). The size of the ICC may also affect the accuracy of the

estimates (Goldstein, 2011). ICCs typically range from 0 to 0.3 in cross-sectional studies, with smaller values tending to be more prevalent (Gulliford, Ukoumunne, & Chinn, 1999). Other authors have suggested ICCs between 0.05 and 0.15 for cross-sectional studies (e.g., Hedges & Hedberg, 2007; Spybrook, Raudenbush, Liu, Congdon, & Martínez, 2006). As such, I examined an ICC of 0.1 to determine how well the imputation methods are likely to fare in cross-sectional studies. In keeping with Enders et al. (2014), I used an ICC of 0.5 to mimic within-subjects data (e.g., data from a longitudinal study or diary data). Though the high ICC condition was intended to mimic within-subjects data, the structure of the level-1 residual covariance matrix was left unchanged (the level-1 residuals were not allowed to covary). The random intercept model used to generate the data could have equivalently been generated using a repeated measures model with a compound symmetric structure. That is, the data for the high ICC condition could be thought of as repeated measurements performed on the same level-2 units without any autoregressive effects. The same ICC was used for all variables in the model.

Non-manipulated aspects of the study. Each of the methods examined in the two simulation studies were used to produce 20 imputations. A simulation study by Graham et al. (2007) showed that 20 imputations produced power similar to that produced by 100 imputations for small fractions of missing information. As such, increasing the number of imputations past 20 would greatly increase the computational demand (and therefore computing time) of the simulations in exchange for only a small increase in power. Consistent with Enders et al. (2014), the imputations were produced using 1,000 burn-in iterations and 500 between-imputation iterations. The 1,000 burn-in

iterations were expected to result in stability for the MCMC algorithm prior to any imputations being produced. The 200 between-imputation iterations were expected to be sufficient to avoid correlations between imputed values across imputations.

Multilevel multiple imputation simulations take far more time per iteration than average studies due to multilevel imputation methods being time-consuming. As such, multilevel imputation simulation studies tend to use 1000 or fewer iterations per simulation condition. For example, van Buuren (2011) used only 100 iterations per condition, but Andridge (2011) and Taljaard et al. (2008) both used 1000 iterations per condition. I performed 1000 iterations per study condition. Each of the two studies contained five imputation methods, two level-2 sample size conditions, two level-1 sample size conditions, and two ICC conditions. Running the two studies with the aforementioned factor values required 40 conditions and 40,000 total replications for each of the two simulation studies.

Analysis model. The analysis model for both simulation studies was a multivariate unconditional model that estimated the grand means and the unstructured covariance matrices at both levels (i.e., the parameters of the joint distribution):

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix} + \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}$$

$$\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \psi_{ww} & \psi_{wx} & \psi_{wy} \\ \psi_{wx} & \psi_{xx} & \psi_{xy} \\ \psi_{wy} & \psi_{wy} & \psi_{yy} \end{bmatrix}\right) \tag{3.29}$$

$$\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{MVN}\left(0, \begin{bmatrix} \sigma_{ww} & \sigma_{wx} & \sigma_{wy} \\ \sigma_{wx} & \sigma_{xx} & \sigma_{xy} \\ \sigma_{wy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix}\right)$$

This model was chosen because it matches the form of the population distributions used to generate the data. As such, differences between the population distribution and the analysis results highlighted biases resulting from the imputation method employed. The analysis of each imputed data set was performed using Mplus 7. Mplus 7 was chosen because it includes functionality for estimating multivariate two-level models. Though Mplus 7 is capable of pooling multiple imputation results and displaying these results in the output, Mplus 7 is not capable of exporting the pooled multiple imputation analysis results to a data set. As such, the pooling of the results was performed in SAS/IML 13.2.

Replicate failures. For any multiple imputation simulation study, it is important to assess whether the imputation step and the analysis/pooling step both complete successfully for each imputation method for each replicate. In addition to recording the parameters estimated for each method for each replicate, I recorded the failure rate for the method, where a failure of the imputation step or the analysis/pooling step for a method counts as a failure for that method. When such a failure occurred, I generated additional

data sets to ensure that estimates were successfully obtained for 1000 replicates in each of the simulation conditions. This allowed me to compare the five imputation methods for the same set of replicates while at the same time ensuring that if a method tended to fail this tendency would be reported.

Outcomes. After the analysis results were pooled, SAS was used to organize the results and to analyze the pooled results in each simulation condition for bias. SAS was chosen for this task because it performs very well for handling and organizing large amounts of data. For each of the simulation conditions, raw bias, corrected bias, and standardized corrected bias were calculated for all of the parameters. I estimated the raw bias for a parameter as the imputed-data estimate of the parameter minus the true population parameter, averaged across all of the replicates in the condition. I calculated corrected bias as the imputed-data estimate of the parameter minus complete-data estimate of the parameter, averaged across all of the replicates in the condition. Consistent with Enders et al. (2014), I calculated the empirical standard error of each parameter as the standard deviation of the parameter estimate from the complete data sets generated in the 1000 replicates within each design cell. I then calculated the standardized corrected bias for the parameter by dividing the corrected bias by the empirical standard error. Standardized corrected bias values greater than 0.4 or less than -0.4 were pointed out, as such extreme values tend to negatively affect statistical inference (Collins, Schafer, & Kam, 2001).

Chapter 4. Results

As stated in the previous chapter, this document aimed to answer the following research questions:

1) Does JM-BD produce expectations that are equivalent to the joint distribution of the data?
2) Does JM-UN produce expectations that are equivalent to the joint distribution of the data?
3) Does JM-Mplus produce expectations that are equivalent to the joint distribution of the data?
4) Does FCS-VB produce expectations that are equivalent to the joint distribution of the data?
5) Does FCS-CK produce expectations that are equivalent to the joint distribution of the data?
6) Do any of the five imputation methods examined in this document produce equivalent expectations? If so, which?

In response to these research questions, I hypothesized that:

1) JM-BD, JM-UN, and FCS-VB produce expectations that are not equivalent to the joint distribution of the data.
2) JM-Mplus and FCS-CK imputation produce expectations that are equivalent to the joint distribution of the data.
3) JM-BD, JM-UN, and FCS-VB do not produce expectations that are equivalent to one another or to JM-Mplus and FCS-CK imputation.
4) JM-Mplus and FCS-CK imputation produce equivalent expectations.

I the following sections, I present the results of the analytic examination of the methods and the simulation studies with the aim of addressing the above research questions and hypotheses.

Analytic Examination Results

This section shows analytically which imputation methods produce expectations that are equivalent to the population model. For methods that are not equivalent to the population model, this section details the conditions (constraints on the population model) necessary for equivalence. This section also shows which methods are equivalent to one another. As a reminder, the analytic work assumes the following population model:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix} + \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}$$

$$\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{N}_3\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{ww} & \psi_{wx} & \psi_{wy} \\ \psi_{wx} & \psi_{xx} & \psi_{xy} \\ \psi_{wy} & \psi_{xy} & \psi_{yy} \end{bmatrix} \right) \tag{4.1}$$

$$\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{N}_3\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{ww} & \sigma_{wx} & \sigma_{wy} \\ \sigma_{wx} & \sigma_{xx} & \sigma_{xy} \\ \sigma_{wy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \right)$$

The population model contains 15 parameters including three means ($\beta^{(w)}$, $\beta^{(x)}$, $\beta^{(y)}$), six level-2 covariance parameters (ψ_{ww}, ψ_{wx}, ψ_{xx}, ψ_{wy}, ψ_{xy}, and ψ_{yy}), and six level-1 covariance parameters (σ_{ww}, σ_{wx}, σ_{xx}, σ_{wy}, σ_{xy}, and σ_{yy}).

Variables *X*, and *Y* were both be treated as incomplete and *W* was treated as complete. Specifying a population model with three variables and treating two of the

three variables as incomplete allows for the examination of the covariances between a complete variable and an incomplete variable, as well as the covariances between a pair of incomplete variables. This allows the analytic work to generalize to a wide range of scenarios, as the conclusions derived here apply equally well to models with more variables. The approach to the analytic work in this section is based on the approach used by Shin and Raudenbush (2007) for equating a two-level population model with a univariate multilevel model, described in Chapter 2. The imputation model for each of the imputation methods is in the form of a multilevel model. In order to compare the population model with the imputation model, it is necessary to transform the unconditional population model in Equation 4.1 above to a conditional distribution in the form of a multilevel model. In the steps below, the terms independent and dependent refer to the roles of the variables in the imputation model.

1. Restate the population joint distribution of the variables with the independent variables (i.e., the variables that are independent in the imputation model) separated into between-cluster and within-cluster components.
2. Determine the distribution of the dependent variables conditional upon the independent variables separated into their within-cluster and between-cluster components.
3. Rewrite the conditional population distribution from step 2 in the form of a multilevel model.

4. Compare the transformed population model from step 3 with the imputation model and state the constraints necessary for equating the imputation model and the population model.

If it is determined in step 4 that the imputation model and the population model do not require constraints to be added to the population model to achieve equivalence, then one can conclude that the imputation model and the population model are equivalent. In this case, I simply equate the parameters from the imputation and population models. If, on the other hand, constraints must be added to the population model to achieve equivalence, I added an extra step:

5. Rewrite the transformed population model from step 3 under the constraints from step 4. That is, show what the population model looks like after applying the constraints from step 4.

The parameters in the constrained population model derived in step 5 can then be equated to the parameters rom the imputation model. Some of the conditions used to equate the population models to the imputation models in this section are based on population constraints suggested by Shin and Raudenbush (2007) for equating a joint multilevel distribution with a univariate multilevel model.

JM-BD. The JM-BD imputation model (joint model with block diagonal level-2 covariance matrix) predicting incomplete variables X and Y from complete variable W is:

$$\begin{bmatrix} X_{ij} & Y_{ij} \end{bmatrix} = \begin{bmatrix} \gamma_0^{(x)} & \gamma_0^{(y)} \end{bmatrix} + \begin{bmatrix} \gamma_1^{(x)} & \gamma_1^{(y)} \end{bmatrix} W_{ij} + \begin{bmatrix} u_j^{(x)} & u_j^{(y)} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(x)} & e_{ij}^{(y)} \end{bmatrix}$$

$$\begin{bmatrix} u_{j,x} \\ u_{j,y} \end{bmatrix} \sim \mathrm{N}_2\left(\mathbf{0}, \boldsymbol{\Psi}\right) = \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{xx} & 0 \\ 0 & \tau_{yy} \end{bmatrix}\right) \tag{4.2}$$

$$\begin{bmatrix} e_{ij,x} \\ e_{ij,y} \end{bmatrix} \sim \mathrm{N}_2\left(\mathbf{0}, \boldsymbol{\Sigma}\right) = \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \varsigma_{xx} & \varsigma_{xy} \\ \varsigma_{xy} & \varsigma_{yy} \end{bmatrix}\right)$$

Notice that the covariance between the level-2 residuals is zero in the psi matrix. The predictor variable is usually treated as fixed when describing an imputation model. To facilitate the comparison between the population and imputation models, however, I treat the predictor variable as random. Because the variables (complete and incomplete) are assumed to be distributed as multivariate normal, the fixed vs. random distinction for the predictor variable has no impact on the imputation procedure. The predictor variable has the following marginal distribution:

$$W_{ij} \sim \mathrm{N}(\beta^{(w)}, \psi_{ww} + \sigma_{ww}) \tag{4.3}$$

The parameters in Equation 4.3 are the same for the imputation model and the population model. Equations 4.2 and 4.3 show that the JM-BD imputation model (including the marginal distribution of the complete variable) contains a total of 12 parameters ($\gamma_0^{(x)}$, $\gamma_0^{(y)}$, $\gamma_1^{(x)}$, $\gamma_1^{(x)}$, τ_{xx}, τ_{yy}, ς_{xx}, ς_{xy}, ς_{yy}, $\beta^{(w)}$, ψ_{ww}, and σ_{ww}).

To transform the unconditional population model in Equation 4.1 to a conditional distribution in the form of a multilevel model, we begin by specifying the joint distribution of the variables:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} \sim \mathrm{N}_3\left(\begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \psi_{ww}+\sigma_{ww} & \psi_{wx}+\sigma_{wx} & \psi_{wy}+\sigma_{wy} \\ \psi_{wx}+\sigma_{wx} & \psi_{xx}+\sigma_{xx} & \psi_{xy}+\sigma_{xy} \\ \psi_{wy}+\sigma_{wy} & \psi_{xy}+\sigma_{xy} & \psi_{yy}+\sigma_{yy} \end{bmatrix}\right) \tag{4.4}$$

Equation 4.4 is equivalent to Equation 4.1, but Equation 4.4 directly describes the distribution of the variables whereas Equation 4.1 described the distributions of the level-1 residuals and level-2 random effects of the variables. Because multilevel models assume that the level-2 random effects of the variables ($b_j^{(\mathrm{w})}$, $b_j^{(x)}$, and $b_j^{(y)}$) are uncorrelated with the level-1 residuals ($\varepsilon_{ij}^{(w)}$, $\varepsilon_{ij}^{(x)}$, and $\varepsilon_{ij}^{(y)}$), the covariance matrix of the variables in Equation 4.4 is simply the sum of the level-1 and level-2 covariance matrices in Equation 4.1.

The multivariate distribution in Equation 4.4 does not recognize the between-cluster portion of the predictor variable, *W*. This is problematic, because obtaining the distribution of *X* and *Y* conditional on *W* would yield the *single-level* regression coefficients predicting *X* and *Y* from *W*. In order to explicitly recognize the between-cluster portion of the predictor variable, *W*, we can add the cluster mean of the predictor, $\overline{W}_j$, as a variable in the distribution. Note that the cluster mean of *W* is equal to the random effect of *W* plus the mean of *W*, $b_j^{(\mathrm{w})} + \beta^{(w)}$. I will refer to the cluster mean of *W* as the between-cluster portion of *W*. Including the between-cluster portion of the predictor variable in the distribution allows us to calculate the conditional distribution $X_{ij}, Y_{ij} \,|\, W_{ij}, \overline{W}_j$, which, as is shown later, yields parameters equivalent to those of a

multivariate multilevel model. In order to simplify the math in this section, it is useful to center W at its cluster mean, $W_{ij} - \bar{W}_j$. The two predictors, $W_{ij} - \bar{W}_j$ and $\bar{W}_j$, now correspond to the within-cluster and between-cluster components of W and are orthogonal to one another. Because each source of variability in the predictor is orthogonal to the other, the regression coefficients for the within- and between-cluster influence of W are very easy to calculate from the covariance matrix. Note that I center W purely as a means for simplifying the calculations below. The final results are not changed as a result of having centered W. The findings of this analytic work are equally valid for centered and uncentered predictors. Software that incorporates cluster means as predictors in the imputation model (i.e., BLImP) does not center the level-1 predictors, but this does not affect the generalizability of the analytic results. However, W needs to be returned to its raw metric in a later step in order to match the JM-BD imputation model (which uses W in its raw metric). The dependent variable cluster means are not added to the model, and the dependent variables are not centered[2]. Equation 4.4 then becomes:

[2] Centering the dependent variables and including their cluster means would result in calculating the conditional distribution $X_{ij} - \bar{X}_j, \bar{X}_j, Y_{ij} - \bar{Y}_j, \bar{Y}_j \mid W_{ij} - \bar{W}_j, \bar{W}_j$, which would be theoretically equivalent to performing separate regressions at level one and level two. Because such a conditional distribution would not be comparable to any of the imputation models, the cluster means of the dependent variables are not included in the distribution and the dependent variables are left in their raw metrics.

$$\begin{bmatrix} W_{ij} - \bar{W}_j \\ \bar{W}_j \\ X_{ij} \\ Y_{ij} \end{bmatrix} \sim \mathrm{N}_4 \left(\begin{bmatrix} 0 \\ \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \sigma_{ww} & 0 & \sigma_{wx} & \sigma_{wy} \\ 0 & \psi_{ww} & \psi_{wx} & \psi_{wy} \\ \sigma_{wx} & \psi_{wx} & \psi_{xx} + \sigma_{xx} & \psi_{xy} + \sigma_{xy} \\ \sigma_{wy} & \psi_{wy} & \psi_{xy} + \sigma_{xy} & \psi_{yy} + \sigma_{yy} \end{bmatrix} \right) \tag{4.5}$$

The level-1 and level-2 covariances between *W* and the two dependent variables, *X* and *Y*, are now represented by different cells in the covariance matrix.

Equation 4.5 can be used to calculate the distribution of *X* and *Y* conditional upon the within-cluster portion *W* and the between-cluster portion of *W*. The regression coefficients are calculated as the inverse of the covariance matrix of the predictors post-multiplied by the matrix of covariances between the independent variables and the dependent variables:

$$\begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix} = \begin{bmatrix} \sigma_{ww} & 0 \\ 0 & \psi_{ww} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{wx} & \sigma_{wy} \\ \psi_{wx} & \psi_{wy} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_{wx}}{\sigma_{ww}} & \frac{\sigma_{wy}}{\sigma_{ww}} \\ \frac{\psi_{wx}}{\psi_{ww}} & \frac{\psi_{wy}}{\psi_{ww}} \end{bmatrix} \tag{4.6}$$

$\alpha_{x.w,1}$ is the within-cluster effect of *W* on *X*. $\alpha_{x.w,2}$ is the between-cluster effect of *W* on *X*. $\alpha_{y.w,1}$ is the within-cluster effect of *W* on *Y*. $\alpha_{y.w,2}$ is the between-cluster effect of *W* on *Y*. The residual covariance matrix is calculated as the covariance matrix of the dependent variables minus the transposed matrix of regression coefficients multiplied by the covariance matrix of the predictors multiplied by the matrix of regression coefficients (i.e., the total variance minus the variance explained by *W*):

$$\begin{bmatrix} \psi_{xx}+\sigma_{xx} & \psi_{xy}+\sigma_{xy} \\ \psi_{xy}+\sigma_{xy} & \psi_{yy}+\sigma_{yy} \end{bmatrix} - \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \sigma_{ww} & 0 \\ 0 & \psi_{ww} \end{bmatrix} \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix}$$
$$= \begin{bmatrix} \left(\psi_{xx} - \alpha_{x.w,2}^2 \psi_{ww}\right) + \left(\sigma_{xx} - \alpha_{x.w,1}^2 \sigma_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \tag{4.7}$$
$$\| \begin{bmatrix} \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\psi_{yy} - \alpha_{y.w,2}^2 \psi_{ww}\right) + \left(\sigma_{yy} - \alpha_{y.w,1}^2 \sigma_{ww}\right) \end{bmatrix}$$

Note that Equation 4.7 displays the result of a matrix operation as two 2 x 1 vectors horizontally concatenated (horizontal concatenation is indicated by ||), rather than as a 2 x 2 matrix. This was done because the 2 x 2 matrix was too large to be placed on a single line. The intercepts are calculated as the means of the dependent variables minus the means of the independent variables post-multiplied by the matrix of regression coefficients:

$$\begin{bmatrix} \beta^{(x)} & \beta^{(y)} \end{bmatrix} - \begin{bmatrix} 0 & \beta^{(w)} \end{bmatrix} \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix} \tag{4.8}$$
$$= \begin{bmatrix} \beta^{(x)} - \alpha_{x.w,2}\beta^{(w)} & \beta^{(y)} - \alpha_{y.w,2}\beta^{(w)} \end{bmatrix}$$

Combining the information in Equations 4.6-4.8, the distribution of *X* and *Y* conditional upon $W_{ij} - \bar{W}_j$ and $\bar{W}_j$ is:

$$X_{ij}, Y_{ij} \mid W_{ij} - \bar{W}_j, \bar{W}_j \sim \mathrm{N}_2\left(\begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w,2}\beta^{(w)}\right) + \alpha_{x.w,1}\left(W_{ij} - \bar{W}_j\right) + \alpha_{x.w,2}\left(\bar{W}_j\right) \\ \left(\beta^{(y)} - \alpha_{y.w,2}\beta^{(w)}\right) + \alpha_{y.w,1}\left(W_{ij} - \bar{W}_j\right) + \alpha_{y.w,2}\left(\bar{W}_j\right) \end{bmatrix}, \right.$$
$$\begin{bmatrix} \left(\psi_{xx} - \alpha^2_{x.w,2}\psi_{ww}\right) + \left(\sigma_{xx} - \alpha^2_{x.w,1}\sigma_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \tag{4.9}$$
$$\left. \Vert \begin{bmatrix} \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\psi_{yy} - \alpha^2_{y.w,2}\psi_{ww}\right) + \left(\sigma_{yy} - \alpha^2_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \right)$$

Note that the covariance term of the distribution in Equation 4.9 was written as two 2 x 1 vectors horizontally concatenated, rather than as a 2 x 2 matrix. The terms in Equation 4.9 correspond to the terms in a multilevel model predicting *X* and *Y* from $W_{ij} - \bar{W}_j$ and $\bar{W}_j$. I centered *W* at its cluster mean in the transformed population model to simplify the calculation of the regression coefficients. However, *W* is not centered in the JM-BD imputation model. In order to equate the transformed population model to the imputation model, it is necessary to return *W* to its raw metric in the transformed population model. This can be done by rearranging terms in Equation 4.9, yielding:

$$X_{ij}, Y_{ij} \mid W_{ij} - \bar{W}_j, \bar{W}_j \sim \mathrm{N}_2\left(\begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w,2}\beta^{(w)}\right) + \alpha_{x.w,1}\left(W_{ij}\right) + \left(\alpha_{x.w,2} - \alpha_{x.w,1}\right)\left(\bar{W}_j\right) \\ \left(\beta^{(y)} - \alpha_{y.w,2}\beta^{(w)}\right) + \alpha_{y.w,1}\left(W_{ij}\right) + \left(\alpha_{y.w,2} - \alpha_{y.w,1}\right)\left(\bar{W}_j\right) \end{bmatrix}, \right.$$
$$\begin{bmatrix} \left(\psi_{xx} - \alpha^2_{x.w,2}\psi_{ww}\right) + \left(\sigma_{xx} - \alpha^2_{x.w,1}\sigma_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \tag{4.10}$$
$$\left. \Vert \begin{bmatrix} \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\psi_{yy} - \alpha^2_{y.w,2}\psi_{ww}\right) + \left(\sigma_{yy} - \alpha^2_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \right)$$

Note that the covariance term of the distribution in Equation 4.10 was written as two 2 x 1 vectors horizontally concatenated, rather than as a 2 x 2 matrix. Equation 4.10 can now be written as a multilevel model. *X* and *Y* are each equal to their conditional means from Equation 4.10 plus their level-2 and level-1 residuals.

$$\begin{aligned}\begin{bmatrix} X_{ij} & Y_{ij} \end{bmatrix} &= \begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w,2}\beta^{(w)}\right) & \left(\beta^{(y)} - \alpha_{y.w,2}\beta^{(w)}\right) \end{bmatrix} \\ &+ \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \end{bmatrix} W_{ij} + \begin{bmatrix} \left(\alpha_{x.w,2} - \alpha_{x.w,1}\right) & \left(\alpha_{y.w,2} - \alpha_{y.w,1}\right) \end{bmatrix} \bar{W}_j \\ &+ \begin{bmatrix} u_j^{(x)} & u_j^{(y)} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(x)} & e_{ij}^{(y)} \end{bmatrix}\end{aligned} \tag{4.11}$$

The conditional covariance matrix elements in Equation 4.10 contain additive terms involving level-1 and level-2 covariance matrix elements, specifically the total (co)variance at a particular level minus the explained (co)variance, i.e., the residual (co)variance at level-1 and level-2. The additive nature of the covariance matrix elements allows the residual covariance matrix in Equation 4.10 to be separated into the level-1 and level-2 residual covariance matrices obtained in a multilevel model.

$$\begin{aligned}\begin{bmatrix} u_{j,x} \\ u_{j,\mathrm{y}} \end{bmatrix} &\sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\psi_{xx} - \alpha_{x.w,2}^2\psi_{ww}\right) & \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) & \left(\psi_{yy} - \alpha_{y.w,2}^2\psi_{ww}\right) \end{bmatrix}\right) \\ \begin{bmatrix} e_{ij,x} \\ e_{ij,\mathrm{y}} \end{bmatrix} &\sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma_{xx} - \alpha_{x.w,1}^2\sigma_{ww}\right) & \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) & \left(\sigma_{yy} - \alpha_{y.w,1}^2\sigma_{ww}\right) \end{bmatrix}\right)\end{aligned} \tag{4.12}$$

The transformed population model in Equation 4.12 cannot be equated to the JM-BD imputation model without adding constraints because it includes a term that captures the unique prediction of the between-cluster portion of the predictor, whereas the imputation model in Equation 4.2 does not. The transformed population model contains distinct estimates of the within- and between-cluster regression coefficients predicting *X* and *Y* from *W*. That is, it contains two coefficients predicting *X* from *W* ($\alpha_{x.w,1}$ and $\alpha_{x.w,2}$) and two coefficients predicting *Y* from *W* ($\alpha_{y.w,1}$ and $\alpha_{y.w,2}$). In contrast, the JM-BD imputation model quantifies the regression of *X* on *W* with a single parameter, $\gamma_1^{(x)}$. Similarly, the regression of *Y* on *W* is also quantified with a single parameter, $\gamma_1^{(y)}$. Using a single parameter to represent two distinct regression coefficients implicitly constrains the coefficients to be equal.

In order to equate the transformed population model and the JM-BD imputation model, we must add constraints to the transformed population model. So, we constrain the level-1 and level-2 coefficients for the regression of each dependent variable on *W* to be equal across levels in the population model.

$$\alpha_{x.w} = \alpha_{x.w,1} = \alpha_{x.w,2} \tag{4.13}$$

$$\alpha_{y.w} = \alpha_{y.w,1} = \alpha_{y.w,2} \tag{4.14}$$

In the above two equations, $\alpha_{x.w}$ and $\alpha_{y.w}$ are labels for the constrained regression parameters. The transformed population distribution in Equation 4.12 also contained a

term representing the level-2 residual covariance, $\psi_{xy} - \frac{\psi_{wx}}{\psi_{ww}}\frac{\psi_{wy}}{\psi_{ww}}\psi_{ww}$. JM-BD constrains the level-2 residual covariance to be equal to zero.

$$\psi_{xy} - \frac{\psi_{wx}}{\psi_{ww}}\frac{\psi_{wy}}{\psi_{ww}}\psi_{ww} = 0 \tag{4.15}$$

Under the constraints in Equations 4.13-4.15, the transformed population model in Equations 4.11 and 4.12 reduces to:

$$\begin{gathered}
\begin{bmatrix} X_{ij} & Y_{ij} \end{bmatrix} = \begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w}\beta^{(w)}\right) & \left(\beta^{(y)} - \alpha_{y.w}\beta^{(w)}\right) \end{bmatrix} + \begin{bmatrix} \alpha_{x.w} & \alpha_{y.w} \end{bmatrix} W_{ij} \\
+ \begin{bmatrix} u_j^{(x)} & u_j^{(y)} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(x)} & e_{ij}^{(y)} \end{bmatrix} \\
\begin{bmatrix} u_{j,x} \\ u_{j,\mathrm{y}} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\psi_{xx} - \alpha_{x.w}^2\psi_{ww}\right) & 0 \\ 0 & \left(\psi_{yy} - \alpha_{y.w}^2\psi_{ww}\right) \end{bmatrix}\right) \\
\begin{bmatrix} e_{ij,x} \\ e_{ij,\mathrm{y}} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma_{xx} - \alpha_{x.w}^2\sigma_{ww}\right) & \left(\sigma_{xy} - \alpha_{x.w}\alpha_{y.w}\sigma_{ww}\right) \\ \left(\sigma_{xy} - \alpha_{x.w}\alpha_{y.w}\sigma_{ww}\right) & \left(\sigma_{yy} - \alpha_{y.w}^2\sigma_{ww}\right) \end{bmatrix}\right)
\end{gathered} \tag{4.16}$$

Notice that the constrained population model has only 12 parameters: three means ($\beta^{(w)}$, $\beta^{(x)}$, $\beta^{(y)}$), three level-2 covariance parameters (ψ_{ww}, ψ_{xx}, and ψ_{yy}), four level-1 covariance parameters (σ_{ww}, σ_{xx}, σ_{xy}, and σ_{yy}), and two parameters quantifying the regression of *X* and *Y* on *W* ($\alpha_{x.w}$ and $\alpha_{y.w}$). Now that the constrained imputation and population models have the same numbers of parameters, the parameters can be equated as follows:

$$\gamma_1^{(x)} = \alpha_{x.w} \tag{4.17}$$

$$\gamma_1^{(y)} = \alpha_{y.w} \tag{4.18}$$

$$\begin{bmatrix} \gamma_0^{(x)} & \gamma_0^{(y)} \end{bmatrix} = \begin{bmatrix} \beta^{(x)} & \beta^{(y)} \end{bmatrix} - \begin{bmatrix} \alpha_{x.w} & \alpha_{y.w} \end{bmatrix} \begin{bmatrix} \beta^{(w)} \end{bmatrix} \tag{4.19}$$

$$\tau_{xx} = \psi_{xx} - \alpha_{x.w}^2 \psi_{ww} \tag{4.20}$$

$$\tau_{yy} = \psi_{yy} - \alpha_{y.w}^2 \psi_{ww} \tag{4.21}$$

$$\begin{bmatrix} \varsigma_{xx} & \varsigma_{xy} \\ \varsigma_{xy} & \varsigma_{yy} \end{bmatrix} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} - \begin{bmatrix} \alpha_{x.w} \\ \alpha_{y.w} \end{bmatrix} \begin{bmatrix} \sigma_{ww} \end{bmatrix} \begin{bmatrix} \alpha_{x.w} & \alpha_{y.w} \end{bmatrix} \tag{4.22}$$

In summary, the answer to the research question "Does JM-BD produce expectations that are equivalent to the joint distribution of the data?" is: no. The JM-BD imputation model forces the level-1 and level-2 regression coefficients to be equal and constrains the level-2 residual covariance to be zero. This finding supports the hypothesis that JM-BD imputation produces expectations that are not equivalent to the joint distribution of the data.

JM-UN. The JM-UN imputation model (joint model with unstructured level-2 covariance matrix) predicting incomplete variables *X* and *Y* from complete variable *W* is:

$$\begin{gathered}\begin{bmatrix} X_{ij} & Y_{ij} \end{bmatrix} = \begin{bmatrix} \gamma_0^{(x)} & \gamma_0^{(y)} \end{bmatrix} + \begin{bmatrix} \gamma_1^{(x)} & \gamma_1^{(y)} \end{bmatrix} W_{ij} + \begin{bmatrix} u_j^{(x)} & u_j^{(y)} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(x)} & e_{ij}^{(y)} \end{bmatrix} \\ \begin{bmatrix} u_j^{(x)} \\ u_j^{(y)} \end{bmatrix} \sim \mathrm{N}_2\left(\mathbf{0}, \boldsymbol{\Psi}\right) = \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{xx} & \tau_{xy} \\ \tau_{xy} & \tau_{yy} \end{bmatrix}\right) \\ \begin{bmatrix} e_{ij}^{(x)} \\ e_{ij}^{(y)} \end{bmatrix} \sim \mathrm{N}_2\left(\mathbf{0}, \boldsymbol{\Sigma}\right) = \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \varsigma_{xx} & \varsigma_{xy} \\ \varsigma_{xy} & \varsigma_{yy} \end{bmatrix}\right)\end{gathered} \tag{4.23}$$

The JM-UN imputation model in Equation 4.23 is the same as the JM-BD imputation model in Equation 4.2, but with the addition of the level-2 residual covariance parameter, τ_{xy}. The predictor variable is usually treated as fixed when describing an imputation model. To facilitate the comparison between the population and imputation models, however, I treated the predictor variable as random. Because the variables (complete and incomplete) are assumed to be distributed as multivariate normal, the fixed vs. random distinction for the predictor variable has no impact on the imputation procedure. The predictor variable has the following marginal distribution:

$$W_{ij} \sim \mathrm{N}(\beta^{(w)}, \psi_{ww} + \sigma_{ww}) \tag{4.24}$$

The parameters in Equation 4.24 are the same for the imputation model and the population model. Equations 4.23 and 4.24 show that the JM-UN imputation model (including the distribution of the complete variable) contains a total of 13 parameters $\left(\gamma_0^{(x)}, \gamma_0^{(y)}, \gamma_1^{(x)}, \gamma_1^{(y)}, \tau_{xx}, \tau_{xy}, \tau_{yy}, \varsigma_{xx}, \varsigma_{xy}, \varsigma_{yy}, \beta^{(w)}, \psi_{ww}, \text{and } \sigma_{ww}\right)$.

The transformation of the unconditional population model in Equation 4.1 to a conditional distribution in the form of a multilevel model is the same for JM-BD and JM-UN. Readers wishing to avoid repetition are advised to skip ahead to the paragraph containing Equations 4.34 and 4.35, which describes the population constraints required by JM-BD. Repeating the information presented for JM-BD, transforming the unconditional population model in Equation 4.1 to a conditional distribution in the form of a multilevel model begins by specifying the joint distribution of the variables:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} \sim \mathrm{N}_3\left(\begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \psi_{ww}+\sigma_{ww} & \psi_{wx}+\sigma_{wx} & \psi_{wy}+\sigma_{wy} \\ \psi_{wx}+\sigma_{wx} & \psi_{xx}+\sigma_{xx} & \psi_{xy}+\sigma_{xy} \\ \psi_{wy}+\sigma_{wy} & \psi_{xy}+\sigma_{xy} & \psi_{yy}+\sigma_{yy} \end{bmatrix}\right) \tag{4.25}$$

Equation 4.25 is equivalent to Equation 4.1, but Equation 4.25 directly describes the distribution of the variables whereas Equation 4.1 described the distributions of the level-1 residuals and level-2 random effects of the variables. Because multilevel models assume that the level-2 random effects of the variables ($b_j^{(\mathrm{w})}$, $b_j^{(x)}$, and $b_j^{(y)}$) are uncorrelated with the level-1 residuals ($\varepsilon_{ij}^{(w)}$, $\varepsilon_{ij}^{(x)}$, and $\varepsilon_{ij}^{(y)}$), the covariance matrix of the variables in Equation 4.25 is simply the sum of the level-1 and level-2 covariance matrices in Equation 4.1.

The multivariate distribution in Equation 4.25 does not recognize the between-cluster portion of the predictor variable, *W*. In the subsection for the JM-BD imputation model I added the cluster mean of *W*, $\bar{W}_j$, to the joint population distribution to account

for the between-cluster portion of *W*. I then centered *W* at its cluster mean, $W_{ij} - \bar{W}_j$. Applying the same steps to Equation 4.25 yields:

$$\begin{bmatrix} W_{ij} - \bar{W}_j \\ \bar{W}_j \\ X_{ij} \\ Y_{ij} \end{bmatrix} \sim \mathrm{N}_4 \left(\begin{bmatrix} 0 \\ \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \sigma_{ww} & 0 & \sigma_{wx} & \sigma_{wy} \\ 0 & \psi_{ww} & \psi_{wx} & \psi_{wy} \\ \sigma_{wx} & \psi_{wx} & \psi_{xx} + \sigma_{xx} & \psi_{xy} + \sigma_{xy} \\ \sigma_{wy} & \psi_{wy} & \psi_{xy} + \sigma_{xy} & \psi_{yy} + \sigma_{yy} \end{bmatrix} \right) \tag{4.26}$$

The level-1 and level-2 covariances between *W* and the two dependent variables, *X* and *Y*, are now represented by different cells in the covariance matrix.

Equation 4.26 can be used to calculate the distribution of *X* and *Y* conditional upon the within-cluster portion *W* and the between-cluster portion of *W*. The regression coefficients are calculated as the inverse of the covariance matrix of the predictors post-multiplied by the matrix of covariances between the independent variables and the dependent variables:

$$\begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix} = \begin{bmatrix} \sigma_{ww} & 0 \\ 0 & \psi_{ww} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{wx} & \sigma_{wy} \\ \psi_{wx} & \psi_{wy} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_{wx}}{\sigma_{ww}} & \frac{\sigma_{wy}}{\sigma_{ww}} \\ \frac{\psi_{wx}}{\psi_{ww}} & \frac{\psi_{wy}}{\psi_{ww}} \end{bmatrix} \tag{4.27}$$

$\alpha_{x.w,1}$ is the within-cluster regression of *X* on *W*. $\alpha_{x.w,2}$ is the between-cluster regression of *X* on *W*. $\alpha_{y.w,1}$ is the within-cluster regression of *Y* on *W*. $\alpha_{y.w,2}$ is the between-cluster regression of *Y* on *W*. The residual covariance matrix is calculated as the

covariance matrix of the dependent variables minus the transposed matrix of regression coefficients multiplied by the covariance matrix of the predictors multiplied by the matrix of regression coefficients (i.e., the total variance minus the variance explained by *W*):

$$\begin{aligned}
&\begin{bmatrix} \psi_{xx}+\sigma_{xx} & \psi_{xy}+\sigma_{xy} \\ \psi_{xy}+\sigma_{xy} & \psi_{yy}+\sigma_{yy} \end{bmatrix} - \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \sigma_{ww} & 0 \\ 0 & \psi_{ww} \end{bmatrix} \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix} \\
&= \begin{bmatrix} \left(\psi_{xx} - \alpha_{x.w,2}^2 \psi_{ww}\right) + \left(\sigma_{xx} - \alpha_{x.w,1}^2 \sigma_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \\
&\| \begin{bmatrix} \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\psi_{yy} - \alpha_{y.w,2}^2 \psi_{ww}\right) + \left(\sigma_{yy} - \alpha_{y.w,1}^2 \sigma_{ww}\right) \end{bmatrix}
\end{aligned} \tag{4.28}$$

Note that Equation 4.28 displays the result of a matrix operation as two 2 x 1 vectors horizontally concatenated (horizontal concatenation is indicated by ||), rather than as a 2 x 2 matrix. The intercepts are calculated as the means of the dependent variables minus the means of the independent variables post-multiplied by the matrix of regression coefficients:

$$\begin{aligned}
&\begin{bmatrix} \beta^{(x)} & \beta^{(y)} \end{bmatrix} - \begin{bmatrix} 0 & \beta^{(w)} \end{bmatrix} \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \\ \alpha_{x.w,2} & \alpha_{y.w,2} \end{bmatrix} \\
&= \begin{bmatrix} \beta^{(x)} - \alpha_{x.w,2}\beta^{(w)} & \beta^{(y)} - \alpha_{y.w,2}\beta^{(w)} \end{bmatrix}
\end{aligned} \tag{4.29}$$

Combining the information in Equations 4.27-4.29, the distribution of *X* and *Y* conditional upon $W_{ij} - \bar{W}_j$ and $\bar{W}_j$ is:

$$X_{ij}, Y_{ij} \mid W_{ij} - \bar{W}_j, \bar{W}_j \sim \mathrm{N}_2\left(\begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w,2}\beta^{(w)}\right) + \alpha_{x.w,1}\left(W_{ij} - \bar{W}_j\right) + \alpha_{x.w,2}\left(\bar{W}_j\right) \\ \left(\beta^{(y)} - \alpha_{y.w,2}\beta^{(w)}\right) + \alpha_{y.w,1}\left(W_{ij} - \bar{W}_j\right) + \alpha_{y.w,2}\left(\bar{W}_j\right) \end{bmatrix}, \right.$$
$$\begin{bmatrix} \left(\psi_{xx} - \alpha^2_{x.w,2}\psi_{ww}\right) + \left(\sigma_{xx} - \alpha^2_{x.w,1}\sigma_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \qquad (4.30)$$
$$\left. \| \begin{bmatrix} \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\psi_{yy} - \alpha^2_{y.w,2}\psi_{ww}\right) + \left(\sigma_{yy} - \alpha^2_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \right)$$

Note that Equation 4.30 displays the a covariance matrix as two 2 x 1 vectors horizontally concatenated (horizontal concatenation is indicated by ||), rather than as a 2 x 2 matrix. The terms in Equation 4.30 correspond to the terms in a multilevel model predicting *X* and *Y* from $W_{ij} - \bar{W}_j$ and $\bar{W}_j$. I centered *W* at its cluster mean in the transformed population model to simplify the calculation of the regression coefficients. However, *W* is not centered in the JM-UN imputation model. In order to equate the transformed population model to the imputation model, it is necessary to return *W* to its raw metric in the transformed population model. This can be done by rearranging terms in Equation 4.30, yielding:

$$X_{ij}, Y_{ij} \mid W_{ij} - \bar{W}_j, \bar{W}_j \sim \mathrm{N}_2\left(\begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w,2}\beta^{(w)}\right) + \alpha_{x.w,1}\left(W_{ij}\right) + \left(\alpha_{x.w,2} - \alpha_{x.w,1}\right)\left(\bar{W}_j\right) \\ \left(\beta^{(y)} - \alpha_{y.w,2}\beta^{(w)}\right) + \alpha_{y.w,1}\left(W_{ij}\right) + \left(\alpha_{y.w,2} - \alpha_{y.w,1}\right)\left(\bar{W}_j\right) \end{bmatrix}, \right.$$
$$\begin{bmatrix} \left(\psi_{xx} - \alpha_{x.w,2}^2\psi_{ww}\right) + \left(\sigma_{xx} - \alpha_{x.w,1}^2\sigma_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \end{bmatrix} \tag{4.31}$$
$$\left. \| \begin{bmatrix} \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) + \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\psi_{yy} - \alpha_{y.w,2}^2\psi_{ww}\right) + \left(\sigma_{yy} - \alpha_{y.w,1}^2\sigma_{ww}\right) \end{bmatrix} \right)$$

Note that Equation 4.31 displays the a covariance matrix as two 2 x 1 vectors horizontally concatenated (horizontal concatenation is indicated by ||), rather than as a 2 x 2 matrix. Equation 4.31 can now be written as a multilevel model. *X* and *Y* are each equal to their conditional means from Equation 4.31 plus their level-2 and level-1 residuals.

$$\begin{bmatrix} X_{ij} & Y_{ij} \end{bmatrix} = \begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w,2}\beta^{(w)}\right) & \left(\beta^{(y)} - \alpha_{y.w,2}\beta^{(w)}\right) \end{bmatrix}$$
$$+ \begin{bmatrix} \alpha_{x.w,1} & \alpha_{y.w,1} \end{bmatrix} W_{ij} + \begin{bmatrix} \left(\alpha_{x.w,2} - \alpha_{x.w,1}\right) & \left(\alpha_{y.w,2} - \alpha_{y.w,1}\right) \end{bmatrix} \bar{W}_j \tag{4.32}$$
$$+ \begin{bmatrix} u_j^{(x)} & u_j^{(y)} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(x)} & e_{ij}^{(y)} \end{bmatrix}$$

The conditional covariance matrix elements in Equation 4.31 contain additive terms involving level-1 and level-2 covariance matrix elements, specifically the total (co)variance at a particular level minus the explained (co)variance, i.e., the residual (co)variance at level-1 and level-2. The additive nature of the covariance matrix elements allows the residual covariance matrix in Equation 4.31 to be separated into the level-1 and level-2 residual covariance matrices obtained in a multilevel model.

$$\begin{bmatrix} u_{j,x} \\ u_{j,\mathrm{y}} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\psi_{xx} - \alpha_{x.w,2}^2 \psi_{ww}\right) & \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w,2}\alpha_{y.w,2}\psi_{ww}\right) & \left(\psi_{yy} - \alpha_{y.w,2}^2 \psi_{ww}\right) \end{bmatrix}\right)$$
$$\begin{bmatrix} e_{ij,x} \\ e_{ij,\mathrm{y}} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma_{xx} - \alpha_{x.w,1}^2 \sigma_{ww}\right) & \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) \\ \left(\sigma_{xy} - \alpha_{x.w,1}\alpha_{y.w,1}\sigma_{ww}\right) & \left(\sigma_{yy} - \alpha_{y.w,1}^2 \sigma_{ww}\right) \end{bmatrix}\right) \quad (4.33)$$

The transformed population model in Equation 4.33 cannot be equated to the JM-UN imputation model without adding constraints because it includes a term that captures the unique prediction of the between-cluster portion of the predictor, whereas the imputation model in Equation 4.23 does not. The transformed population model contains distinct estimates of the within- and between-cluster regression coefficients predicting *X* and *Y* from *W*. That is, it contains two coefficients predicting *X* from *W* ($\alpha_{x.w,1}$ and $\alpha_{x.w,2}$) and two coefficients predicting *Y* from *W* ($\alpha_{y.w,1}$ and $\alpha_{y.w,2}$). In contrast, the JM-UN imputation model quantifies the regression of *X* on *W* with a single parameter, $\gamma_1^{(x)}$. Similarly, the regression of *Y* on *W* is also quantified with a single parameter, $\gamma_1^{(y)}$. Using a single parameter to represent two distinct regression coefficients implicitly constrains the coefficients to be equal.

In order to equate the transformed population model and the JM-UN imputation model, we must add constraints to the transformed population model. So, we constrain the level-1 and level-2 coefficients for the regressions of the dependent variables on *W* to be equal across levels in the population model.

$$\alpha_{x.w} = \alpha_{x.w,1} = \alpha_{x.w,2} \quad (4.34)$$

$$\alpha_{y.w} = \alpha_{y.w,1} = \alpha_{y.w,2} \tag{4.35}$$

In the above two equations, $\alpha_{x.w}$ and $\alpha_{y.w}$ are labels for the constrained regression parameters.

Under the constraints in Equations 4.34 and 4.35, the transformed population model in Equations 4.32 and 4.33 reduces to:

$$\begin{aligned}
\begin{bmatrix} X_{ij} & Y_{ij} \end{bmatrix} &= \begin{bmatrix} \left(\beta^{(x)} - \alpha_{x.w}\beta^{(w)}\right) & \left(\beta^{(y)} - \alpha_{y.w}\beta^{(w)}\right) \end{bmatrix} + \begin{bmatrix} \alpha_{x.w} & \alpha_{y.w} \end{bmatrix} W_{ij} \\
&\quad + \begin{bmatrix} u_j^{(x)} & u_j^{(y)} \end{bmatrix} + \begin{bmatrix} e_{ij}^{(x)} & e_{ij}^{(y)} \end{bmatrix} \\
\begin{bmatrix} u_{j,x} \\ u_{j,\mathrm{y}} \end{bmatrix} &\sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\psi_{xx} - \alpha_{x.w}^2\psi_{ww}\right) & \left(\psi_{xy} - \alpha_{x.w}\alpha_{y.w}\psi_{ww}\right) \\ \left(\psi_{xy} - \alpha_{x.w}\alpha_{y.w}\psi_{ww}\right) & \left(\psi_{yy} - \alpha_{y.w}^2\psi_{ww}\right) \end{bmatrix} \right) \\
\begin{bmatrix} e_{ij,x} \\ e_{ij,\mathrm{y}} \end{bmatrix} &\sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma_{xx} - \alpha_{x.w}^2\sigma_{ww}\right) & \left(\sigma_{xy} - \alpha_{x.w}\alpha_{y.w}\sigma_{ww}\right) \\ \left(\sigma_{xy} - \alpha_{x.w}\alpha_{y.w}\sigma_{ww}\right) & \left(\sigma_{yy} - \alpha_{y.w}^2\sigma_{ww}\right) \end{bmatrix} \right)
\end{aligned} \tag{4.36}$$

Notice that the constrained population model has only 13 parameters: three means ($\beta^{(w)}$, $\beta^{(x)}$, $\beta^{(y)}$), four level-2 covariance parameters (ψ_{ww}, ψ_{xx}, ψ_{xy}, and ψ_{yy}), four level-1 covariance parameters (σ_{ww}, σ_{xx}, σ_{xy}, and σ_{yy}), and two parameters quantifying the regressions of *X* and *Y* on *W* ($\alpha_{x.w}$ and $\alpha_{y.w}$). Now that the constrained imputation and population models have the same numbers of parameters, the parameters can be equated as follows:

$$\gamma_1^{(x)} = \alpha_{x.w} \tag{4.37}$$

$$\gamma_1^{(y)} = \alpha_{y.w} \tag{4.38}$$

$$\begin{bmatrix} \gamma_0^{(x)} & \gamma_0^{(y)} \end{bmatrix} = \begin{bmatrix} \beta^{(x)} & \beta^{(y)} \end{bmatrix} - \begin{bmatrix} \alpha_{x.w} & \alpha_{y.w} \end{bmatrix} \begin{bmatrix} \beta^{(w)} \end{bmatrix} \tag{4.39}$$

$$\begin{bmatrix} \tau_{xx} & \tau_{xy} \\ \tau_{xy} & \tau_{yy} \end{bmatrix} = \begin{bmatrix} \psi_{xx} & \psi_{xy} \\ \psi_{xy} & \psi_{yy} \end{bmatrix} - \begin{bmatrix} \alpha_{x.w} \\ \alpha_{y.w} \end{bmatrix} \begin{bmatrix} \psi_{ww} \end{bmatrix} \begin{bmatrix} \alpha_{x.w} & \alpha_{y.w} \end{bmatrix} \tag{4.40}$$

$$\begin{bmatrix} \varsigma_{xx} & \varsigma_{xy} \\ \varsigma_{xy} & \varsigma_{yy} \end{bmatrix} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} - \begin{bmatrix} \alpha_{x.w} \\ \alpha_{y.w} \end{bmatrix} \begin{bmatrix} \sigma_{ww} \end{bmatrix} \begin{bmatrix} \alpha_{x.w} & \alpha_{y.w} \end{bmatrix} \tag{4.41}$$

In summary, the answer to the research question “Does JM-UN produce expectations that are equivalent to the joint distribution of the data?” is: no. The JM-UN imputation model forces the level-1 and level-2 regression coefficients to be equal. This finding supports the hypothesis that JM-UN imputation produces expectations that are not equivalent to the joint distribution of the data. As a reminder, including the complete-variable cluster means in the imputation model would result in JM-UN being equivalent to the joint distribution of the data. This was not done here, however, because the complete-variable cluster means are not included in JM-UN by default and the majority of analysts tend to use the default options for analysis procedures.

JM-Mplus. The JM-Mplus imputation model treating *W*, *X*, and *Y* as response variables is:

$$
\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix} + \begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix}
$$

$$
\begin{bmatrix} b_j^{(\mathrm{w})} \\ b_j^{(\mathrm{x})} \\ b_j^{(y)} \end{bmatrix} \sim \mathrm{N}_3 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{ww} & \psi_{wx} & \psi_{wy} \\ \psi_{wx} & \psi_{xx} & \psi_{xy} \\ \psi_{wy} & \psi_{wy} & \psi_{yy} \end{bmatrix} \right) \tag{4.42}
$$

$$
\begin{bmatrix} \varepsilon_{ij}^{(w)} \\ \varepsilon_{ij}^{(x)} \\ \varepsilon_{ij}^{(\mathrm{y})} \end{bmatrix} \sim \mathrm{N}_3 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{ww} & \sigma_{wx} & \sigma_{wy} \\ \sigma_{wx} & \sigma_{xx} & \sigma_{xy} \\ \sigma_{wy} & \sigma_{xy} & \sigma_{yy} \end{bmatrix} \right)
$$

Notice that JM-Mplus imputation model in Equation 4.42 is identical to the population model in Equation 4.1. There is a one-to-one correspondence between the imputation model parameters and the population model parameters. As such, the JM-Mplus imputation model produces expectations that are equivalent to the joint distribution of the data. In summary, the answer to the research question "Does JM-Mplus produce expectations that are equivalent to the joint distribution of the data?" is: yes. This finding supports the hypothesis that JM-Mplus imputation produces expectations that are equivalent to the joint distribution of the data.

FCS-VB. The FCS-VB imputation method uses a separate univariate imputation model for each incomplete variable. The imputation model predicting *X* from *W* and *Y* is:

$$
\begin{aligned} X_{ij} &= \gamma_0^{(x)} + \gamma_1^{(x)} W_{ij} + \gamma_2^{(x)} Y_{ij} + u_j^{(x)} + e_{ij}^{(x)} \\ u_j^{(x)} &\sim \mathrm{N}\left(0, \tau_{xx}\right) \\ e_{ij}^{(x)} &\sim \mathrm{N}\left(0, \varsigma_{xx}\right) \end{aligned} \tag{4.43}
$$

The predictor variables are usually treated as fixed when describing an imputation model. To facilitate the comparison between the population and imputation models, however, I treated the predictor variables as random. Because the variables (complete and incomplete) are assumed to be distributed as multivariate normal, the fixed vs. random distinction for the predictor variables has no impact on the imputation procedure. The predictor variables have the following marginal distribution:

$$\begin{bmatrix} W_{ij} \\ Y_{ij} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} \beta^{(w)} \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \psi_{ww}+\sigma_{ww} & \psi_{wy}+\sigma_{wy} \\ \psi_{wy}+\sigma_{wy} & \psi_{yy}+\sigma_{yy} \end{bmatrix}\right) \tag{4.44}$$

The parameters in Equation 4.44 are the same for the imputation model and the population model. Equations 4.43 and 4.44 show that the FCS-VB imputation model (including the distribution of the predictors, *W* and *Y*) contains a total of 13 parameters $\left(\gamma_0^{(x)},\ \gamma_1^{(x)},\ \gamma_2^{(x)},\ \tau_{xx},\ \varsigma_{xx},\ \beta^{(w)},\ \beta^{(y)},\ \psi_{ww},\ \psi_{wy},\ \psi_{yy},\ \sigma_{ww},\ \sigma_{wy},\ \text{and}\ \sigma_{yy}\right)$.

The imputation model predicting *Y* from *W* and *X* is:

$$\begin{aligned} Y_{ij} &= \gamma_0^{(y)} + \gamma_1^{(y)} W_{ij} + \gamma_2^{(y)} X_{ij} + u_j^{(y)} + e_{ij}^{(y)} \\ u_j^{(y)} &\sim \mathrm{N}\left(0, \tau_{xx}\right) \\ e_{ij}^{(y)} &\sim \mathrm{N}\left(0, \varsigma_{xx}\right) \end{aligned} \tag{4.45}$$

The predictor variables are usually treated as fixed when describing an imputation model. To facilitate the comparison between the population and imputation models, however, I treated the predictor variables as random. Because the variables (complete and incomplete) are assumed to be distributed as multivariate normal, the fixed vs. random distinction for the predictor variables has no impact on the imputation procedure. The predictor variables have the following marginal distribution:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \end{bmatrix}, \begin{bmatrix} \psi_{ww} + \sigma_{ww} & \psi_{wx} + \sigma_{wx} \\ \psi_{wx} + \sigma_{wx} & \psi_{xx} + \sigma_{xx} \end{bmatrix}\right) \tag{4.46}$$

Equations 4.45 and 4.46 show that the FCS-VB imputation model (including the distribution of the predictors, *W* and *X*) contains a total of 13 parameters $\left(\gamma_0^{(y)}, \gamma_1^{(y)}, \gamma_2^{(y)}, \tau_{xx}, \varsigma_{xx}, \beta^{(w)}, \beta^{(x)}, \psi_{ww}, \psi_{wx}, \psi_{xx}, \sigma_{ww}, \sigma_{wx}, \text{and } \sigma_{xx}\right)$.

Equations 4.43-4.46 show that the number of parameters is the same for each of the univariate imputation models employed by FCS-VB. Further, the structure of each imputation model is the same (one incomplete variable predicted from all other variables). As such, equivalence (or lack thereof) between the FCS-VB imputation model for *Y* and the population model would also imply the same between the FCS-VB imputation model for *X* and the population model. Because the findings for one imputation model should be generalizable to the other, I examined only the FCS-VB imputation model for *Y* in this section. This serves to streamline the section and avoid

confusion that might arise from providing two sets of similar but slightly different equations.

To transform the unconditional population model in Equation 4.1 to a conditional distribution in the form of a multilevel model, we begin by specifying the joint distribution of the variables:

$$\begin{bmatrix} W_{ij} \\ X_{ij} \\ Y_{ij} \end{bmatrix} \sim \mathrm{N}_3\left(\begin{bmatrix} \beta^{(w)} \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \psi_{ww}+\sigma_{ww} & \psi_{wx}+\sigma_{wx} & \psi_{wy}+\sigma_{wy} \\ \psi_{wx}+\sigma_{wx} & \psi_{xx}+\sigma_{xx} & \psi_{xy}+\sigma_{xy} \\ \psi_{wy}+\sigma_{wy} & \psi_{xy}+\sigma_{xy} & \psi_{yy}+\sigma_{yy} \end{bmatrix}\right) \tag{4.47}$$

Equation 4.47 is equivalent to Equation 4.1, but Equation 4.47 directly describes the distribution of the variables whereas Equation 4.1 described the distributions of the level-1 residuals and level-2 random effects of the variables. Because multilevel models assume that the level-2 random effects of the variables ($b_j^{(\mathrm{w})}$, $b_j^{(x)}$, and $b_j^{(y)}$) are uncorrelated with the level-1 residuals ($\varepsilon_{ij}^{(w)}$, $\varepsilon_{ij}^{(x)}$, and $\varepsilon_{ij}^{(y)}$), the covariance matrix of the variables in Equation 4.47 is simply the sum of the level-1 and level-2 covariance matrices in Equation 4.1.

The multivariate distribution in Equation 4.47 does not recognize the between-cluster portions of the predictor variables, *W* and *X*. In the subsection for the JM-BD imputation model I added the cluster mean of *W*, $\bar{W}_j$, to the joint population distribution to account for the between-cluster portion of *W*. I then centered *W* at its cluster mean, $W_{ij} - \bar{W}_j$. Applying the same steps to Equation 4.47 for both predictors, *W* and *X*, yields:

$$\begin{bmatrix} W_{ij} - \bar{W}_j \\ \bar{W}_j \\ X_{ij} - \bar{X}_j \\ \bar{X}_j \\ Y_{ij} \end{bmatrix} \sim \mathrm{N}_5 \left(\begin{bmatrix} 0 \\ \beta^{(w)} \\ 0 \\ \beta^{(x)} \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \sigma_{ww} & 0 & \sigma_{wx} & 0 & \sigma_{wy} \\ 0 & \psi_{ww} & 0 & \psi_{wx} & \psi_{wy} \\ \sigma_{wx} & 0 & \sigma_{xx} & 0 & \sigma_{xy} \\ 0 & \psi_{wx} & 0 & \psi_{xx} & \psi_{xy} \\ \sigma_{wy} & \psi_{wy} & \sigma_{xy} & \psi_{xy} & \psi_{yy} + \sigma_{yy} \end{bmatrix} \right) \tag{4.48}$$

The level-1 and level-2 covariances between the two predictors, *W* and *X*, and the dependent variable, *Y*, are now represented by different cells in the covariance matrix.

Equation 4.48 can be used to calculate the distribution of *Y* conditional upon the within-cluster portions of *W* and *X* and the between-cluster portions of *W* and *X*. The regression coefficients are calculated as the inverse of the covariance matrix of the predictors post-multiplied by the vector of covariances between the independent variables and the dependent variables:

$$\begin{bmatrix} \alpha_{y.w,1} \\ \alpha_{y.w,2} \\ \alpha_{y.x,1} \\ \alpha_{y.x,2} \end{bmatrix} = \begin{bmatrix} \sigma_{ww} & 0 & \sigma_{wx} & 0 \\ 0 & \psi_{ww} & 0 & \psi_{wx} \\ \sigma_{wx} & 0 & \sigma_{xx} & 0 \\ 0 & \psi_{wx} & 0 & \psi_{xx} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{wy} \\ \psi_{wy} \\ \sigma_{xy} \\ \psi_{xy} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_{wy}}{\sigma_{ww}} \\ \frac{\psi_{wy}}{\psi_{ww}} \\ \frac{\sigma_{xy}}{\sigma_{xx}} \\ \frac{\psi_{xy}}{\psi_{xx}} \end{bmatrix} \tag{4.49}$$

$\alpha_{y.w,1}$ is the within-cluster regression of *Y* on *W*. $\alpha_{y.w,2}$ is the between-cluster regression of *Y* on *W*. $\alpha_{y.x,1}$ is the within-cluster regression of *Y* on *X*. $\alpha_{y.x,2}$ is the between-cluster

regression of *Y* on *X*. The residual variance of *Y* is calculated as the variance of *Y* minus the transposed matrix of regression coefficients multiplied by the covariance matrix of the predictors multiplied by the matrix of regression coefficients (i.e., the total variance minus the variance explained by *W* and *X*):

$$\begin{aligned}\left(\psi_{yy}+\sigma_{yy}\right)-&\begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.w,2}\\ \alpha_{y.x,1}\\ \alpha_{y.x,2}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\sigma_{ww} & 0 & \sigma_{wx} & 0\\ 0 & \psi_{ww} & 0 & \psi_{wx}\\ \sigma_{wx} & 0 & \sigma_{xx} & 0\\ 0 & \psi_{wx} & 0 & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.w,2}\\ \alpha_{y.x,1}\\ \alpha_{y.x,2}\end{bmatrix}\\ &=\left(\psi_{yy}-\begin{bmatrix}\alpha_{y.w,2}\\ \alpha_{y.x,2}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\psi_{ww} & \psi_{wx}\\ \psi_{wx} & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,2}\\ \alpha_{y.x,2}\end{bmatrix}\right)\\ &+\left(\sigma_{yy}-\begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.x,1}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\sigma_{ww} & \sigma_{wx}\\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.x,1}\end{bmatrix}\right)\end{aligned} \tag{4.50}$$

The intercepts are calculated as the mean of the dependent variable minus the means of the independent variables post-multiplied by the matrix of regression coefficients:

$$\beta^{(y)}-\begin{bmatrix}0 & \beta^{(x)} & 0 & \beta^{(w)}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.w,2}\\ \alpha_{y.x,1}\\ \alpha_{y.x,2}\end{bmatrix}=\beta^{(y)}-\beta^{(x)}\alpha_{y.w,2}-\beta^{(w)}\alpha_{y.x,2} \tag{4.51}$$

Combining the information in Equations 4.49-4.51, the distribution of *Y* conditional on $W_{ij}-\bar{W}_j$, $\bar{W}_j$, $X_{ij}-\bar{X}_j$, and $\bar{X}_j$ is:

$$
\begin{aligned}
&Y_{ij} \mid W_{ij} - \bar{W}_j, \bar{W}_j, X_{ij} - \bar{X}_j, \bar{X}_j \sim \\
&\mathrm{N}\left(\begin{array}{l}\left(\beta^{(y)} - \beta^{(x)}\alpha_{y.w,2} - \beta^{(w)}\alpha_{y.x,2}\right) \\ +\alpha_{y.w,1}\left(W_{ij} - \bar{W}_j\right) + \alpha_{y.w,2}\left(\bar{W}_j\right) + \alpha_{y.x,1}\left(X_{ij} - \bar{X}_j\right) + \alpha_{y.x,2}\left(\bar{X}_j\right),\end{array}\right. \\
&\left.\left(\psi_{yy} - \begin{bmatrix}\alpha_{y.w,2} \\ \alpha_{y.x,2}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\psi_{ww} & \psi_{wx} \\ \psi_{wx} & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,2} \\ \alpha_{y.x,2}\end{bmatrix}\right) + \left(\sigma_{yy} - \begin{bmatrix}\alpha_{y.w,1} \\ \alpha_{y.x,1}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\sigma_{ww} & \sigma_{wx} \\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,1} \\ \alpha_{y.x,1}\end{bmatrix}\right)\right)
\end{aligned}
\tag{4.52}
$$

The terms in Equation 4.52 correspond to the terms in a multilevel model predicting *Y* from $W_{ij} - \bar{W}_j$, $\bar{W}_j$, $X_{ij} - \bar{X}_j$, and $\bar{X}_j$. I centered *W* and *X* at their cluster means in the transformed population model to simplify the calculation of the regression coefficients. However, *W* and *X* are not centered in the FCS-VB imputation model. In order to equate the transformed population model to the imputation model, it is necessary to return *W* and *X* to their raw metrics in the transformed population model. This can be done by rearranging terms in Equation 4.52, yielding:

$$
\begin{aligned}
&Y_{ij} \mid W_{ij}, \bar{W}_j, X_{ij}, \bar{X}_j \sim \\
&\mathrm{N}\Big(\Big(\beta^{(y)} - \beta^{(x)}\alpha_{y.w,2} - \beta^{(w)}\alpha_{y.x,2}\Big) \\
&+\alpha_{y.w,1}\left(W_{ij}\right) + \left(\alpha_{y.w,2} - \alpha_{y.w,1}\right)\left(\bar{W}_j\right) + \alpha_{y.x,1}\left(X_{ij}\right) + \left(\alpha_{y.x,2} - \alpha_{y.x,1}\right)\left(\bar{X}_j\right), \\
&\left(\psi_{yy} - \begin{bmatrix}\alpha_{y.w,2} \\ \alpha_{y.x,2}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\psi_{ww} & \psi_{wx} \\ \psi_{wx} & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,2} \\ \alpha_{y.x,2}\end{bmatrix}\right) \\
&+\left(\sigma_{yy} - \begin{bmatrix}\alpha_{y.w,1} \\ \alpha_{y.x,1}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\sigma_{ww} & \sigma_{wx} \\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,1} \\ \alpha_{y.x,1}\end{bmatrix}\right)
\end{aligned}
\tag{4.53}
$$

Equation 4.53 can now be written as a multilevel model. *Y* is equal to its conditional mean from Equation 4.53 plus its level-2 and level-1 residuals.

$$\begin{aligned} Y_{ij} &= \left(\beta^{(y)} - \beta^{(x)}\alpha_{y.w,2} - \beta^{(w)}\alpha_{y.x,2}\right) \\ &+\alpha_{y.w,1}\left(W_{ij}\right) + \left(\alpha_{y.w,2} - \alpha_{y.w,1}\right)\left(\bar{W}_j\right) + \alpha_{y.x,1}\left(X_{ij}\right) + \left(\alpha_{y.x,2} - \alpha_{y.x,1}\right)\left(\bar{X}_j\right) \\ &+u_j^{(y)} + e_{ij}^{(y)} \end{aligned} \quad (4.54)$$

The conditional variance in Equation 4.53 contains additive terms involving level-1 and level-2 variance elements, specifically the total variance at a particular level minus the explained variance, i.e., the residual variance at level-1 and level-2. The additive nature of the covariance matrix elements allows the residual variance in Equation 4.53 to be separated into the level-1 and level-2 residual covariance matrices obtained in a multilevel model.

$$\begin{aligned} u_{j,\mathrm{y}} &\sim \mathrm{N}\left(0, \psi_{yy} - \begin{bmatrix}\alpha_{y.w,2}\\ \alpha_{y.x,2}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\psi_{ww} & \psi_{wx}\\ \psi_{wx} & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,2}\\ \alpha_{y.x,2}\end{bmatrix}\right) \\ e_{ij,\mathrm{y}} &\sim \mathrm{N}\left(0, \sigma_{yy} - \begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.x,1}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\sigma_{ww} & \sigma_{wx}\\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.x,1}\end{bmatrix}\right) \end{aligned} \quad (4.55)$$

The transformed population model in Equations 4.54 and 4.55 cannot be equated to the FCS-VB imputation model without adding constraints because they include a term that captures the unique prediction of the between-cluster portion of the predictor, whereas the imputation model in Equation 4.45 does not. The transformed population

model contains distinct estimates of the within- and between-cluster regression coefficients predicting *Y* from *W* and *X*. That is, it contains two coefficients predicting *Y* from *W* $\left(\alpha_{y.w,1}\right.$ and $\left.\alpha_{y.w,2}\right)$ and two coefficients predicting *Y* from *X* $\left(\alpha_{y.x,1}\right.$ and $\left.\alpha_{y.x,2}\right)$. In contrast, the FCS-VB imputation predicting *Y* from *W* and *X* model quantifies the regression of *Y* on *W* with a single parameter, $\gamma_1^{(y)}$. Similarly, the regression of *Y* on *X* is also quantified with a single parameter, $\gamma_2^{(y)}$. Using a single parameter to represent two distinct regression coefficients implicitly constrains the coefficients to be equal.

In order to equate the transformed population model and the FCS-VB imputation model, we must add constraints to the transformed population model. So, we constrain the level-1 and level-2 regressions of *Y* on the independent variables to be equal across levels in the population model.

$$\alpha_{y.w} = \alpha_{y.w,1} = \alpha_{y.w,2} \tag{4.56}$$

$$\alpha_{y.x} = \alpha_{y.x,1} = \alpha_{y.x,2} \tag{4.57}$$

In the above two equations, $\alpha_{y.w}$ and $\alpha_{y.x}$ are labels for the constrained regression parameters.

Under the constraints in Equations 4.56 and 4.57, the transformed population model in Equations 4.54 and 4.55 reduces to:

$$
\begin{aligned}
Y_{ij} &= \left(\beta^{(y)} - \beta^{(x)}\alpha_{y.w} - \beta^{(w)}\alpha_{y.x}\right) + \alpha_{y.w}\left(W_{ij}\right) + \alpha_{y.x}\left(X_{ij}\right) + u_j^{(y)} + e_{ij}^{(y)} \\
u_{j,\mathrm{y}} &\sim \mathrm{N}\left(0, \psi_{yy} - \begin{bmatrix}\alpha_{y.w}\\ \alpha_{y.x}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\psi_{ww} & \psi_{wx}\\ \psi_{wx} & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w}\\ \alpha_{y.x}\end{bmatrix}\right) \\
e_{ij,\mathrm{y}} &\sim \mathrm{N}\left(0, \sigma_{yy} - \begin{bmatrix}\alpha_{y.w}\\ \alpha_{y.x}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\sigma_{ww} & \sigma_{wx}\\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w}\\ \alpha_{y.x}\end{bmatrix}\right)
\end{aligned}
\tag{4.58}
$$

Notice that the constrained population model has only 13 parameters: three means ($\beta^{(w)}$, $\beta^{(x)}$, $\beta^{(y)}$), four level-2 covariance parameters (ψ_{ww}, ψ_{xx}, ψ_{xy}, and ψ_{yy}), four level-1 covariance parameters (σ_{ww}, σ_{xx}, σ_{xy}, and σ_{yy}), and two parameters quantifying the slopes coefficients for the regression of Y on W and X ($\alpha_{y.w}$ and $\alpha_{y.x}$). Now that the imputation and constrained population models have the same numbers of parameters, the parameters can be equated as follows:

$$\gamma_1^{(y)} = \alpha_{y.w} \tag{4.59}$$

$$\gamma_2^{(y)} = \alpha_{y.x} \tag{4.60}$$

$$\gamma_0^{(y)} = \beta^{(y)} - \begin{bmatrix}\beta^{(w)} & \beta^{(x)}\end{bmatrix}\begin{bmatrix}\alpha_{y.w}\\ \alpha_{y.x}\end{bmatrix} \tag{4.61}$$

$$\tau_{yy} = \psi_{yy} - \begin{bmatrix}\alpha_{y.w}\\ \alpha_{y.x}\end{bmatrix}\begin{bmatrix}\psi_{ww} & \psi_{ww}\\ \psi_{wx} & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w} & \alpha_{y.x}\end{bmatrix} \tag{4.62}$$

$$\varsigma_{yy} = \sigma_{yy}\begin{bmatrix}\sigma_{xx} & \sigma_{xy}\\ \sigma_{xy} & \sigma_{yy}\end{bmatrix} - \begin{bmatrix}\alpha_{y.w}\\ \alpha_{y.x}\end{bmatrix}\begin{bmatrix}\sigma_{ww} & \sigma_{ww}\\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w} & \alpha_{y.x}\end{bmatrix} \tag{4.63}$$

In summary, the answer to the research question "Does FCS-VB produce expectations that are equivalent to the joint distribution of the data?" is: no. The FCS-VB imputation model forces the level-1 and level-2 regression coefficients to be equal. This finding supports the hypothesis that FCS-VB imputation produces expectations that are not equivalent to the joint distribution of the data.

FCS-CK. The FCS-CK imputation method uses a separate univariate imputation model for each incomplete variable. The imputation model predicting X from W and Y is:

$$\begin{gathered} X_{ij} = \gamma_0^{(x)} + \gamma_1^{(x)}\left(W_{ij} - \overline{W}_j\right) + \gamma_2^{(x)}\overline{W}_j \\ +\gamma_3^{(x)}\left(Y_{ij} - \overline{Y}_j\right) + \gamma_4^{(x)}\left(\overline{Y}_j\right) + u_j^{(x)} + e_{ij}^{(x)} \\ u_j^{(x)} \sim \mathrm{N}\left(0, \tau_{xx}\right) \\ e_{ij}^{(x)} \sim \mathrm{N}\left(0, \varsigma_{xx}\right) \end{gathered} \tag{4.64}$$

The predictor variables are usually treated as fixed when describing an imputation model. To facilitate the comparison between the population and imputation models, however, I treated the predictor variables as random. Because the variables (complete and incomplete) are assumed to be distributed as multivariate normal, the fixed vs. random distinction for the predictor variables has no impact on the imputation procedure. The predictor variables have the following marginal distribution:

$$\begin{bmatrix} W_{ij} - \overline{W}_j \\ \overline{W}_j \\ Y_{ij} - \overline{Y}_j \\ \overline{Y}_j \end{bmatrix} \sim \mathrm{N}_4 \left(\begin{bmatrix} 0 \\ \beta^{(w)} \\ 0 \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \sigma_{ww} & 0 & \sigma_{wy} & 0 \\ 0 & \psi_{ww} & 0 & \psi_{wy} \\ \sigma_{wy} & 0 & \sigma_{yy} & 0 \\ 0 & \psi_{wy} & 0 & \psi_{yy} \end{bmatrix} \right) \tag{4.65}$$

The parameters in Equation 4.65 are the same for the imputation model and the population model. Equations 4.64 and 4.65 show that the FCS-CK imputation model (including the distribution of the predictors, *W* and *Y*) contains a total of 15 parameters $\left(\gamma_0^{(x)}, \gamma_1^{(x)}, \gamma_2^{(x)}, \gamma_3^{(x)}, \gamma_4^{(x)}, \tau_{xx}, \varsigma_{xx}, \beta^{(w)}, \beta^{(y)}, \psi_{ww}, \psi_{wy}, \psi_{yy}, \sigma_{ww}, \sigma_{wy}, \text{and } \sigma_{yy}\right)$.

The imputation model predicting *Y* from *W* and *X* is:

$$\begin{gathered} Y_{ij} = \gamma_0^{(y)} + \gamma_1^{(y)} \left(W_{ij} - \overline{W}_j \right) + \gamma_2^{(y)} \overline{W}_j \\ + \gamma_3^{(y)} \left(X_{ij} - \overline{X}_j \right) + \gamma_4^{(y)} \overline{X}_j + u_j^{(y)} + e_{ij}^{(y)} \\ u_j^{(y)} \sim \mathrm{N}\left(0, \tau_{xx}\right) \\ e_{ij}^{(y)} \sim \mathrm{N}\left(0, \varsigma_{xx}\right) \end{gathered} \tag{4.66}$$

The predictor variables are usually treated as fixed when describing an imputation model. To facilitate the comparison between the population and imputation models, however, I treated the predictor variables as random. Because the variables (complete and incomplete) are assumed to be distributed as multivariate normal, the fixed vs. random distinction for the predictor variables has no impact on the imputation procedure. The predictor variables have the following marginal distribution:

$$\begin{bmatrix} W_{ij} - \overline{W}_j \\ \overline{W}_j \\ X_{ij} - \overline{X}_j \\ \overline{X}_j \end{bmatrix} \sim \mathrm{N}_4 \left(\begin{bmatrix} 0 \\ \beta^{(w)} \\ 0 \\ \beta^{(y)} \end{bmatrix}, \begin{bmatrix} \sigma_{ww} & 0 & \sigma_{wx} & 0 \\ 0 & \psi_{ww} & 0 & \psi_{wx} \\ \sigma_{wx} & 0 & \sigma_{xx} & 0 \\ 0 & \psi_{wx} & 0 & \psi_{xx} \end{bmatrix} \right) \tag{4.67}$$

Equations 4.66 and 4.67 show that the FCS-CK imputation model (including the distribution of the predictors, *W* and *X*) contains a total of 15 parameters $\left(\gamma_0^{(y)}, \gamma_1^{(y)}, \gamma_2^{(y)}, \gamma_3^{(y)}, \gamma_4^{(y)}, \tau_{xx}, \varsigma_{xx}, \beta^{(w)}, \beta^{(x)}, \psi_{ww}, \psi_{wx}, \psi_{xx}, \sigma_{ww}, \sigma_{wx}, \text{and } \sigma_{xx}\right)$.

Equations 4.64-4.67 show that the number of parameters is the same for each of the univariate imputation models employed by FCS-CK. Further, the structure of each imputation model is the same (one incomplete variable predicted from all other variables). As such, equivalence (or lack thereof) between the FCS-CK imputation model for *Y* and the population model would also imply the same between the FCS-CK imputation model for *X* and the population model. Because the findings for one imputation model should be generalizable to the other, I examine only the FCS-CK imputation model for *Y* in this section. This serves to streamline the section and avoid confusion that might arise from providing two sets of similar but slightly different equations.

The process of transforming the joint distribution of the variables into a univariate multilevel model predicting *Y* from *W* and *X* was already shown in the subsection describing FCS-VB. To avoid repetition, the process is not described here. The transformed population model (already shown in Equations 4.54 and 4.55) is:

$$
\begin{aligned}
Y_{ij} &= \left(\beta^{(y)} - \beta^{(x)}\alpha_{y.w,2} - \beta^{(w)}\alpha_{y.x,2}\right) \\
&+\alpha_{y.w,1}\left(W_{ij}\right) + \left(\alpha_{y.w,2} - \alpha_{y.w,1}\right)\left(\bar{W}_j\right) + \alpha_{y.x,1}\left(X_{ij}\right) + \left(\alpha_{y.x,2} - \alpha_{y.x,1}\right)\left(\bar{X}_j\right) \\
&+u_j^{(y)} + e_{ij}^{(y)}
\end{aligned} \tag{4.68}
$$

$$
u_{j,\mathrm{y}} \sim \mathrm{N}\left(0, \psi_{yy} - \begin{bmatrix}\alpha_{y.w,2}\\ \alpha_{y.x,2}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\psi_{ww} & \psi_{wx}\\ \psi_{wx} & \psi_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,2}\\ \alpha_{y.x,2}\end{bmatrix}\right)
$$

$$
e_{ij,\mathrm{y}} \sim \mathrm{N}\left(0, \sigma_{yy} - \begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.x,1}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\sigma_{ww} & \sigma_{wx}\\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}\begin{bmatrix}\alpha_{y.w,1}\\ \alpha_{y.x,1}\end{bmatrix}\right)
$$

Both the population model and the FCS-CK imputation model have 15 parameters. Because the two models have the same numbers of parameters, the population model and the FCS-CK imputation model for *Y* are equivalent. The parameters describing the distribution of the predictors for the imputation model predicting *Y* have a one to one correspondence with parameters in the population distribution. The remaining seven parameters from the imputation model for *Y* ($\gamma_0^{(x)}$, $\gamma_1^{(x)}$, $\gamma_2^{(x)}$, $\gamma_3^{(x)}$, $\gamma_4^{(x)}$, τ_{xx}, ς_{yy}) can be calculated from the population parameters:

$$
\begin{bmatrix}\gamma_1^{(y)}\\ \gamma_3^{(y)}\end{bmatrix} = \begin{bmatrix}\sigma_{ww} & \sigma_{wx}\\ \sigma_{wx} & \sigma_{xx}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{wy}\\ \sigma_{xy}\end{bmatrix} \tag{4.69}
$$

$$
\begin{bmatrix}\gamma_2^{(x)}\\ \gamma_4^{(x)}\end{bmatrix} = \begin{bmatrix}\psi_{ww} & \psi_{wx}\\ \psi_{wx} & \psi_{xx}\end{bmatrix}^{-1}\begin{bmatrix}\psi_{wy}\\ \psi_{xy}\end{bmatrix} \tag{4.70}
$$

$$\gamma_0^{(y)} = \beta^{(y)} - \begin{bmatrix} \beta^{(w)} & \beta^{(x)} \end{bmatrix} \begin{bmatrix} \psi_{ww} & \psi_{wx} \\ \psi_{wx} & \psi_{xx} \end{bmatrix}^{-1} \begin{bmatrix} \psi_{wy} \\ \psi_{xy} \end{bmatrix} \tag{4.71}$$

$$\tau_{yy} = \psi_{yy} - \begin{bmatrix} \psi_{wy} & \psi_{xy} \end{bmatrix} \begin{bmatrix} \psi_{ww} & \psi_{wx} \\ \psi_{wx} & \psi_{xx} \end{bmatrix}^{-1} \begin{bmatrix} \psi_{wy} \\ \psi_{xy} \end{bmatrix} \tag{4.72}$$

$$\varsigma_{yy} = \sigma_{yy} - \begin{bmatrix} \sigma_{wy} & \sigma_{xy} \end{bmatrix} \begin{bmatrix} \sigma_{ww} & \sigma_{wx} \\ \sigma_{wx} & \sigma_{xx} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{wy} \\ \sigma_{xy} \end{bmatrix} \tag{4.73}$$

The ability of the FCS-CK method to preserve the covariances in the population model without constraints can be shown by demonstrating that the parameters of the population model can be calculated from the FCS-CK univariate imputation model predicting *Y* from *W* and *X* (or from the imputation model predicting *X* from *W* and *Y*, not shown). Using the distribution of the predictors from Equation 4.67 and the equivalencies laid out in Equations 4.69 through 4.73, the population parameters can be calculated from the parameters in the imputation model for *Y* as follows:

$$\begin{bmatrix} \sigma_{wy} \\ \psi_{wy} \\ \sigma_{xy} \\ \psi_{xy} \end{bmatrix} = \begin{bmatrix} \sigma_{ww} & 0 & \sigma_{wx} & 0 \\ 0 & \psi_{ww} & 0 & \psi_{wx} \\ \sigma_{wx} & 0 & \sigma_{xx} & 0 \\ 0 & \psi_{wx} & 0 & \psi_{xx} \end{bmatrix} \begin{bmatrix} \gamma_1^{(y)} \\ \gamma_2^{(y)} \\ \gamma_3^{(y)} \\ \gamma_4^{(y)} \end{bmatrix} \tag{4.74}$$

$$\beta^{(y)} = \gamma_0^{(y)} + \gamma_2^{(y)} \begin{bmatrix} \beta^{(w)} \end{bmatrix} + \gamma_4^{(y)} \begin{bmatrix} \beta^{(x)} \end{bmatrix} \tag{4.75}$$

$$\psi_{yy} = \tau_{yy} + \begin{bmatrix} \gamma_2^{(y)} & \gamma_4^{(y)} \end{bmatrix} \begin{bmatrix} \psi_{ww} & \psi_{wx} \\ \psi_{wx} & \psi_{xx} \end{bmatrix} \begin{bmatrix} \gamma_2^{(y)} \\ \gamma_4^{(y)} \end{bmatrix} \tag{4.76}$$

$$\sigma_{yy} = \varsigma_{yy} + \begin{bmatrix} \gamma_1^{(y)} & \gamma_3^{(y)} \end{bmatrix} \begin{bmatrix} \sigma_{ww} & \sigma_{wx} \\ \sigma_{wx} & \sigma_{xx} \end{bmatrix} \begin{bmatrix} \gamma_1^{(y)} \\ \gamma_3^{(y)} \end{bmatrix} \quad (4.77)$$

In summary, the answer to the research question "Does FCS-CK produce expectations that are equivalent to the joint distribution of the data?" is: yes. This finding supports the hypothesis that FCS-CK imputation produces expectations that are equivalent to the joint distribution of the data.

Equivalence of imputation methods. The previous subsections compared each of the five imputation methods to the population distribution to determine the conditions under which each method would produce expectations that are equivalent to the population distribution. This subsection examines whether the five imputation methods are equivalent to one another. Imputation methods JM-Mplus and FCS-CK produce expectations that are equivalent to the joint distribution. Because JM-Mplus and FCS-CK both produce expectations that are equivalent to the joint distribution, their expectations must also be equivalent to one another. This supports the hypothesis that JM-Mplus and FCS-CK imputation methods produce equivalent expectations.

Imputation methods JM-BD, JM-UN, and FCS-VB produce expectations that are equivalent to the joint model only under specific conditions. JM-BD, JM-UN, and FCS-VB should produce expectations equivalent to JM-Mplus and FCS-CK only under conditions where JM-BD, JM-UN, and FCS-VB are equivalent to the population distribution. These conditions were stated previously. This supports the hypothesis that JM-BD, JM-UN, and FCS-VB do not produce expectations that are equivalent to JM-Mplus and FCS-CK imputation.

Imputation methods JM-BD, JM-UN, and FCS-VB differ from one another in the conditions required to produce expectations that are equivalent to the joint model. The three methods differ in the assumptions that they make regarding covariances between incomplete variables. FCS-VB makes the assumption that the regression of an incomplete variable on any other incomplete variable is the same at level-1 and level-2. JM-BD makes the assumption that the level-2 relationship between a pair of incomplete variables must be completely explained by the regression of the incomplete variables on the set of complete variables (i.e., the residual level-2 covariance between the variables is assumed to be zero). If the imputation model contains no complete variables, JM-BD assumes that the incomplete variables have zero covariances at level 2. JM-UN does not make any assumptions about the relationships between pairs of incomplete variables, conditional upon the complete variables, and uses separate parameters to represent the residual level-1 and level-2 covariances. The fact that FCS-VB, JM-BD, and JM-UN differ in their assumptions about incomplete variables supports the hypothesis that FCS-VB, JM-BD, and JM-UN do not produce equivalent expectations.

Summary of analytic results. The analytic examination of the five imputation methods showed that JM-Mplus and FCS-CK were the only methods whose expectations were equivalent to the population model. JM-BD, JM-UN, and FCS-VB produce expectations that are not equivalent to the population model, generally. JM-BD, JM-UN, and FCS-VB are equivalent to the population model under certain conditions, however. JM-Mplus and FCS-CK are equivalent to one another. JM-BD, JM-UN, and FCS-VB are not equivalent to one another or to JM-Mplus and FCS-CK.

Simulation Results

The simulation portion of this document consisted of two simulation studies. The first study generated data under a population model where the level-2 correlations between variables differed from the level-1 correlations between the variables. Study 1 was intended to examine how well the imputation methods preserve the level-1 and level-2 covariances from the population distribution. The second study generated data under a population model where the correlations between variables were identical across levels. Study 2 was intended to determine whether the more general multilevel imputation methods (JM-Mplus and FCS-CK imputation) perform well even when the models are overparameterized in representing the population model. Note that it was not my 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.

Complete data analyses. Multilevel modeling can sometimes produce biased estimates, even for complete data. Given these findings, it is helpful to examine the bias of the complete-data estimates in each of the study conditions prior to examining the findings from the imputed-data analyses. Figures 1 and 2 show the standardized bias for the level-1 covariance parameters in each condition in studies 1 and 2. Figures 3 and 4 show the standardized bias for the mean parameters in each condition in studies 1 and 2. Figures 5 and 6 show the standardized bias for the level-2 covariance parameters in each condition in studies 1 and 2.

The absolute values of the level-1 covariance parameter standardized biases were less than .2. This means that the complete data estimates were within .2 standard errors of the population parameters. The absolute values of the mean parameter standardized

biases were less than .05. The level-2 covariance parameters showed the greatest bias, with the absolute value exceeding .2 for multiple conditions. The level-2 variance of *W* had a standardized bias of -0.22 for study 2 in the condition containing an ICC of .5, 30 clusters, and 5 observations per cluster. The level-2 covariance between *W* and *Y* had a standardized bias of 0.33 for study 1 in the condition containing an ICC of .1, 30 clusters, and 5 observations per cluster. The level-2 covariance between X and Y had large standardized biases for two conditions in study 1. The condition containing an ICC of .1, 30 clusters, and 5 observations per cluster had a standardized bias of 0.41. The condition containing an ICC of .1, 100 clusters, and 5 observations per cluster had a standardized bias of 0.23. Finally, the level-2 variance of *Y* for study 1 in the condition containing an ICC of .1, 30 clusters and 5 observations per cluster had a standardized bias of 0.22. In summary, the condition containing an ICC of .1, 30 clusters and 5 observations per cluster in study 1 produced the two largest standardized bias values (for the level-2 covariance between *W* and *Y* and the level-2 covariance between *X* and *Y*), making it the most problematic condition for the complete data.

C**onvergence failures**. In order for an imputation method's bias to be evaluated, the method must first produce a data set that can be analyzed. Table 2 shows, for each method and condition, the number of iterations (out of 1000) that produced imputed data sets that could not be analyzed. If any of the 20 imputed data sets in an iteration could not be analyzed (i.e., the analysis failed to converge), then that iteration was counted as a failed iteration. Table 2 shows the number of failed iterations for each combination of imputation method and condition.

Iteration failures for JM-BD, JM-UN, JM-Mplus, and FCS-VB occurred exclusively for the conditions with ICC values of .1. In contrast, FCS-CK had iteration failures for an ICC = .5 condition (in study 1 with 5 observations per cluster and 30 clusters). Iteration failures appeared to be more common for conditions with 5 observations per cluster than for conditions with 30 observations per cluster. Very few iteration failures occurred overall. Even in the worst condition, few iteration failures occurred (11 out of 1000). The majority of conditions had no iteration failures.

The highest failure rates occurred for the ICC = .1, observations per cluster = 5, number of clusters = 100 condition. This is counterintuitive, as one would expect the failure rate to be higher in the ICC = .1, observations per cluster = 5, number of clusters = 30 condition. That is, estimation problems are typically expected to decrease as the number of clusters increases, rather than the reverse. Re-running the high failure rate condition with new sample data produced similar failure rates. Further investigation revealed that each failed iteration occurred due to a convergence failure in the analysis phase rather than in the imputation phase. In each case, Mplus reported that the estimated between-cluster covariance matrix was not positive definite. I is important to note that the analysis model used in this study is rather complex. Mplus treats all three variables as endogenous two-level variables that are allowed to have different covariances across levels. However, in each failed analysis at least one of the three variables did not have significant level-2 variance. In each case, treating any variables with nonsignificant level-2 variance as within-cluster variables (variables that have variance only at level-1) allowed the analysis model to successfully converge. Switching to a contextual analysis model in which one variable (*Y*) was treated a dependent variable and the other two

variables (*W* and *X*) were treated as fixed predictors also resolved the non-convergence problem. This information indicates that the convergence failures are not very problematic, as they can be avoided by using a simpler analysis model or by treating variables without significant level-2 variance as within-cluster variables. I do not have an explanation for why the failure rate was higher for 100 clusters than for 30 clusters,

Largest biases. Recall that standardized corrected bias (SCB) is calculated as the imputed-data estimate minus the complete-data estimate, divided by the standard deviation of the complete-data estimates across all iterations. Tables 3 through 17 show the SCB values produced by each method for each parameter and condition. SCB values greater than 0.4 or less than -0.4 were be pointed out, as such extreme values tend to negatively affect statistical inference (Collins et al., 2001). This subsection does not directly address the study hypotheses. Rather, it points out which combinations of methods and conditions may cause the largest problems for statistical inference and which parameters are most affected. Method/condition combinations that are not listed here as generating a large amount of bias do not necessarily provide unbiased parameter estimates. Rather, they were simply not found to produce bias exceeding the aforementioned cutoff at the missing data rate of 20%. These method/condition combinations may produce very biased estimates for larger missing data rates.

Of the five imputation methods, FCS-VB was the only method to produce absolute SCB values exceeding .4 (Tables 12 and 13). These biases occurred only for the level-1 and level-2 covariances between *X* and *Y*, the two incomplete variables, and only in study 1 (in which the level-1 and level-2 covariances had opposite signs). In conditions where bias was found, the level-2 covariance between *X* and *Y* was positively biased and

the level-1 covariance between *X* and *Y* was negatively biased. These biases support this document's hypotheses. However, these biases did not occur for every condition. The two conditions that displayed high SCB for the level-1 covariance both had ICC values of .5 and both used 5 observations per cluster, but the conditions differed in the number of clusters. As such, it appears that the combination of a high ICC and low number of observations per cluster tends to result in high degrees of bias for FCS-VB. For the level-2 covariance between *X* and *Y*, no clear pattern emerged for the high levels of SCB, which appeared for both levels of ICC, both levels of number of clusters, and both levels of observations per cluster.

Evaluating hypotheses. The imputation methods were hypothesized to differ on four parameters (the level-1 and level-2 covariances between *W* and *Y* and between *X* and *Y*). The analytic work at the beginning of this chapter indicated that these parameters should be biased for some methods but not others. I used a series of between-subjects ANOVAs to determine whether corrected bias (imputed data estimate minus complete data estimate) varied substantively across imputation methods and to determine whether differences were affected by study conditions. Note that imputation method was treated as a between-group effect despite the fact that imputation method was a within-subjects factor to simplify the analyses. The factors of interest were imputation method (JM-BD, JM-UN, JM-Mplus, FCS-VB, and FCS-CK), ICC (.5 and .1), number of clusters (30 and 100), and observations per cluster (5 and 30).

Methodologists recommend using generalized eta-squared rather than eta-squared (Bakeman, 2005; Olejnik & Algina, 2003). Unfortunately, major software packages (e.g., SAS and SPSS) do not output all of the components required for computing generalized

eta-squared. I also attempted to use a macro for calculating generalized eta-squared (de Gil et al., 2013), but the macro failed due to the sample size being too large. Because the estimation of generalized eta-squared would be impractical for my data, I used eta-squared values to judge whether interactions and main effects were worth discussing in this document. I treated method as a between-subjects factor in order to facilitate computation of eta-squared. Effects with eta square values equal to or greater than .01 (corresponding to Cohen's cutoff for a small effect size) were judged to be large enough for discussion.

Level-1 covariance between W and Y in study 1. Tables 3, 6, 9, 12, and 15 show the mean standardized corrected bias values for the level-1 covariance between *W* and *Y* for all cells in the design. The analytic work indicates that the estimated level-1 covariance between *W* and *Y* should be biased toward the population level-2 covariance between *W* and *Y* for JM-BD, JM-UN, and FCS-VB in study one. This bias was theorized to occur because the aforementioned methods constrain the level-1 and level-2 coefficients for the regression of *Y* on *W* to be equal. Because the level-1 covariance parameter is positive in the population and the level-2 covariance parameter is negative in the population, constraining the regression coefficients to be equal across levels in the imputation model should bias the level-1 covariance estimate toward the level-2 covariance parameter in the analysis, resulting in negative bias for JM-BD, JM-UN, and FCS-VB in study one.

The eta-squared values for the 4-way, 3-way, and 2-way interactions between factors were all below .01. As such, only main effects are discussed here. The main effects with eta-squared values exceeding the cutoff were: observations per cluster =

.021, and method = .021. Larger biases (more negative) were found with 5 observations per cluster than with 30 observations per cluster. This effect does not warrant further discussion because it is not central to my hypotheses. As the eta-squared values indicated differences across observations per cluster conditions and across methods, the mean standardized corrected bias values were examined to compare the conditions further. Averaging across the other factors, the mean standardized corrected bias values for the methods were as follows: JM-BD = -0.142, JM-UN = -0.15, JM-Mplus = -0.007, FCS-VB = -0.125, FCS-CK = -0.008. These results can be interpreted as Cohen's d values that compare the mean parameter values to the average complete-data estimate. Comparison across methods can be obtained by subtraction. For example, on average, the bias of the JM-BD, JM-UN, and FCS-VB methods was about 0.132 more negative than the bias of the JM-Mplus and FCS-CK methods. These findings support the analytic work.

Level-1 covariance between W and Y in study 2. Tables 3, 6, 9, 12, and 15 show the mean standardized corrected bias values for the level-1 covariance between *W* and *Y* for all cells in the design. The analytic work does not imply any bias for the imputation methods for the level-1 covariance between *W* and *Y* in study two. The eta-squared values for the 4-way, 3-way, and 2-way interactions between factors as well as for the main effects were all below .01. These results match expectations.

Level-2 covariance between W and Y in study 1. Tables 4, 7, 10, 13, and 16 show the mean standardized corrected bias values for the level-2 covariance between *W* and *Y* for all cells in the design. The analytic work indicates that the estimated level-2 covariance between *W* and *Y* should be biased toward the population level-1 covariance between *W* and *Y* for JM-BD, JM-UN, and FCS-VB in study one. This bias was theorized

to occur because the aforementioned methods constrain the level-1 and level-2 coefficients for the regression of *Y* on *W* to be equal. Because the level-2 covariance parameter is negative in the population and the level-1 covariance parameter is positive in the population, constraining the regression coefficients to be equal across levels in the imputation model should bias the level-2 covariance estimate toward the level-1 covariance parameter in the analysis, resulting in positive bias for JM-BD, JM-UN, and FCS-VB in study one.

The eta-squared values for the 4-way and 3-way interactions between factors were all below .01. Of the two-way interactions, only the interaction between method and number of observations per cluster had an eta-squared above .01 ($\eta^2 = 0.016$). Figure 7 shows this interaction. The lines in Figure 7 are included only as visual aids (this use of lines in graphing an ANOVA interaction is consistent with Seltman, 2014). Because the factor on the horizontal axis (imputation method) is categorical, the lines should not be used to interpolate between levels of the variable. From the figure we see that, in both the 5 observations per cluster and 30 observations per cluster conditions, the magnitude of bias is greater for JM-BD, JM-UN, and FCS-VB than for JM-Mplus and FCS-CK. Further, the magnitude of bias for JM-BD, JM-UN, and FCS-VB is greater in the 5 observations per cluster condition than in the 30 observations per cluster condition. These findings support the analytic work.

Level-2 covariance between W and Y in study 2. Tables 4, 7, 10, 13, and 16 show the mean standardized corrected bias values for the level-2 covariance between *W* and *Y* for all cells in the design. The analytic work does not imply any bias for the imputation methods for the level-2 covariance between *W* and *Y* in study two. The eta-squared values

for the 4-way, 3-way, and 2-way interactions between factors as well as for the main effects were all below .01. These results match expectations.

Level-1 covariance between X and Y in study 1. Tables 3, 6, 9, 12, and 15 show the mean standardized corrected bias values for the level-1 covariance between *X* and *Y* for all cells in the design. The analytic work indicates that the estimated level-1 covariance between *X* and *Y* should be biased toward the population level-2 covariance between *X* and *Y* for FCS-VB in study one. This bias was theorized to occur because FCS-VB constrains the level-1 and level-2 coefficients for the regression of *Y* on *X* to be equal. Because the level-1 covariance parameter is positive in the population and the level-2 covariance parameter is negative in the population, constraining the regression coefficients to be equal across levels in the imputation model should bias the level-1 covariance estimate toward the level-2 covariance parameter in the analysis, resulting in negative bias for FCS-VB in study one.

The eta-squared values for the four-way and three-way interactions between factors were less than .01. Of the two-way interactions, only the interaction between method and number of observations per cluster had an eta-squared above .01 ($\eta^2 = 0.029$). Figure 8 shows this interaction. From the figure we see that for both observations per cluster conditions, FCS-VB was more negatively biased than all of the other imputation methods. The magnitude of the negative bias for FCS-VB was largest for the observations per cluster = 5 condition. These findings support the analytic work.

Level-1 covariance between X and Y in study 2. Tables 3, 6, 9, 12, and 15 show the mean standardized corrected bias values for the level-1 covariance between *X* and *Y* for all cells in the design. The analytic work does not imply any bias for the imputation

methods for the level-2 covariance between *X* and *Y* in study two. The eta-squared values for the 4-way, 3-way, and 2-way interactions between factors as well as for the main effects were all below .01. These results match expectations.

Level-2 covariance between X and Y in study 1. Tables 4, 7, 10, 13, and 16 show the mean standardized corrected bias values for the level-2 covariance between *X* and *Y* for all cells in the design. The analytic work indicates that the estimated level-2 covariance between *X* and *Y* should be biased toward the population level-1 covariance between *X* and *Y* for FCS-VB in study one and should be biased toward zero for JM-BD in study 1. The bias for FCS-VB was theorized to occur because FCS-VB constrains the level-1 and level-2 slope coefficients for the regression of *Y* on *X* to be equal. Because the level-2 covariance parameter is negative in the population and the level-1 covariance parameter is positive in the population, constraining the slope coefficients to be equal across levels in the imputation model should bias the level-2 covariance estimate toward the level-1 covariance parameter in the analysis, resulting in positive bias for FCS-VB in study one. The bias for JM-BD was theorized to occur because JM-BD constrains the level-2 residual covariance between *X* and *Y* to be zero. Because the level-2 covariance parameter is negative in the population, constraining the level-2 residual covariance between the variables to be zero should bias the level-2 covariance estimate toward zero, resulting in positive bias for JM-BD in study one. In summary, the analytic work implies that the estimated level-2 covariance between *X* and *Y* should be positively biased for both FCS-VB and JM-BD in study one.

The eta-squared values for the four-way and most of the three-way interactions between factors were less than .01. The eta-squared value of the three-way interaction

between method, ICC, and observations per cluster was .016. This interaction is shown in Figure 9. For ICC equal to .5 (the right panel of Figure 9), JM-BD produced a large amount of bias for the 5 observations per cluster condition but not for the 30 observations per cluster condition. For ICC equal to .1, JM-BD a large amount of bias in both the 5 observations per cluster and 30 observations per cluster conditions. The bias was larger in the 30 observations per cluster condition than in the 5 observations per cluster condition. These findings support the analytic work.

Level-2 covariance between X and Y in study 2. Tables 4, 7, 10, 13, and 16 show the mean standardized corrected bias values for the level-2 covariance between *X* and *Y* for all cells in the design. The analytic work indicates that the estimated level-2 covariance between *X* and *Y* should be biased toward zero for JM-BD in study 2. The bias for JM-BD was theorized to occur because JM-BD constrains the level-2 residual covariance between *X* and *Y* to be zero. Because the level-2 covariance parameter is positive in the population, constraining the level-2 residual covariance between the variables to be zero should bias the level-2 covariance estimate toward zero, resulting in negative bias for JM-BD in study two. In summary, the analytic work implies that the estimated level-2 covariance between *X* and *Y* should be negatively biased for JM-BD in study two.

The eta-squared values for the 4-way and 3-way interactions between factors were all below .01. The two-way interaction between method and ICC ($\eta^2 = 0.01$) and the two-way interaction between method and number of observations per cluster ($\eta^2 = 0.021$) were the only two-way interactions whose eta-squared values exceeded the cutoff. Figure 10 shows the interaction between method and ICC. Figure 11 shows the

interaction between method and number of observations per cluster. Figure 10 shows us that JM-BD displays more negative bias than the other methods. The magnitude of the negative bias for JM-BD is greater for the ICC = .1 condition than the ICC = .5 condition. Figure 11 shows a very similar pattern for the interaction between method and number of observations per cluster. Again, JM-BD displays more negative bias than the other methods. The magnitude of the negative bias for JM-BD is greater for the 5 observations per cluster condition than for the 30 observations per cluster condition. These findings support the analytic work.

Summary of simulation results. The simulation findings solidly support the analytic work. Every bias implied by the analytic work was found, and the direction of bias was as expected. The magnitude of bias was affected by the number of observations per cluster and the ICC for some of the parameters. In general, biases tended to be more extreme for five observations per cluster than for thirty observations per cluster. Biases were also more extreme for the ICC = .1 condition than for the ICC = .5 condition. The number of clusters did not substantially affect bias for any of the parameters.

In study one, where correlations differed across levels, JM-Mplus and FCS-CK did not produce biased estimates. JM-BD, JM-UN, and FCS-VB all produced biased estimates for multiple parameters. FCS-VB was the only method to produce large standardized corrected biases (greater than .4 or less than -.4). In study two, JM-UN, JM-Mplus, FCS-VB, and FCS-CK produced unbiased estimates for all parameters. JM-BD produced biased estimates, but only for the level-2 covariance between *X* and *Y*.

Chapter 5. Discussion

Multiple imputation methods can generally be divided into two broad categories: joint model (JM) imputation (Rubin & Schafer, 1990; Schafer, 1997) and fully conditional specification (FCS) imputation (Raghunathan et al., 2001; van Buuren et al., 2006). JM draws missing values simultaneously for all incomplete variables using a multivariate distribution (e.g., multivariate normal). FCS, on the other hand, imputes variables one at a time, drawing missing values from a series of univariate distributions. In the single-level context, these two approaches have been shown to be equivalent with multivariate normal data (Hughes et al., 2014). However, less is known about the similarities and differences of these two approaches with multilevel data, and the methodological literature provides no insight into the situations under which the approaches would produce identical results.

Three variations of JM imputation have been proposed for multilevel data (Asparouhov & Muthén, 2010a, 2010f; Schafer, 2001; Schafer & Yucel, 2002). Although these methods share much in common, they possess subtle differences. No previous methodological research has investigated the differences among the JM models, in particular their ability to produce imputations that preserve characteristics of the population distribution. Thus, one of the overarching goals for this project was to examine the situations under which the three JM methods reproduced (or preserved) the mean and covariance structure of a population random intercept model with multivariate normal data. To date, two FCS models have been proposed in the literature, only one of which is currently implemented in statistical software (Carpenter & Kenward, 2012, p. 221; van Buuren & Groothuis-Oudshoorn, 2011). Like JM, no previous methodological

research investigated the differences between the FCS models or their ability to preserve a multilevel data structure. Thus, the second overarching goal for this project was to examine the situations under which FCS imputation reproduced the mean and covariance structure of a population random intercept model with multivariate normal data. The analytic work for these two goals also provided insight into the situations where JM and FCS were equivalent.

Summary of Findings

As a reminder, all of the analytic and simulation work in this document assumed a two-level population model that contained random intercepts but no random slopes. The population model used for the analytic work allowed the correlations among variables to differ between level one and level two. For the simulation work, study one used a population model in which the correlations among the variables differed between level one and level two and study two used a population model in which the correlations among the variables did not differ between levels.

The analytic examination of the five imputation methods showed that JM-Mplus and FCS-CK were the only methods whose expectations were equivalent to the particular population model that I examined. JM-BD, JM-UN, and FCS-VB produced expectations that were not equivalent to the population model, generally. JM-BD was equivalent to the population model only when the regression of each incomplete variable on each complete variable was the same across levels and the residual level-2 covariances between the incomplete variables after controlling for the complete variables were equal to zero. JM-UN was equivalent to the population model only when the regressions of the incomplete variables on the complete variable were the same at both level one and level two. FCS-

VB was equivalent to the population model only when the regression of each incomplete variable on every other variable was the same at level one and level two. JM-Mplus and FCS-CK were equivalent to one another. JM-BD, JM-UN, and FCS-VB were not equivalent to one another or to JM-Mplus and FCS-CK.

The simulation findings solidly supported the analytic work. Every bias implied by the analytic work was found, and the direction of bias was as expected. The magnitude of bias was affected by the number of observations per cluster and the ICC for some of the parameters. In general, biases tended to be more extreme for five observations per cluster than for thirty observations per cluster and were more extreme for the ICC = .1 condition than for the ICC = .5 condition. This pattern was likely due to the fact that cluster size and ICC both affect the reliability of cluster means. Specifically, the reliability of the mean of a variable in group *j* can be calculated from the ICC for that variable and the size of the group (Snijders & Bosker, 2012):

$$reliability_j = \frac{n_j * ICC}{1 + (n_j - 1) * ICC} \quad 5.1$$

Because all three variables had the same ICC in each condition, their group means were equally reliable. The condition with the worst reliability (ICC = .1, observations per cluster = 5) tended to have the largest bias, and the condition with the best reliability (ICC = .5, observations per cluster = 30) tended to have the lowest bias. This combination of conditions is known to be problematic for level-2 estimation in the complete-data context (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Lüdtke et al.,

2008). Thus, it is not a complete surprise that imputation might make the problem worse. The number of clusters did not substantially affect bias for any of the parameters.

In study one, where correlations differed across levels, JM-Mplus and FCS-CK did not produce biased estimates. This was expected because, as shown in the analytic work, JM-Mplus and FCS-CK were both congenial with an analysis model identical to the random intercept population model (as a reminder, an analysis model and an imputation model are said to be uncongenial if the analysis model cannot be derived from the imputation model, or vice versa; Meng, 1994; Schafer, 1997, 2003). JM-BD, JM-UN, and FCS-VB all produced biased estimates for multiple parameters. This was expected because, as shown in the analytic work, JM-BD, JM-UN, and FCS-VB were not congenial with an analysis model identical to the random intercept population model. FCS-VB was the only method to produce large standardized corrected biases (greater than .4 or less than -.4). This was likely due to the fact that FCS-VB misspecified the most parameters in the model. That is, FCS-VB misspecified the level-1 and level-2 covariances between *W* and *Y*, the level-1 and level-2 covariances between *W* and *X*, and the level-1 and level-2 covariances between *X* and *Y* (6 parameters total). In contrast, JM-BD misspecified the level-1 and level-2 covariances between *W* and *Y*, the level-1 and level-2 covariances between *W* and *X*, and the level-2 covariances between *X* and *Y* (5 parameters total). JM-UN misspecified the level-1 and level-2 covariances between *W* and *Y* and the level-1 and level-2 covariances between *W* and *X*, (4 parameters total). In study two, JM-UN, JM-Mplus, FCS-VB, and FCS-CK produced unbiased estimates for all parameters. JM-BD produced biased estimates, but only for the level-2 covariance between *X* and *Y*.

Integration with Existing Literature

As this document has mentioned previously, an analysis model and an imputation model are said to be uncongenial if the analysis model cannot be derived from the imputation model, or vice versa (Meng, 1994; Schafer, 1997, 2003). The analytic work in this document examined whether each imputation model could be derived from a random intercept population model where associations differed between level one and level two. The analytic work can be thought of as checking whether each imputation model is congenial with an analysis model that is equivalent to the random intercept population model. If an effect that is present in the analysis model is left out of (or constrained to zero in) the imputation model, the estimate of the effect is expected to be biased toward zero. If two parameters are constrained to equality, the estimate of each parameter is biased toward the other parameter. Consistent with the literature on congeniality, effects that were constrained to equality in the imputation models were biased toward one another in the analysis models. This finding matches a fact that is well established in the single-level imputation literature.

JM-BD was suggested by Joseph Schafer in order to obtain a stable estimate for the level-2 residual covariance matrix when the number of random effects is large (a situation that is problematic for JM-UN; Schafer, 2001). Schafer wrote that JM-BD "assumes that the random effects for each response are independent of those for any other response." He then stated that "unless the correlations among the random effects for some pairs of responses are unusually strong, the potential biases incurred by using a block-diagonal Ψ rather than an unstructured Ψ tend to be minor." The findings in the current document do not contradict Schafer, as his statement allowed for the possibility that large

correlations between random effects would generate problematic levels of bias. However, I suggest that it is not actually unusual for correlations among random effects to be strong enough to produce problematic levels of bias. For example, if all variables in a data set are incomplete, the level-2 covariance matrix for the random effects in a JM-UN imputation model is equal to the population level-2 covariance matrix for the variables (except for differences due to missing values). If the level-2 covariance for any pair of variables differs much from zero in the population, use of a block diagonal level-2 residual covariance matrix in the imputation model (i.e., using JM-BD) would produce problematic levels of bias.

The analytic work showed that the default imputation model used by JM-UN is not equivalent to the joint distribution and can produce biased results. That is, the JM-UN imputation model is not congenial with an analysis model that allows covariances between complete and incomplete variables to differ between level one and level two. Carpenter and Kenward (2012) wrote that complete-variable cluster means should be included in JM imputation to match the multilevel joint distribution. Though this dissertation did not evaluate JM-UN imputation with complete-variable cluster means, it does agree with the assessment that the default form of JM-UN does not match the multilevel joint distribution.

The analytic and simulation work indicated that JM-Mplus is congenial with an analysis model that allows covariances between complete and incomplete variables to differ between level one and level two. This makes logical sense, as the JM-Mplus imputation model is identical to the joint distribution without performing any transformations on the parameters. A simulation study conducted by Enders et al. (2014)

examined the ability of JM-Mplus to preserve regression coefficients in a population model that allowed level-1 and level-2 regression coefficients to differ. Specifically, the level-1 slope was positive ($\beta_1 = .34$), whereas the level-2 slope was negative ($\beta_2 = -.4$). The probability of missingness was set to .4, the missing data mechanism was MCAR, and the ICCs were set to .5. Consistent with this dissertation, the authors found that the level-1 and level-2 regression coefficients were both unbiased.

The analytic and simulation work in this document indicated that FCS-VB is not equivalent to a random intercept population model when the level-1 and level-2 correlations differ. This means that FCS-VB is not congenial with a random intercept analysis model containing different parameters for level one and level two correlations. Carpenter and Kenward (2012) wrote that cluster means should be included in FCS imputation to match the multilevel joint distribution. When cluster means are added to the FCS imputation model (FCS-CK), the imputation model is equivalent to a random intercept population model when the level-1 and level-2 correlations differ. That is, FCS-CK is congenial with a random intercept population model containing different parameters for level one and level two correlations.

A simulation study conducted by Enders et al. (2014) examined the ability of FCS-VB to preserve regression coefficients in a population model that allowed level-1 and level-2 slope coefficients to differ. Specifically, the level-1 slope was positive ($\beta_1 = .34$), whereas the level-2 slope was negative ($\beta_2 = -.4$). The probability of missingness was set to .4, the missing data mechanism was MCAR, and the ICCs were set to .5. Consistent with this dissertation, the authors found that the level-1 and level-2 regression coefficients were both pulled toward a common value. The authors also ran the

same simulation with probability of missingness set to .2. FCS-VB was also biased for this condition, but the bias was not as extreme as the high rate of missingness condition. The biases found by Enders et al. (2014) for FCS-VB were not as extreme as those found in this dissertation. This difference can be attributed to the authors' of an ICC of .5 (this dissertation found greater bias for an ICC of .1 than for an ICC of .5) and a cluster size of 20 (this dissertation found bias to be exacerbated by small cluster sizes). The fact that the authors found FCS-VB to produce less extreme levels of bias under these conditions is consistent with this dissertation.

Practical Recommendations

A common approach to handling missing data is to perform multiple imputation once and later perform many analyses on the imputed data sets. If a multiply imputed data set is to be used in multiple analyses, the imputation model must be congenial with all of the analysis models to be used. This has led researchers to recommend a "kitchen-sink" approach to multiple imputation in which all variables and effects that may later be of interest are included in the imputation process. For example, Schafer and Olsen (1998) stated that "a rich imputation model that preserves a large number of associations is desirable because it may be used for a variety of post-imputation analyses." Rubin (1996) stated that "the press to include all possibly relevant predictors is demanding in practice, but it is generally a worthy goal." This issue has received little attention in the multilevel imputation literature, and researchers currently have few (if any) recommendations for choosing among competing multilevel imputation methods. The results from this study have practical implications for substantive research.

In the context of traditional multilevel models, there is a literature on contextual effects, whereby a predictor's influence differs between level-1 and level-2 (e.g., the influence of school-average SES on achievement differs from the influence of individual SES on achievement; the influence of daily pain fluctuations on positive affect differs from the influence of average or chronic pain on positive affect). "For theoretically important variables in multilevel studies, it is the rule rather than the exception that within-group regression coefficients differ from between-group regression coefficients" (Snijders & Bosker, 2012). Another situation where associations can differ across levels is multilevel structural equation modeling. For example, in a multilevel confirmatory factor analysis model, the loadings and other estimates (even the structure itself) can differ between level-1 and level-2.

Given the fact that level-1 and level-2 correlations between variables can, and often do, differ from one another, attempting to include all possible relationships between variables in a multilevel data set in an imputation model requires using an imputation method that allows level-1 and level-2 correlations to differ. Researchers have several choices of imputation method (JM-BD, JM-UN, JM-Mplus, FCS-VB, and FCS-CK), and these methods differ in their ability to separately model level-1 and level-2 correlations between variables. The analytic work and simulations clearly show that JM-Mplus and FCS-CK are the methods of choice for research questions involving differential correlations at level one and level two (for analysis models that contain random intercepts but no random slopes).

The previous paragraph's recommendation to use an imputation method that allows level-1 and level-2 correlations to differ (e.g., JM-Mplus or FCS-CK) could

theoretically be disregarded if an analyst knew that the level-1 and level-2 correlations between variables were identical in the population. This is an unlikely situation, as an analyst typically is not collecting and analyzing a data set if she already knows the population model. A second situation in which the recommendation to use an imputation method that allows level-1 and level-2 correlations to differ could be disregarded is where the analyst knows that all subsequent analysis models will constrain the level-1 and level-2 correlations (or regression coefficients, etc.) to be equal. For example, an analyst may not care about the distinct within- and between-cluster regression coefficients and may instead want to obtain only overall regression coefficients. In such a case, the imputation and analysis models would be congenial. Though a final analysis model may constrain level-1 and level-2 regression coefficients to be equal, it is recommended that an analyst test for whether level-1 and level-2 regression coefficients are equal instead of assuming them to be equal (at least for predictors central to one's research question). In order to test for equality of level-1 and level-2 regression coefficients, the analyst should use an imputation model that allows regression coefficients to differ between level one and level two.

As a third example, an analyst may be interested in only the within-cluster regression of one variable on another. That is, the analyst may group-mean-center all of the variables to be used as predictors in a multilevel model. The output from the model will consist of only within-cluster regression coefficients. Because the analyst is not simultaneously estimating within-cluster and between-cluster regression coefficients, the analyst might think that it is not necessary to employ a rich imputation model and may choose to instead use a model with fewer parameters, such as FCS-VB. Contrary to the

analyst's expectations, using an imputation method that constrains the level-1 and level-2 regression coefficients to be equal in the imputation model could still bias the within-cluster regression coefficients in the analysis toward the population parameters for the between-cluster regression coefficients if the within-cluster and between-cluster regression coefficients differ in the population model. Although I did not examine this possibility in this dissertation, it is worthy of future investigation.

Given the above considerations, the findings in this document lead me to make the following recommendations (for analysis models that contain random intercepts but no random slopes):

1. Avoid JM-BD. If the level-1 and level-2 coefficients for regressions of the complete variables on the incomplete variables differ, JM-BD produces biased estimates. Though not examined in this document, the aforementioned bias for slope coefficients between complete and incomplete variables should be alleviated if the user adds complete-variable cluster means. However, JM-BD also produces biased estimates for the level-2 residual covariances between incomplete variables. Though not tested in this document, the bias for level-2 residual covariances between incomplete variables should persist even if complete-variable cluster means are added as predictors.
2. Avoid FCS-VB if the intended analyses will allow regression coefficients to differ across levels or the intended analyses will examine within-cluster regression coefficients only or between-cluster regression coefficients only. If regression coefficients differ between levels, FCS-VB produces biased estimates of the regression coefficients (though the amount of bias depends on multiple factors,

including the population model, the missing data rate, and the number of observations per cluster). Adding complete-variable cluster means fixes this problem for regressions of incomplete variables on complete variables, but not for regressions of incomplete variables on other incomplete variables. Though FCS-VB works well if level-1 and level-2 slope coefficients are equal (demonstrated in Study 2), it may be difficult or even impossible to determine whether the level-1 and level-2 slope coefficients differ prior to handling missing data. FCS-VB would also work well if all subsequent analysis models constrain the level-1 and level-2 slope coefficients to be equal. This means that the analyst would not be able to analyze a model containing only level-1 regression coefficients or a model containing only level-2 regression coefficients. In each case, the analysis model would not constrain the level-1 and level-2 regression coefficients to be equal, so the imputation and analysis models would not be congenial. Additionally, the analyst would not be able to test whether the regression coefficients are equal across levels because FCS-VB would bias the estimates of the level-1 and level-2 regression coefficients toward one another. This would artificially improve the fit of an analysis model that constrains the level-1 and level-2 regression coefficients to be equal. It is recommended that an analyst test whether level-1 and level-2 regression coefficients are equal rather than assuming that they are equal. In summary, FCS-VB can be used without biasing one's results if the analyst knows the population prior to analysis or if the analyst restricts her analysis models to a subset of all possible analysis models. Because of this, I recommend avoiding FCS-VB.

3. Use JM-UN with caution. JM-UN produces biased estimates if slope coefficients for a regression of the incomplete variables on the complete variables differ between level-1 and level-2. As with FCS-VB, this is not a problem if the analyst knows that the slope coefficients are the same across levels in the population or if the analyst uses only analysis models in which the slope coefficients are constrained to be equal across levels. As explained for FCS-VB, the analyst is unlikely to know the population and restricting one's analyses can be problematic. Adding complete-variable cluster means to the JM-UN imputation model should remove the bias produced by JM-UN, but this claim was not examined in this document.
4. JM-Mplus and FCS-CK are recommended. JM-Mplus and FCS-CK produce unbiased estimates regardless of whether level-1 and level-2 slope coefficients (or covariances) are equal in the population. Because of this, an analyst can safely use JM-Mplus or FCS-CK without prior knowledge about the population and without restricting her analysis models. Neither method requires additional work or knowledge on the part of the user.

Limitations

This document compared the five multilevel imputation methods under circumstances that do not represent all possible circumstances in which users might employ these methods. The remainder of this section addresses six limitations of the studies in this document. First, the studies did not examine the performance of the methods for three-level data. Second, the smallest cluster size examined was five observations per cluster (in contrast, dyadic data have two observations per cluster).

Third, study 1 of the simulation work assumed a very large difference between the level-1 and level-2 correlations among variables. Fourth, JM-UN was examined without cluster means incorporated as predictors. Fifth, the population and analysis models included random intercepts but not random slopes. Sixth, FCS-CK includes observed cluster means as predictors in the imputation model, ignoring the possibility that the observed cluster means may be unreliable measures of the unobserved group means (this is explained in greater detail later in this section). Finally, the analytic and simulation results reported in this dissertation apply only to *current* software implementations of multilevel imputation. Changing these aspects analytic and/or simulation work would likely change the relative performance of the imputation methods.

The studies in this document focused exclusively on two-level data. No attempt was made to address how the five imputation methods would perform for data with three or more levels. Data with three or more levels are fairly common (e.g., many longitudinal studies examine repeated measures nested within people nested within groups). As such, it would be useful to determine how well the imputation methods perform for data with three or more levels. However, it seems likely that the main findings/patterns found in this dissertation would also apply to three-level models.

The simulation studies in this dissertation employed two cluster size conditions: 5 observations per cluster and 30 observations per cluster. Biases produced by the imputation methods appeared to be exacerbated by the small cluster size condition. This begs the question: how much worse will be bias be for even smaller cluster sizes? It is not uncommon for studies in the social sciences to examine small groups, which may contain fewer than five participants per group. Dyadic studies, for example, have only two level-1

units per cluster. Given that this study found much larger bias for the 5 observations per cluster condition than for the 30 observations per cluster condition, it seems likely that bias may greatly increase for studies with fewer than five level-1 units per cluster.

The difference between the level-1 and level-2 correlations in study 1 of the simulation was rather extreme. Specifically, the level-1 and level-2 correlations were equal in magnitude but opposite in sign. The large difference between level one and level two correlations highlighted the ability (or lack thereof) of the imputation methods to handle such differences (which was the purpose of this document). Such a large difference between levels is unlikely to be found in real data. It is possible that the imputation methods examined in this document would produce little/no bias for small differences between level-1 and level-2 correlations. It would be useful to conduct a follow-up simulation study with small differences between level-1 and level-2 correlations to determine whether each method is useful (produces little/no bias) for realistic data.

Implementations of JM-UN (e.g., PAN in R and Splus) do not automatically add complete variable cluster means as predictors in an imputation model. Because analysts tend to accept software defaults, this document evaluated JM-UN without complete variable cluster means included as predictors. An analyst well-versed in multiple imputation theory, however, might calculate the complete variable cluster means prior to imputation and then add these means to the imputation model. Doing this would likely remove any biases produced by JM-UN. As such, JM-UN should be re-evaluated with complete-variable cluster means included as predictors in the imputation model.

The simulations and analytic work in this document (and hence the recommendations in this chapter) focused exclusively on population and imputation models that contained random intercepts but no random slopes. Some of the imputation models are very limited in their ability to include random slopes. The JM-Mplus imputation model does not allow random slopes to be included. That is, JM-Mplus models all level-1 covariances using a level-1 covariance matrix that is not allowed to differ across clusters. JM-BD and JM-UN allow random slopes to be added between complete variables and incomplete variables, but not between pairs of incomplete variables. That is, the level-1 covariance between incomplete variables, after controlling for complete variables, is determined by the level-1 residual covariance matrix. Because this matrix is the same for all clusters, the level-1 covariances between incomplete variables must be the same for all clusters and the imputation model cannot be congenial with an analysis model containing random slopes between incomplete variables. Recai Yucel proposed an adaptation of JM-UN which uses heterogeneous level-1 residual covariance matrices (Yucel, 2011). This adaptation of JM-UN has yet to be implemented in publicly available software. Of the five methods examined in this document, FCS-VB is the only publicly available method able to incorporate random slopes between all variables. FCS-CK is also able to handle random slopes between all variables, but the software implementation of FCS-CK, BLImP, is not yet publicly available. To my knowledge, FCS-CK is not implemented in any other software package.

Much of the justification for incorporating random slopes into FCS multilevel imputation relies on what works in practice, rather than analytical justification. For example, van Buuren (2011) noted that simply adding random slopes to the FCS model

does not generate good imputations if predictors in the analysis model are incomplete. van Buuren suggests that the FCS imputation model should be adjusted to incorporate heterogeneous level-1 residual variances, stating that this leads to a "considerable advance in imputation quality." However, van Buuren did not provide analytical evidence that heterogeneous level-1 residual variances are needed to replicate the joint distribution. Further, the method proposed by van Buuren should be compared against competing methods, such as Recai Yucel's adaptation of JM-UN.

Level-2 variables that are cluster aggregates of level-1 variables are called contextual variables. The FCS-CK imputation model predicting *Y* from *W* and *X* also includes the cluster means of *W* and *X* as predictors. The cluster means ($\overline{W}_j$ and $\overline{X}_j$) are contextual variables. That is, $\overline{W}_j$ is calculated as the average of *W* for cluster *j* and $\overline{X}_j$ is calculated as the average of *X* for cluster *j*. A model that includes the same variable at both levels is called a contextual analysis model (Firebaugh, 1978; Raudenbush & Bryk, 2002). So, the FCS-CK imputation model is a contextual model. Lüdtke et al. (2008) labeled contextual analysis models that use only the observed clusters means (such as the FCS-CK imputation model) as multilevel manifest covariate (MMC) models. MMC models assume that the contextual variables are measured without error. If this assumption is violated, the contextual model may produce biased estimates of the difference between the level-1 and level-2 regression coeficients (Lüdtke et al., 2008).

In some cases, contextual variables may be unreliable measures of the unobserved group averages (O'Brien, 1990; Raudenbush, Rowan, & Kang, 1991), thus violating the assumption of an error-free contextual variable. The error in a contextual variable may

stem from measurement error in assessing a latent construct. For example, consider student ratings of an instructor's teaching ability. Each student rating consists of the true score for instructor's teaching ability plus some measurement error. Error in a contextual variable may also stem from taking a finite sample from a large or potentially infinite population. Consider a sample of 20 high school students from a school containing 2,000 students. The average height of the 20 students can be used as an estimate of the average height of students in the school. Though height might be assessed with almost no measurement error, an average based on 20 students will provide a poor measure of the average height of the 2,000 students due to sampling error.

Lüdtke et al. (2008) suggested that if contextual variables are unreliable measures of the unobserved group variable, they should be treated as latent rather than observed. The authors referred to this approach as the multilevel latent covariate (MLC) approach. Treating the contextual variables as latent allows the model to incorporate contextual variable error into the model, producing unbiased coefficients and standard errors. The MLC approach is not always necessary, however, and may introduce error where there should be none. Consider small-group research where each group consists of five participants. A researcher might be interested in the proportion of each group that is female. The average of a binary variable representing gender (1=female, 0=male) can be used as an estimate of the proportion of students in the group that are female. Because the size of the group sample (n=5) equals the size of the group, the contextual variable does not suffer from sampling error. Gender should be measured with little/no error, so the contextual variable does not suffer from measurement error. Because the contextual variable should be error free (or very close to it), the MMC approach could be used

without violating the error-free contextual variable assumption. On the other hand, the MLC approach would inappropriately treat the contextual variable as being unreliable. The JM-Mplus imputation method employs an MLC approach. That is, JM-Mplus treats complete-variable group means as latent, rather than observed. It bears mentioning that analysis of multilevel models with the MLC approach tends to yield standard errors that are too large when the number of level-1 units within each level-2 unit is small and the ICC is low (Lüdtke et al., 2011).

As previously mentioned, FCS-CK employs an MMC approach to handling complete-variable group means whereas JM-Mplus employs an MLC approach to handling complete-variable group means. It is possible that, in situations where the contextual variables are unreliable measures of the unobserved group variables, FCS-CK may run into problems due to unreliable group means whereas JM-Mplus may work quite well. In contrast, in situations where the contextual variables are reliable measures of the group variables, FCS-CK may perform well whereas JM-Mplus may perform poorly. To my knowledge, no published research has explored the impact of the reliability of contextual variables as measures of group variables on the efficacy of multilevel imputation methods.

The analytic and simulation results reported in this dissertation apply only to *current* software implementations of multilevel imputation. JM-BD and JM-UN are available for R and S-Plus in packages called PAN (Schafer, 2001; Schafer & Yucel, 2002). JM-BD is available in SAS as a macro called MMI_IMPUTE (Mistler, 2013). JM-BD and JM-UN imputation can also be performed in a standalone software package called REALCOM-IMPUTE (Carpenter et al., 2011) or the latent variable modeling

package Mplus (Asparouhov & Muthén, 2010f). JM-Mplus is available in Mplus (Asparouhov & Muthén, 2010f). FCS-VB is implemented in the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011) and also in a standalone software package called BLImP (Keller & Enders, 2014). As of the writing of this dissertation, BLImP is not yet available to the public. FCS-CK is implemented in BLImP (Enders et al., 2014). This dissertation should not be assumed to apply to future versions of the aforementioned software, as their implementations of multilevel imputation may change.

Summary

This document examined the equivalence of five multilevel imputation methods (JM-BD, JM-UN, JM-Mplus, FCS-VB, and FCS-CK) to one another and to the joint distribution for a random intercept population model. JM-Mplus and FCS-CK were found to be equivalent to the joint distribution and to one another. JM-BD, JM-UN, and FCS-VB were found to be different from the joint distribution and from one another. Due to the mismatch between these three methods and the joint distribution, JM-BD, JM-UN, and FCS-VB were each found to produce biased parameter estimates for some parameters in some situations. Biases tended to be greater for smaller numbers of observations per cluster and for smaller ICCs. In contrast, JM-Mplus and FCS-CK produced unbiased estimates for all of the parameters in all situations.

References

Andridge, R. R. (2011). Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biometrical Journal, 53*(1), 57-74.

Asparouhov, T., & Muthén, B. (2010a). Bayesian analysis using Mplus: Technical implementation (Version 3).

Asparouhov, T., & Muthén, B. (2010f). Multiple imputation with Mplus.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods, 37*(3), 379-384.

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. (2012). *Multiple Imputation and its Application*. West Sussex, UK: John Wiley & Sons.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330.

de Gil, P. R., Pham, T., Rasmussen, P., Kellermann, A., Romano, J., Chen, Y.-H., & Kromrey, J. D. (2013). *GEN_ETA2: A SAS® Macro for Computing the Generalized Eta-Squared Effect Size Associated with Analysis of Variance Models*. Paper presented at the SAS Global Forum, San Francisco, CA.

Eisenhart, C. (1947). The Assumptions Underlying the Analysis of Variance. *Biometrics, 3*(1), 1-21.

Enders, C. K., Mistler, S. A., & Keller, B. T. (2014). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. [Manuscript submitted for publication].

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American sociological review*, 557-572.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association, 57*, 369-375.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). West Sussex, UK: John Wiley & Sons.

Goldstein, H., Carpenter, J. R., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling, 9*(3), 173-197.

Goodnight, J. H. (1979). A Tutorial on the SWEEP Operator. *The American Statistician, 33*(3), 149-158.

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*(3), 206-213.

Gulliford, M. C., Ukoumunne, O. C., & Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. *American Journal of Epidemiology, 149*(9), 876-883.

Harville, D. A. (1976a). Confidence intervals and sets for linear combinations of fixed and random effects. *Biometrics*, 403-407.

Harville, D. A. (1976c). Extension of the Gauss-Markov theorem to include the estimation of random effects. *The Annals of Statistics*, 384-395.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87.

Henderson, C. R. (1950). Estimation of genetic parameters. *Biometrics, 6*(2).

Henderson, C. R. (1963). Selection index and expected genetic advance. *Statistical genetics and plant breeding, 982*, 141-163.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

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*(1), 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, 23*(2), 93-116.

Keller, B. T., & Enders, C. (2014). *A Latent Variable Chained Equations Approach for Multilevel Multiple Imputation*. Paper presented at the Modern Modeling Methods Conference, Storrs, CT.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*: Sage.

Kreft, I. G. G., De Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30*(1), 1-21.

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

Lüdtke, 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, 16*(4), 444.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*(3), 203-229. doi: 10.1037/a0012869

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science, 9*(4), 538-558.

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

O'Brien, R. M. (1990). Estimating the reliability of aggregate-level variables based on individual-level characteristics. *Sociological Methods & Research, 18*(4), 473-504.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods, 8*(4), 434.

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*(1), 85-96.

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

Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to US high-school data. *Journal of Educational and Behavioral Statistics, 16*(4), 295-330.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, New Jersey: John Wiley & Sons.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91*(434), 473-489.

Rubin, D. B., & Schafer, J. L. (1990). *Efficiently creating multiple imputations for incomplete multivariate normal data.* Paper presented at the Proceedings of the Statistical Computing Section of the American Statistical Association.

SAS Institute Inc. (2014). SAS/IML 13.2 User's Guide. Cary, NC: SAS Institute Inc.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall/CRC.

Schafer, J. L. (2001). Multiple imputation with PAN.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica neerlandica, 57*(1), 19-35.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*(2), 147.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*(4), 545-571.

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*(2), 437-457.

Seltman, H. J. (2014). *Experimental Design and Analysis* Retrieved from http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf

Shin, Y., & Raudenbush, S. W. (2007). Just-identified versus overidentified two-level hierarchical linear models with missing data. *Biometrics, 63*(4), 1262-1268.

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

Spybrook, J., Raudenbush, S. W., Liu, X.-f., Congdon, R., & Martínez, A. (2006). Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software. *Survey Research Center of the Institute of Social Research at University of Michigan*.

Taljaard, M., Donner, A., & Klar, N. (2008). Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biometrical Journal, 50*(3), 329-345.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research, 16*(3), 219-242. doi: 10.1177/0962280206074463

van Buuren, S. (2011). Multiple imputation of multilevel data. In J. K. Roberts & J. J. Hox (Eds.), *The Handbook of Advanced Multilevel Analysis* (pp. 173-196). New York: Routledge.

van Buuren, S. (2012). *Flexible imputation of missing data*. New York: 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*(12), 1049-1064.

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

Wicklin, R. (2013). *Simulating Data with SAS*. Cary, NC: SAS Institute Inc.

Yucel, R. M. (2008). Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366*(1874), 2389-2403.

Yucel, R. M. (2011). Random covariances and mixed-effects models for imputing multivariate multilevel continuous data. *Statistical Modelling, 11*(4), 351-370.

Yucel, R. M., Schenker, N., & Raghunathan, T. E. (2006). *Multiple imputation for incomplete multilevel data with SHRIMP*. Paper presented at the annual conference on New Methods for the Analysis of Family and Dyadic Processes, Amherst, MA.

Table 1

Software Implementations of Multilevel Imputaion Methods

Method	Software	Categorical	Description
JM-BD	Splus and R	No	Available as package "PAN." Imputes level-1 variables only.
	SAS	No	Available in SAS macro "MMI_IMPUTE."
	Mplus	Yes	Can be specified by user as an imputation model.
JM-UN	Splus and R	No	Available as package "PAN." Imputes level-1 variables only.
	Mplus	Yes	Can be specified by user as an imputation model.
	REALCOM-IMPUTE	Yes	Standalone software; designed to interface with Stata and MLwiN.
JM-Mplus	Mplus	Yes	Default imputation model.
FCS-VB	R	No	Available as package "MICE."
	BLImP	Yes	Standalone software.
FCS-CK	BLImP	Yes	Standalone software.

Note: The "Categorical" column refers to whether the software is able to handle discrete variables (e.g., nominal). All software in the table can perform imputation at both level one and level two unless otherwise stated.

Table 2

Number of failed iterations (out of 1000) for each method, by condition

ICC	Clusters	Obs per cluster	JM-BD	JM-UN	JM-Mplus	FCS-VB	FCS-CK
			Study 1				
0.5	30	5	0	0	0	0	2
0.5	30	30	0	0	0	0	0
0.5	100	5	0	0	0	0	0
0.5	100	30	0	0	0	0	0
0.1	30	5	5	2	2	4	0
0.1	30	30	3	2	0	1	0
0.1	100	5	9	11	8	4	9
0.1	100	30	0	0	0	0	0
			Study 2				
0.5	30	5	0	0	0	0	0
0.5	30	30	0	0	0	0	0
0.5	100	5	0	0	0	0	0
0.5	100	30	0	0	0	0	0
0.1	30	5	2	1	0	2	3
0.1	30	30	0	0	0	0	0
0.1	100	5	4	3	2	3	5
0.1	100	30	0	0	0	0	0

Note: if the analysis failed to converge for any of the imputations within an iteration, the iteration is categorized as failed.

Table 3

Standardized corrected bias for level-1 covariance parameters for JM-BD, by study and by condition

ICC	Clusters	Observations per Cluster	SIGMA_WW	SIGMA_WX	SIGMA_XX	SIGMA_WY	SIGMA_XY	SIGMA_YY
				Study 1				
0.5	30	5	0.00	0.04	0.12	-0.22	-0.03	-0.02
0.5	30	30	0.00	0.02	0.00	-0.07	-0.02	-0.02
0.5	100	5	0.00	0.03	0.05	-0.34	-0.09	-0.14
0.5	100	30	0.00	-0.02	0.03	-0.14	0.00	-0.03
0.1	30	5	0.02	0.06	0.06	-0.07	0.00	0.03
0.1	30	30	0.00	0.01	0.05	-0.06	0.01	-0.02
0.1	100	5	0.02	0.09	0.04	-0.14	-0.05	-0.06
0.1	100	30	0.00	0.02	0.02	-0.09	-0.03	-0.04
				Study 2				
0.5	30	5	0.00	0.00	0.17	0.00	0.11	0.13
0.5	30	30	0.00	-0.01	0.04	0.00	-0.02	0.01
0.5	100	5	0.00	-0.02	0.15	0.00	0.16	0.12
0.5	100	30	0.00	-0.01	0.04	-0.02	-0.02	0.01
0.1	30	5	0.01	-0.01	0.07	0.00	0.05	0.08
0.1	30	30	0.00	-0.01	0.03	-0.02	0.01	0.01
0.1	100	5	0.00	0.00	0.10	0.00	0.16	0.08
0.1	100	30	0.00	0.01	0.03	0.00	0.07	0.07

Study 1: Level-specific associations

Study 2: Common associations.

Table 4

Standardized corrected bias for level-2 covariance parameters for JM-BD, by study and by condition

ICC	Clusters	Observations per Cluster	PSI_WW	PSI_WX	PSI_XX	PSI_WY	PSI_XY	PSI_YY
				Study 1				
0.5	30	5	0.00	-0.02	0.01	0.08	0.12	0.04
0.5	30	30	0.00	0.00	0.00	0.01	0.02	0.01
0.5	100	5	0.00	-0.03	0.00	0.17	0.24	0.09
0.5	100	30	0.00	0.00	0.01	0.02	0.04	0.01
0.1	30	5	-0.04	-0.11	0.11	0.12	0.08	0.12
0.1	30	30	-0.01	-0.01	0.01	0.07	0.10	0.07
0.1	100	5	-0.04	-0.14	0.07	0.24	0.22	0.16
0.1	100	30	-0.01	-0.03	0.02	0.14	0.21	0.11
				Study 2				
0.5	30	5	0.00	0.01	-0.02	0.00	-0.15	0.00
0.5	30	30	0.00	0.00	0.00	0.00	-0.03	0.00
0.5	100	5	0.00	0.00	-0.05	-0.01	-0.33	-0.05
0.5	100	30	0.00	0.00	0.00	0.00	-0.06	0.00
0.1	30	5	-0.02	-0.01	0.07	-0.02	-0.21	0.05
0.1	30	30	0.00	-0.01	-0.01	0.00	-0.17	-0.02
0.1	100	5	0.00	0.00	-0.04	-0.01	-0.35	-0.06
0.1	100	30	-0.01	0.00	-0.03	-0.01	-0.31	-0.04

Study 1: Level-specific associations

Study 2: Common associations.

Table 5

Standardized corrected bias for mean parameters for JM-BD, by study and by condition

ICC	Clusters	Observations per Cluster	BETA_W	BETA_X	BETA_Y
			Study 1		
0.5	30	5	0.00	0.00	0.00
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	-0.01	0.00
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.00	-0.01
0.1	30	30	0.00	0.00	0.00
0.1	100	5	0.00	0.00	0.02
0.1	100	30	0.00	-0.01	0.01
			Study 2		
0.5	30	5	0.00	0.00	0.01
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	-0.01	0.01
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.02	0.03
0.1	30	30	0.00	0.01	0.00
0.1	100	5	0.00	0.00	0.02
0.1	100	30	0.00	-0.01	0.01

Study 1: Level-specific associations

Study 2: Common associations.

Table 6

Standardized corrected bias for level-1 covariance parameters for JM-UN, by study and by condition

ICC	Clusters	Observations per Cluster	SIGMA_WW	SIGMA_WX	SIGMA_XX	SIGMA_WY	SIGMA_XY	SIGMA_YY
				Study 1				
0.5	30	5	0.00	-0.05	0.13	-0.20	-0.01	0.00
0.5	30	30	0.00	-0.02	0.00	-0.08	-0.03	-0.02
0.5	100	5	0.00	-0.11	0.09	-0.36	-0.03	-0.11
0.5	100	30	0.00	-0.12	0.04	-0.17	-0.03	-0.04
0.1	30	5	0.02	0.05	0.05	-0.05	0.00	0.03
0.1	30	30	0.00	-0.02	0.04	-0.07	0.01	-0.02
0.1	100	5	0.02	0.07	0.03	-0.15	0.01	-0.05
0.1	100	30	0.00	-0.03	0.03	-0.11	-0.01	-0.03
				Study 2				
0.5	30	5	0.00	-0.02	0.13	0.02	0.01	0.11
0.5	30	30	0.00	-0.01	0.04	-0.01	-0.02	0.01
0.5	100	5	0.00	-0.01	0.06	0.00	0.01	0.05
0.5	100	30	0.00	-0.01	0.02	-0.03	-0.03	0.00
0.1	30	5	0.01	-0.03	0.03	0.00	-0.02	0.03
0.1	30	30	0.00	-0.01	0.02	-0.03	-0.01	0.00
0.1	100	5	0.00	0.00	0.02	0.00	0.02	0.01
0.1	100	30	0.00	0.01	0.00	-0.01	0.01	0.04

Study 1: Level-specific associations

Study 2: Common associations.

Table 7

Standardized corrected bias for level-2 covariance parameters for JM-UN, by study and by condition

ICC	Clusters	Observations per Cluster	PSI_WW	PSI_WX	PSI_XX	PSI_WY	PSI_XY	PSI_YY
				Study 1				
0.5	30	5	0.00	0.02	0.01	0.08	-0.02	0.04
0.5	30	30	0.00	0.01	0.00	0.01	0.00	0.01
0.5	100	5	0.00	0.04	-0.01	0.18	0.00	0.08
0.5	100	30	0.00	0.02	0.01	0.03	0.01	0.01
0.1	30	5	-0.04	-0.10	0.17	0.11	0.01	0.18
0.1	30	30	0.00	0.03	0.01	0.08	0.00	0.08
0.1	100	5	-0.03	-0.09	0.08	0.25	-0.01	0.17
0.1	100	30	-0.01	0.05	0.00	0.16	0.00	0.10
				Study 2				
0.5	30	5	0.00	0.01	0.01	0.00	0.00	0.02
0.5	30	30	0.00	0.00	0.00	0.00	-0.01	0.00
0.5	100	5	0.00	0.00	0.00	0.00	-0.01	-0.01
0.5	100	30	0.00	0.00	0.00	0.00	0.00	0.00
0.1	30	5	-0.02	0.00	0.17	-0.02	-0.01	0.16
0.1	30	30	0.00	0.00	0.03	0.00	-0.02	0.01
0.1	100	5	0.00	-0.01	0.10	-0.02	-0.01	0.06
0.1	100	30	-0.01	0.00	0.02	-0.01	0.00	0.01

Study 1: Level-specific associations

Study 2: Common associations.

Table 8

Standardized corrected bias for mean parameters for JM-UN, by study and by condition

ICC	Clusters	Observations per Cluster	BETA_W	BETA_X	BETA_Y
			Study 1		
0.5	30	5	0.00	-0.01	0.00
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	0.00	0.01
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	-0.01	-0.01
0.1	30	30	0.00	0.00	0.00
0.1	100	5	0.00	0.01	0.02
0.1	100	30	0.00	-0.01	0.00
			Study 2		
0.5	30	5	0.00	-0.01	0.01
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	0.00	0.00
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.00	0.02
0.1	30	30	0.00	0.02	0.00
0.1	100	5	0.00	0.00	0.01
0.1	100	30	0.00	-0.01	0.01

Study 1: Level-specific associations

Study 2: Common associations.

Table 9

Standardized corrected bias for level-1 covariance parameters for JM-Mplus, by study and by condition

ICC	Clusters	Observations per Cluster	SIGMA_WW	SIGMA_WX	SIGMA_XX	SIGMA_WY	SIGMA_XY	SIGMA_YY
				Study 1				
0.5	30	5	0.00	0.01	0.10	-0.03	-0.01	0.06
0.5	30	30	0.00	0.02	0.00	0.00	-0.03	0.03
0.5	100	5	0.00	-0.02	0.08	0.01	-0.04	0.05
0.5	100	30	0.00	-0.03	0.04	-0.01	0.01	0.03
0.1	30	5	0.01	0.03	0.04	-0.03	-0.01	0.02
0.1	30	30	0.00	0.00	0.05	-0.01	0.01	0.01
0.1	100	5	0.01	0.01	0.02	-0.01	-0.05	-0.01
0.1	100	30	0.00	0.00	0.03	0.01	0.00	0.03
				Study 2				
0.5	30	5	0.00	0.00	0.11	0.00	0.02	0.09
0.5	30	30	0.00	-0.01	0.04	-0.01	-0.02	0.02
0.5	100	5	0.00	-0.02	0.06	0.00	-0.01	0.04
0.5	100	30	0.00	-0.01	0.03	-0.03	-0.03	-0.01
0.1	30	5	0.00	-0.02	0.00	-0.01	-0.01	0.03
0.1	30	30	0.00	-0.01	0.02	-0.03	-0.01	0.01
0.1	100	5	0.00	-0.02	0.01	0.01	-0.01	0.01
0.1	100	30	0.00	0.01	0.01	0.01	0.02	0.05

Study 1: Level-specific associations

Study 2: Common associations.

Table 10

Standardized corrected bias for level-2 covariance parameters for JM-Mplus, by study and by condition

ICC	Clusters	Observations per Cluster	PSI_WW	PSI_WX	PSI_XX	PSI_WY	PSI_XY	PSI_YY
				Study 1				
0.5	30	5	0.00	-0.01	0.03	-0.01	-0.01	0.02
0.5	30	30	0.00	0.00	0.01	0.00	0.00	0.00
0.5	100	5	0.00	-0.01	0.00	-0.01	0.00	0.01
0.5	100	30	0.00	0.00	0.01	0.00	0.00	0.00
0.1	30	5	-0.01	-0.05	0.21	0.03	0.02	0.18
0.1	30	30	0.00	0.01	0.03	-0.01	0.00	0.05
0.1	100	5	-0.01	0.00	0.13	0.00	0.06	0.12
0.1	100	30	-0.01	-0.01	0.01	-0.01	-0.01	0.01
				Study 2				
0.5	30	5	0.00	0.01	0.02	0.00	0.01	0.03
0.5	30	30	0.00	0.00	0.00	0.00	-0.01	0.00
0.5	100	5	0.00	0.00	0.01	0.00	-0.01	0.00
0.5	100	30	0.00	0.00	0.00	0.00	0.00	0.00
0.1	30	5	0.00	-0.01	0.24	-0.02	-0.01	0.20
0.1	30	30	0.00	-0.01	0.03	0.00	-0.01	0.02
0.1	100	5	0.00	0.00	0.12	-0.01	0.00	0.10
0.1	100	30	-0.01	0.00	0.03	-0.01	0.00	0.01

Study 1: Level-specific associations

Study 2: Common associations.

Table 11

Standardized corrected bias for mean parameters for JM-Mplus, by study and by condition

ICC	Clusters	Observations per Cluster	BETA_W	BETA_X	BETA_Y
			Study 1		
0.5	30	5	0.00	0.01	0.00
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	0.01	-0.01
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.01	-0.02
0.1	30	30	0.00	0.00	-0.01
0.1	100	5	0.00	0.01	-0.01
0.1	100	30	0.00	-0.01	0.00
			Study 2		
0.5	30	5	0.00	0.00	0.01
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	0.01	-0.01
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.02	0.01
0.1	30	30	0.00	0.01	-0.01
0.1	100	5	0.00	0.01	-0.01
0.1	100	30	0.00	-0.01	0.01

Study 1: Level-specific associations

Study 2: Common associations.

Table 12

Standardized corrected bias for level-1 covariance parameters for FCS-VB, by study and by condition

ICC	Clusters	Observations per Cluster	SIGMA_WW	SIGMA_WX	SIGMA_XX	SIGMA_WY	SIGMA_XY	SIGMA_YY
				Study 1				
0.5	30	5	0.00	0.09	-0.11	-0.19	-0.50	-0.18
0.5	30	30	0.00	0.06	-0.11	-0.05	-0.16	-0.05
0.5	100	5	0.00	0.14	-0.27	-0.29	-0.94	-0.32
0.5	100	30	0.00	0.04	-0.12	-0.10	-0.24	-0.09
0.1	30	5	0.02	0.09	0.01	-0.09	-0.18	0.00
0.1	30	30	0.00	0.04	-0.05	-0.06	-0.11	-0.06
0.1	100	5	0.03	0.14	-0.08	-0.14	-0.36	-0.11
0.1	100	30	0.00	0.07	-0.10	-0.08	-0.23	-0.09
				Study 2				
0.5	30	5	0.00	-0.01	0.10	0.01	0.05	0.08
0.5	30	30	0.00	-0.01	0.02	-0.01	-0.01	0.01
0.5	100	5	0.00	-0.01	0.07	-0.01	0.05	0.06
0.5	100	30	0.00	-0.01	0.03	-0.03	-0.01	0.00
0.1	30	5	0.01	-0.03	0.03	-0.01	-0.01	0.04
0.1	30	30	0.00	-0.01	0.01	-0.03	-0.02	0.00
0.1	100	5	0.00	0.00	0.07	0.00	0.03	0.06
0.1	100	30	0.00	0.01	0.01	0.00	0.03	0.05

Study 1: Level-specific associations

Study 2: Common associations.

Table 13

Standardized corrected bias for level-2 covariance parameters for FCS-VB, by study and by condition

ICC	Clusters	Observations per Cluster	PSI_WW	PSI_WX	PSI_XX	PSI_WY	PSI_XY	PSI_YY
				Study 1				
0.5	30	5	0.00	-0.05	0.06	0.07	0.25	0.06
0.5	30	30	0.00	-0.01	0.02	0.01	0.03	0.01
0.5	100	5	0.00	-0.08	0.13	0.15	0.47	0.16
0.5	100	30	0.00	-0.01	0.04	0.02	0.06	0.02
0.1	30	5	-0.03	-0.16	0.00	0.15	0.43	-0.02
0.1	30	30	-0.01	-0.05	0.07	0.06	0.21	0.09
0.1	100	5	-0.04	-0.22	0.16	0.21	0.81	0.18
0.1	100	30	-0.01	-0.10	0.20	0.11	0.41	0.17
				Study 2				
0.5	30	5	0.00	0.01	-0.04	-0.01	0.01	-0.02
0.5	30	30	0.00	0.00	-0.01	0.00	0.00	-0.01
0.5	100	5	0.00	0.00	-0.02	0.00	0.02	-0.03
0.5	100	30	0.00	0.00	0.00	0.00	0.00	0.00
0.1	30	5	-0.02	0.00	-0.05	-0.02	-0.01	-0.06
0.1	30	30	0.00	0.00	-0.04	0.00	0.00	-0.04
0.1	100	5	0.00	0.00	-0.08	-0.01	0.00	-0.10
0.1	100	30	-0.01	0.00	-0.01	-0.01	0.03	-0.02

Study 1: Level-specific associations

Study 2: Common associations.

Table 14

Standardized corrected bias for mean parameters for FCS-VB, by study and by condition

ICC	Clusters	Observations per Cluster	BETA_W	BETA_X	BETA_Y
			Study 1		
0.5	30	5	0.00	0.00	0.00
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	0.00	0.00
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.00	-0.02
0.1	30	30	0.00	0.00	0.00
0.1	100	5	0.00	0.01	0.00
0.1	100	30	0.00	-0.01	0.00
			Study 2		
0.5	30	5	0.00	-0.01	0.01
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	-0.01	0.00
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.01	0.00
0.1	30	30	0.00	0.01	0.00
0.1	100	5	0.00	0.00	0.00
0.1	100	30	0.00	-0.01	0.01

Study 1: Level-specific associations

Study 2: Common associations.

Table 15

Standardized corrected bias for level-1 covariance parameters for FCS-CK, by study and by condition

ICC	Clusters	Observations per Cluster	SIGMA_WW	SIGMA_WX	SIGMA_XX	SIGMA_WY	SIGMA_XY	SIGMA_YY
				Study 1				
0.5	30	5	0.00	0.01	0.04	-0.02	0.00	-0.01
0.5	30	30	0.00	0.02	-0.03	0.01	-0.03	0.01
0.5	100	5	0.00	-0.02	0.03	0.01	-0.04	0.03
0.5	100	30	0.00	-0.02	0.02	0.00	-0.01	0.02
0.1	30	5	-0.01	0.02	0.03	-0.03	0.00	0.03
0.1	30	30	0.00	0.00	0.01	-0.02	0.01	-0.01
0.1	100	5	-0.01	0.00	0.02	-0.02	-0.02	-0.01
0.1	100	30	0.00	0.01	0.02	0.01	0.00	0.01
				Study 2				
0.5	30	5	0.00	0.00	0.08	0.01	0.06	0.07
0.5	30	30	0.00	-0.01	0.02	-0.01	-0.01	0.01
0.5	100	5	0.00	-0.01	0.07	0.00	0.05	0.06
0.5	100	30	0.00	-0.02	0.03	-0.02	-0.01	-0.01
0.1	30	5	-0.02	-0.02	0.00	-0.01	-0.01	0.02
0.1	30	30	0.00	-0.01	0.01	-0.02	0.00	0.00
0.1	100	5	-0.01	0.00	0.04	0.00	0.01	0.04
0.1	100	30	0.00	0.01	0.01	0.00	0.03	0.04

Study 1: Level-specific associations

Study 2: Common associations.

Table 16

Standardized corrected bias for level-2 covariance parameters for FCS-CK, by study and by condition

ICC	Clusters	Observations per Cluster	PSI_WW	PSI_WX	PSI_XX	PSI_WY	PSI_XY	PSI_YY
				Study 1				
0.5	30	5	0.00	0.00	0.00	-0.01	-0.03	-0.01
0.5	30	30	0.00	0.00	0.00	0.00	-0.01	0.00
0.5	100	5	0.00	-0.01	-0.01	-0.01	-0.03	0.00
0.5	100	30	0.00	0.00	0.01	0.00	-0.01	0.00
0.1	30	5	0.02	-0.03	0.03	0.05	0.06	0.00
0.1	30	30	0.00	0.01	-0.01	0.00	-0.01	0.01
0.1	100	5	0.01	0.02	-0.01	0.01	0.08	0.00
0.1	100	30	-0.01	-0.01	-0.01	-0.01	-0.03	-0.01
				Study 2				
0.5	30	5	0.00	0.00	-0.01	0.00	0.02	0.00
0.5	30	30	0.00	0.00	-0.01	0.00	0.00	0.00
0.5	100	5	0.00	0.00	0.00	-0.01	0.02	-0.01
0.5	100	30	0.00	0.00	0.00	0.00	0.00	0.00
0.1	30	5	0.03	-0.01	0.09	0.00	0.04	0.05
0.1	30	30	0.00	-0.01	-0.01	0.00	0.00	-0.02
0.1	100	5	0.02	0.00	0.00	-0.02	0.06	-0.03
0.1	100	30	-0.01	0.00	0.01	-0.01	0.03	-0.01

Study 1: Level-specific associations

Study 2: Common associations.

Table 17

Standardized corrected bias for mean parameters for FCS-CK, by study and by condition

ICC	Clusters	Observations per Cluster	BETA_W	BETA_X	BETA_Y
			Study 1		
0.5	30	5	0.00	0.00	0.00
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	0.00	-0.01
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	-0.01	-0.02
0.1	30	30	0.00	-0.01	0.00
0.1	100	5	0.00	0.00	0.00
0.1	100	30	0.00	-0.01	0.01
			Study 2		
0.5	30	5	0.00	0.00	0.00
0.5	30	30	0.00	0.00	0.00
0.5	100	5	0.00	-0.01	0.00
0.5	100	30	0.00	0.00	0.00
0.1	30	5	0.00	0.01	0.00
0.1	30	30	0.00	0.02	0.00
0.1	100	5	0.00	0.00	0.00
0.1	100	30	0.00	-0.01	0.01

Study 1: Level-specific associations

Study 2: Common associations.

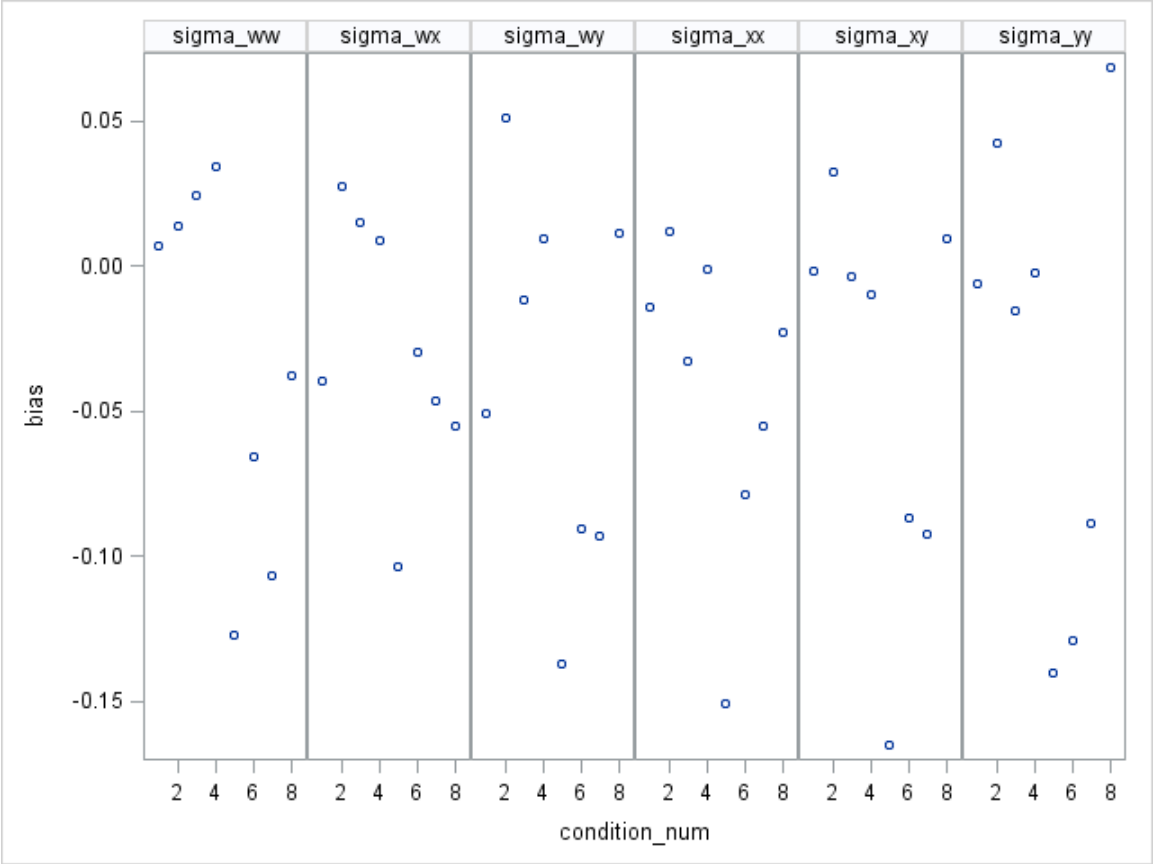


Figure 1. Standardized bias for level-1 covariance terms in study 1. The graph displays the mean standardized bias (the difference between the average complete-data estimate versus the population parameter expressed as a z-score) for the level-1 covariance parameters in study 1 (level-1 specific associations), broken down by condition. The figure shows that the complete-data estimates displayed some bias, and that this bias tended to be greatest for condition 5.
Condition 1: ICC = .5; nclusters = 30; nobs per cluster = 5.
Condition 2: ICC = .5; nclusters = 30; nobs per cluster = 30.
Condition 3: ICC = .5; nclusters = 100; nobs per cluster = 5.
Condition 4: ICC = .5; nclusters = 100; nobs per cluster = 30.
Condition 5: ICC = .1; nclusters = 30; nobs per cluster = 5.
Condition 6: ICC = .1; nclusters = 30; nobs per cluster = 30.
Condition 7: ICC = .1; nclusters = 100; nobs per cluster = 5.
Condition 8: ICC = .1; nclusters = 100; nobs per cluster = 30.

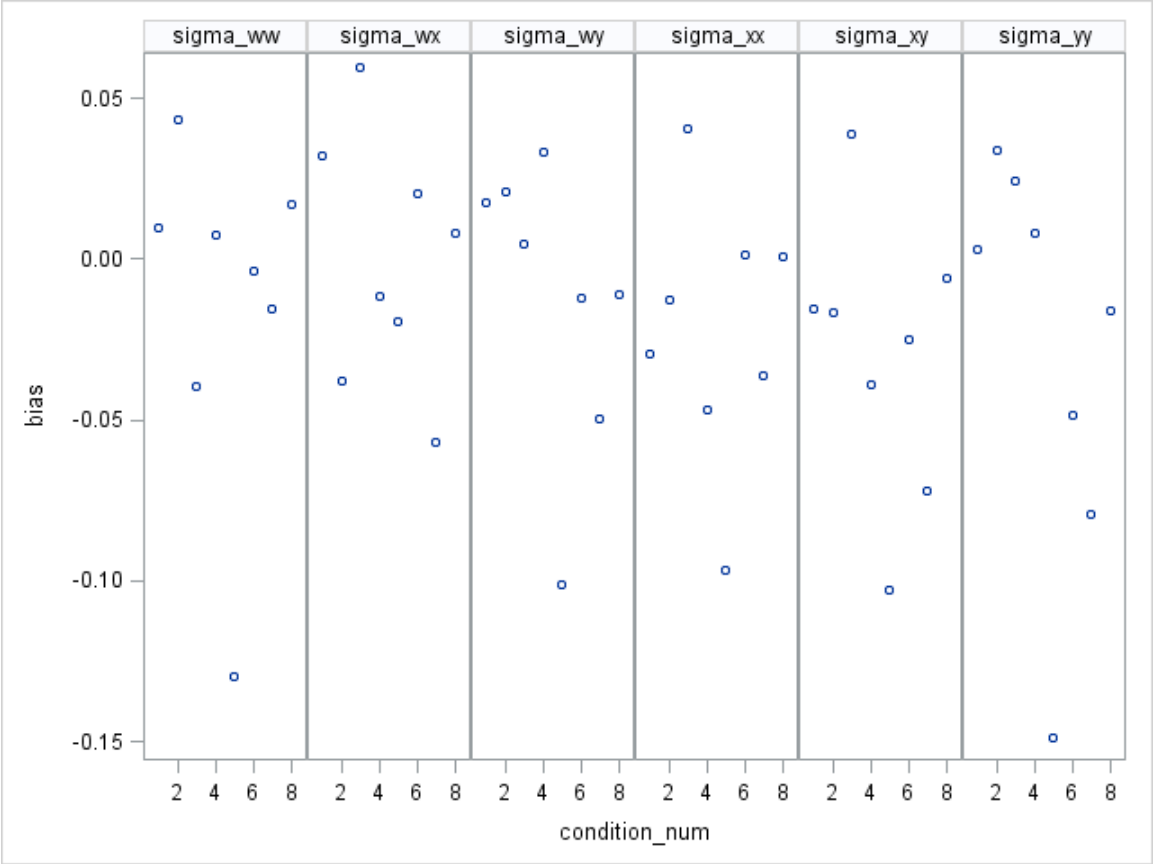


Figure 2. Standardized bias for level-1 covariance terms in study 2. The graph displays the mean standardized bias (the difference between the average complete-data estimate versus the population parameter expressed as a z-score) for the level-1 covariance parameters in study 2 (equal associations), broken down by condition. The figure shows that the complete-data estimates displayed some bias, and that this bias tended to be greatest for condition 5.
Condition 1: ICC = .5; nclusters = 30; nobs per cluster = 5.
Condition 2: ICC = .5; nclusters = 30; nobs per cluster = 30.
Condition 3: ICC = .5; nclusters = 100; nobs per cluster = 5.
Condition 4: ICC = .5; nclusters = 100; nobs per cluster = 30.
Condition 5: ICC = .1; nclusters = 30; nobs per cluster = 5.
Condition 6: ICC = .1; nclusters = 30; nobs per cluster = 30.
Condition 7: ICC = .1; nclusters = 100; nobs per cluster = 5.
Condition 8: ICC = .1; nclusters = 100; nobs per cluster = 30.

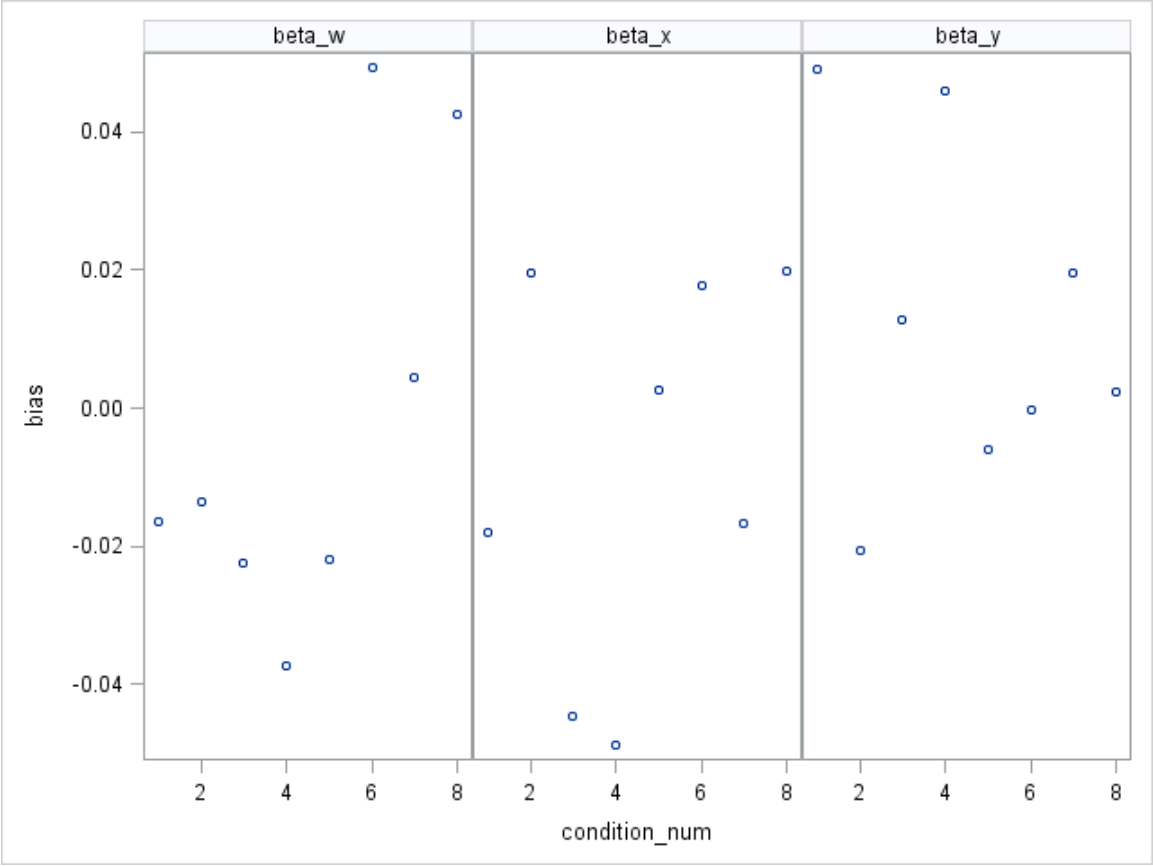


Figure 3. Standardized bias for mean terms in study 1. The graph displays the mean standardized bias (the difference between the average complete-data estimate versus the population parameter expressed as a z-score) for the mean parameters in study 1 (level-1 specific associations), broken down by condition. The figure shows that the complete-data estimates displayed no apparent bias.
Condition 1: ICC = .5; nclusters = 30; nobs per cluster = 5.
Condition 2: ICC = .5; nclusters = 30; nobs per cluster = 30.
Condition 3: ICC = .5; nclusters = 100; nobs per cluster = 5.
Condition 4: ICC = .5; nclusters = 100; nobs per cluster = 30.
Condition 5: ICC = .1; nclusters = 30; nobs per cluster = 5.
Condition 6: ICC = .1; nclusters = 30; nobs per cluster = 30.
Condition 7: ICC = .1; nclusters = 100; nobs per cluster = 5.
Condition 8: ICC = .1; nclusters = 100; nobs per cluster = 30.

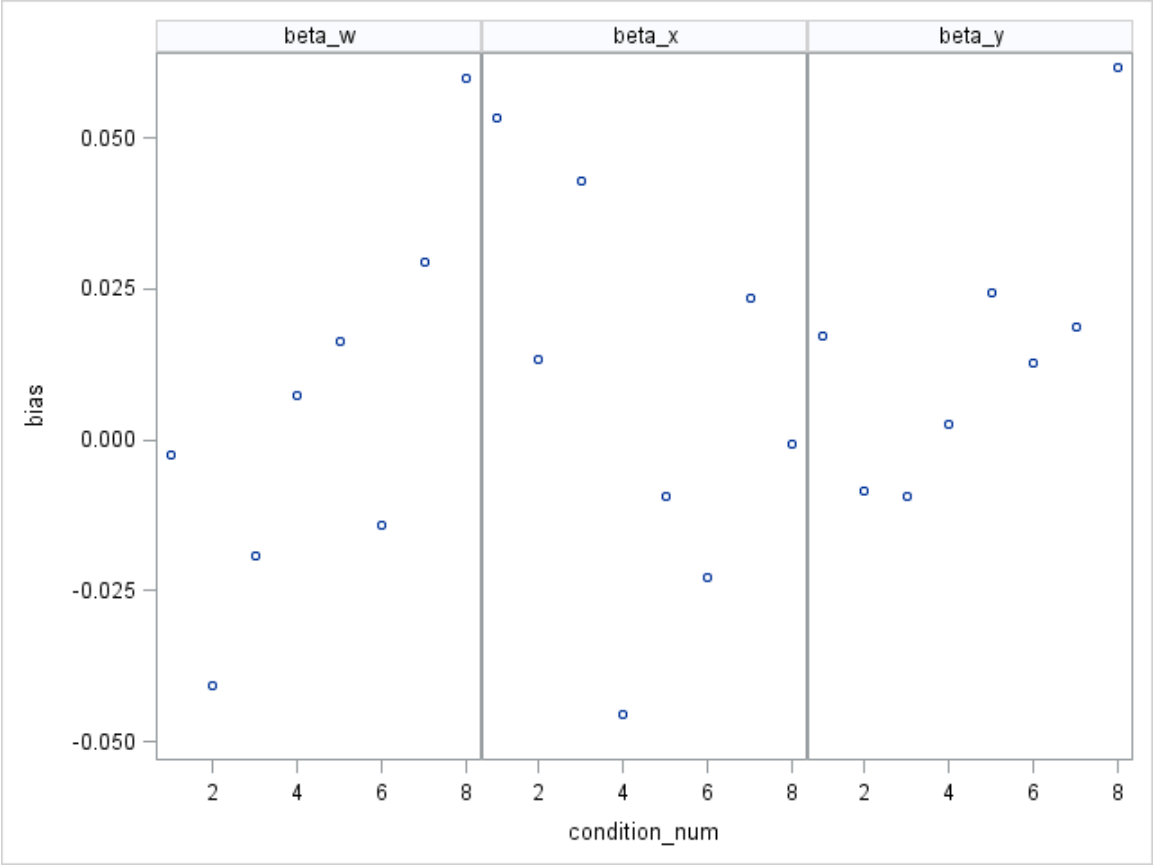


Figure 4. Standardized bias for mean terms in study 2. The graph displays the mean standardized bias (the difference between the average complete-data estimate versus the population parameter expressed as a z-score) for the mean parameters in study 2 (equal associations), broken down by condition. The figure shows that the complete-data estimates displayed no apparent bias.
Condition 1: ICC = .5; nclusters = 30; nobs per cluster = 5.
Condition 2: ICC = .5; nclusters = 30; nobs per cluster = 30.
Condition 3: ICC = .5; nclusters = 100; nobs per cluster = 5.
Condition 4: ICC = .5; nclusters = 100; nobs per cluster = 30.
Condition 5: ICC = .1; nclusters = 30; nobs per cluster = 5.
Condition 6: ICC = .1; nclusters = 30; nobs per cluster = 30.
Condition 7: ICC = .1; nclusters = 100; nobs per cluster = 5.
Condition 8: ICC = .1; nclusters = 100; nobs per cluster = 30.

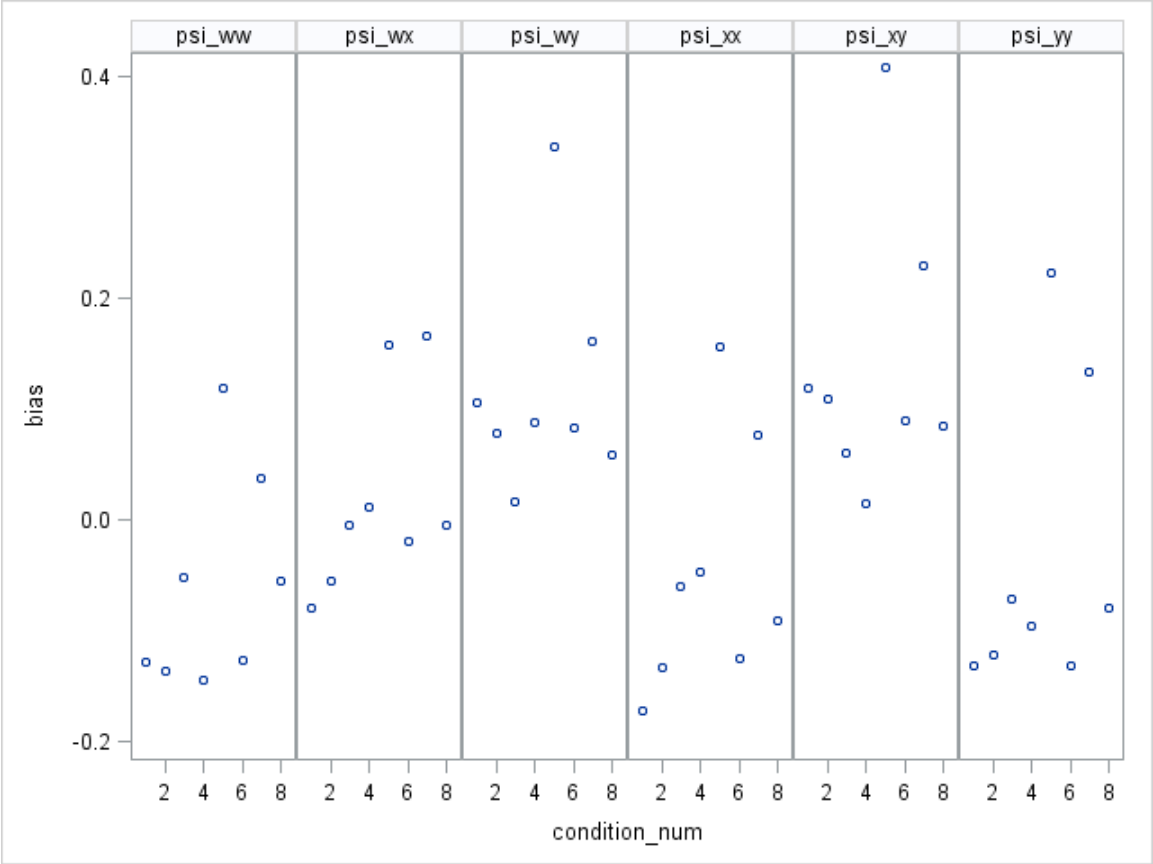


Figure 5. Standardized bias for level-2 covariance terms in study 1. The graph displays the mean standardized bias (the difference between the average complete-data estimate versus the population parameter expressed as a z-score) for the level-2 covariance parameters in study 1 (level-1 specific associations), broken down by condition. The figure shows that the complete-data estimates displayed some bias, and that this bias tended to be greatest for condition 5.
Condition 1: ICC = .5; nclusters = 30; nobs per cluster = 5.
Condition 2: ICC = .5; nclusters = 30; nobs per cluster = 30.
Condition 3: ICC = .5; nclusters = 100; nobs per cluster = 5.
Condition 4: ICC = .5; nclusters = 100; nobs per cluster = 30.
Condition 5: ICC = .1; nclusters = 30; nobs per cluster = 5.
Condition 6: ICC = .1; nclusters = 30; nobs per cluster = 30.
Condition 7: ICC = .1; nclusters = 100; nobs per cluster = 5.
Condition 8: ICC = .1; nclusters = 100; nobs per cluster = 30.

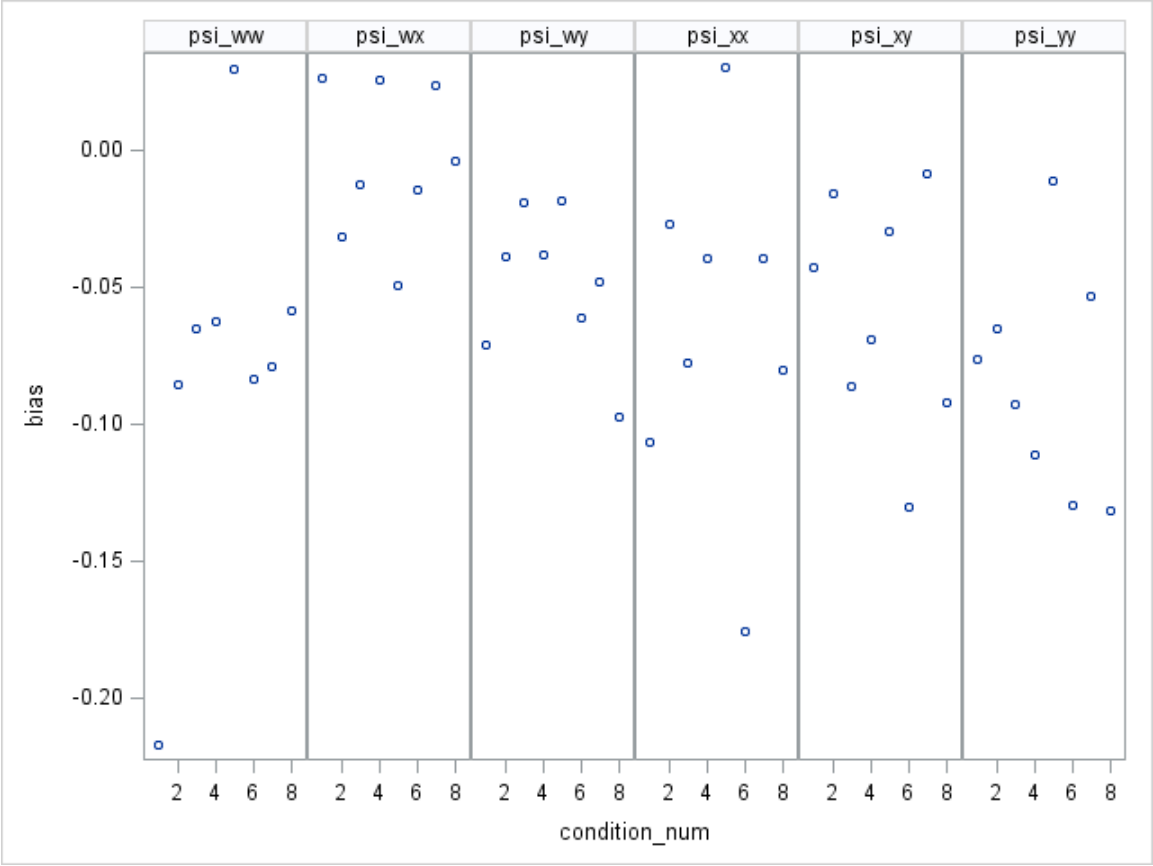


Figure 6. Standardized bias for level-2 covariance terms in study 2. The graph displays the mean standardized bias (the difference between the average complete-data estimate versus the population parameter expressed as a z-score) for the level-2 covariance parameters in study 2 (equal associations), broken down by condition. The figure shows that the complete-data estimates displayed some bias, and that this bias tended to be greatest for condition 5.
Condition 1: ICC = .5; nclusters = 30; nobs per cluster = 5.
Condition 2: ICC = .5; nclusters = 30; nobs per cluster = 30.
Condition 3: ICC = .5; nclusters = 100; nobs per cluster = 5.
Condition 4: ICC = .5; nclusters = 100; nobs per cluster = 30.
Condition 5: ICC = .1; nclusters = 30; nobs per cluster = 5.
Condition 6: ICC = .1; nclusters = 30; nobs per cluster = 30.
Condition 7: ICC = .1; nclusters = 100; nobs per cluster = 5.
Condition 8: ICC = .1; nclusters = 100; nobs per cluster = 30.

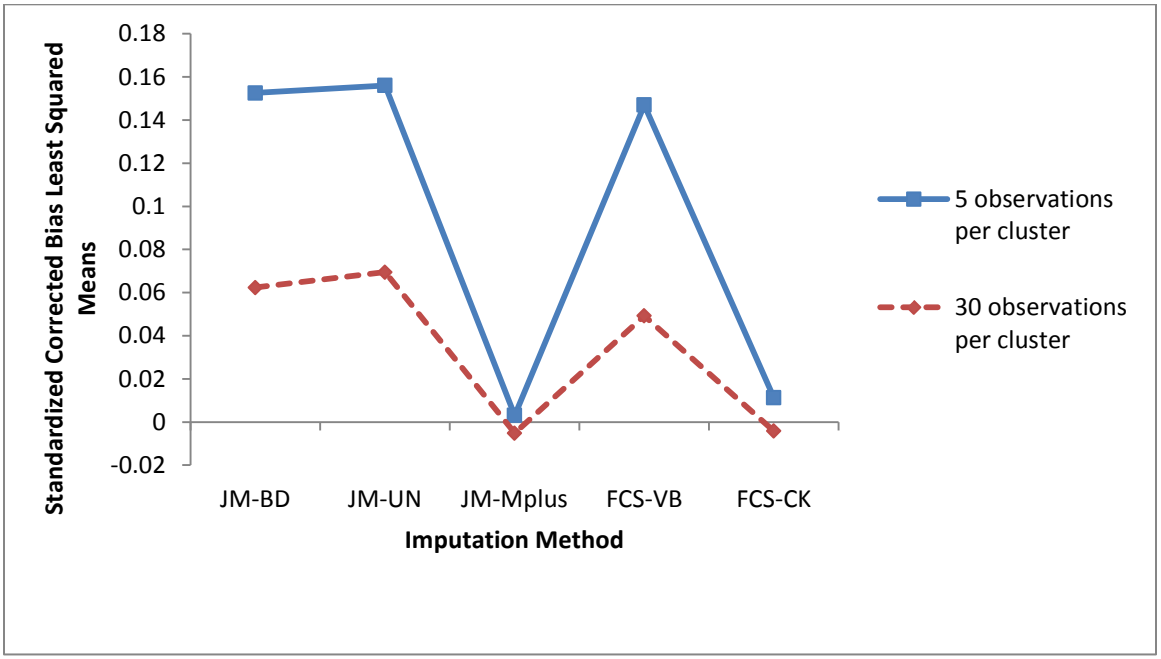


Figure 7. Standardized corrected bias least squared means for level-2 covariance between W and Y in study 1, by method and number of observations per cluster. The graph displays the mean standardized corrected bias (the difference between the average imputation estimate versus the average complete-data estimate expressed as a z-score) for the level-1 covariance between W and Y for the five imputation methods examined in study 1 (level-1 specific associations), broken down by number of observations per cluster. The figure shows that JM-Mplus and FCS-CK estimates were relatively free of bias, whereas the remaining three methods produced biased estimates of the level-1 covariance. Additionally, the graph shows that bias was greater in conditions with 5 observations per cluster. 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.

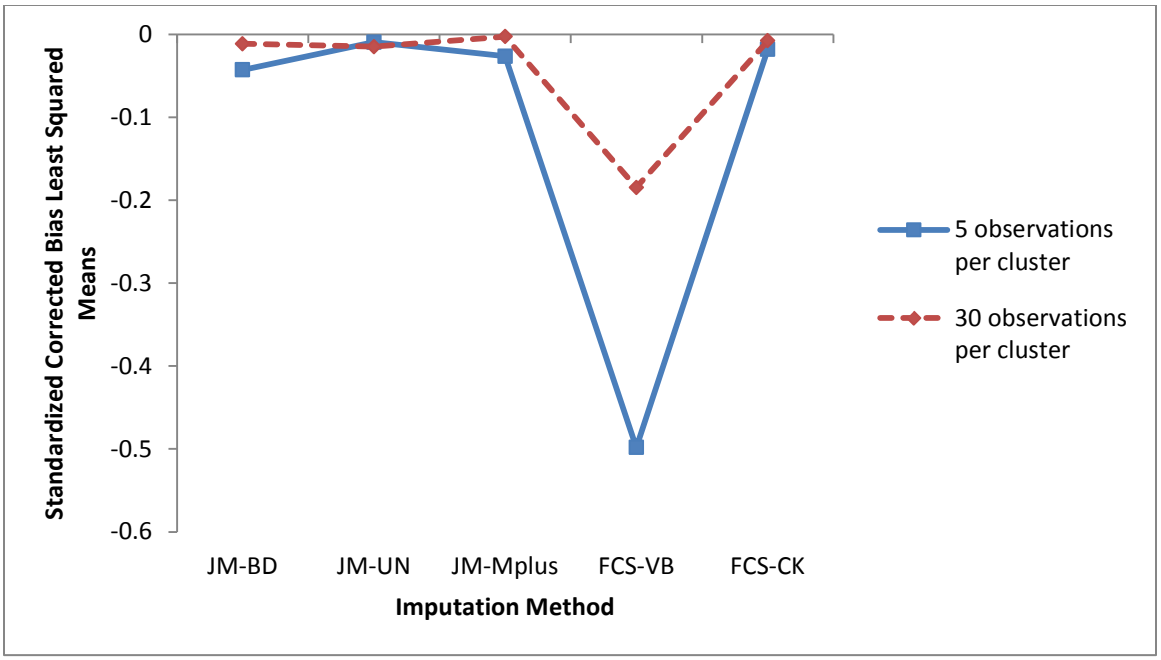


Figure 8. Standardized corrected bias least squared means for the level-1 covariance between X and Y in study 1, by method and number of observations per cluster. The graph displays the mean standardized corrected bias (the difference between the average imputation estimate versus the average complete-data estimate expressed as a z-score) for the level-1 covariance between X and Y for the five imputation methods examined in study 1 (level-1 specific associations), broken down by number of observations per cluster. The figure shows that JM-BD, JM-UN, JM-Mplus, and FCS-CK estimates were relatively free of bias, whereas FCS-VB produced biased estimates of the level-1 covariance. Additionally, the graph shows that bias was greater in conditions with 5 observations per cluster. 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.

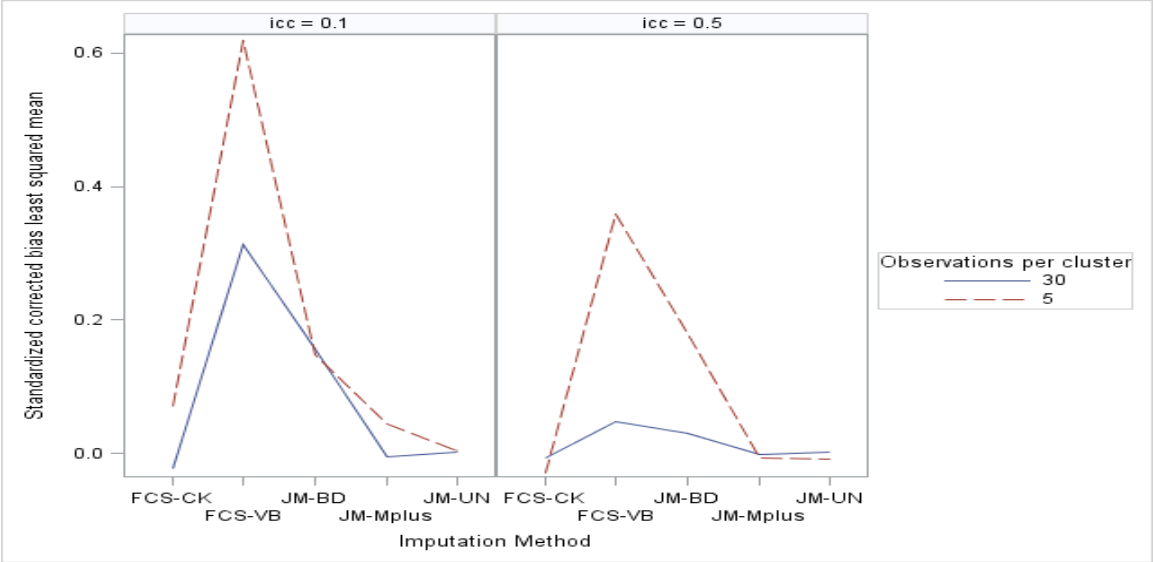


Figure 9. Standardized corrected bias least squared means for the level-2 covariance between X and Y in study 1, by method, ICC, and number of observations per cluster. The graph displays the mean standardized corrected bias (the difference between the average imputation estimate versus the average complete-data estimate expressed as a z-score) for the level-2 covariance between X and Y for the five imputation methods examined in study 2 (equal associations), broken down by number of observations per cluster. The figure shows that JM-UN, JM-Mplus, and FCS-CK estimates were relatively free of bias, whereas JM-BD and FCS-VB produced biased estimates of the level-2 covariance. Additionally, the graph shows that bias was greater in conditions with 5 observations per cluster and that this effect was magnified in the ICC=.1 condition. 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.

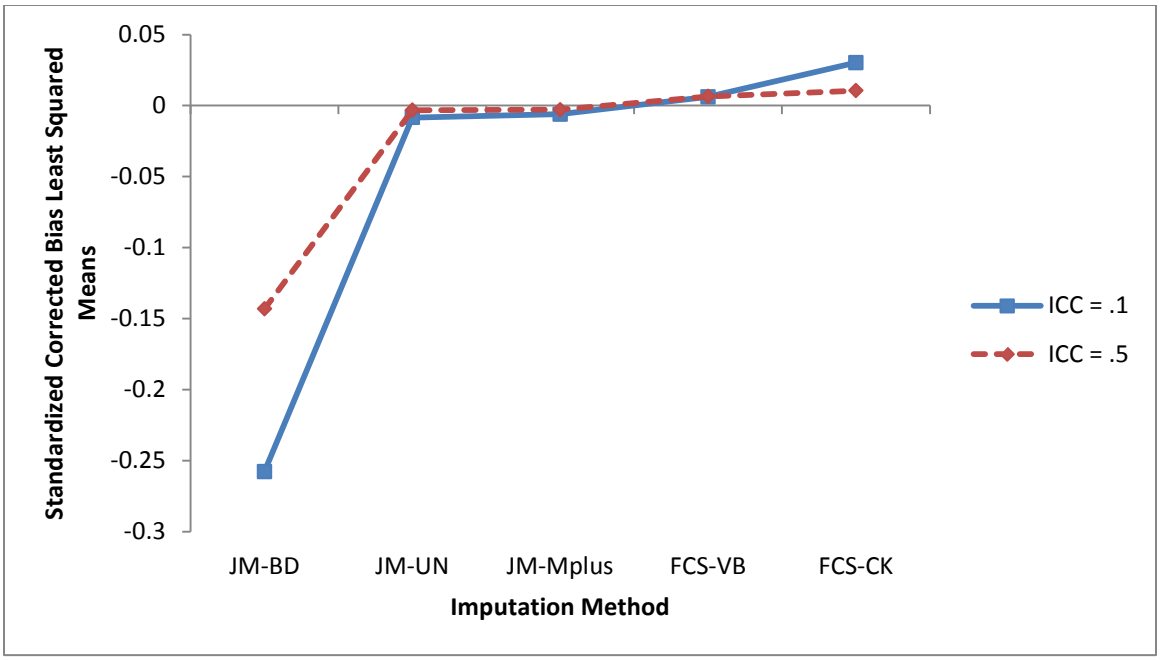


Figure 10. Standardized corrected bias least squared means for the level-2 covariance between X and Y in study 2, by method and ICC. The graph displays the mean standardized corrected bias (the difference between the average imputation estimate versus the average complete-data estimate expressed as a z-score) for the level-2 covariance between X and Y for the five imputation methods examined in study 2 (equal associations), broken down by ICC. The figure shows that JM-UN, JM-Mplus, FCS-VB, and FCS-CK estimates were relatively free of bias, whereas JM-BD produced biased estimates of the level-2 covariance. Additionally, the graph shows that bias was greater in conditions with ICC=.1. 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.

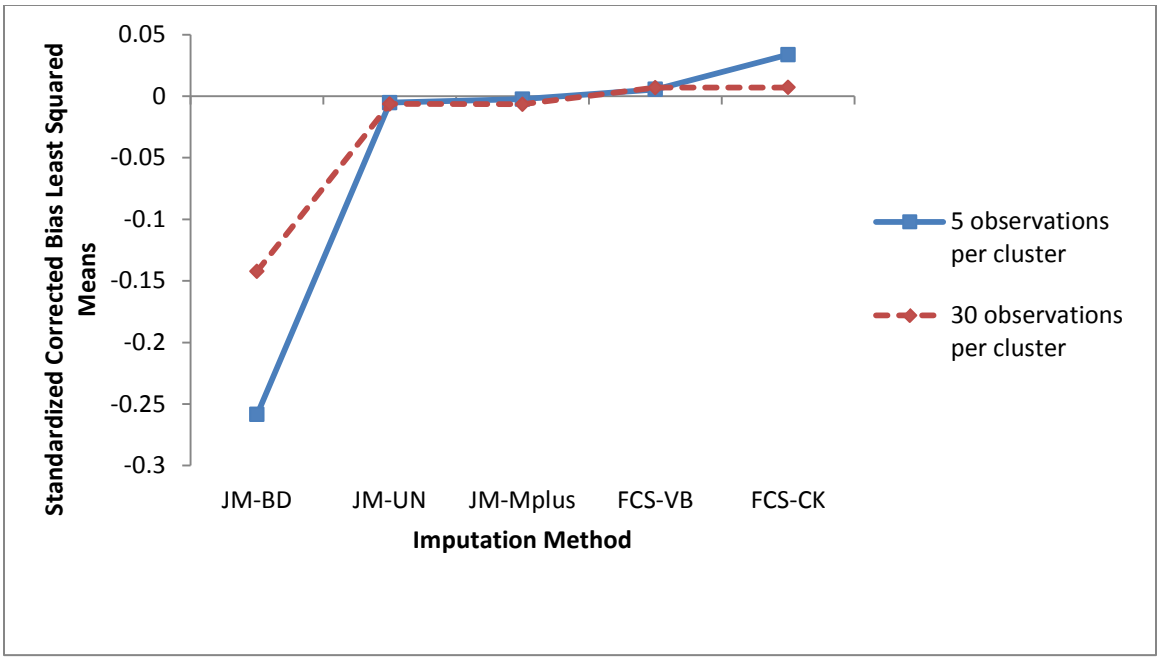


Figure 11. Standardized corrected bias least squared means for the level-2 covariance between X and Y in study 2, by method and number of observations per cluster. The graph displays the mean standardized corrected bias (the difference between the average imputation estimate versus the average complete-data estimate expressed as a z-score) for the level-2 covariance between X and Y for the five imputation methods examined in study 2 (equal associations), broken down by number of observations per cluster. The figure shows that JM-UN, JM-Mplus, FCS-VB, and FCS-CK estimates were relatively free of bias, whereas JM-BD produced biased estimates of the level-2 covariance. Additionally, the graph shows that bias was greater in conditions with 5 observations per cluster. 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.