*Structural Equation Modeling*, 18:35–54, 2011
Copyright © Taylor & Francis Group, LLC
ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2011.532695

Psychology Press
Taylor & Francis Group

# Multiple Imputation Strategies for Multiple Group Structural Equation Models

Craig K. Enders and Amanda C. Gottschall
*Arizona State University*

Although structural equation modeling software packages use maximum likelihood estimation by default, there are situations where one might prefer to use multiple imputation to handle missing data rather than maximum likelihood estimation (e.g., when incorporating auxiliary variables). The selection of variables is one of the nuances associated with implementing multiple imputation, because the imputer must take special care to preserve any associations or special features of the data that will be modeled in the subsequent analysis. For example, this article deals with multiple group models that are commonly used to examine moderation effects in psychology and the behavioral sciences. Special care must be exercised when using multiple imputation with multiple group models, as failing to preserve the interactive effects during the imputation phase can produce biased parameter estimates in the subsequent analysis phase, even when the data are missing completely at random or missing at random. This study investigates two imputation strategies that have been proposed in the literature, product term imputation and separate group imputation. A series of simulation studies shows that separate group imputation adequately preserves the multiple group data structure and produces accurate parameter estimates.

Techniques for analyzing missing data have been widely studied in the methodological literature during the last decade, and two so-called modern missing data methods, maximum likelihood estimation and multiple imputation, are currently considered to be the "state of the art" (Schafer & Graham, 2002). These modern methods are attractive on the basis of statistical theory, because they require the relatively relaxed missing at random (MAR) assumption (Rubin, 1976) where missingness on a variable is related to other measured variables in the data. In contrast, most traditional approaches for dealing with missing data (e.g., deletion methods) require the more stringent missing completely at random (MCAR) condition (i.e., missingness is unrelated to any measured variables), and will yield biased estimates when MCAR does not hold. Numerous empirical studies have demonstrated the relative accuracy of maximum likelihood and multiple imputation (e.g., Arbuckle, 1996; Enders, 2001; Enders & Bandalos, 2001; Graham & Schafer, 1999), and these are the missing data handling techniques that are

Correspondence should be addressed to Craig K. Enders, Department of Psychology, Arizona State University, P. O. Box 871104, Tempe, AZ 85287–1104, USA. E-mail: Craig.Enders@asu.edu

generally recommended in the methodological literature. A number of accessible introductions to missing data are available in the literature (Allison, 2002; Enders, 2006a, 2006b; Schafer & Graham, 2002), as are a number of more in-depth treatments (Enders, 2010; Little & Rubin, 2002; Schafer, 1997).

## MULTIPLE IMPUTATION

Multiple imputation is a simulation-based approach that handles missing data in two distinct phases: the imputation phase and the analysis phase. The goal of the imputation phase is to generate multiple copies of a data set (e.g., $m = 20$), each of which is imputed with slightly different estimates of the missing values. The imputation phase utilizes an iterative data augmentation procedure that repeatedly cycles between an imputation step and a posterior step (the I- and P-step, respectively). The basic idea behind the I-step is to use an estimate of the mean vector and the covariance matrix to build a set of regression equations that predict the incomplete variables from the observed variables. Substituting the observed data into these equations generates predicted scores for the missing values. The predicted scores fall directly on a regression surface, so the I-step restores variability to the data by adding a normally distributed residual term to each predicted value.

The ultimate goal of the imputation phase is to generate *m* complete data sets, each of which contains unique estimates of the missing values. Creating multiple imputations requires different estimates of the regression coefficients at each I-step, and the purpose of the P-step is to generate these estimates. The P-step uses the filled-in data from the preceding I-step to define the posterior distributions of the mean vector and the covariance matrix and then uses Monte Carlo simulation to draw new parameter values from these distributions. The resulting parameter values are randomly different from those that generated the regression coefficients in the preceding I-step. The subsequent I-step uses these updated parameter values to construct a new set of regression coefficients and a different set of imputations. The new imputations carry forward to the next P-step, where the algorithm creates another set of plausible parameter estimates. The data augmentation procedure is repeated a large number of times, and imputed data sets are saved at specified intervals in the sequence (e.g., after every 200th I-step).

Having generated *m* imputed data sets, the analysis phase is relatively straightforward. The analysis of interest is performed on each of the *m* complete data sets, and a single set of point estimates is obtained by averaging the *m* sets of parameter estimates. Combining standard errors is slightly more complicated than taking an arithmetic average, but the computations are straightforward using rules given by Rubin (1987). A number of accessible introductions to multiple imputation are available to readers who are interested in a more detailed account of the procedure (Allison, 2002; Enders, 2006a; Schafer, 1997, 1999; Schafer & Graham, 2002; Schafer & Olsen, 1998; Sinharay, Stern, & Russell, 2001).

When multiple imputation is used in structural equation modeling (SEM), an interesting question arises about how to best preserve interactive effects in a multiple group model. Specifically, an ideal imputation model should preserve any group differences that exist in the mean structure, the covariance structure, or both. Failing to preserve these group differences in the imputation phase can distort subsequent tests of invariance. Two approaches for preserving interactive effects have been proposed in the literature. Product term imputation (PTI) introduces a series of dummy variables and product terms in the imputation regression equations, whereas separate group imputation (SGI) imputes the data separately within each subpopulation. From an analytical perspective, we had strong reasons to expect that SGI would perform better than PTI, and the details of this reasoning are discussed later. However, the literature provides little empirical support for choosing between these two approaches, and their relative performance becomes less predictable in the SEM context. Consequently, the purpose of this article is to assess the relative performance of PTI and SGI in multiple group SEMs.

Given that maximum likelihood estimation routines are readily available in every commercial SEM software program, it is reasonable to ask whether the extra hassle of imputation is even warranted. We believe that there are practical reasons to prefer multiple imputation, particularly in multiple group structural equation models. First, the methodological literature typically recommends a so-called inclusive analysis strategy that includes auxiliary variables that are correlates of missingness or correlates of the incomplete analysis variables. Although it is possible to include auxiliary variables in a maximum likelihood analysis (Graham, 2003), it is arguably easier to do so in the context of multiple imputation (Enders, 2010). Second, researchers typically use multiple group models to assess the item-level measurement invariance of a set of questionnaire items. Frequently, the ultimate goal is to compute a scale score that can be used in a subsequent analysis, and the multiple group model is often a precursor to computing the scale score. Because maximum likelihood does not produce a filled-in data set, there is no effective way to compute a scale score from item-level data (except to leave the entire scale score missing, which wastes data and lowers power). Multiple imputation is ideally suited for item-level analyses such as this, so there are good reasons to study its performance in multiple group structural equation models.

## SELECTING VARIABLES FOR THE IMPUTATION PHASE

The separation of the imputation and analysis phase is arguably one of the strengths of multiple imputation because a single set of imputed data sets can be used for a number of different statistical analyses, all of which can be conducted using standard complete-data techniques. Also, the separation of the imputation and analysis phase allows for auxiliary variables (i.e., variables that are included to improve estimation and satisfy the MAR assumption, but are not of theoretical interest) to easily be incorporated at the imputation phase, as opposed to the more difficult incorporation of auxiliary variables in an analysis where maximum likelihood estimation is used to handle missing data. However, separating the missing data handling from the analysis is also a potential Achilles heel because the regression model used in the imputation phase must preserve any associations or special features of the data that will be modeled in the subsequent analysis phase. For example, this article deals with multiple group models that are commonly used to examine moderation effects in psychology and the behavioral sciences. These models are characterized by the presence of interactive effects, such that $G$ qualitatively different subpopulations (e.g., males and females) have different mean structures, covariance structures, or both. Special care must be exercised when using multiple imputation because failing to preserve these interactive effects during the imputation phase can produce biased parameter estimates in the subsequent analysis phase, even when the data are MCAR or MAR.

Schafer (1997) discussed the selection of variables for the imputation phase, and suggested that the imputer should formulate a model that “is general enough to preserve any associations among the variables (two-, three-, or even higher-way associations) that may be the target of subsequent analyses” (p. 143). It is well known that omitting a variable from the imputation phase will attenuate its association with other variables in the subsequent statistical analysis (e.g., correlations involving the omitted variable will be biased toward zero). However, the literature provides little to no guidance on preserving higher order effects that might be present in the data. The purpose of this article is to outline two procedures that have been proposed in the literature and to evaluate their performance using a series of computer simulation studies.

## TWO METHODS FOR PRESERVING INTERACTIVE EFFECTS

Allison (2002) outlines two imputation strategies that can be used to preserve group differences in the mean and covariance structure, product term imputation and separate group imputation. These strategies are both quite general and can be used to formulate imputation models for a variety of multiple group analyses (e.g., multiple group regression models, factor analysis models, growth models). To make the ensuing discussion more concrete, consider a simple research scenario where it is of interest to examine whether the association between $X$ and $Y$ differed between two groups. The challenge of using multiple imputation in this context is to specify an imputation model that preserves the interactive effect that is potentially present in the data (i.e., group differences in the mean vector and the covariance matrice).

### Product Term Imputation

PTI preserves interactive effects by incorporating dummy codes and product terms into the imputation model (in the context of interactions with continuous variables, von Hippel [2009] refers to PTI as *transform, then impute*). Returning to the bivariate example, assume that a subgroup of cases has missing $Y$ values but complete data on $X$. The missing $Y$ values would be imputed from a regression equation that includes the observed $X$ values, a dummy-coded group membership variable, and a product term between $X$ and the code variable, as follows:

$$Y_{im}^{*} = B_{0m} + B_{1m}X_i + B_{2m}D_i + B_{3m}(X_i)(D_i) + z_{im}, \tag{1}$$

where $Y_{im}^{*}$ is the imputed value for case $i$ in data set $m$, the $B_{km}$ terms are regression coefficients, $D_i$ is a dummy code for group membership, and $z_{im}$ is a random residual term sampled from a normal distribution. The code variable in Equation 1 produces imputed values that preserve the group mean differences on $Y$, and the product term preserves group differences in the covariance.

It is not immediately obvious, but PTI makes the potentially restrictive assumption that the variance of $X$ and $Y$ is constant across groups. The reason behind this assumption lies in the fact that the random residuals (i.e., the $z_{im}$ term in Equation 1) for both groups are sampled from the same normal distribution with mean zero and variance $\sigma_z^2$. When the population variances are equal, PTI will appropriately preserve the multiple group data structure. However, if the population variances are unequal, the PTI model will distort the variance estimates for each group. Although this might not impact the regression coefficients from the subsequent analysis, it is likely to bias residual variance estimates and estimates of effect size such as $R^2$. Testing the equality of group variances is potentially of interest in multiple group structural equation models, but using PTI can increase the likelihood of concluding that there is invariance when the population variances are truly different. PTI is also more complicated to implement in multiple imputation because of the necessity to explicitly include product terms in the imputation phase (for some multiple group structural equation models, the required product terms are not necessarily obvious). Finally, including incomplete product terms in an imputation procedure violates the multivariate normality assumption (required by multiple imputation [MI]) because the product of two variables is not normal, even when the components of the product are normal. In the context of a multiple group model, one or more of the product terms (e.g., a product involving the grouping variable and one of the incomplete analysis model variables) would frequently be missing, making the use of a normality-based imputation model suspect.

## Separate Group Imputation

As its name implies, SGI imputes the data separately for each of the $G$ subgroups (von Hippel [2009] describes a similar procedure as *stratify, then impute*). This requires cases to have information on the grouping variable. If a case is missing on the grouping variable then this is not simply a missing data issue and the case has to be deleted because their group membership is undefined. Returning to the previous regression example, the imputation model for a subgroup of cases with missing $Y$ values and complete data on $X$ is

$$Y^*_{img} = B_{0mg} + B_{1mg}X_{ig} + z_{img}. \tag{2}$$

The addition of the $g$ subscript in Equation 2 denotes the fact that each group's data are imputed from a distinct regression equation. This allows the mean and covariance structure to naturally vary across the $G$ subgroups without the need to introduce potentially cumbersome code and product variables. Importantly, SGI allows the variance of the residual terms to vary across groups. Consequently, SGI places no restrictions on the covariance structure of the imputed values.

Like PTI, SGI also has some noteworthy limitations. First, SGI is limited to applications where the moderator is a nominal variable. Obviously, the imputation phase cannot be performed separately for each level of the moderator when it is a continuous variable. In this article, we only consider applications where cases can be classified into distinct and mutually exclusive groups. However, there are many research questions involving the interaction of two continuous variables and in those situations SGI is not possible (see von Hippel [2009] for a discussion of interactions between continuous variables). A second difficulty is that SGI requires a sufficient number of cases in each group. Although simulation studies suggest that multiple imputation can perform well in relatively small samples (Graham & Schafer, 1999), applying multiple imputation to a small sample can sometimes result in convergence failures, particularly when the number of variables in the imputation model is large (Enders, 2010). Finally, SGI uses a saturated model to impute the data for each group, so SGI requires far more parameters than PTI. This lack of parsimony could result in problems such as larger sampling fluctuation and wider confidence intervals.

## WHAT DOES THE METHODOLOGICAL LITERATURE RECOMMEND?

The methodological literature does not provide clear guidance about which approach to use and when. Some authors recommend the PTI method to preserve interactive effects (King, Honaker, Joseph, & Scheve, 2001; Schafer & Olsen, 1998), whereas others recommend SGI (Graham, 2009; Schafer & Graham, 2002). Importantly, these recommendations come with little to no empirical evidence to support them. Allison (2002) presented the results from a single data analysis that compared PTI and SGI. He found that the two methods produced similar results and suggested that either method is acceptable. Finally, von Hippel (2009) used analytic methods to study the impact of different imputation models for interaction terms (as noted previously, he uses the phrase *transform, then impute* to refer to PTI and uses *stratify, then impute* to denote SGI). The primary purpose of the von Hippel study was to determine whether it was better to incorporate the interaction into the imputation process or to form the interaction after imputing the data based on the component parts of the interaction (e.g., impute with $X$ and $Z$ then form $XZ$ from the imputed data). Results from this study clearly indicate that the imputation procedure needs to explicitly incorporate information about the interaction, but the study did not compare PTI and SGI.

Although it is reasonable to expect that SGI will provide more accurate parameter estimates than PTI, the relative performance of the two methods might be less predictable in multiple group structural equation models because the variance of any single variable is usually decomposed across several model parameters. Consequently, the purpose of this article is to assess the relative performance of PTI and SGI in multiple group structural equation models. We designed a series of simulation studies to address this issue.

## METHOD

A series of small simulation studies was used to compare PTI and SGI and to illustrate the bias that can result when group differences in the mean and covariance structure are ignored during imputation. We begin with a multiple group path analysis model and then present simulation results for a multiple group confirmatory factor analysis (CFA) and a multiple group latent growth model (LGM). These analysis models are certainly not exhaustive, but we chose these models because they are representative of a wide class of multiple group models that researchers use to examine moderation effects in psychology and the behavioral sciences.

The simulation studies were not intended to be comprehensive in nature, but it was of interest to determine whether sample size and the missing data rate had an impact on the performance of PTI and SGI. Consequently, we manipulated two sample size values ($n_1 = 50$ and $n_2 = 250$) and three MCAR missing data rates (5%, 15%, and 25%). Although multiple imputation requires the less stringent MAR mechanism, we chose to use MCAR data to illustrate the point that misspecifying the imputation model (e.g., by failing to account for the multiple group data structure) can produce bias, even when the missing values are completely benign. However, we did conduct a series of MAR simulations, and the results from the MCAR simulations in this article generalize to MAR missingness as well.

We used M*plus* 4.21 (Muthén & Muthén, 1998–2010) to generate 1,000 multivariate normal data sets within each design cell and used the MODEL MISSING command to create the desired proportion of MCAR missing values. Note that missing values were imposed on all model variables. The artificial data sets were imputed using PROC MI in SAS version 9.2 and $m = 10$ imputed data sets were generated for each sample replicate. The multiple group models were subsequently fit to the imputed data using M*plus,* and the resulting parameter estimates and standard errors were combined using the SAS MIANALYZE procedure. The primary purpose of the simulations was to examine parameter estimate bias, so the average estimates from each design cell were compared to the corresponding population parameters. In addition to reporting raw bias values, we also expressed bias on an effect size metric by dividing the raw bias values (i.e., the average estimate minus the parameter value) by the standard deviation of each estimate within a particular design cell (the imputation strategies produced very similar standard deviations, so the pooled estimate was used as the divisor). The standard deviation of the estimates represents the true sampling error, so the standardized bias values can be interpreted in standard error units (e.g., a value of .50 means that the bias was equivalent to one half of a standard error). Collins, Schafer, and Kam (2001) used the same measure of standardized bias and suggested that values that exceed $\pm.40$ are practically significant. We give additional details on the specific population models in subsequent sections.

As an aside, recent methodological research suggests that using more than 10 imputations is generally a good idea (e.g., Bodner, 2008; Graham, Olchowski, & Gilreath, 2007). The rationale for using a larger number of imputations is that it improves power. For example, Graham et al. (2007) suggested that analyzing $m = 20$ imputations generally yields power that is comparable to a full information maximum likelihood analysis. Under an MAR mechanism, any single data set from a multiple imputation procedure yields unbiased estimates of the population parameters (Schafer, 1997), so the number of imputations has no relation to bias. Because our simulations were only concerned with the bias of the two imputation strategies, the number of imputations is essentially arbitrary and has no material impact on the results. Nevertheless, it is important to put our design choices in the context of recent recommendations from the methodological literature.

## STUDY 1: MULTIPLE GROUP PATH ANALYSIS

A typical application of multiple group path analysis involves testing whether the structural regression coefficients differ across $G$ qualitatively different subgroups. The path analysis model in the multiple group case is

$$\mathbf{Y}_g = \mathbf{B}_g\mathbf{Y}_g + \mathbf{\Gamma}_g\mathbf{X}_g + \boldsymbol{\zeta}_g, \tag{3}$$

where $\mathbf{B}_g$ is a matrix that contains the structural regression coefficients among the endogenous variables for group $g$, $\mathbf{\Gamma}_g$ is a matrix that contains the regressions relating the exogenous and endogenous variables, and $\boldsymbol{\zeta}_g$ is a vector of disturbance (i.e., residual) terms. The mean structure is typically not of interest in path models, so the vector of regression intercepts is omitted from Equation 3 and the remainder of this section.

The model-implied covariances for group $g$ are given by

$$\mathbf{\Sigma}_{XYg} = \mathbf{\Phi}_g\mathbf{\Gamma}'_g(\mathbf{I} - \mathbf{B}_g)^{-1'}, \tag{4}$$

where $\mathbf{\Phi}_g$ is the covariance matrix of the exogenous variables for group $g$, and $\mathbf{I}$ is an identity matrix. The model-implied variances of the endogenous variables are

$$\mathbf{\Sigma}_{YYg} = (\mathbf{I} - \mathbf{B}_g)^{-1}(\mathbf{\Gamma}_g \mathbf{\Phi}_g \mathbf{\Gamma}'_g + \mathbf{\Psi}_g)(\mathbf{I} - \mathbf{B}_g)^{-1'}, \tag{5}$$

where $\mathbf{\Psi}_g$ is the covariance matrix of the endogenous disturbance terms for group $g$. Finally, the model-implied covariance matrix of the endogenous variables is $\mathbf{\Phi}_g$, which contains the sample estimates of the variances and covariances. Collectively, the matrices in Equations 4 and 5 show that the $G$ groups will share a common covariance matrix if *all* regression coefficients, residual variances, and exogenous variable variances are identical (i.e., invariant) across groups.

The goal of a multiple group path analysis is to determine whether the model parameters differ across groups. Equations 4 and 5 express the covariance matrix as a function of the model parameters, but these equations also imply that group differences in the parameters will be evidenced by group differences in the covariance structure. Consequently, the imputation phase that precedes a multiple group path analysis should preserve group differences in the covariance matrix. The mean structure of these models is typically not of interest and is not considered here. Nevertheless, PTI and SGI will adequately preserve group mean differences, as illustrated in the CFA and LGM examples given later in the article.

To illustrate, consider a two-group mediation model (i.e., moderated mediation; MacKinnon, 2007) where the association between $X$ and $Y$ is completely mediated by a variable $M$. A graphical depiction of this model is shown in Figure 1. As seen in Figure 1, the unstandardized regression coefficients differ across groups, as do the residual variance terms for $M$ and $Y$. The population regression model parameters produced group differences in the variance of $M$ and $Y$, the covariance between $X$ and $M$, and the covariance between $M$ and $Y$. Preserving group differences in the covariance matrix is straightforward using SGI because it is only necessary to perform a separate imputation phase for each group. The procedure is particularly

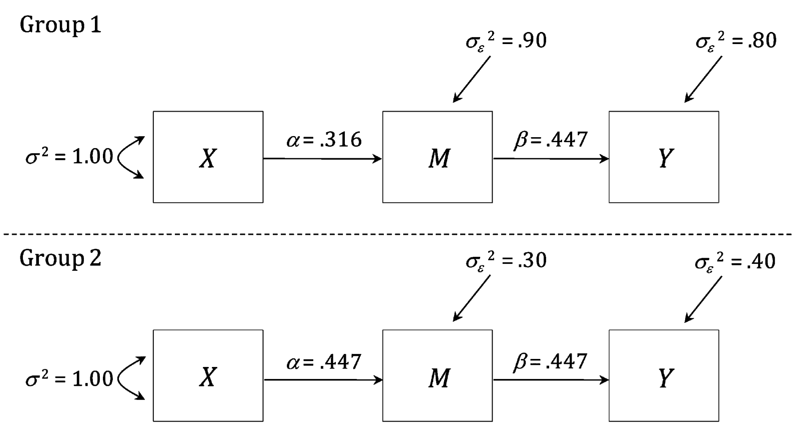


FIGURE 1 Multiple group mediation path model used in Study 1. Note that the variance of $M$ and $Y$ differed between groups. In the first group, $X$ explained 10% of the variance in $M$, and $M$ explained 20% of the variance in $Y$. In the second group, $X$ explained 40% of the variance in $M$, and $M$ explained 20% of the variance in $Y$.

easy to implement in SAS, as it is only necessary to specify the grouping variable on the BY subcommand of PROC MI after first sorting the data by the grouping variable.

Preserving group differences using PTI is slightly more complicated. The imputation model for each variable would include a code variable and two product terms involving the dummy code and the remaining variables (e.g., the imputation model for $Y$ would include $X$, $M$, $D$, $DX$, and $DM$). As such, the imputation model would include a total of six variables: the three manifest variables, the code variable, and the product terms involving the code variable and the manifest variables. Fortunately, multiple imputation software packages do not require the imputer to specify the exact regression model for each variable. Rather, the user specifies the set of variables that will be used during imputation (e.g., by listing the seven variables on the VAR subcommand of PROC MI), and the software selects the variables that are appropriate for a given missing data pattern.

A simulation study was conducted using the mediation model in Figure 1, and the average parameter estimates are shown in Table 1. For the sake of brevity, results from the $n = 250$ and 15% missing data rate condition are given in Table 1, and the corresponding results from the $n = 50$ condition are available on request. To facilitate interpretation, the parameter estimates that produced standardized bias values larger than .40 (i.e., 4/10 of a standard error unit) are shown in bold type. This is a threshold that Collins et al. (2001) suggested is practically significant.

As seen in Table 1, the imputation model based on the three analysis variables (i.e., an imputation model that ignored the interactive effect by imputing all cases together and failing to

TABLE 1
Average Estimates from the Mediation Simulation ($n$ = 250, 15% Missing Data)

| | | *Model Variables Only* | | | *PTI* | | | *SGI* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Population Parameter* | | *Est.* | *Raw Bias* | *Std. Bias* | *Est.* | *Raw Bias* | *Std. Bias* | *Est.* | *Raw Bias* | *Std. Bias* |
| Group 1 | | | | | | | | | | |
| M on X (α) | 0.316 | 0.34 | 0.03 | 0.36 | 0.33 | 0.02 | 0.26 | 0.31 | 0.00 | −0.02 |
| Y on B (β) | 0.447 | 0.47 | 0.02 | 0.36 | 0.47 | 0.02 | 0.31 | 0.44 | −0.01 | −0.09 |
| Mediation (αβ) | 0.141 | **0.16** | **0.02** | **0.50** | 0.16 | 0.01 | 0.39 | 0.14 | 0.00 | −0.07 |
| M residual ($\sigma^2_M$) | 0.90 | **0.86** | **−0.04** | **−0.52** | **0.83** | **−0.07** | **−0.81** | 0.90 | 0.00 | −0.03 |
| Y residual ($\sigma^2_Y$) | 0.80 | 0.77 | −0.03 | −0.35 | **0.73** | **−0.07** | **−0.98** | 0.80 | 0.00 | −0.06 |
| $R^2_M$ | 0.10 | **0.12** | **0.02** | **0.55** | 0.12 | 0.02 | 0.36 | 0.10 | 0.00 | 0.08 |
| $R^2_Y$ | 0.20 | 0.22 | 0.02 | 0.38 | **0.22** | **0.02** | **0.44** | 0.20 | 0.00 | −0.01 |
| Group 2 | | | | | | | | | | |
| M on X (α) | 0.447 | **0.41** | **−0.03** | **−0.87** | 0.44 | −0.01 | −0.15 | 0.45 | 0.00 | −0.05 |
| Y on B (β) | 0.447 | 0.44 | 0.00 | −0.07 | 0.43 | −0.02 | −0.28 | 0.45 | 0.00 | 0.07 |
| Mediation (αβ) | 0.200 | **0.18** | **−0.02** | **−0.56** | 0.19 | −0.01 | −0.34 | 0.20 | 0.00 | 0.03 |
| M residual ($\sigma^2_M$) | 0.30 | **0.37** | **0.07** | **2.21** | **0.34** | **0.04** | **1.15** | 0.30 | 0.00 | −0.05 |
| Y residual ($\sigma^2_Y$) | 0.40 | **0.44** | **0.04** | **1.13** | **0.43** | **0.03** | **0.76** | 0.40 | 0.00 | −0.08 |
| $R^2_M$ | 0.40 | **0.31** | **−0.09** | **−1.67** | **0.37** | **−0.03** | **−0.56** | 0.40 | 0.00 | −0.04 |
| $R^2_Y$ | 0.20 | 0.19 | −0.01 | −0.12 | 0.19 | −0.01 | −0.22 | 0.21 | 0.01 | 0.13 |

*Note.* Std. bias is standardized bias, or raw bias expressed in standard error units. Estimates with bias that exceeded 4/10 of a standard error unit are shown in bold type. PTI = product term imputation; SGI = separate group imputation.

include dummy codes and product terms) produced biased estimates of most model parameters, most notably the residual variance terms, mediation effect, and $R^2$ values. This result is not surprising, as there was a clear mismatch between the imputation and analysis model. Note that the direction of the bias has important implications for tests of invariance, as failing to preserve the multiple group structure attenuated group differences (i.e., the estimates for each group were biased toward the mean parameter value), and would lead to an increased Type II error rate (i.e., the researcher concludes that the mediation model parameters are invariant, when they actually differ in the population).

PTI produced moderately biased estimates of most model parameters (bias values of .30 were typical), but the bias was most pronounced in the residual variance and $R^2$ values. The magnitude of the raw bias values in the $n = 50$ condition were comparable, but the standardized bias values were somewhat lower due to the large sampling fluctuation. In contrast, SGI produced estimates that were quite accurate, even in the $n = 50$ condition. For example, the largest standardized bias value in the $n = 250$ simulation condition was .13 (i.e., slightly larger than 1/10 of a standard error unit), and the largest bias value from the $n = 50$ condition was .26. Perhaps not surprisingly, the missing data rate played a significant role in the performance of PTI and SGI. The two procedures gave very similar estimates in the 5% missing data rate condition (results not presented), but the bias in the PTI estimates increased as the missing data rate increased (SGI estimates were relatively constant).

## STUDY 2: MULTIPLE GROUP CONFIRMATORY FACTOR ANALYSIS

A typical application of multiple group CFA involves invariance tests, where it is of interest to determine whether the measurement model parameters (e.g., factor loadings, measurement intercepts) differ across $G$ qualitatively different subgroups. The CFA measurement model in the multiple group case is

$$\mathbf{Y}_g = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\xi}_g + \boldsymbol{\varepsilon}_g \tag{6}$$

where $\mathbf{Y}$ is the vector of manifest indicators, $\boldsymbol{\tau}$ is a vector of measurement intercepts, $\boldsymbol{\Lambda}$ is the matrix of factor loadings, and $\boldsymbol{\varepsilon}$ is a vector of residuals. The $g$ subscript is used to denote group membership, and implies that the model parameters can take on different values for each group.

The model-implied covariance matrix for group $g$ is

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g \boldsymbol{\Phi}_g \boldsymbol{\Lambda}'_g + \boldsymbol{\Theta}_g, \tag{7}$$

where $\boldsymbol{\Phi}$ is the latent variable covariance matrix for group $g$, and $\boldsymbol{\Theta}$ is the residual covariance matrix for that group. Equation 7 shows that the covariance matrices for the $G$ groups will only be identical if all of the factor loadings, factor variances and covariances, and residual variances are identical (i.e., invariant) across groups. In this case, the $g$ subscript vanishes, and the groups share a common covariance matrix.

Moving to the mean structure, the expectation (i.e., mean) of the manifest indicators is written as

$$\mathbf{E}(\mathbf{Y}_g) = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\kappa}_g \tag{8}$$

where $\boldsymbol{\kappa}$ is the latent variable mean for group $g$. Consistent with the covariance structure, the means for the $G$ groups are only identical if the measurement intercepts, factor loadings, and latent variable means are invariant across groups.

Of course, the purpose of the multiple group CFA analysis is to determine whether the measurement model parameters are invariant (i.e., the same) across groups, so these unknown parameters must be estimated from the observed data. Equations 7 and 8 express the covariance matrix and mean vector as a function of the model parameters, but these equations also imply that a lack of invariance (i.e., the need for parameters to vary across groups) will result when the mean or covariance structure differs across groups. As such, the imputation phase that precedes a multiple group CFA should preserve group differences in the mean and covariance structure.

To illustrate, consider a two-group CFA model with six manifest indicators loading on a single factor, a diagram of which is shown in Figure 2. As seen in Figure 2, the $Y_3$ loading differs between groups ($\lambda_3 = 1$ and .50 for Groups 1 and 2, respectively), as does the $Y_2$ intercept ($\tau_2 = 0$ and 1 for Groups 1 and 2, respectively) and the $Y_2$ residual variance ($\theta_{22} = .25$ and .40 for Groups 1 and 2, respectively). The lack of invariance in the CFA model parameters implies group differences in the mean and covariance structure of the data. Preserving these group differences using SGI is straightforward, as it is only necessary to sort the data by the grouping variable and perform the imputation procedure separately for each group—the SAS imputation syntax for this example is given in Appendix A, and the M*plus* syntax for fitting the CFA model to the multiply imputed data sets is given in Appendix B. Preserving group differences using PTI is somewhat more complicated. Extending the logic of Equation 1, the imputation phase would include a code variable for group membership and a set of product terms involving the code variable and the manifest indicators (i.e., the imputation model would include the six manifest variables, a single code variable, and six product terms involving the indicators and the code variable). As described earlier, the code variable preserves group mean differences, and the product terms account for the possibility that the pairwise associations between $Y_j$ and $Y_k$ differ between groups.

A simulation study was conducted based on the population model shown in Figure 2, and the mean estimates for selected parameters are given in Table 2. For the sake of brevity, results from the $n = 250$ and 15% missing data rate condition are given in Table 2, and the corresponding results from the $n = 50$ condition are available on request. To facilitate interpretation, the parameter estimates that produced standardized bias values larger than .40 (i.e., 4/10 of a standard error unit) are shown in bold type.

Not surprisingly, the imputation model based solely on the analysis variables (i.e., the six manifest indicators) produced biased estimates of the noninvariant parameters (i.e., $\lambda_3$, $\tau_2$, and $\theta_{22}$). This was true even in the 5% missing data rate condition (results not presented), and the magnitude of the bias increased as the proportion of missing values increased. The bias values associated with the noninvariant parameters were generally quite substantial, and exceeded one standard error unit in some cases. Consistent with the mediation simulation, the direction of the bias attenuated group differences, and would likely lead to an increased Type II error rate for any invariance tests (i.e., the researcher concludes that the measurement model parameters are invariant, when they actually differ in the population).

The performance of PTI improved when moving to a latent variable model, most likely due to the fact that the magnitude of the group differences in the variances was somewhat smaller in the CFA simulation (see the Discussion section). As seen in Table 2, PTI produced unbiased

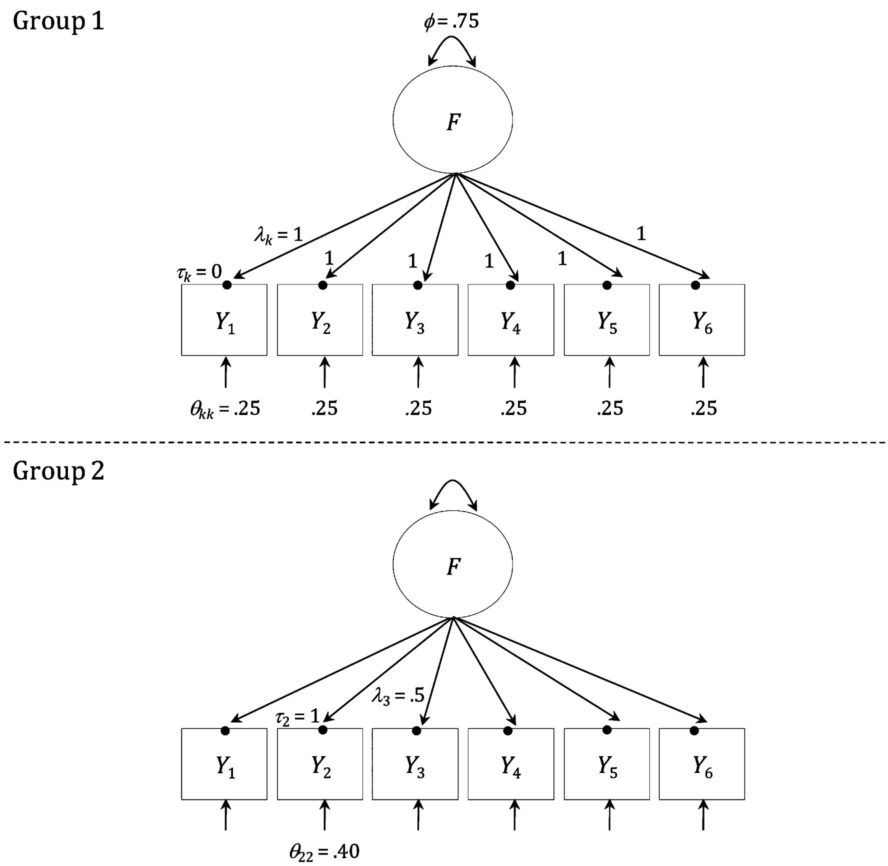


FIGURE 2 Multiple group confirmatory factor analysis model used in Study 2. The parameters for Group 2 were identical to those for Group 1, unless otherwise noted. Note that the measurement intercepts are represented by black dots on the observed variables. All measurement intercepts were set at zero, unless otherwise noted. Also, the latent variable means took on values of zero in both groups, but were omitted from the figure to reduce clutter.

estimates of all mean structure parameters, and the bias in the covariance structure portion of the model was isolated to residual variance terms. Consistent with the mediation simulation, SGI produced unbiased estimates of all model parameters, even with a group size of 50; the largest standardized bias value in $n = 250$ condition was .08, and the largest standardized bias value from the corresponding $n = 50$ condition was .12. As before, the PTI and SGI were quite similar with 5% missing data, and the two methods diverged as the missing data rate increased.

## STUDY 3: MULTIPLE GROUP LATENT GROWTH MODEL

A single group linear latent growth curve model can be parameterized as a two-factor CFA

TABLE 2
Average Estimates from the Confirmatory Factor Analysis Simulation ($n = 250$, 15% Missing Data)

| | | *Model Variables Only* | | | *PTI* | | | *SGI* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Population Parameter* | | *Est.* | *Raw Bias* | *Std. Bias* | *Est.* | *Raw Bias* | *Std. Bias* | *Est.* | *Raw Bias* | *Std. Bias* |
| Group 1 | | | | | | | | | | |
| Loading 2 ($\lambda_2$) | 1.00 | 0.99 | −0.01 | −0.11 | 1.00 | 0.00 | −0.01 | 1.00 | 0.00 | −0.01 |
| Loading 3 ($\lambda_3$) | 1.00 | **0.95** | **−0.05** | **−0.79** | 1.00 | 0.00 | 0.02 | 1.00 | 0.00 | 0.02 |
| Intercept 2 ($\tau_2$) | 0.00 | **0.07** | **0.07** | **1.12** | 0.00 | 0.00 | −0.03 | 0.00 | 0.00 | −0.03 |
| Intercept 3 ($\tau_3$) | 0.00 | −0.01 | −0.01 | −0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Factor variance ($\phi$) | 0.75 | 0.76 | 0.01 | 0.07 | 0.75 | 0.00 | −0.03 | 0.75 | 0.00 | −0.02 |
| Residual 2 ($\theta_{22}$) | 0.25 | **0.34** | **0.09** | **2.81** | **0.23** | **−0.02** | **−0.81** | 0.25 | 0.00 | −0.08 |
| Residual 3 ($\theta_{33}$) | 0.25 | **0.27** | **0.02** | **0.58** | **0.23** | **−0.02** | **−0.81** | 0.25 | 0.00 | −0.03 |
| Group 2 | | | | | | | | | | |
| Loading 2 ($\lambda_2$) | 1.00 | 1.00 | 0.00 | −0.06 | 1.00 | 0.00 | −0.02 | 1.00 | 0.00 | −0.01 |
| Loading 3 ($\lambda_3$) | 0.50 | **0.54** | **0.04** | **0.93** | 0.50 | 0.00 | 0.00 | 0.50 | 0.00 | 0.00 |
| Intercept 2 ($\tau_2$) | 1.00 | **0.91** | **−0.09** | **−1.51** | 1.00 | 0.00 | −0.04 | 1.00 | 0.00 | −0.03 |
| Intercept 3 ($\tau_3$) | 0.00 | 0.00 | 0.00 | −0.02 | 0.00 | 0.00 | −0.05 | 0.00 | 0.00 | −0.05 |
| Latent mean ($\kappa$) | 0.00 | 0.01 | 0.01 | 0.18 | 0.00 | 0.00 | 0.06 | 0.00 | 0.00 | 0.06 |
| Factor variance ($\phi$) | 0.75 | 0.74 | −0.01 | −0.09 | 0.75 | 0.00 | 0.03 | 0.75 | 0.00 | 0.00 |
| Residual 2 ($\theta_{22}$) | 0.40 | **0.46** | **0.06** | **1.38** | 0.38 | −0.02 | −0.33 | 0.40 | 0.00 | 0.01 |
| Residual 3 ($\theta_{33}$) | 0.25 | **0.26** | **0.01** | **0.42** | 0.25 | 0.00 | 0.01 | 0.25 | 0.00 | −0.04 |

*Note*. Std. bias is standardized bias, or raw bias expressed in standard error units. Estimates with bias that exceeded 4/10 of a standard error unit are shown in bold type. The following parameters were fixed for identification purposes: $\lambda_1 = 1$ (both groups), $\kappa = 0$ (Group 1), $\tau_1 = 0$ (Group 2). PTI = product term imputation; SGI = separate group imputation.

model by fixing the loadings in a manner that is consistent with the specified growth form (e.g., a linear growth model with six repeated measures might be represented with loadings of 0, 1, 2, 3, 4, and 5). The linear model describes the average developmental trajectory using two latent variable means (e.g., the intercept factor mean, and the slope factor mean), and variation around the average growth curve is captured by the factor variance estimates. The basic model can be expanded to included higher-order growth forms (e.g., quadratic) by adding additional latent variables. It is beyond the scope of this article to provide a detailed overview of the latent growth curve model, but a number of excellent resources are available to interested readers (e.g., Bollen & Curran, 2006; Hancock & Lawrence, 2006; Singer & Willett, 2003).

In the longitudinal context it is frequently of interest to determine whether two or more groups have different developmental trajectories. A multiple group growth model is arguably the most flexible technique for this problem, as it allows for group comparisons on all model parameters (e.g., growth rates, intercept variance estimates, residual variance estimates). The model-implied covariance matrix for the multiple group LGM is identical to Equation 7, except that the elements in $\boldsymbol{\Lambda}_g$ are fixed at values that are consistent with the hypothesized growth form. Similarly, the model-implied mean vector is identical to Equation 8, except that the measurement intercepts in $\boldsymbol{\tau}_g$ are fixed at zero. Consistent with the CFA model, the $G$ covariance matrices will only be equal if the growth factor variances and residual variances are invariant across groups, and the $G$ mean vectors will only be the same if the growth factor means (i.e.,

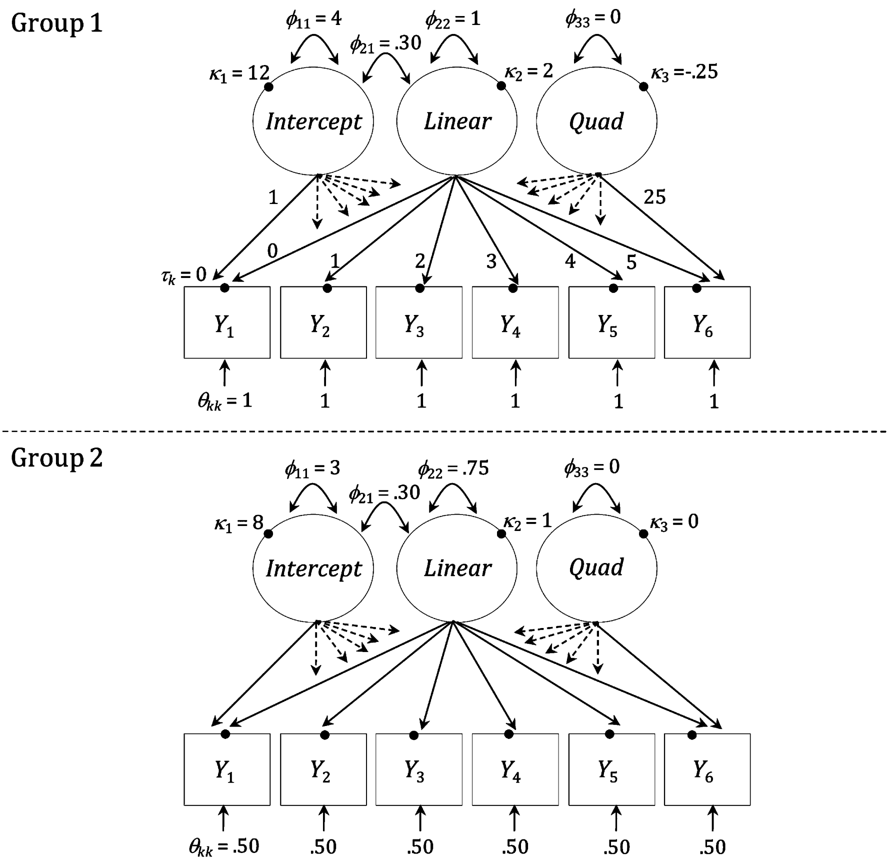


FIGURE 3 Multiple group latent growth curve model used in Study 3. The parameters for Group 2 were identical to those for Group 1, unless otherwise noted. The intercept and quadratic loadings are shown with dotted lines to reduce visual clutter. All intercept loadings were fixed at one, and the quadratic loadings were fixed at the squared values of the linear loadings. Note that the latent variable means and measurement intercepts are represented by black dots, and all measurement intercepts were fixed at zero. Finally, the covariances with the quadratic growth factor were fixed at zero and were omitted from the graphic because the variance of the latent variable is zero.

the elements in $\boldsymbol{\kappa}_g$) are identical across groups. As before, the imputation phase that precedes the LGM analysis should preserve group differences in both the mean and covariance structure.

To illustrate, consider a two-group LGM with six repeated measures, a graphic of which is shown in Figure 3. As seen in Figure 3, the average growth trajectory for the first group is described by a quadratic trend where scores increase and level off over time (i.e., the linear growth factor mean is positive, and the quadratic growth factor has a negative mean). The second group has a slightly lower intercept and has a trajectory that is flatter and linear in shape (i.e., the quadratic growth factor has a mean of zero). Finally, both groups have the same degree of heterogeneity in the intercepts and slopes, but the residual variance parameters for

TABLE 3
Average Estimates from the Latent Growth Model Simulation ($n$ = 250, 15% Missing Data)

| | | *Model Vars* | | | *PTI* | | | *SGI* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Population Parameter* | | *Est.* | *Raw Bias* | *Std. Bias* | *Est.* | *Raw Bias* | *Std. Bias* | *Est.* | *Raw Bias* | *Std. Bias* |
| Group 1 | | | | | | | | | | |
| Intercept mean ($\kappa_1$) | 12.00 | 12.00 | 0.00 | −0.02 | 11.99 | −0.01 | −0.06 | 11.99 | −0.01 | −0.06 |
| Linear mean ($\kappa_2$) | 2.00 | **1.96** | **−0.04** | **−0.49** | 2.00 | 0.00 | −0.01 | 2.00 | 0.00 | −0.01 |
| Quadratic mean ($\kappa_3$) | −0.25 | **−0.24** | **0.01** | **0.93** | −0.25 | 0.00 | 0.02 | −0.25 | 0.00 | 0.02 |
| Intercept variance ($\phi_{11}$) | 4.00 | 4.13 | 0.13 | 0.32 | 3.94 | −0.06 | −0.15 | 3.96 | −0.04 | −0.10 |
| Linear variance ($\phi_{22}$) | 1.00 | 1.00 | 0.00 | 0.02 | 0.99 | −0.01 | −0.07 | 1.00 | 0.00 | −0.02 |
| Covariance ($\phi_{21}$) | 0.30 | 0.26 | −0.04 | −0.28 | 0.31 | 0.01 | 0.04 | 0.30 | 0.00 | −0.02 |
| Residual ($\theta_k$) | 1.00 | **0.98** | **−0.03** | **−0.53** | **0.90** | **−0.10** | **−2.08** | 1.01 | 0.01 | 0.20 |
| Group 2 | | | | | | | | | | |
| Intercept mean ($\kappa_1$) | 8.00 | 7.99 | −0.01 | −0.07 | 8.00 | 0.00 | −0.02 | 8.00 | 0.00 | −0.02 |
| Linear mean ($\kappa_2$) | 1.00 | **1.04** | **0.04** | **0.64** | 1.00 | 0.00 | 0.04 | 1.00 | 0.00 | 0.04 |
| Quadratic mean ($\kappa_3$) | 0.00 | **−0.01** | **−0.01** | **−1.26** | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.04 |
| Intercept variance ($\phi_{11}$) | 3.00 | 3.09 | 0.09 | 0.30 | 3.03 | 0.03 | 0.08 | 3.00 | 0.00 | 0.00 |
| Linear variance ($\phi_{22}$) | 0.75 | 0.75 | 0.00 | −0.03 | 0.75 | 0.00 | 0.04 | 0.75 | 0.00 | −0.01 |
| Covariance ($\phi_{21}$) | 0.30 | 0.28 | −0.02 | −0.16 | 0.30 | 0.00 | −0.04 | 0.30 | 0.00 | 0.04 |
| Residual ($\theta_k$) | 0.50 | **0.58** | **0.08** | **3.15** | **0.55** | **0.05** | **1.92** | 0.50 | 0.00 | 0.13 |

*Note*. Std. bias is standardized bias, or raw bias expressed in standard error units. Estimates with bias that exceeded 4/10 of a standard error unit are shown in bold type. PTI = product term imputation; SGI = separate group imputation.

the second group are smaller in magnitude, owing to the fact that the individual trajectories fit a straight line reasonably well.

Interestingly, the process of implementing PTI and SGI is identical to the CFA model described earlier. This might seem somewhat counterintuitive given that the first group has a quadratic growth trajectory, but a powered term is not required in the imputation model. Each of the repeated measure variables is imputed from a different regression equation, so the fact that each equation has a different intercept term implies that the means are free to vary at each time point. The fact that the two groups have different trajectory shapes also has no bearing on the imputation process. SGI takes this into account using a separate imputation phase for each group, and PTI preserves the "group by time" interaction via the regression intercept and the coefficient attached to the code variable.

A simulation study was conducted based on the population model shown in Figure 3, and the mean estimates for selected parameters are given in Table 3. For the sake of brevity, results from the $n = 250$ and 15% missing data rate condition are given in Table 3, and the corresponding results from the $n = 50$ condition are available on request. To facilitate interpretation, the parameter estimates that produced standardized bias values larger than .40 (i.e., 4/10 of a standard error unit) are shown in bold type.

The LGM simulation results were largely the same as those from the CFA simulation. Specifically, the imputation model based solely on the six analysis variables produced biased parameter estimates, and the direction of the bias attenuated group differences. As before, PTI produced very accurate estimates of mean structure parameters, and the bias in the covariance structure was isolated to the residual variance terms. SGI again produced estimates that were quite accurate, even with a group size of $n = 50$.

## A CAUTIONARY NOTE ON LATENT CATEGORICAL VARIABLES

Thus far we have only considered models where group membership is observed. However, there are a number of popular statistical models that treat group membership as a latent categorical variable—finite mixtures (Lubke & Muthén, 2005; McLachlan & Peel, 2000; Muthén, 2001, 2004) and latent class models (McCutcheon, 1987) are two such examples. These models are important to consider because they are becoming increasingly common in the psychological and behavioral science literature. These models are relevant to this study because they are essentially multiple group models where group membership must be inferred from the data. This means that it is not possible to use PTI or SGI (or any standard multiple imputation scheme) to preserve group differences in the mean and covariance structure across the hidden latent classes. We include a discussion of mixture models to underscore the point that multiple imputation is inappropriate for these analyses.

Based on the previous simulation results, it is reasonable to expect that multiple imputation will yield biased parameter estimates in models where the moderator variable is a latent categorical variable. To examine this possibility, the data from the CFA and LGM simulations were reanalyzed using a factor mixture and growth mixture model, respectively. Factor and growth mixture models are conceptually similar to their CFA and LGM counterparts. Like multiple group models, mixture model parameter estimates can vary across latent classes. The key difference is that the number of latent classes is typically unknown in advance, and must be inferred from the data in a stepwise fashion that mimics the extraction of factors in an exploratory factor analysis (Tofighi & Enders, 2007). A more detailed description of factor mixtures can be found in Lubke and Muthén (2005), and Muthén (2001, 2004) discussed growth mixture models.

The factor mixture and growth mixture simulations used the raw data from the previous CFA and LGM studies. Although the data were generated from two known groups, the subsequent analyses treated group membership as a latent class variable in the mixture modeling framework implemented in M*plus*. The models are identical to the CFA and LGM but contain an additional parameter (a latent class intercept, or logit) that quantifies the relative probability of membership in the two classes. It is beyond the scope of this article to investigate how imputation affects the class enumeration process, so the correctly specified two-class mixture was fit to each sample replicate, and the population parameter values were used as starting values. The imputer would have no a priori knowledge of class membership, so the imputation phase was carried out using only the manifest variables.

The mixture simulations were more limited in scope, and only included the $n = 250$ and 15% missing data rate conditions. The raw bias values from the mixture simulations were quite comparable to those in Tables 2 and 3 where the imputation model was based solely on the analysis model variables, so there is no need to present further tabular results (interested readers can contact the first author for a copy of the tables). The severely biased estimates underscore the point that standard multiple imputation schemes should be avoided altogether when estimating latent class or mixture models. Fortunately, mixture models with missing data can generally be estimated using maximum likelihood.

## DISCUSSION

Multiple imputation is currently viewed as a “state of the art” missing data technique (Schafer & Graham, 2002), and is one of two procedures recommended in the methodological literature (maximum likelihood being the other). The selection of variables for the imputation phase is one of the nuances associated with implementing multiple imputation, because the imputer must take special care to preserve any associations or special features of the data that will be modeled in the subsequent analysis. When using multiple imputation with multiple group models, failing to preserve the interactive effects during the imputation phase can produce biased parameter estimates in the subsequent analysis phase, even when the data are MCAR or MAR. This study outlines two simple imputation strategies introduced in the literature (PTI and SGI) that are applicable to multiple group models and that can be implemented without the need for specialized software.

The PTI approach preserves group differences in the mean and covariance structure using a set of code and product variables in the imputation model. From an analytical perspective, PTI is potentially problematic because it implicitly uses a homoscedastic model to impute the missing values. In some sense, the simulations in this article were “rigged” against PTI because the models always implied group differences in the population variances. When such differences do exist, the simulation results indicated that the PTI model will not preserve the covariance structure of the data, and can yield biased estimates of certain model parameters. At a minimum, residual variance estimates were distorted, but the bias also propagated into parameters that are of key substantive interest in the observed variable model (e.g., the $R^2$ values).

To put the potential for bias in a practical perspective, we used the population parameters to derive the model-implied covariance matrices from the three simulations and then used the group-specific variances to form F-max statistics (i.e., the ratio of the largest to the smallest group variance) for each manifest variable. In the mediation model, the F-max values were 1.66 and 2.00 for $Y$ and $M$, respectively (the variances of $X$ were the same for both groups). The group differences in the variances were somewhat smaller in the latent variable models. Specifically, the CFA model had F-max values between 1.10 and 1.13, and the LGM had F-max values between 1.31 and 1.43. From a practical standpoint, many researchers would consider the differences in these variances to be relatively small and well within the range of differences that would routinely appear in applied research. Interestingly, failing to preserve these relatively small variance differences during the imputation process can lead to substantial bias in the multiple group model parameter estimates. This bias could readily lead a researcher to conclude that certain parameters are invariant across groups when, in fact, they are not.

In contrast, implementing SGI turned out to be a highly effective strategy. Unlike the PTI strategy, SGI places no restrictions on the structure of the imputation model, so all components of the mean and covariance structure are allowed to vary across groups. SGI has the added benefit of being easy to implement, and does not require the use of potentially cumbersome code and product variables. The simulation results unequivocally favored SGI over PTI, and the procedure performed surprisingly well with group sizes as low as $n = 50$. A group size of 50 is probably small enough to make many people think twice about estimating a CFA or LGM, so these simulation results are encouraging (note that I would not advocate an $n = 50$ rule of thumb based on these results, however). However, it is important to reiterate that SGI requires a complete grouping variable (this would virtually always be the case in a multiple group model) and is only applicable to the situation where one of the interacting variables is categorical.

Although the simulation models used in this study were relatively straightforward, it is reasonable to expect that PTI will yield biased estimates in more complex models, as increasing the number of variables would almost certainly increase the likelihood that the variances are not homogeneous across groups. To be fair, PTI works quite well when the population variances are equal across groups. Although not shown in this article, a preliminary set of simulations showed no material differences between SGI and PTI when the population variances are equal. However, the potential for bias suggests that researchers should avoid the PTI strategy, particularly when a less awkward and more accurate procedure is available.

In sum, researchers who are interested in examining moderation effects using multiple group models need to specify an imputation model that appropriately preserves the interactive effects that are potentially present in the data. The simulation studies presented in this article suggest that interactive effects can be preserved by performing the imputation separate for each group, and researchers are encouraged to adopt this strategy when formulating imputation models for multiple group analyses. Finally, researchers should avoid using standard multiple imputation schemes (e.g., PROC MI in SAS) when the model includes a latent categorical variable, as there is no way to preserve between-class differences when group membership is an unobserved variable.

## REFERENCES

Allison, P. D. (2002). *Missing data.* Newbury Park, CA: Sage.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates.

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling, 15,* 651–675.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach.* New York: Wiley.

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

Enders, C. K. (2001). The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods, 6,* 352–370.

Enders, C. K. (2006a). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *A second course in structural equation modeling* (pp. 313–344). Greenwich, CT: Information Age.

Enders, C. K. (2006b). A primer on the use of modern missing data methods in psychosomatic medicine research. *Psychosomatic Medicine, 68,* 427–436.

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

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8,* 430–457.

Graham, J. W. (2003). Adding missing-data relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10,* 80–100.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8,* 206–213.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1–29). Thousand Oaks, CA: Sage.

Hancock, G. R., & Lawrence, F. R. (2006). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 171–196). Greenwood, CT: Information Age.

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review, 95*(1), 49–69.

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

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10,* 21–39.

MacKinnon, D. P. (2007). *Introduction to statistical mediation analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

McCutcheon, A. L. (1987). *Latent class analysis.* Newbury Park, CA: Sage.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, L. K., & Muthén, B. O. (1998–2010). *M*plus *user's guide* (6th ed.). Los Angeles: Muthén & Muthén.

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

Rubin, D. B. (1987). *Multiple imputation for non-response in surveys.* New York: Wiley.

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

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

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

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal analysis.* New York: Oxford University Press.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6,* 317–329.

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in a growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Greenwich, CT: Information Age.

von Hippel, P. T. (2009). How to impute interactions, squares, and other transformed variables. *Sociological Methodology, 39,* 265–291.

## APPENDIX A
## SAS MULTIPLE IMPUTATION PROGRAM FOR SGI PROCEDURE

```
datacfa;
infile ‘c:\CFA\rep1. dat’;
input y1 y2 y3 y4 y5 y6 group;
/* CHANGE MISSING CODE 999 TO . */
array y[6] y1 y2 y3 y4 y5 y6;
do i = 1 to 6;
if y[i] = 999 then y[i] = . ;
end;
drop i;
/* SORT BY GROUP */
procsort data = cfa;
by group;
/* PROC MI */
```

```
procmi data = cfa out = imputedcfanimpute = 10;
var y1 y2 y3 y4 y5 y6;
mcmcnbiter = 200 niter = 200;
by group;
/* OUTPUT STACKED FILE TO 10 SEPARATE FILES */
%macroimpute;
%do j = 1 %to 10;
dataoutmi;
setimputedcfa (where = (_Imputation_ = &j));
file "c:\Imputations\imp&j..dat";
put @01 (y1) (8.6)
    @11 (y2) (8.6)
    @21 (y3) (8.6)
    @31 (y4) (8.6)
    @41 (y5) (8.6)
    @51 (y6) (8.6)
    @61 (group) (3.0);
%end;
%mend;
%impute;
run;
```

## APPENDIX B
## M*PLUS* PROGRAM FOR ANALYZING MULTIPLY IMPUTED DATA SETS

```
data:
file = ‘c:\Imputations\replist.dat’;
type = imputation;
variable:
names = y1 y2 y3 y4 y5 y6 group;
grouping = group (1 = g1 2 = g2);
analysis:
type = meanstructure;
model:
f1 by y1@1;
f1 by y2 - y6*1;
   f1*.75;
   y1 - y6*.25;
   [y1 - y6*0];
   [f1@0];
model g2:
f1 by y2*1;
f1 by y4 - y6*1;
f1 by y3*.5;
   [y2 - y6*0];
   [y2*1];
   [f1*0];
   [y1@0];
output:
standardized;
```