AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

2021, Vol. 26, No. 4, 466–485
https://doi.org/10.1037/met0000381

A Comparison of Full Information Maximum Likelihood and Multiple Imputation in Structural Equation Modeling With Missing Data

Taehun Lee[1] and Dexin Shi[2]
[1] Department of Psychology, Chung-Ang University
[2] Department of Psychology, University of South Carolina

Abstract

This article compares two missing data procedures, full information maximum likelihood (FIML) and multiple imputation (MI), to investigate their relative performances in relation to the results from analyses of the original *complete data* or the hypothetical data available before missingness occurred. By expressing the FIML estimator as a special MI estimator, we predicted the expected patterns of discrepancy between the two estimators. Via Monte Carlo simulation studies where we have access to the original *complete data*, we compare the performance of FIML and MI estimators to that of the *complete data* maximum likelihood (ML) estimator under a wide range of conditions, including differences in sample size, percent of missingness, and degrees of model misfit. Our study confirmed well-known knowledge that the two estimators tend to yield essentially equivalent results to each other and to those from analysis of *complete data* when the postulated model is correctly specified. However, some noteworthy patterns of discrepancies were found between the FIML and MI estimators when the hypothesized model does not hold exactly in the population: MI-based parameter estimates, comparative fit index (CFI), and the Tucker Lewis index (TLI) tend to be closer to the counterparts of the *complete data* ML estimates, whereas FIML-based chi-squares and root mean square error of approximation (RMSEA) tend to be closer to the counterparts of the *complete data* ML estimates. We explained the observed patterns of discrepancy between the two estimators as a function of the interplay between the parsimony and accuracy of the imputation model. We concluded by discussing practical and methodological implications and issues for further research.

Translational Abstract

In this article, two classes of modern missing data procedures, maximum likelihood (ML) and multiple imputation (MI), are systematically compared. Although it has been argued that the two classes of missing data procedures are essentially equivalent, we showed that they may not produce equivalent results as practiced empirically under realistic conditions where researchers work with imperfect models. Following a review of relevant estimation theory, we have made specific a priori predictions on the expected patterns of the discrepancy between FIML and MI estimators, with respect to parameter estimates, their associated sampling variabilities, and the goodness of fit indices. Via Monte Carlo simulation studies where we have access to the original *complete data*, we showed that the two classes of procedures exhibit subtle but important differences in their performance in relation to the results from *complete data* analysis. Based on our theoretical predictions and the observed patterns, we provided a few practical recommendations and directions for future research.

Keywords: full information maximum likelihood (FIML), multiple imputation (MI), structural equation modeling (SEM), model misspecification, average relative increase in variance (ARIV)

Supplemental materials: https://doi.org/10.1037/met0000381.supp

Incomplete data can occur for many reasons. Surveys may be incomplete because some respondents refused to answer certain questions because of fears that their anonymity would not be protected. Some respondents might not answer questions near the end of a long survey because of slow reading. Data values might simply be lost during the data collection or data storage process because of equipment malfunctions (Graham, 2012). In planned missing data designs such as matrix sampling designs, some ques-

This article was published Online First January 28, 2021.

Taehun Lee https://orcid.org/0000-0001-8261-701X

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) [2017R1C1B2012424] and the Advanced Support for Innovative Research Excellence Grant [13580-17-44758] funded by the Office of the Vice President for Research at the University of South Carolina.

Correspondence concerning this article should be addressed to Taehun Lee, Department of Psychology, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul 157-756, Korea. Email: lee.taehun@gmail.com

tions are never asked in the first place to reduce the burden on respondents (Graham et al., 2006). For example, in large-scale educational assessments, such as the National Assessment of Educational Progress, not every test item is administered to every student (Kaplan, 1995), yielding an incomplete data matrix.

In applications of structural equation modeling (SEM), it would be fair to say that virtually every data set has some missing data, rendering missing data the norm rather than the exception. According to a study that examined the prevalence of missing data in psychological research, missing data were found in about two thirds of studies in a random sample of empirical research journal articles from the PsycINFO database for the year 1999 (Bodner, 2006). Peng et al. (2006) surveyed 11 major education and psychological journals and found that 48% of the articles involved missing data and about 16% of articles did not clearly report missing data information.

A variety of methods for handling missing data are available, but in the context of SEM, the most widely available and arguably the most popular method for addressing missing data is the full information maximum likelihood (FIML; Allison, 1987; Arbuckle, 1996; Finkbeiner, 1979; Wilks, 1932). Multiple imputation (MI) procedures, first proposed in Rubin (1976, 1987), have recently begun to emerge as a viable alternative solution among applied SEM researchers (Lee & Cai, 2012; Little & Rubin, 2002).

In practical applications of SEM, the FIML and MI procedures have now become the standard methods for handling missing data and estimating the structural parameters of interest specified in the model. MI takes a two-step approach, first constructing a database by replacing nonresponses with simulated values and then fitting the postulated model to the constructed database (Little & Rubin, 2002; Schafer, 1997). The database constructors and the ultimate users can be distinct entities (Rubin, 1996). By contrast, FIML allows the researcher to handle missing data and the model estimation all in a single step (Graham, 2009). Applied researchers using SEM may find it easier to implement FIML, especially because of the all-in-one-step property of FIML. In addition, incorporating auxiliary information[1] to improve the performance of the estimator in terms of power, bias, and efficiency (Graham, 2003), which has long been straightforward in MI, is now automated for FIML in the leading SEM software such as Mplus (Muthén & Muthén, 1998–2012) and the lavaan package in R (R Core Team, 2018). However, endeavors for enhancing users' convenience have also been undertaken on the MI side as well. For example, Mplus and the lavaan package in R offer simple options for researchers for automating the two steps and producing the outputs in a single step. Therefore, from a practical viewpoint, it appears that there is currently no reason to prefer one method over the other.

From a purely statistical point of view, it has been said that there is no theoretical reason to prefer MI to FIML or vice versa, because FIML and MI procedures will produce essentially equivalent results when implemented in comparable ways, in the sense defined by Collins et al. (2001) (e.g., using the same set of input data, under equivalent distributional assumptions to the variables, and with a sufficiently large number of imputations). In fact, perhaps due in part to this widely accepted view among SEM researchers, there has been very little empirical research comparing relative performance between FIML and MI procedures in SEM. Rather, empirical studies have focused on either FIML or MI, and compared their performance with simple ad hoc methods (e.g., Enders, 2001; Enders & Bandalos, 2001; Graham & Schafer, 1999). One important exception was the work by Enders and Mansolf (2018), which compared MI and FIML estimators with respect to their fit assessment of structural equation models. However, the focus of Enders and Mansolf (2018) was to investigate whether MI-based chi-square statistics and other practical fit indexes are well calibrated to FIML-based statistics.

Purposes of This Article

A primary objective of the p resent study is to compare FIML and MI in the context of SEM, to determine whether or not the two procedures exhibit noticeable differences in their performances (i.e., parameter estimates, sampling variability of the estimates, and model fit indexes) in realistic settings that applied SEM researchers routinely encounter. In the present study, the two procedures are to be compared in comparable ways in the sense defined by Collins et al. (2001), but under a more realistic condition than in previous studies that have compared various missing data procedures: that in which the model under consideration does not hold exactly in the population (Box, 1979; MacCallum, 2003; Shi et al., 2018).

In this article, we investigated which procedure, FIML or MI, tends to produce results closer to those from a hypothetical *complete data* analysis. We investigated relative performances of FIML and MI estimators in relation to the results from *complete data* analysis because the latter is the analysis planned at the outset (Longford, 2006, p. 54). Moreover, when the data model under consideration does not hold in the population, there may not be parameters corresponding to those specified in the posited model, which in turn makes it even more natural to interpret the results in relation to the results from *complete data* analysis. In this article, by expressing the FIML estimator as a special MI estimator, we could make general theoretical predictions about the expected patterns of discrepancy between the two estimators in relation to the results from the *complete data* analysis, which were confirmed via both subsequent Monte Carlo simulation studies and empirical data analysis. Therefore, a unique contribution of the present article is that it provides both theoretical and empirical evidence concerning which procedure, FIML or MI, would tend to produce results closer to those from the analysis of *complete data* with respect to parameter estimates, sampling variability of the estimates, and model fit statistics. Based on our findings, we discussed practical implications and issues for further research.

Some Theoretical Basis

In this article we consider an $N \times p$ *complete data* matrix Y where N and p represent the sample size and the number of manifest variables, respectively. Then the i-th row of Y, denoted by, Y_i^T, is assumed to be a vector of observations drawn from a p-dimensional multivariate normal distribution with mean vector μ_0 and covariance matrix Σ_0. In SEM, it is assumed that there exists a $q \times 1$ vector ϑ_0 in the parameter space θ such that $\mu_0 =$

[1] Auxiliary information comes from variables that are not part of the main analysis model but are used to improve the performance of the missing data estimator (Collins et al., 2001).

$\mu(\vartheta_0)$ and $\Sigma_0 = \Sigma(\vartheta_0)$, where $\mu(\vartheta_0)$ and $\Sigma(\vartheta_0)$ represent the model-implied mean vector and the model-implied covariance structure matrix, respectively.

If there are no missing data in Y, under multivariate normality, we can obtain the ML estimates for ϑ_0 by maximizing the *complete data* log-likelihood $\log L(\vartheta \mid Y)$, which is defined as

$$\log L(\vartheta \mid Y) = -\frac{N}{2}\{\ln|\Sigma(\vartheta)| + tr[S\Sigma^{-1}(\vartheta)] + [\bar{Y} - \mu(\vartheta)]^T\Sigma^{-1}(\vartheta)[\bar{Y} - \mu(\vartheta)]\}, \quad (1)$$

where $\bar{Y}$ and S represent the sample mean vector and sample covariance matrix, respectively. Thus, the ML estimator can be formally defined as

$$\hat{\vartheta}_{ML} = \arg\max_{\vartheta \in \Theta} \log L(\vartheta \mid Y) \quad (2)$$

Under appropriate conditions, it is known that $\hat{\vartheta}_{ML}$ is consistent, asymptotically normal, and asymptotically efficient. The asymptotic covariance matrix for $\hat{\vartheta}_{ML}$ can be obtained from the inverse of the *complete data* observed information matrix

$$\hat{V}_{ML} = -\left[\frac{\partial^2 \log L(\vartheta \mid Y)}{\partial \vartheta\, \partial \vartheta'}\right]^{-1}_{\vartheta = \hat{\vartheta}_{ML}}, \quad (3)$$

which is the inverse of the negative second order partial derivative matrix of the *complete data* log-likelihood function. Let the saturated estimate of the mean vector be denoted by $\hat{\mu}$ and the saturated covariance matrix be $\hat{\Sigma}$. Then, with $\hat{\omega}_{ML} = (vec(\hat{\mu})', vech(\hat{\Sigma})')'$ the statistic

$$T_{ML} = -2[\log L(\hat{\vartheta}_{ML} \mid Y) - \log L(\hat{\omega}_{ML} \mid Y)], \quad (4)$$

follows asymptotically a central chi-square distribution with $d = p(p + 1)/2 + p - q$ degrees of freedom if the model is correctly specified.

When only part of the *complete data* Y is observed for various reasons, none of the *complete data* statistics given by Equations 2, 3, or 4 can be calculated. Let Y_i be partitioned as $Y_i = (Y_{i,obs}, Y_{i,mis}')$, where $Y_{i,obs}$ and $Y_{i,mis}$ denote the observed and missing part of Y_i, respectively. Assuming that the missing data mechanism is ignorable,[2] inferences about ϑ can be drawn based on the *observed-data* log-likelihood function given as

$$\log L(\vartheta \mid Y_{obs}) = \sum_{i=1}^{N} \log L_i(\vartheta \mid Y_{i,obs}), \quad (5)$$

where Y_{obs} is the set of observed values of the entire sample and $L_i(\vartheta \mid Y_{i,obs}) = \int L_i(\vartheta \mid Y_{i,obs}, Y_{i,mis}) dY_{i,mis}$ represent individual likelihoods. The integral in the individual likelihoods is taken with respect to the joint marginal distribution of the missing data. In SEM with normally distributed variables, the closed form of $\log L(\vartheta \mid Y_{obs})$ can be derived using the standard probability theory, which can be written as follows:

$$\log L(\vartheta \mid Y_{obs}) = \sum_{i=1}^{N} \log L_i(\vartheta \mid Y_{i,obs}) = \sum_{i=1}^{N} -\frac{1}{2}\{\ln|\Sigma_i(\vartheta)| + [Y_{i,obs} - \mu_i(\vartheta)]^T\Sigma_i^{-1}(\vartheta)[Y_{i,obs} - \mu_i(\vartheta)]\}, \quad (6)$$

where $\mu_i(\vartheta)$ and $\Sigma_i(\vartheta)$ represent the subvector of the model-implied mean vector and the submatrix of the model-implied covariance matrix corresponding to $Y_{i,obs}$ (see Arbuckle, 1996 for more details). Just as the *complete data* inference, a consistent and efficient estimate of ϑ_0 and its associated variance-covariance matrix can be found from the minimizer of $\log L(\vartheta \mid Y_{obs})$ and from the inverse of the negative second derivative of $\log L(\vartheta \mid Y_{obs})$, respectively (Kenward & Molenberghs, 1998; Little & Rubin, 2002; Savalei, 2010). In SEM, the inferences about ϑ_0 based on the observed-data log-likelihood has been known as the FIML (Allison, 1987; Arbuckle, 1996; Finkbeiner, 1979; Wilks, 1932). Therefore, we can formally define the FIML estimator as

$$\hat{\vartheta}_{FIML} = \arg\max_{\vartheta \in \Theta} \log L(\vartheta \mid Y_{obs}), \quad (7)$$

and

$$\hat{V}_{FIML} = -\left[\frac{\partial^2 \log L(\vartheta \mid Y_{obs})}{\partial \vartheta\, \partial \vartheta'}\right]^{-1}_{\vartheta = \hat{\vartheta}_{FIML}}. \quad (8)$$

The *observed data* test statistic corresponding to Equation 4, which we denote T_{FIML}, can be obtained by substituting $\hat{\vartheta}_{FIML}$ and $\hat{\omega}_{FIML}$ into Equation 4. In large samples, T_{FIML} is distributed as a chi-square variable with d degrees of freedom.

As an alternative to maximizing the *observed-data* log-likelihood obtained by integrating the missing data out over the joint distribution of the missing data (as in FIML), inferences about ϑ can be based on the *completed data* where the missing entries of the incomplete data sets are multiply imputed or filled in with multiple sets of draws from the posterior predictive distribution. In the most commonly adopted multiple imputation (MI; Rubin, 1987) method, the missing data, Y_{mis}, are replaced by M sets of draws from the posterior predictive distribution of the missing data, $\{Y_{mis}^{(l)}, l = 1, 2, \ldots, M\}$, resulting in M sets of *completed data*. In this article we will denote the *completed data* sets as follows:

$$\{Y_*^{(l)}; l = 1, \ldots, M\} = \{(Y_{obs}, Y_{mis}^{(l)}); l = 1, \ldots, M\}$$

where $Y_{mis}^{(l)} \sim p(Y_{mis} \mid Y_{obs})$. The M sets of *completed data* are then analyzed by the standard *complete data* statistical methods, producing M sets of parameter estimates, $\hat{\vartheta}_{*1}, \hat{\vartheta}_{*2}, \ldots, \hat{\vartheta}_{*M}$, and the associated variance-covariance matrices, $\hat{V}_{*1}, \hat{V}_{*2}, \ldots, \hat{V}_{*M}$. The ML estimator of the parameter vector and its asymptotic variance-covariance matrix for the lth imputed data set can be formally defined as

$$\hat{\vartheta}_{*l} = \arg\max_{\vartheta \in \Theta} \log L(\vartheta \mid Y_*^{(l)}) \quad (9)$$

and

$$\hat{V}_{*l} = -\left[\frac{\partial^2 \log L(\vartheta \mid Y_*^{(l)})}{\partial \vartheta\, \partial \vartheta'}\right]^{-1}_{\vartheta = \hat{\vartheta}_{*l}} \quad (10)$$

[2] The ignorability of the missing-data mechanism involves two conditions: missing at random (MAR) and distinctness of parameters (Little & Rubin, 2002; Rubin, 1987). An informal definition of MAR is that the probability that an observation is missing depends only on some of the observed data but not on the missing data. The distinctness of parameters condition is that the parameters of the data model and the parameters of the missing data mechanism are distinct. If both MAR and distinctness hold true, then the missing-data mechanism is said to be ignorable.

It should be noted that $\log L(\vartheta \mid Y_*^{(l)})$ takes the same form as the *complete data* log-likelihood function in Equation 1 except the fact that $\bar{Y}$ and S are replaced by $\bar{Y}_*^{(l)}$ and $S_*^{(l)}$ that are computed based on the l-th imputed dataset $Y_*^{(l)}$. The missing data are typically generated from the posterior predictive distributions $p(Y_{mis} \mid Y_{obs})$ using the data augmentation algorithm by Tanner and Wong (1987).[3]

According to "Rubin's rule" (Rubin, 1987, Chapter 3), the MI estimator for ϑ can be based on the arithmetic average of $\{\hat{\vartheta}_{*l}, l = 1, 2, \ldots, M\}$ and the associated asymptotic variance-covariance matrix can be obtained from a combination of the arithmetic average of $\{\hat{V}_{*l}, l = 1, 2, \ldots, M\}$ and the variance of $\{\hat{\vartheta}_{*l}, l = 1, 2, \ldots, M\}$. Let $\hat{\vartheta}_{\mathrm{MI}}$ and $\hat{V}_{\mathrm{MI}}$ be the MI point estimator and its asymptotic variance-covariance matrix. They can then be formally defined as

$$\hat{\vartheta}_{MI} = \frac{1}{M}\sum_{l=1}^{M} \hat{\vartheta}_{*l} \tag{11}$$

and

$$\hat{V}_{MI} = \bar{V}_* + (1 + M^{-1})B_M \tag{12}$$

with

$$\bar{V}_* = \frac{1}{M}\sum_{l=1}^{M} \hat{V}_{*l}$$

and

$$B_M = \frac{1}{M-1}\sum_{l=1}^{M} (\hat{\vartheta}_{*l} - \hat{\vartheta}_{MI})(\hat{\vartheta}_{*l} - \hat{\vartheta}_{MI}).$$

Following the pooling procedure proposed in Meng and Rubin (1992), the overall goodness-of-fit testing can be conducted using the test statistic, T_{MI}, given as

$$T_{MI} = \frac{\tilde{T}}{1 + r_3}, \tag{13}$$

with

$$r_3 = \frac{M+1}{d(M-1)}(\bar{T} - \tilde{T}) \tag{14}$$

$$\tilde{T} = \frac{1}{M}\sum_{l=1}^{M} -2\left[\log L(\hat{\vartheta}_{MI} \mid Y_*^{(l)}) - \log L(\hat{\omega}_{MI} \mid Y_*^{(l)})\right] \tag{15}$$

and

$$\bar{T} = \frac{1}{M}\sum_{l=1}^{M} -2\left[\log L(\hat{\vartheta}_{*l} \mid Y_*^{(l)}) - \log L(\hat{\omega}_{*l} \mid Y_*^{(l)})\right]. \tag{16}$$

The two likelihood ratio statistics in Equations 15 and 16 represent the average of the likelihood ratio statistics evaluated at the pooled estimates and the average of the likelihood ratio statistics evaluated at the imputation-specific estimates, respectively. Conceptually, $\tilde{T}$ is a direct counterpart to the *complete data* likelihood ratio statistic in Equation 4, which is evaluated at $\hat{\vartheta}_{MI}$ and averaged across *completed data* sufficient statistics.[4] The pooled likelihood ratio statistic in Equation 13 is constructed so as to adjust $\tilde{T}$ for its uncertainty due to missing data using r_3 in Equation 14, which Rubin (1987) calls the average relative increase in variance (ARIV) due to nonresponse (Meng & Rubin, 1992; see Appendix for more details). The p value associated with T_{MI} can be obtained by referencing the observed value of the pooled test statistic to a chi-square distribution with d degrees of freedom (Asparouhov & Muthén, 2010) or by referencing T_{MI}/d to F distribution (Meng & Rubin, 1992).

General Expectations

For the purpose of comparing the FIML estimator and MI estimator and their relationships with the standard ML estimator, it would be advantageous to give an alternative definition of the FIML estimator as follows:

$$\tilde{\vartheta}_{FIML} = \frac{1}{M}\sum_{l=1}^{M} arg\ \max_{\vartheta \in \Theta} \sum_{i=1}^{N} \log L_i(\vartheta \mid Y_{i,obs},\ Y_{i,mis}^{(l)}) \tag{17}$$

where $\{Y_{i,mis}^{(l)}; l = 1, \ldots, M\}$ are drawn from $p(Y_{mis} \mid Y_{obs}, \hat{\vartheta}_{FIML})$, which is a multivariate normal distribution with the mean and the variance given as

$$\mathrm{E}(Y_{i,mis} \mid Y_{i,obs},\ \hat{\vartheta}_{FIML}) = \hat{\mu}_{mis} + \hat{\Sigma}_{mo}\hat{\Sigma}_{oo}^{-1}(Y_{i,obs} - \hat{\mu}_{obs})$$

$$\mathrm{Var}(Y_{mis} \mid Y_{obs},\ \hat{\vartheta}_{FIML}) = \hat{\Sigma}_{mm} - \hat{\Sigma}_{mo}\hat{\Sigma}_{oo}^{-1}\hat{\Sigma}_{om}$$

where $\hat{\mu}_{mis}$ and $\hat{\mu}_{obs}$ contain the elements of the model-implied mean vector, $\mu(\hat{\vartheta}_{FIML})$ corresponding to missing and observed values, respectively, and $\hat{\Sigma}_{mm}$ $\hat{\Sigma}_{oo}$ and $\hat{\Sigma}_{mo}$ contain rows and columns of the model-implied variance-covariance matrix, $\Sigma(\hat{\vartheta}_{FIML})$ corresponding to observed and missing values, respectively (Arbuckle, 1996, p. 274). Therefore, at a conceptual level, FIML can also be understood as an improper multiple imputation (Little & Rubin, 2002, Chapter 10; Rubin, 1987, Chapter 3) where plausible values of missing data are generated from $p(Y_{mis} \mid Y_{obs}, \hat{\vartheta}_{FIML})$[5].

On the other hand, the MI estimator can be rewritten as

$$\hat{\vartheta}_{MI} = \frac{1}{M}\sum_{l=1}^{M} arg\ \max_{\vartheta \in E} \sum_{i=1}^{N} \log L_i(\vartheta \mid Y_{i,obs},\ Y_{i,mis}^{(l)}) \tag{18}$$

where $\{Y_{i,mis}^{(l)}; l = 1, \ldots, m\}$ are drawn from $p(Y_{mis} \mid Y_{obs})$. Notice that $p(Y_{mis} \mid Y_{obs})$ may be written as

$$p(Y_{mis} \mid Y_{obs}) = \int p(Y_{mis} \mid Y_{obs}, \omega)p(\omega \mid Y_{obs})d\omega. \tag{19}$$

[3] Data augmentation consists of an iterative sampling scheme. First, given a current parameter estimate, $\vartheta^{(t)}$, draw a value of the missing data from its conditional predictive distribution: $Y_{mis}^{(t+1)} \sim p(Y_{mis} \mid Y_{obs}, \vartheta^{(t)})$. Second, given $Y_{mis}^{(t+1)}$, draw a value of ϑ from its complete data posterior: $\vartheta^{(t+1)} \sim p(\vartheta \mid Y_{obs}, Y_{mis}^{(t+1)})$. After repeating the two steps from a starting value of $\vartheta^{(0)}$ for a suitably large numbers (e.g., t = 5,000), we can regard $\{Y_{mis}^{(t)}: t \geq 5{,}000\}$ as an approximate draw from $p(Y_{mis} \mid Y_{obs})$. Interested readers can gain further information about data augmentation in the context of handling missing data from Schafer (1997, Chapter 4), Little and Rubin (2002, Chapter 10), and Enders (2010, Chapter 7).

[4] In cases where the *complete data* density is from an exponential family and thus the log-likelihood is linear in the *complete data* sufficient statistics, $\bar{T}$ takes the same form as the *complete data* log-likelihood statistic evaluated at the pooled parameter estimates and at the averaged *complete data* sufficient statistics (Meng & Rubin, 1992).

[5] In fact, Arbuckle (1996, pp. 271–275) pointed out that missing data can be generated from $p(Y_{mis} \mid Y_{obs}, \hat{\vartheta}_{FIML})$. However, the author suggested the idea as optional steps that can be performed after the model is fitted. On the other hand, Jamshidian and Bentler (1999) proposed an EM algorithm where the missing values are replaced by $\mathrm{E}(y_{i,m} \mid y_{i,o}, \vartheta^{(t)})$ with $\vartheta^{(t)}$ being continually replaced during the iteration process.

In the above equation, it is important to note that ω is not necessarily equal to the parameters specified in the posited structural equation model, that is, $\omega \neq \vartheta$. Moreover, $p(Y_{mis} \mid Y_{obs}, \omega)$ is not necessarily the assumed probability distribution underlying the postulated SEM model. In SEM with continuous variables, the most popular model for imputation is the multivariate normal model with ω equal to the saturated mean vector and covariance matrix or $\omega^* = (vec(\mu^*)', vech(\Sigma^*)')'$, which is drawn from the posterior distribution. In this case, $\{Y_{i,mis}^{(l)}; l = 1, \ldots, M\}$ are random draws from a multivariate normal distribution with the mean and the variance given as

$$E(Y_{i,mis} \mid Y_{i,obs}, \omega^*) = \mu_{mis} + \Sigma_{mo}\Sigma_{00}^{-1}(Y_{i,obs} - \mu_{obs})$$

$$Var(Y_{mis} \mid Y_{obs}, \omega^*) = \Sigma_{mm} - \Sigma_{mo}\Sigma_{oo}^{-1}\Sigma_{mo},$$

where μ_{mis} and μ_{obs} contain the elements of the saturated mean vector, μ^*, corresponding to missing and observed values, respectively, and Σ_{mm}, Σ_{oo} and Σ_{mo} contain rows and columns of the saturated variance-covariance matrix, Σ^*, corresponding to observed and missing values.

It is interesting to notice that, when the model is correctly specified under multivariate normality, both $p(Y_{mis} \mid Y_{obs}, \hat{\vartheta}_{FIML})$ and $p(Y_{mis} \mid Y_{obs}, \omega)$ have a multivariate normal distribution with $\omega = [vec^T{}_h(\mu(\vartheta)), vec^T(\Sigma(\vartheta))]^T$. Therefore, the two probability distributions $p(Y_{mis} \mid Y_{obs}, \hat{\vartheta}_{FIML})$ and $p(Y_{mis} \mid Y_{obs}, \omega)$ are identical except that $\hat{\vartheta}_{FIML}$ is fixed at the FIML estimates whereas ω has a posterior distribution given the observed data, $p(\omega \mid Y_{obs})$[6]. As such, it can be expected that $Y_{mis,MI}^{(l)}$ drawn from $p(Y_{mis} \mid Y_{obs})$ are susceptible to greater variability than $Y_{mis,FIML}^{(l)}$ drawn from $p(Y_{mis} \mid Y_{obs}, \hat{\vartheta}_{FIML})$ because of the additional source of variability in ω stemming from its prior distribution. A greater degree of variability in $Y_{mis,MI}^{(l)}$ may lead to a larger loss of efficiency for $\hat{\vartheta}_{MI}$ than for $\hat{\vartheta}_{FIML}$ relative to $\hat{\vartheta}_{ML}$ as the percent of missingness increases in the data. However, as the size of the data increases, the degree of uncertainty in estimating ω, reflected in its posterior distribution $p(\omega \mid Y_{obs})$, diminishes, and so does the uncertainty propagated to the likelihood function in Equation 18 or $p(Y_{mis} \mid Y_{obs}, \omega)$. Therefore, statistical properties of $\hat{\vartheta}_{MI}$ and $\hat{\vartheta}_{FIML}$ based on $Y_{mis,MI}^{(l)}$ and $Y_{mis,FIML}^{(l)}$, respectively, will tend to be asymptotically equivalent to each other. Such results have been repeatedly observed in previous studies (e.g., Collins et al., 2001; Graham, 2009; Schafer & Graham, 2002).

In practice, in all analyses of real data, a statistical model is only an approximation to reality, and thus some departures from modeling assumptions are inevitable. In the context of SEM with missing data, misspecification of the postulated model implies that $\omega \neq [vec(\mu(\vartheta)), vech(\Sigma(\vartheta))]^T$, which in turn suggests that the two probability distributions $p(Y_{mis} \mid Y_{obs}, \hat{\vartheta}_{FIML})$ and $p(Y_{mis} \mid Y_{obs}, \omega)$ are not identical even when the multivariate normality assumption holds. Therefore, $Y_{mis,FIML}^{(l)}$ would be different from $Y_{mis,MI}^{(l)}$ in their statistical properties.

Specifically, $Y_{mis,FIML}^{(l)}$ would be more likely to be consistent with the posited model because $Y_{mis,FIML}^{(l)}$ are based on $\hat{\vartheta}_{FIML}$ or the parameter estimates specified in the posited model. On the other hand $Y_{mis,MI}^{(l)}$ would be more likely to be consistent with the important features of the observed data as a whole such as the means, variances, and covariances because $Y_{mis,MI}^{(l)}$ are based on ω or the saturated mean vector and covariance matrix specified in the multivariate normal imputation model.

With this line of reasoning, it can be expected that the FIML point estimates, $\hat{\vartheta}_{FIML}$ tend to be affected by the postulated model, especially when it involves misspecifications because the estimation is based on $\{(Y_{obs}, Y_{mis,FIML}^{(l)}); l = 1, \ldots, M\}$. On the other hand, the MI point estimates, $\hat{\vartheta}_{MI}$, tend to remain unaffected by the types of the postulated model because the estimation is based on $\{(Y_{obs}, Y_{mis,MI}^{(l)}); l = 1, \ldots, M\}$, which is independent of the postulated model. It can also be expected that the MI point estimates tend to be closer to the standard ML estimates[7] than the FIML estimates to the extent that the saturated mean vector and covariance matrix or ω^* based on $\{(Y_{obs}, Y_{mis,MI}^{(l)}); l = 1, \ldots, M\}$ are similar to $\bar{y}$ and S based on the original/hypothetical *complete data*.

With respect to the sampling variability of the parameter estimates, it would be very difficult to make precise predictions on the expected patterns of discrepancy between FIML and MI estimators when working with imperfect models. In general, however, an estimator for a parameter tends to show larger variability across repeated sampling when the model involves specification errors than it does when the model is correctly specified. In addition, parameter estimates specified in the complex model tend to show larger sampling variability than do parameter estimates specified in the more parsimonious model (Myung, 2000; Zucchini, 2000). In this sense, FIML and MI estimators have different mechanisms operating for determining the magnitudes of their sampling variabilities. That is, FIML estimators are based on $\{(Y_{obs}, Y_{mis,FIML}^{(l)}); l = 1, \ldots, M\}$ where $Y_{mis,FIML}^{(l)}$ are drawn from the more parsimonious postulated model with specification errors,[8] whereas MI estimators are based on $\{(Y_{obs}, Y_{mis,MI}^{(l)}); l = 1, \ldots, M\}$ where $Y_{mis,MI}^{(l)}$ are drawn from the saturated model with no specification errors, yet more complex than the postulated model. Thus, what is certain is that the relative efficiency between the two estimators will be dependent upon the interplay between the parsimony and accuracy of the model employed in the imputation stage. Again, although it would be difficult to make a precise prediction on the expected patterns of relative efficiency, it would be likely that the MI estimator tends to show larger sampling variability than the FIML estimator due to the use of a more complex saturated model for imputation and the additional source of variability stemming from the prior distribution. However, it might be the case that the FIML

[6] In the iterative data augmentation (DA) algorithm, the conventional noninformative Jeffrey's prior distribution is generally assumed for the mean and covariance matrix, that is, $p(\mu^*, \Sigma^*) \propto |\Sigma|^{-\frac{p+1}{2}}$. In DA, the saturated covariance matrix and the mean vector are sampled from their *complete data* posterior distributions: $\Sigma^{*(t+1)} \sim p(\Sigma^* \mid Y_{obs}, Y_{mis}^{(t+1)}, \mu^{*(t)})$ and $\mu^{*(t+1)} \sim p(\mu^* \mid Y_{obs}, Y_{mis}^{(t+1)}, \Sigma^{*(t+1)})$. The interested reader is referred to Little and Rubin (2002, Chapter 10) and Enders (2010, Chapter 7) for detailed descriptions of the posterior step in DA.

[7] In the present article, the standard ML estimates represent the ML estimate in Equation 2 that maximizes the *complete data* log-likelihood in Equation 1, where the term *complete data* refers to the original dataset before missing values were generated by any missing data mechanism. The listwise deleted dataset is not treated as *complete data* in this article.

[8] There is no drawing of missing values for the FIML estimator according to its original definition in Equation 7. However, actual drawing of missing values is involved in our alternative definition of the FIML estimator given in Equation 17. The two definitions are equivalent at the conceptual level, but we have also shown empirically that the two FIML estimates according to Equations 7 and 17 are the same up to the third decimal point (see Footnote 14).

estimator may show as large as or an even larger sampling variability than the MI estimator once the degree of the postulated model's misfit passes a certain magnitude. It remains to be seen whether the percent of missingness increases or decreases the impact of model accuracy and/or model parsimony on the sampling variability of the two estimators.

Model fit indexes have even more difficulty predicting expected patterns of relative performance between MI and FIML because they are based on chi-square test statistics that may not be asymptotically equivalent for mis-specified hypothesized SEM models as well as the null/independence models (Enders & Mansolf, 2018). According to our general expectations about the parameter estimates of the two estimates in relation to the *complete data* ML estimates, one may expect that the MI-based chi-squares should tend to be closer (i.e., larger) to the *complete data* chi-squares than the FIML-based chi-squares when the postulated model does not hold exactly in the population. A close examination of T_{MI} in Equation 13, however, suggests that the term r_3 will also be at play in determining the performance of T_{MI}.

The term r_3 is an estimate of the average relative increase in variance (ARIV) and attenuates the test statistic to account for missing data uncertainty (Rubin, 1987; Schafer, 1997; see Appendix). Conceptually, the relative increase in variance (RIV) due to nonresponse is defined as the proportional increase in sampling variability that is due to the missing data. A formal definition of RIV can be given as

$$\nabla = V_{com}^{-1}(V_{obs} - V_{com})$$

where V_{com} and V_{obs} represent the asymptotic covariance matrix for the *complete data* estimates and the *observed-data* estimates, respectively (Li et al. (1991); see Appendix). The ARIV can be obtained from the average of the diagonal elements in ∇. Specifically, ARIV for FIML and MI estimates are defined as, respectively

$$\bar{\nabla}_{FIML} = \frac{\mathrm{tr}\left[V_{ML}^{-1}(V_{FIML} - V_{ML})\right]}{q} \quad (20)$$

$$\bar{\nabla}_{MI} = \frac{\mathrm{tr}\left[V_{ML}^{-1}(V_{MI} - V_{ML})\right]}{q} \quad (21)$$

with V_{ML}, V_{FIML}, and V_{MI} defined earlier in Equations 3, 8, and 12, respectively.

For FIML, the missing data uncertainty is incorporated into the chi-square test statistic by the missing-data patterns, that is, by directly summing individual log-likelihoods constructed based on the observed data points for each case as in Equation 6. On the other hand, MI chi-squares employ an estimate of $\bar{\nabla}_{MI}$ or r_3 to incorporate missing data uncertainty in the likelihood ratio tests as in Equation 13. Therefore, what can be expected from the relative performance of T_{FIML} and T_{MI} is that the degree of bias in r_3 will have an impact on the performance of the pooled likelihood ratio test statistic under MI, which eventually affect model fit indexes such as the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger, 1989, 1990), the comparative fit index (CFI; Bentler, 1990), and the Tucker Lewis index (TLI; Tucker & Lewis, 1973). RMSEA is a function of the chi-square statistics for the postulated model, whereas incremental fit indexes, such as TLI and CFI, are functions of chi-square test statistics for both the postulated model and the baseline model. The degree of bias in r_3 can also be non-negligible for the baseline model. Therefore, it may be the case that absolute fit indexes such as RMSEA, and incremental fit indexes such as TLI and CFI, show different patterns of relative performance between FIML and MI.

Monte Carlo Simulations

In the present study, we would like to provide empirical evidence concerning distinctive patterns of the inferences about ϑ and model fit evaluations between FIML and MI estimators. To complement and verify our (imperfect) theoretical conjectures just described, we conducted Monte Carlo simulation studies designed to provide more concrete evidence about relative performances of FIML and MI estimators in relation to the ML estimators.

To empirically examine relative performance between FIML and MI estimators, a series of simulation studies were conducted under a wide range of conditions in the context of three-wave longitudinal confirmatory factor analysis (CFA) models. Thus, *complete data* were generated based on CFA models with three correlated factors. Each factor was measured by five continuous indicators, which resulted in 15 observed variables (X1–X15).

For the true models used to generate *complete data* (see Figure 1 and Table 1), the population factor loadings were set to be different across the three waves based on the predetermined simulation conditions (as discussed below). In addition, the unique factors of the first indicators for each of the three factors (i.e., X1, X6, and X11) are allowed to be correlated in the population model. With the specified population parameter values, incomplete data were created by generating missing values on the indicators for the second and third factors (i.e., X6–X15) according to the missing data mechanisms (as described below).

The model misspecifications were introduced in the analysis models by imposing equality constraints on the factor loadings of indicators administered repeatedly across the three waves (e.g., $\lambda_{21} = \lambda_{72} = \lambda_{12,2}$; $\lambda_{51} = \lambda_{10,2} = \lambda_{15,2}$), and ignoring the covariances among unique factors. Such misspecification scenarios can occur in practice when SEM researchers test the weak factorial invariance (Meredith, 1993) in longitudinal data analysis. The specific simulation conditions under which FIML and MI were compared to each other and to standard ML included four variables: mechanism of missingness, percentage of missingness, level of model misspecification, and sample size.

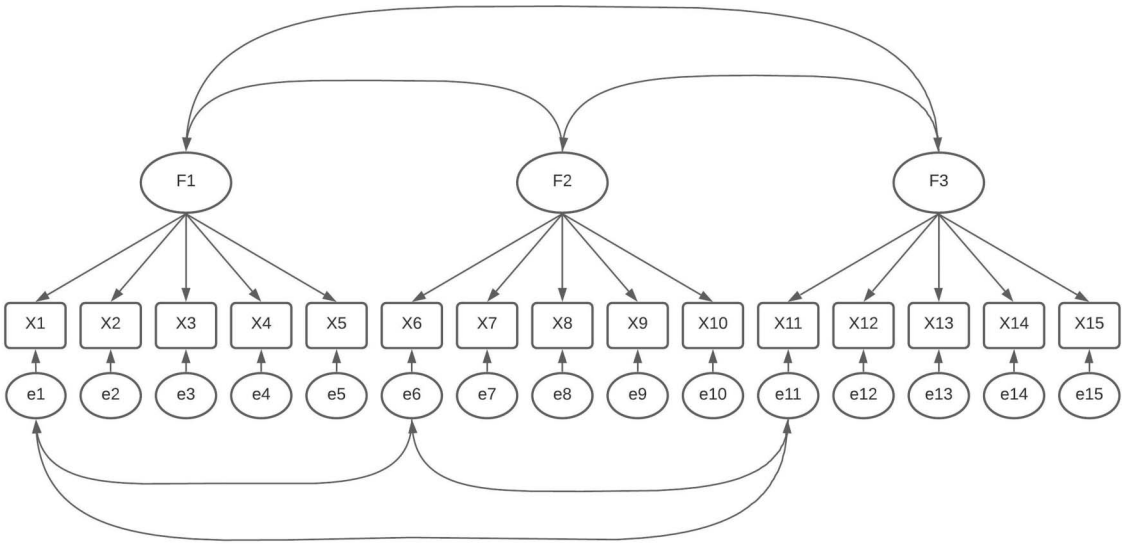
Mechanism of Missingness

Both missing completely at random (MCAR) and missing at random (MAR) were considered. In terms of MCAR, the occurrences of missingness were randomly determined, and had no correlation with any variables within the dataset. For cases of MAR, missing data were generated by the MAR-linear mechanism described in Collins et al. (2001). The probabilities of missingness on the last 10 observed variables (i.e., X6–X15) depended on quartiles of the sum of the first five variables (i.e., X1–X5). Specific distributions of missingness can be found in Table 2.

Percentage of Missingness

Under the MCAR condition, 15%, 25%, and 50% of data for the last 10 observed variables were randomly deleted. Under the MAR

Figure 1
Path Diagram of the True Data-Generating Model



condition, the probabilities of missingness on the last 10 observed variables (i.e., X6–X15) depended on quartiles of the sum of the first five variables (i.e., X1–X5) as presented in Table 2. The three distributions of missingness probabilities were decided so as to yield 15%, 25%, and 50% of overall percent of missingness in the data.

Level of Model Misspecification

Five levels of model misspecification were considered by manipulating population RMSEAs of the CFA model under consideration, including 0.00 (i.e., a correctly specified model), 0.05, 0.07, 0.09, and 0.12. As mentioned earlier, model misspecifications were introduced by imposing (incorrect) equality constraints on factor loadings and ignoring residual correlations. The five different levels of population RMSEA were determined by manipulating true population parameter values on factor loadings and residual correlations. Table 1 presents the population parameter values used for data generation across different levels of misspecification.

Sample Size

Sample sizes included 100, 200, 500, and 1,000. In total, 120 ($2 \times 3 \times 5 \times 4$) different conditions were considered in this study, with each condition consisting of 500 replications. The complete data sets were generated with Mplus 7.11 (Muthén & Muthén, 1998–2012). SAS 9.3 was used to create missing values according to the aforementioned missing data mechanisms.

Analysis

The misspecified model was fit to the simulated data sets with missing values addressed by FIML and MI methods. For MI, the missing observations were imputed by the saturated-normal-model; the number of imputations was 100. Furthermore, the identical misspecified model was fit to the *complete data*—the original data set before missing values were generated—using the standard ML method. By comparing the behaviors of FIML and MI estimators with those of the standard ML estimator, direct comparisons could be made concerning which method, FIML or MI, produced results closer to those that would have been obtained had we access to the *complete data*. Specifically, the performance of FIML and MI estimators were compared across simulation conditions to that of the standard ML estimator with respect to parameter estimates, sampling variability of the estimates, and model fit statistics. All data analyses were conducted using Mplus 7.11 (Muthén & Muthén, 1998–2012). R (R Core Team, 2018) was used to calculate and organize the outcome variables.

Outcome Variables

We first obtained the parameter estimates for all freely estimated parameters. For each simulated condition, the average parameter estimates and their standard deviations across replications were computed. We compared the results from FIML, MI, and those from *complete data* analysis. To better compare the behaviors of FIML- and MI-based parameter estimates, we calculated the relative differences (RD) as

$$RD = \frac{\bar{\theta}_{incomp} - \bar{\theta}_{comp}}{\bar{\theta}_{comp}}$$

where $\bar{\theta}_{incomp}$ represents the average value in each of the simulation conditions for parameter estimates from incomplete data analysis by FIML or MI estimator, and $\bar{\theta}_{comp}$ represents the corresponding average value from *complete data* analysis by the standard ML estimator.

Considering both bias and sampling variance of an estimator, we also computed the root mean squared errors (RMSE) for the MI estimates and the FIML estimates around the standard ML estimates as

$$RMSE = \sqrt{(\bar{\theta}_{incomp} - \bar{\theta}_{comp})^2 + var(\theta_{incomp})},$$

where $var(\theta_{incomp})$ represents the empirical variance of the incom-

Table 1
Population Values Used for Data Generation

Parameters	Population RMSEA				
	0.00	0.05	0.07	0.09	0.12
Factor loadings					
F1 by					
X1 (anchor item)	1.00	1.00	1.00	1.00	1.00
X2	0.70	0.70	0.80	0.80	0.90
X3	0.70	0.70	0.80	0.80	0.90
X4	0.70	0.70	0.75	0.80	0.90
X5	0.70	0.70	0.75	0.80	0.90
F2 by					
X6 (anchor item)	1.00	1.00	1.00	1.00	1.00
X7	0.70	0.60	0.60	0.60	0.60
X8	0.70	0.60	0.60	0.60	0.60
X9	0.70	0.60	0.60	0.60	0.60
X10	0.70	0.60	0.60	0.60	0.60
F3 by					
X11 (anchor item)	1.00	1.00	1.00	1.00	1.00
X12	0.70	0.40	0.40	0.40	0.30
X13	0.70	0.50	0.50	0.40	0.30
X14	0.70	0.50	0.50	0.30	0.20
X15	0.70	0.50	0.50	0.35	0.20
Factor variances					
F1/F2/F3	1.00/1.20/1.50	1.00/1.20/1.50	1.00/1.20/1.50	1.00/1.20/1.50	1.00/1.20/1.50
Factor covariances					
F1 with F2	0.50	0.50	0.50	0.50	0.50
F2 with F3	0.50	0.50	0.50	0.50	0.50
F1 with F3	0.30	0.30	0.30	0.30	0.30
Residual variances					
X1–X15	0.30	0.30	0.30	0.30	0.30
Residual covariances					
X1 with X6	0.00	0.10	0.15	0.20	0.27
X6 with X11	0.00	0.10	0.15	0.20	0.27
X1 with X11	0.00	0.05	0.05	0.10	0.07

Note. RMSEA = root mean square error of approximation.

plete data analysis (by FIML or MI estimator) across replications. All other notations remain as defined above.

In terms of the goodness of model fit, we computed the average estimates of the fit indices across all simulation conditions. The average estimates from FIML and MI were compared to results obtained from the *complete data* analysis. To explain the patterns observed in the fit indices (and variability in parameter estimates), the average relative increase in variance (ARIV) under the fitted model and baseline model were computed for both FIML (Equation 20) and MI estimates (Equation 21). Finally, when using MI, r_3 (Equation 14) was computed under both the fitted model and baseline model.

Table 2
Distributions of Missingness Under MAR

Overall percent of missingness	Quartile of sum of X1–X5			
	First	Second	Third	Fourth
15%	1.5%	10.5%	19.5%	28.5%
25%	2.5%	17.5%	32.5%	47.5%
50%	5%	35%	65%	95%

Note. MAR = missing at random.

Results

All replications converged for all conditions.[9] We observed that the overall patterns of the results were the same for all parameters, and therefore, to simplify the presentation, we have presented the results only for selected parameters under MAR conditions (i.e., the point estimates and standard deviations for the factor loadings of X2 and X3). The results for all other parameters under MCAR and MAR conditions are available in the online supplementary material.

The Parameter Estimates

The results from parameter estimates are shown in Table 3. For both MI and FIML, cases yielding RD larger than 10% (in absolute values) are highlighted in bold in the table. In general, the parameter estimates based on MI using the saturated-normal-model were fairly close to the standard ML estimates across all conditions. Specifically, the MI estimates for all four factor loadings were

[9] For MI, all 100 imputations converged across all conditions. There was no convergence issue. For FIML, only a few out of 500 replications failed to converge under the conditions of 50% MAR and 50% MCAR with N = 100. We removed those cases and ran the simulation until we had a total of 500 replications under those conditions. The results were based on the converged solutions only.

Table 3
Average Parameter Estimates Under Missing at Random (MAR)

		F by X2							F by X3						
			15% missing		25% missing		50% missing			15% missing		25% missing		50% missing	
SS	Pop RMSEA	C	FIML	MI	FIML	MI	FIML	MI	C	FIML	MI	FIML	MI	FIML	MI
100	0.00	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
	0.05	0.57	0.58	0.57	0.59	0.57	**0.62**	0.57	0.61	0.62	0.61	0.63	0.61	0.65	0.61
	0.07	0.61	0.63	0.61	0.65	0.61	**0.69**	0.62	0.65	0.67	0.65	0.68	0.65	**0.72**	0.66
	0.09	0.66	0.68	0.67	0.70	0.66	**0.73**	0.67	0.66	0.69	0.67	0.70	0.67	**0.74**	0.67
	0.12	0.76	0.78	0.76	0.80	0.76	**0.84**	0.76	0.76	0.78	0.76	0.80	0.76	**0.84**	0.77
200	0.00	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
	0.05	0.57	0.58	0.57	0.59	0.57	**0.62**	0.57	0.61	0.62	0.61	0.63	0.61	0.65	0.61
	0.07	0.61	0.63	0.61	0.64	0.61	**0.69**	0.61	0.65	0.67	0.65	0.68	0.65	**0.72**	0.66
	0.09	0.66	0.68	0.66	0.70	0.66	**0.73**	0.67	0.66	0.68	0.67	0.70	0.66	**0.73**	0.67
	0.12	0.75	0.78	0.75	0.79	0.75	**0.83**	0.76	0.76	0.78	0.76	0.80	0.76	**0.84**	0.76
500	0.00	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
	0.05	0.57	0.58	0.57	0.59	0.57	**0.62**	0.57	0.61	0.62	0.61	0.63	0.61	0.64	0.61
	0.07	0.61	0.63	0.61	0.65	0.61	**0.69**	0.61	0.65	0.67	0.65	0.68	0.65	**0.72**	0.65
	0.09	0.66	0.68	0.66	0.70	0.66	**0.73**	0.67	0.66	0.68	0.66	0.70	0.66	**0.73**	0.67
	0.12	0.75	0.78	0.76	0.79	0.76	**0.83**	0.76	0.76	0.78	0.76	0.80	0.76	**0.83**	0.76
1,000	0.00	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
	0.05	0.57	0.58	0.57	0.59	0.57	**0.62**	0.57	0.61	0.62	0.61	0.62	0.61	0.64	0.61
	0.07	0.61	0.63	0.61	0.64	0.61	**0.69**	0.61	0.65	0.67	0.65	0.68	0.65	**0.72**	0.65
	0.09	0.66	0.68	0.66	0.70	0.66	**0.73**	0.66	0.66	0.68	0.66	0.70	0.66	**0.73**	0.66
	0.12	0.75	0.78	0.75	0.79	0.75	**0.83**	0.76	0.76	0.78	0.76	0.79	0.76	**0.83**	0.76

Note. SS = sample size; Pop RMSEA = population RMSEA; C = complete data results; FIML = full information maximum likelihood results; MI = multiple imputation results; RMSEA = root mean square error of approximation. We also computed the misspecified population parameters. The misspecified population parameters are identical to the complete data results when $N = 1{,}000$ (to the third decimal places). Values of the population parameters under the misspecified model are underlined in the table. Cases yielding relative differences larger than 10% (in absolute values) are highlighted in bold.

essentially identical to the *complete data* ML estimates up to the second decimal point.

On the other hand, an interesting pattern emerged for the FIML estimates. That is, Table 3 shows an interesting pattern as functions of population RMSEA values and percent of missingness. That is, the FIML estimates were, on average, higher than the *complete data* ML estimates, except for the condition of correct model specification (i.e., population RMSEA = 0). Such patterns became more and more pronounced as the percent of missingness increased from 15% to 25% to 50% and the degree of model misspecification increased. Under somewhat extreme conditions of 50% missingness, the FIML estimates were noticeably larger than the *completed data* ML estimates. The RD for FIML estimates could go above 10% under the condition of 50% missingness and a population RMSEA greater than .05.

The Variability of Parameter Estimates

The sampling variabilities of individual parameter estimates under FIML or MI approach are typically inflated because of the added uncertainty of missing data. Table 4 shows the standard deviations (*SD*) of the *complete data* ML estimates, the FIML estimates, and the MI estimates for the factor loadings across conditions. It can easily be seen that the *SD*s under FIML and MI are higher than the *SD*s under ML with higher inflation as the percent of missingness increases.

In Table 4, we also computed and showed the ratio of the variance for the FIML estimates for the factor loadings to the variance for the MI estimates under each condition. As expected, the variability of the MI estimates tends to be larger than the variability of the FIML estimates under no model misfit conditions (i.e., population RMSEA = 0.00). However, it is interesting to notice that, under the conditions of nonzero model misfit and 15% and 25% of missingness, the variance ratio could be larger than 1.0, suggesting that the sampling variability of the MI estimates for the factor loadings is *not always* larger than that of the FIML estimates. The variance ratio tends to get noticeably smaller than 1.00 under the condition that the percent of missingness is 50% regardless of the levels of misfit, suggesting that the impact of the amount of missingness on sampling variability is larger for the MI estimates than for the FIML estimates. Although not presented in a table, the sampling variabilities of factor variances and covariances were larger under MI than those under FIML across all conditions.

Table 5 shows the ARIV for MI estimates (Equation 20) and ARIV for FIML estimates (Equation 21) across all conditions. It can easily be seen that both $ARIV_{FIML}$ and $ARIV_{MI}$ are positive and increase with the percent of missingness, suggesting that the variances of parameter estimates become increasingly inflated because of increasing missingness. It can also be seen that, with only a few exceptions, $ARIV_{MI}$ are higher than $ARIV_{FIML}$ in general, suggesting that MI estimates tend to vary to a greater extent than FIML estimates over repeated sampling.

Root Mean Squared Error

Table 6 show the RMSE for the MI estimates and the FIML estimates around the standard ML estimates for the four factor

Table 4
Empirical Standard Deviations of Parameter Estimates Under Missing at Random (MAR)

		F by X2										F by X3									
			15% missing			25% missing			50% missing				15% missing			25% missing			50% missing		
SS	Pop RMSEA	C	FIML	MI	R	FIML	MI	R	FIML	MI	R	C	FIML	MI	R	FIML	MI	R	FIML	MI	R
100	0.00	0.038	0.040	0.041	0.95	0.044	0.047	0.91	0.051	0.057	0.79	0.035	0.038	0.039	0.98	0.042	0.044	0.88	0.049	0.059	0.71
	0.05	0.044	0.047	0.046	1.03	0.051	0.052	0.94	0.058	0.065	0.80	0.041	0.045	0.045	1.01	0.049	0.048	1.04	0.056	0.063	0.80
	0.07	0.049	0.053	0.052	1.05	0.057	0.057	1.00	0.062	0.070	0.78	0.046	0.051	0.049	1.07	0.056	0.055	1.04	0.061	0.067	0.83
	0.09	0.054	0.059	0.059	1.00	0.060	0.061	0.96	0.065	0.074	0.78	0.052	0.056	0.056	1.02	0.059	0.059	0.98	0.064	0.073	0.77
	0.12	0.062	0.065	0.065	1.01	0.067	0.067	1.01	0.071	0.082	0.74	0.061	0.064	0.063	1.03	0.067	0.066	1.01	0.071	0.078	0.82
200	0.00	0.026	0.029	0.029	0.97	0.031	0.031	0.98	0.034	0.040	0.76	0.026	0.029	0.029	0.96	0.030	0.031	0.96	0.034	0.038	0.82
	0.05	0.029	0.032	0.031	1.02	0.034	0.034	0.99	0.039	0.042	0.88	0.029	0.031	0.031	1.02	0.034	0.033	1.03	0.037	0.041	0.82
	0.07	0.033	0.035	0.035	1.05	0.039	0.038	1.05	0.042	0.046	0.83	0.032	0.035	0.035	1.03	0.036	0.035	1.07	0.042	0.044	0.90
	0.09	0.035	0.038	0.037	1.03	0.040	0.039	1.01	0.044	0.049	0.81	0.036	0.038	0.038	1.02	0.041	0.041	1.01	0.045	0.049	0.85
	0.12	0.041	0.043	0.043	1.02	0.044	0.045	0.97	0.048	0.051	0.88	0.042	0.045	0.045	1.02	0.045	0.045	1.01	0.048	0.052	0.88
500	0.00	0.017	0.019	0.019	0.97	0.020	0.020	0.93	0.022	0.026	0.76	0.017	0.019	0.019	0.98	0.020	0.021	0.96	0.022	0.025	0.80
	0.05	0.020	0.021	0.021	1.00	0.023	0.023	1.00	0.025	0.029	0.79	0.019	0.021	0.020	1.03	0.022	0.022	1.02	0.025	0.027	0.84
	0.07	0.022	0.023	0.023	1.03	0.024	0.024	0.99	0.028	0.030	0.87	0.021	0.022	0.022	1.03	0.024	0.023	1.04	0.027	0.028	0.90
	0.09	0.023	0.025	0.025	1.02	0.025	0.026	0.98	0.028	0.031	0.78	0.023	0.025	0.024	1.03	0.026	0.026	1.01	0.029	0.031	0.84
	0.12	0.026	0.027	0.027	1.03	0.027	0.028	1.00	0.029	0.032	0.83	0.027	0.028	0.027	1.02	0.029	0.029	0.99	0.030	0.032	0.90
1,000	0.00	0.013	0.014	0.014	0.98	0.014	0.014	0.98	0.016	0.018	0.80	0.013	0.013	0.014	0.98	0.014	0.014	0.95	0.016	0.018	0.77
	0.05	0.014	0.015	0.015	1.00	0.016	0.016	1.04	0.018	0.019	0.89	0.014	0.015	0.015	1.01	0.015	0.015	1.00	0.017	0.019	0.86
	0.07	0.016	0.017	0.016	1.06	0.018	0.018	1.02	0.020	0.021	0.87	0.016	0.017	0.016	1.04	0.017	0.017	1.00	0.019	0.020	0.88
	0.09	0.017	0.018	0.017	1.01	0.018	0.018	1.04	0.020	0.021	0.88	0.017	0.018	0.017	1.02	0.019	0.018	1.03	0.019	0.021	0.87
	0.12	0.019	0.020	0.019	1.01	0.020	0.020	1.04	0.022	0.023	0.91	0.019	0.020	0.020	1.03	0.021	0.020	1.04	0.022	0.022	0.92

Note. SS = sample size; Pop RMSEA = population RMSEA; C = complete data results; FIML = full information maximum likelihood results; MI = multiple imputation results; R = ratio of the FIML estimates to the MI estimates; RMSEA = root mean square error of approximation.

Table 5
The Behaviors of the Average Relative Increase in Variance (ARIV) Under Missing at Random (MAR)

		The fitted model						The baseline model					
		15% missing		25% missing		50% missing		15% missing		25% missing		50% missing	
SS	Pop RMSEA	FIML	MI	FIML	MI	FIML	MI	FIML	MI	FIML	MI	FIML	MI
100	0.00	0.12	0.15	0.24	0.34	1.83	1.59	0.29	0.17	0.54	0.36	1.35	1.97
	0.05	0.12	0.15	0.24	0.32	1.16	1.60	0.23	0.17	0.43	0.34	1.16	1.88
	0.07	0.13	0.16	0.25	0.34	1.19	1.64	0.24	0.17	0.43	0.34	1.23	1.96
	0.09	0.12	0.15	0.23	0.31	1.20	1.52	0.24	0.17	0.43	0.34	1.12	1.93
	0.12	0.12	0.14	0.26	0.31	1.58	1.54	0.24	0.16	0.46	0.38	1.20	1.94
200	0.00	0.13	0.15	0.24	0.29	0.76	1.34	0.30	0.15	0.50	0.32	1.22	1.66
	0.05	0.12	0.14	0.24	0.29	0.75	1.31	0.24	0.15	0.41	0.31	1.01	1.62
	0.07	0.12	0.13	0.25	0.29	0.75	1.30	0.23	0.14	0.42	0.30	1.04	1.62
	0.09	0.13	0.14	0.24	0.29	0.76	1.30	0.20	0.14	0.44	0.31	1.00	1.59
	0.12	0.13	0.13	0.25	0.27	0.76	1.24	0.22	0.15	0.44	0.33	1.03	1.63
500	0.00	0.13	0.14	0.25	0.29	0.76	1.25	0.29	0.15	0.54	0.31	1.24	1.56
	0.05	0.11	0.12	0.24	0.27	0.75	1.21	0.23	0.14	0.41	0.30	1.05	1.51
	0.07	0.12	0.13	0.25	0.28	0.74	1.20	0.24	0.15	0.42	0.29	1.06	1.52
	0.09	0.12	0.13	0.24	0.26	0.73	1.15	0.23	0.14	0.44	0.31	1.00	1.48
	0.12	0.12	0.12	0.25	0.26	0.76	1.13	0.23	0.14	0.39	0.28	1.08	1.50
1,000	0.00	0.12	0.13	0.24	0.27	0.77	1.21	0.30	0.14	0.51	0.29	1.21	1.54
	0.05	0.12	0.13	0.24	0.27	0.74	1.18	0.22	0.14	0.40	0.29	1.08	1.52
	0.07	0.12	0.13	0.24	0.27	0.75	1.16	0.23	0.14	0.45	0.32	1.10	1.51
	0.09	0.12	0.12	0.24	0.26	0.74	1.15	0.24	0.15	0.41	0.30	0.98	1.43
	0.12	0.12	0.12	0.24	0.25	0.76	1.14	0.24	0.13	0.43	0.28	1.06	1.39

Note. SS = sample size; Pop RMSEA = population RMSEA; MI = multiple imputation results; FIML = full information maximum likelihood results; RMSEA = root mean square error of approximation.

loadings. Considering that RMSE is a measure of balance between bias and variance of an estimator, the patterns of RMSEs we observed suggest that, across repeated sampling, the MI estimates tend to fluctuate around the value that is fairly close to the standard ML estimates under each condition, whereas the FIML estimates tend to fluctuate to a slightly lesser degree than MI estimates but they vary around a value that is different from the standard ML estimates under nonzero misfit conditions. As a result, RMSEs for the MI estimates tend to be smaller than those of the FIML estimates across all conditions of nonzero population RMSEA.

The Chi-Squared Statistics and Model Fit Indexes

The behaviors of the chi-squared statistics, RMSEA, CFI, and TLI under missing data are reported in Tables 7 and 8. There were no noticeable differences between MAR and MCAR across all simulated conditions.

As can be seen in Table 7, the FIML method tended to yield larger chi-square values than the MI method, which made FIML-based chi-squares closer to the standard ML chi-square statistics from complete data sets than MI. The MI approach tended to produce smaller chi-square statistics, especially when the percent of missingness was high. As an absolute measure of fit based on chi-square statistics, the RMSEA estimates showed the same pattern as the chi-square statistics.

With respect to CFI and TLI, as can be seen in Table 8, the estimates under MI tends to be smaller, which made the MI-based estimates closer to the estimates under ML than the FIML estimates. Such differences between the two approaches became more pronounced as the percent of missingness increased and the level of misspecification became larger. For example, when 50% of the data were missing at random, the sample size was 1,000, and the population RMSEA was 0.12, the CFI and TLI using complete data were .85 and .83, respectively. The average estimates using FIML were overestimated (i.e., CFI = .91; TLI = .90), whereas the average estimates using MI were the same (at the second decimal point) as the complete data results.

Table 9 shows r_3 values for the fitted model and the baseline model,[10] respectively. Table 5 shows $\bar{\nabla}_{MI}$ in Equation 21 for the fitted model and the baseline model, respectively. Inspection of these tables shows that r_3 overestimates $\bar{\nabla}_{MI}$, resulting in overadjustment of the chi-square statistics under MI, suggesting that the MI-based chi-square test has lower power than the FIML-based chi-square test to reject misspecified models. As for incremental fit indexes, such as CFI and TLI, overadjustment of the chi-squares due to overestimation of $\bar{\nabla}_{MI}$ by r_3 applied both to the fitted model and the baseline model, which in turn happened to yield values closer to those based on T_{ML}.

Illustration Using Empirical Data: Fitting Factor Analysis Models to the Students' Life Satisfaction Scale

In this section, we used empirical data and compared the two missing data approaches in the context of fitting factor analysis models. We used part of a larger dataset that had been analyzed in

[10] The baseline model in this study was the null or independence model wherein the covariances among the observed variables were fixed to zero.

Table 6
Root Mean Squared Errors (RMSE) Under Missing at Random (MAR)

		F by X2						F by X3					
		15% missing		25% missing		50% missing		15% missing		25% missing		50% missing	
SS	Pop RMSEA	FIML	MI	FIML	MI	FIML	MI	FIML	MI	FIML	MI	FIML	MI
100	0.00	0.04	0.04	0.04	0.05	0.05	0.06	0.04	0.04	0.04	0.04	0.05	0.06
	0.05	0.05	0.05	0.06	0.05	0.08	0.06	0.05	0.04	0.05	0.05	0.07	0.06
	0.07	0.06	0.05	0.07	0.06	0.10	0.07	0.05	0.05	0.06	0.05	0.09	0.07
	0.09	0.06	0.06	0.07	0.06	0.10	0.07	0.06	0.06	0.07	0.06	0.10	0.07
	0.12	0.07	0.06	0.08	0.07	0.11	0.08	0.07	0.06	0.08	0.07	0.11	0.08
200	0.00	0.03	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.03	0.03	0.04
	0.05	0.03	0.03	0.04	0.03	0.07	0.04	0.03	0.03	0.04	0.03	0.05	0.04
	0.07	0.04	0.03	0.05	0.04	0.09	0.05	0.04	0.03	0.05	0.03	0.08	0.04
	0.09	0.04	0.04	0.05	0.04	0.08	0.05	0.04	0.04	0.05	0.04	0.08	0.05
	0.12	0.05	0.04	0.06	0.04	0.09	0.05	0.05	0.04	0.06	0.05	0.09	0.05
500	0.00	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.03
	0.05	0.02	0.02	0.03	0.02	0.06	0.03	0.02	0.02	0.03	0.02	0.05	0.03
	0.07	0.03	0.02	0.04	0.02	0.08	0.03	0.03	0.02	0.04	0.02	0.07	0.03
	0.09	0.03	0.02	0.04	0.03	0.07	0.03	0.03	0.02	0.04	0.03	0.07	0.03
	0.12	0.04	0.03	0.05	0.03	0.08	0.03	0.04	0.03	0.05	0.03	0.08	0.03
1,000	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.02
	0.05	0.02	0.02	0.03	0.02	0.05	0.02	0.02	0.01	0.02	0.02	0.04	0.02
	0.07	0.03	0.02	0.04	0.02	0.08	0.02	0.02	0.02	0.04	0.02	0.07	0.02
	0.09	0.03	0.02	0.04	0.02	0.07	0.02	0.03	0.02	0.04	0.02	0.07	0.02
	0.12	0.03	0.02	0.04	0.02	0.08	0.02	0.03	0.02	0.04	0.02	0.08	0.02

Note. SS = sample size; Pop RMSEA = population RMSEA; FIML = full information maximum likelihood results; MI = multiple imputation results; RMSEA = root mean square error of approximation.

a published study (Jiang et al., 2020).[11] Our purpose lay in confirming the patterns of differences between the two missing-data handling methods observed in our simulation studies via real data analyses and drawing the practical implications for applied users of SEM with missing data.

The scale of interest was the *Students' Life Satisfaction Scale* (SLSS; Huebner, 1991), which was designed to measure global life satisfaction in children and adolescents. Respondents rate seven items on a 6-point Likert scale ranging from 1 (i.e., *strongly disagree*) to 6 (i.e., *strongly agree*). The original dataset has no missing values, with the size being 991. The SLSS was designed to measure a unidimensional construct. Therefore, we fitted a total of three single-factor models that were created by imposing nested constraints successively from the least restrictive single-factor model. In the first and the least restrictive model (Model 1), four residual covariances were added to the single-factor model, for illustration purposes only, to improve the overall goodness-of-fit to a clearly acceptable level. In the second model (Model 2), three of the four residual covariances specified in Model 1 were constrained to be zero. The residual covariance between the two negatively worded items, X_3 and X_4, remained as a free parameter in Model 2. In the third model (Model 3), equality constraints were imposed on the factor loadings to Model 2, resulting in an essentially tau-equivalent model. The results of fitting Model 1, Model 2, and Model 3 to the original complete dataset are presented in Table 10.

To illustrate the patterns of results from the two incomplete data estimators, we created missing values under both MCAR and MAR mechanisms. The planned missing data design (Graham et al., 2006) makes the MCAR assumption entirely reasonable, and thus we created a variant of the three-form design from the SLSS. Specifically, Form 1 consisted of six items (out of seven) without X_4, Form 2 consisted of five items without X_3 and X_7, and Form 3 consisted of six items without X_6, resulting in approximately 20% missingness from the entire observations. Each of the three forms was administered to a randomly selected third of the entire sample. Following the general advice in the planned missing data designs literature,[12] we included the three indicators that turned out to be of the highest loadings in the complete-data analysis (i.e., X1, X2, X5) in the X set. Readers who are not familiar with the three-form design can simply imagine a situation where the researcher would like to collect data on all seven items, but without presenting each subject with all of the items in the scale. For detailed descriptions of the planned missing data designs, interested readers are referred to Graham et al. (1996, 2006).

For the MAR mechanism, we conditioned the missingness on the total scale score, or $Y = X_1 + X_2 + \ldots + X_7$. Specifically, we created an incomplete dataset with approximately 50% missingness by randomly removing 25% (75%) of observations from each of the seven indicators if Y is greater than or equal to (smaller than) its median.

[11] We the authors appreciate Scott Huebner and Xu Jiang for their generosity and willingness to provide the data to be used in our illustration of empirical data analysis.

[12] Specifically, Graham et al. (2006, p. 326) and Graham (2012, p. 291) state that variables involved in the most important study hypotheses would normally be included in the X set. In studying factorial structure of psychological scales, it would be a reasonable choice to include the items of the highest loadings or the most reliable items in the X set to maximize the information available in the complete data.

Table 7
The Behaviors of the Chi-Square Test Statistics and RMSEA Under Missing at Random (MAR)

		Chi-square test statistics							RMSEA						
			15% missing		25% missing		50% missing			15% missing		25% missing		50% missing	
SS	Pop RMSEA	C	FIML	MI	FIML	MI	FIML	MI	C	FIML	MI	FIML	MI	FIML	MI
100	0.00	102.99	106.10	103.19	109.74	102.74	128.18	90.99	0.03	0.03	0.03	0.03	0.02	0.05	0.01
	0.05	127.04	126.36	122.57	125.12	116.85	138.10	100.63	0.06	0.05	0.05	0.05	0.04	0.06	0.02
	0.07	149.22	142.66	137.21	140.09	128.60	146.31	105.78	0.07	0.07	0.06	0.07	0.06	0.07	0.03
	0.09	179.25	166.32	159.49	158.65	145.13	156.38	112.68	0.09	0.09	0.08	0.08	0.07	0.08	0.04
	0.12	238.27	208.95	198.36	193.31	173.60	179.22	125.18	0.12	0.11	0.10	0.10	0.09	0.09	0.05
200	0.00	99.58	100.64	100.32	102.45	100.93	105.60	97.69	0.01	0.02	0.02	0.02	0.02	0.02	0.01
	0.05	147.95	139.51	137.24	136.28	132.01	128.31	116.50	0.05	0.05	0.05	0.05	0.04	0.04	0.03
	0.07	192.42	174.78	170.65	165.70	158.44	146.69	129.72	0.07	0.06	0.06	0.06	0.06	0.05	0.04
	0.09	253.64	221.52	217.50	202.50	194.50	168.43	149.02	0.09	0.08	0.08	0.07	0.07	0.06	0.05
	0.12	371.69	309.91	305.18	273.77	262.47	211.97	184.75	0.12	0.11	0.10	0.10	0.09	0.08	0.07
500	0.00	96.48	96.32	96.08	97.26	97.08	97.33	95.34	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	0.05	218.13	195.46	192.23	181.63	175.96	155.60	143.90	0.05	0.05	0.04	0.04	0.04	0.04	0.03
	0.07	329.45	284.63	278.89	257.61	247.32	206.39	186.55	0.07	0.06	0.06	0.06	0.06	0.05	0.04
	0.09	482.53	402.36	398.70	355.09	346.61	262.12	241.88	0.09	0.08	0.08	0.07	0.07	0.06	0.06
	0.12	777.36	623.04	625.29	533.99	530.96	367.61	344.73	0.12	0.11	0.11	0.10	0.10	0.08	0.07
1,000	0.00	95.90	95.72	95.69	95.89	95.82	97.72	97.17	0.00	0.01	0.01	0.01	0.01	0.01	0.01
	0.05	339.34	293.47	288.13	265.36	255.68	211.50	193.42	0.05	0.05	0.04	0.04	0.04	0.03	0.03
	0.07	561.80	473.19	463.82	419.06	402.17	309.33	277.39	0.07	0.06	0.06	0.06	0.06	0.05	0.04
	0.09	868.81	705.61	701.46	610.63	601.19	423.66	395.62	0.09	0.08	0.08	0.07	0.07	0.06	0.06
	0.12	1458.81	1146.54	1159.18	969.46	975.94	639.17	614.81	0.12	0.11	0.11	0.10	0.10	0.08	0.07

Note. SS = sample size; Pop RMSEA = population RMSEA; C = complete data results; FIML = full information maximum likelihood results; MI = multiple imputation results; RMSEA = root mean square error of approximation.

To analyze the incomplete data sets created under the 20% MCAR and 50% MAR mechanisms, we applied both FIML and MI methods. To be consistent with our simulation study, we set the number of imputations to be 100 for MI estimator. The results from the analysis of MCAR and MAR data are presented in Table 10. Overall, we found that the results are consistent with the findings from our simulation studies. More specifically, under both MCAR and MAR conditions, the FIML-based parameter estimates become more discrepant from the ML estimates as the level of model misfit increases. In Table 10, the parameter estimates that are closer to the complete-data ML estimates are underscored. It can be seen that the number of underscored estimates under FIML decreases as goodness-of-fit deteriorates from Model 1 to Model 3.

The patterns for the fit indexes in Table 10 are also consistent with the patterns observed in our simulation studies. For all three models, FIML-based chi-squares and RMSEA estimates tend to be larger and closer to their counterparts from the complete-data analysis under both MCAR and MAR conditions. On the other hand, MI-based CFI/TLI values tend to be smaller and thus closer to those from the complete-data analysis.

The patterns we observed consistently in both the simulated and empirical data can have important practical implications for applied researchers who routinely evaluate global fitness of the postulated models. In our empirical example, we happened to know, due to the availability of the original complete data, that Model 3 is clearly unacceptable according to the widely adopted guidelines for interpreting fit indexes (Brown, 2015; Browne & Cudeck, 1993; MacCallum et al., 1996). However, the MI-based RMSEA estimate suggests an "adequate/fair" fit under MAR condition (i.e., $0.05 < \text{RMSEA} \leq 0.08$) and "mediocre" under MCAR condition (i.e., $0.08 < \text{RMSEA} \leq 0.10$), which may not necessarily lead the researcher to reject or consider modifications of Model 3. On the other hand, FIML-based RMSEA estimate suggests a "mediocre" fit under MAR and a "poor" fit (i.e., $\text{RMSEA} \geq 0.10$) under MCAR, which would lead the researcher more likely to cast doubt onto the viability of Model 3. The incremental fit indexes such as CFI and TLI tell us a slightly different story. That is, FIML-based CFI/TLI values are indicative of an "acceptable" fit (i.e., $0.90 < \text{CFI/TLI} \leq 0.95$; Bentler, 1990; Brown, 2015) under both MCAR and MAR conditions, whereas MI-based CFI/TLI values would lead the researcher to strongly suspect the reasonability of Model 3 (i.e., $\text{CFI/TLI} < 0.90$; Bentler, 1990; Brown, 2015).

Based on our findings, we recommend that applied researchers employ both FIML and MI methods for handling missing data in a given dataset and take the distinct patterns of parameter estimates and fit indexes from the two missing data approaches into account when evaluating and interpreting a substantive model under consideration. According to our findings, when the fitness of a model to the complete data falls into "close fit" ranges, it is likely that the researcher would find little to no difference between FIML and MI results, which can be seen in our simulation results under the conditions of $\text{RMSEA} \leq 0.05$ and from the results for Model 1 in our empirical example. As the fitness of the model to the complete data falls into "fair fit" or "mediocre fit" ranges, the researcher can start to notice discrepancies between FIML and MI results, which will get more pronounced as the adequacy of the model deteriorates to "unacceptable/poor fit" ranges. Such patterns were observed both in our simulation studies under the conditions of

Table 8
The Behaviors of CFI and TLI Under Missing at Random (MAR)

		CFI							TLI						
			15% missing		25% missing		50% missing			15% missing		25% missing		50% missing	
SS	Pop RMSEA	C	FIML	MI	FIML	MI	FIML	MI	C	FIML	MI	FIML	MI	FIML	MI
100	0.00	0.99	0.99	0.99	0.98	0.98	0.95	0.99	0.99	0.99	0.99	0.98	0.99	0.94	1.02
	0.05	0.96	0.96	0.96	0.96	0.96	0.92	0.97	0.96	0.95	0.95	0.95	0.95	0.91	0.98
	0.07	0.94	0.94	0.94	0.94	0.94	0.91	0.96	0.94	0.93	0.93	0.93	0.93	0.90	0.96
	0.09	0.90	0.91	0.90	0.91	0.90	0.89	0.93	0.89	0.90	0.89	0.90	0.89	0.88	0.93
	0.12	0.84	0.86	0.85	0.86	0.85	0.86	0.89	0.82	0.84	0.83	0.85	0.84	0.85	0.88
200	0.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	0.99	1.00	0.99	1.00
	0.05	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.97	0.97	0.97	0.97	0.96	0.96	0.96
	0.07	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.94	0.94	0.95	0.94	0.95	0.94
	0.09	0.91	0.92	0.91	0.92	0.91	0.93	0.91	0.90	0.91	0.90	0.91	0.90	0.93	0.90
	0.12	0.84	0.86	0.85	0.88	0.85	0.90	0.86	0.83	0.85	0.83	0.86	0.83	0.89	0.84
500	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.05	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
	0.07	0.95	0.95	0.95	0.95	0.95	0.96	0.95	0.94	0.95	0.94	0.95	0.94	0.96	0.94
	0.09	0.91	0.92	0.91	0.92	0.91	0.94	0.91	0.90	0.91	0.90	0.92	0.90	0.93	0.90
	0.12	0.85	0.86	0.85	0.88	0.85	0.91	0.85	0.83	0.85	0.83	0.86	0.83	0.90	0.83
1,000	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.05	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97
	0.07	0.95	0.95	0.95	0.96	0.95	0.96	0.95	0.94	0.95	0.94	0.95	0.94	0.96	0.94
	0.09	0.91	0.92	0.91	0.93	0.91	0.94	0.91	0.90	0.91	0.90	0.92	0.90	0.93	0.90
	0.12	0.85	0.87	0.85	0.88	0.85	0.91	0.85	0.83	0.85	0.83	0.87	0.83	0.90	0.83

Note. SS = sample size; Pop RMSEA = population RMSEA; C = complete data results; FIML = full information maximum likelihood results; MI = multiple imputation results; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker Lewis index.

RMSEA $\geq$ 0.07 and from the results for Model 2 and Model 3 in our empirical data analysis.

In practice, when both FIML and MI methods are applied to analysis of a given incomplete data set, it is likely that the researcher would obtain patterns of model fit indexes such that FIML-based fit indexes, whether absolute (e.g., RMSEA) or relative (CFI/TLI), tend to be larger than the MI-based values. The discrepancy in results between the two methods would get larger as the degree of model misfit increases. Thus, in a sense, a noticeable discrepancy between the FIMI and MI results can lend some insight as to the researcher's evaluation of the adequacy of the model. Such insight will not be gained from the results provided by either the FIML or MI method alone, because an individual method may not produce a clear indication of severe model misspecification for a clearly "unacceptable" model (e.g., Model 3). As mentioned earlier, the leading SEM software (e.g., Mplus and the lavaan package in R) offers simple options for automating the intermediate steps to produce outputs in a single step. From an applied researcher's perspective, instead of choosing one method over the other, it will be more beneficial to apply both methods and look into the patterns of the results as a whole.

Table 9
The Behaviors of r_3 Under Missing at Random (MAR)

		The fitted model			The baseline model		
SS	Pop RMSEA	15% missing	25% missing	50% missing	15% missing	25% missing	50% missing
100	0.00	0.30	0.62	1.86	0.31	0.63	1.96
	0.05	0.30	0.62	1.87	0.31	0.62	1.91
	0.07	0.31	0.63	1.88	0.31	0.62	1.93
	0.09	0.31	0.62	1.89	0.31	0.62	1.91
	0.12	0.32	0.64	1.92	0.31	0.62	1.92
200	0.00	0.28	0.55	1.63	0.29	0.57	1.76
	0.05	0.28	0.55	1.62	0.28	0.56	1.71
	0.07	0.28	0.55	1.64	0.28	0.56	1.72
	0.09	0.28	0.56	1.65	0.28	0.56	1.69
	0.12	0.29	0.58	1.68	0.28	0.56	1.69
500	0.00	0.27	0.52	1.52	0.28	0.54	1.66
	0.05	0.27	0.52	1.52	0.27	0.53	1.61
	0.07	0.27	0.53	1.53	0.27	0.54	1.61
	0.09	0.27	0.53	1.55	0.27	0.53	1.60
	0.12	0.28	0.54	1.59	0.27	0.53	1.60
1,000	0.00	0.26	0.51	1.50	0.27	0.54	1.63
	0.05	0.26	0.51	1.50	0.27	0.52	1.59
	0.07	0.27	0.52	1.51	0.27	0.53	1.60
	0.09	0.27	0.52	1.52	0.27	0.52	1.58
	0.12	0.28	0.54	1.56	0.27	0.53	1.57

Note. SS = sample size; Pop RMSEA = population RMSEA; RMSEA = root mean square error of approximation.

What constitutes a "noticeable" discrepancy is a difficult question that requires more systematic investigations. What can be said based on our findings is that if the discrepancy is large enough for the researchers to draw substantively different conclusions, that would certainly constitute a "large" discrepancy. Further study for developing a formal test or empirical guidelines is worth the attention in the missing data literature and in the SEM literature in general. We have illustrated R code examples for obtaining FIML and MI estimates in the online supplementary material.

Discussion

The primary question posed in the present study can be stated as follows: Which method of *incomplete data* analysis, FIML or

Table 10
The Parameter Estimates and Fit Indexes for FIML and MI: A Numeric Illustration

Parameter label	Model 1 (df = 10)					Model 2 (df = 13)					Model 3 (df = 19)				
		MCAR (20%)		MAR (50%)			MCAR (20%)		MAR (50%)			MCAR (20%)		MAR (50%)	
	Standard ML	FIML	MI	FIML	MI	Standard ML	FIML	MI	FIML	MI	Standard ML	FIML	MI	FIML	MI
							Parameter estimates								
F1 by X1	1.087	1.098	1.102	1.097	1.104	1.133	1.140	1.131	1.103	1.124	1.090	1.115	1.093	1.060	1.090
F1 by X2	1.233	1.251	1.256	1.223	1.234	1.311	1.321	1.319	1.250	1.281	1.090	1.115	1.093	1.060	1.090
F1 by X3	0.687	0.655	0.662	0.685	0.690	0.683	0.643	0.651	0.680	0.683	1.090	1.115	1.093	1.060	1.090
F1 by X4	0.879	0.839	0.834	0.803	0.798	0.874	0.834	0.828	0.797	0.787	1.090	1.115	1.093	1.060	1.090
F1 by X5	1.175	1.159	1.155	1.177	1.172	1.132	1.118	1.126	1.164	1.153	1.090	1.115	1.093	1.060	1.090
F1 by X6	1.009	1.028	1.022	1.025	1.035	0.956	0.954	0.970	1.019	0.994	1.090	1.115	1.093	1.060	1.090
F1 by X7	0.620	0.659	0.663	0.635	0.615	0.619	0.641	0.665	0.670	0.624	1.090	1.115	1.093	1.060	1.090
X1~~X1	0.526	0.502	0.494	0.546	0.534	0.424	0.408	0.428	0.525	0.488	0.465	0.432	0.466	0.557	0.516
X2~~X2	0.766	0.723	0.709	0.711	0.696	0.570	0.543	0.548	0.631	0.578	0.724	0.689	0.709	0.753	0.707
X3~~X3	2.951	3.057	3.056	3.052	3.056	2.957	3.072	3.070	3.061	3.067	3.045	3.191	3.175	3.108	3.144
X4~~X4	2.851	2.996	2.993	3.139	3.201	2.860	3.007	3.002	3.150	3.217	2.862	3.027	3.026	3.146	3.235
X5~~X5	0.465	0.503	0.512	0.461	0.472	0.565	0.596	0.579	0.493	0.517	0.546	0.562	0.556	0.579	0.523
X6~~X6	1.268	1.317	1.325	1.232	1.215	1.369	1.441	1.425	1.245	1.296	1.310	1.402	1.371	1.185	1.250
X7~~X7	2.224	2.146	2.149	2.204	2.229	2.224	2.164	2.146	2.159	2.218	2.313	2.248	2.203	2.156	2.304
X3~~X4	1.241	1.481	1.459	1.116	1.150	1.248	1.495	1.471	1.125	1.163	1.275	1.550	1.518	1.139	1.203
X1~~X2	0.188	0.156	0.146	0.140	0.123	—	—	—	—	—	—	—	—	—	—
X6~~X7	0.302	0.288	0.309	0.265	0.290	—	—	—	—	—	—	—	—	—	—
X1~~X6	−0.120	−0.143	−0.147	−0.122	−0.143	—	—	—	—	—	—	—	—	—	—
							Fit indices								
Chi-square	33.789	19.386	14.264	11.528	11.214	120.487	62.545	49.791	53.197	39.286	329.362	232.359	157.745	170.969	120.696
RMSEA	0.049	0.031	0.021	0.012	0.011	0.091	0.062	0.053	0.056	0.045	0.128	0.106	0.086	0.090	0.073
CFI	0.992	0.996	0.997	0.999	0.999	0.964	0.980	0.974	0.976	0.973	0.895	0.913	0.899	0.910	0.897
TLI	0.983	0.992	0.994	0.998	0.997	0.941	0.967	0.957	0.961	0.957	0.884	0.904	0.890	0.900	0.886

Note. Standard ML columns show the parameter estimates and fit indexes obtained from the original complete data. The underscored parameter estimates represent the values between FIML and MI that are closer to the standard ML estimates. FIML = full information maximum likelihood results; MI = multiple imputation results; MCAR = missing completely at random; MAR = missing at random; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker Lewis index.

MI, will produce results more similar to those from analysis of *complete data* when the model under consideration does not hold exactly in the population? Understanding the statistical properties of *incomplete data* handling methods in relation to the *complete data* method is important because the results from analysis of *complete data* is typically the analysis planned at the outset (Longford, 2006, p. 54). In practical applications of SEM, researchers are highly likely to work with imperfect models whose parameters may not have counterparts in the hypothetical true generating model in the population. This makes it even more important to understand the statistical properties of *incomplete data* handling methods in relation to the *complete data* method.[13]

The overall patterns of the results regarding the parameter estimates, sampling variability, and model fit indexes, in general, turned out to be consistent with our general expectations stated earlier. First, when comparing the estimates for factor loadings under FIML and MI with those under ML, an interesting pattern emerged. As the percent of missingness or degree of misfit or RMSEA increases, the discrepancy between FIML and MI estimates becomes larger. Interestingly, the FIML estimates become larger than the ML estimates as the degree of misfit increases, while MI estimates are nearly equivalent to the standard ML estimates regardless of the level of misfit. It was also observed that such patterns of discrepancy tended to become more pronounced as the percent of missingness increased.

We the authors believe that one reason that the FIML estimates became larger and more discrepant from the ML estimates is closely related to the patterns of the factor loadings in the population shown in Table 1 and is related to the missing

[13] An anonymous reviewer raised the issue of which estimator would best approximate the true population values. This is certainly an important problem theoretically, but in the present article, we do not address this issue, because not only do we concur with the primary goal of missing data analysis stated by Longford (2006), and our research question is equally as important in practical applications of SEM with missing data and with imperfect models, but also, as should become clear in the section on Monte Carlo Simulations, it often times becomes ambiguous to define what would be the true population value for a given parameter estimate specified in the postulated model with specification errors. For example, in our simulation study, a factor loading estimate generated by (erroneously) imposing equality constraints on multiple loadings, by definition, does not have a unique population value. If the lack of fit were, as defined in Cudeck and Browne (1992), due only to a nonstochastic residual matrix and the parametric structure specified in the model still holds exactly in the population, it could be meaningful to compare how well the given missing data procedures recover the population parameters. This is certainly an interesting direction for future research, which is further discussed in the Discussion section. Another anonymous reviewer suggested using the population parameter values of the incorrect model (i.e., the estimates obtained by fitting the incorrect model to the population) as the reference. Following this suggestion, we conducted a simulation with the population parameter values of the incorrect model as the reference. The results are identical to those from the condition of $N = 1{,}000$ up to the third decimal places.

data patterns resulting from the mechanism of missingness employed in our simulation studies. That is, according to the missing data mechanism in our study, missingness occurred only for the last 10 observed variables (i.e., X6–X15) under both MCAR and MAR. Therefore, when the posited model was trying to produce common estimates for the associated factor loadings (e.g., a common factor loading for X2, X7, and X12), the portion of the data corresponding to X1–X5 made a larger contribution to the *observed data* log-likelihood, which in turn pulled the FIML estimates for the factor loadings toward the factor loadings for the first five variables.[14] Further study on the generalizability of our observations is warranted, but the point is that the MI estimates remain unaffected and close to the ML estimates regardless of the level of model misfit, while the FIML estimates become more discrepant from the ML estimates as the level of model misfit increases.

Second, as can readily be expected, it was observed that the sampling variability of the parameter estimates from the two *incomplete data* methods was larger than the sampling variability of the standard ML estimates due to the increased uncertainty from missingness across all conditions. Although the sampling variability of the factor loading estimates were essentially indistinguishable between FIML and MI, we could observe a recognizable pattern that the sampling variance of the MI estimates for factor loadings could become smaller (to a very small degree, though) than that of the FIML estimates under the conditions of nonzero misfit and 15% missingness. On average, however, the relative increase in sampling variability for MI estimates for all parameter estimates specified in the posited model were higher than the relative increase in sampling variability for MI estimates, that is, $\text{ARIV}_{MI} > \text{ARIV}_{FIML}$ in general. This pattern of results suggests that the loss of efficiency for the MI estimates may become larger at a faster rate due to the use of a more complex saturated normal model for imputations along with posterior uncertainties than the loss of efficiency for the FIML estimates due to the use of a misspecified yet parsimonious model. Such patterns became more obvious as the percent of missingness increased.

The combined measure of bias and variance in parameter estimates or RMSEs revealed that RMSEs for the MI estimates tended to be essentially identical to or smaller than those for the FIML estimates across all conditions. Further study seems to be warranted regarding the effect of the interplay between the parsimony and accuracy of the imputation model on the relative increase in sampling variability for the two *incomplete data* methods.

With respect to the chi-square test statistics, we observed a consistent pattern showing that the MI-based chi-squares were smaller than the FIML-based chi-squares. Similar observations have been reported in Enders and Mansolf (2018, Table 4, Table 5) and Lee and Cai (2012, Table 2), where the pooled chi-square statistics exhibited a slightly lower empirical power than the FIML-based chi-squares. Enders and Mansolf (2018) conjectured that the lower power of the pooled chi-squares was due to the loss of efficiency from the use of a more complex (saturated) model for imputations under MI. However, they also called for further research to identify additional factors at play.

According to our theoretical predictions and the observed patterns reported in Table 3, the MI parameter estimates tend to be closer to the ML parameter estimates than FIML, and thus it would be natural for one to expect that the MI-based chi-squares should tend to be closer (i.e., larger) than the FIML-based chi-squares. However, this expectation was not verified in our stimulation study. One plausible reason could be the overestimation of ARIV by r_3 that is used as an adjustment term for T_{MI} in Equation 13.[15] A close inspection of Table 9 and Table 5 reveals that as the percent of missingness increases, the degree of overestimation of ARIV for MI by r_3 tends to increase, which, in turn, tends to yield the MI-based chi-squares smaller than the FIML-based chi-squares. It is important to note that the replacement of r_3 with ARIV in Equation 13 would result in the increment of T_{MI} to the extent that the MI-based chi-squares become larger than the FIML-based chi-squares, and therefore closer to T_{ML}.[16]

As such, further research is warranted to develop a more accurate estimate of ARIV than r_3. In fact, r_3 was suggested as an alternative to $r_1 = (1 + M^{-1})tr(B_M\bar{V}_*^{-1})/q$ (Schafer, 1997; see Appendix) with B_M and $\bar{V}_*$ as defined in Equation 12. Therefore, r_1 might be a more accurate estimate of ARIV under MI. The use of r_1 for T_{ML}, however, would overturn the very reason for the development of r_3 (Meng & Rubin, 1992). Moreover, it should be kept in mind that r_1 assumes that the fraction of missing information for all parameters are equal (Schafer, 1997). Therefore, there is no reason to expect that r_1 would be a better choice than r_3 for T_{ML}. In a different context where the residual-based test statistic under MI, T_{BM} (Lee & Cai, 2012) is under consideration, it would be worthwhile to investigate the performance and utility of r_1 as an adjustment term because T_{BM} actually requires the computations of B_M and $\bar{V}_*$.

Interestingly enough, incremental fit indices such as CFI and TLI exhibited such patterns that MI-based estimates were closer to the *complete data* estimates than the FIML estimates. However, researchers should not hasten to conclude that MI-based CFI or TLI should be used instead of FIML estimates, because we the authors believe that the patterns we observed just happened to occur because the bias in r_3

[14] To verify our reasoning, we conducted two simulation studies where a single-factor structure for three indicators was considered. In the first study, the three-factor loadings were .9, .6, and .3, respectively, and the unique variances were .19, .64, and .91, respectively. The factor variance was fixed to 1.0. Under both MCAR and MAR, missingness occurred in X3 only. The hypothesized model was the parallel-tests model where the equality constraints were imposed both on the three factor loadings and on the unique variances. A random sample of size equal to 500 ($N = 500$) was generated, based on which the complete data ML estimate for the factor loading came out as 0.548. Under this condition of 50% missingness, the MI estimate was 0.546 with 100 imputations. The two-stage EM estimate (Savalei & Bentler, 2009), which can be considered a deterministic version of MI estimation, was 0.545. The FIML estimates were obtained in two ways: Using Equations 7 and 17 with $M = 1{,}000$, the FIML estimates were 0.599 and 0.601, respectively. In the second study, the magnitudes of the three factor loadings for X1, X2, and X3 were reversed and they were .3, .6, and .9, respectively. Under the same condition as in the first study, the *complete data* ML estimate, the MI estimate and the two-stage EM estimate were 0.529, 0.522 and 0.523, respectively. The two FIML estimates based on Equations 7 and 17 with $M = 1{,}000$ were 0.486 and 0.486, respectively. In the two simulation studies, it was observed that the MI or two-stage estimates remain relatively unaffected, while the FIML estimates went up or down depending on the sizes of the factor loadings for the indicators without missingness.

[15] Of course, this will not be the sole factor that affects the underpowered nature of MI-based chi-square overall goodness-of-fit testing, and Enders and Mansolf (2018) offered an explanation of when T_{MI} can become nonsensical in an SEM context.

[16] We appreciate two anonymous reviewers for making this comment, which led the authors to have a deeper understanding and offer a coherent explanation of the patterns observed in Tables 7 and 10.

somehow cancelled out the underestimation of chi-squares for both the posited model and the baseline model. This offers another reason for finding a less biased estimate of ARIV under MI.

Our study is not without limitations. First, we only focused on fit indices based on the chi-square test statistics (i.e., RMSEA, CFI, and TLI). It would be interesting to examine the behaviors of other SEM fit indices under missing data, such as the standardized root mean squared residual (SRMR; Bentler, 1995; Jöreskog & Sörbom, 1988; Maydeu-Olivares, 2017). The SRMR is preferable to other unstandardized measures of model misfit (e.g., RMSEA) as it is a standardized effect size measure and can be crudely interpreted as the average standardized residual covariance (Maydeu-Olivares, 2017; Shi et al., 2018). However, methodologically, it is not clear what would be the best approach to compute the SRMR under missing data. Future studies are needed to investigate this issue further.

In addition, our findings are based on normally distributed data and on a rather narrow notion of imperfect models: *incompleteness* in representing the full complexities of the true generating models (MacCallum, 2003; Meehl, 1990). A more general notion of imperfect models has been referred to as "model error" (MacCallum & Tucker, 1991), "approximation discrepancy" (Cudeck & Henly, 1991), or "adventitious error" (Wu and Browne, 2015), all of which represent fundamental contradictions between the model and the real world. This general lack of fit could manifest itself, for example, when adopting linear models to represent nonlinear relationships and/or when invoking a normality assumption to describe non-normal data (MacCallum, 2003; Meehl, 1990). Implications of the existence of such model errors for our use of models (in analysis of *complete data*) have been systematically studied by Briggs and MacCallum (2003); MacCallum et al. (2001); Tucker et al. (1969), and more recently by Wu and Browne (2015). For future research, systematic examinations of relative performance of missing data techniques are warranted in SEM under more general conditions with non-normally distributed data (cf. Yuan et al., 2012) and categorical indicators (Jia & Wu, 2019; Shi et al., 2020).

Notwithstanding these limitations, our work can provide a few recommendations. It is well known among applied researchers that FIML and MI would produce essentially equivalent results when both procedures are implemented in comparable ways. However, fewer researchers are aware that the equivalence between FIML and MI depends on the important assumption that the model is correctly specified. Under more realistic scenarios with misspecified SEM models, which approach should be used?

We recommend using both. As mentioned earlier, there is currently no reason to prefer one method over the other from an applied user's convenience viewpoint. Both methods can handle missing data and model estimation in a single step. Moreover, by having the results obtained from the two estimators, one can look into the patterns of the results as a whole and gain further insight in the process of model fit evaluation. Specifically, our results indicated that the difference between FIML and MI in terms of the parameter estimates become more pronounced as the level of model misspecification increases, especially when the percentage of missingness is large (e.g., ≥25%). Therefore, if the results obtained from FIML and MI are close to each other, the researchers gain confidence that the findings are robust to the presence of missing data. On the other hand, when FIML and MI results diverge to a large extent, the discrepancy could imply that the fitted model is likely to be severely misspecified, and thus, researchers should be cautious in interpreting the results.

If an applied user must decide for some reason to choose one estimator over the other, they would be well-advised to be aware of our findings that FIML-based parameter estimates tend to be affected by the level of misfit of the posited model as well as the missing data patterns, while MI-based estimates tend to stay close to the standard ML estimates regardless of the level of model misfit and/or missing data patterns. Our findings also show that RMSEs around the standard ML estimates tend to be smaller for the MI-based estimates across all nonzero misfit conditions. Therefore, it can be said that the MI estimator is more recommendable than the FIML estimator, at least for the purpose of obtaining parameter estimates that tend to be closer to the *complete data* estimates.

In practice, applied users of SEM typically inspect model fit indexes such as RMSEA, CFI, and TLI before interpreting the parameter estimates of interest. Our findings showed that the two estimators exhibited different behaviors in model fit evaluations in relation to the *complete data* analysis. That is, absolute fit indexes such as chi-squares and RMSEA based on MI appear to paint a rosier picture of the plausibility of the model (i.e., smaller than FIML), whereas relative fit indexes such as CFI and TLI based on FIML appear to paint a rosier picture of the plausibility of the model (i.e., larger than MI). Therefore, if, for some reason, an applied user decides to choose one and only one method over the other, our findings recommend that FIML users should take FIML-based RMEA more seriously than FIML-based CFI/TLI, while MI users should take MI-based CFI/TLI more seriously than MI-based RMSEA in the process of evaluating the goodness-of-fit of posited models. It may well be the case that one can be tempted to pay selective attention to certain fit indexes within a given estimator. However, knowing that MI-based CFI/TLI and FIML-based RMSEA are closer to the *complete data* results would (and should) certainly help the applied researcher to evaluate the model fit with more objective criteria.

References

Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological Methodology*, *17*(1), 71–103. https://doi.org/10.2307/271029

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

Asparouhov, T., & Muthén, B. (2010). *Chi-square statistics with multiple imputation* [Technical report]. http://www.statmodel2.com/download/MI7.pdf

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (1995). EQS 5 [Computer program]. Multivariate Software Inc.

Bodner, T. E. (2006). Missing data: Prevalence and reporting practices. *Psychological Reports*, *99*(3), 675–680. https://doi.org/10.2466/PR0.99.3.675-680

Box, G. E. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, *74*(365), 1–4. https://doi.org/10.1080/01621459.1979.10481600

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation.

Multivariate Behavioral Research, *38*(1), 25–56. https://doi.org/10.1207/S15327906MBR3801_2

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Press Publications.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Cai, L., & Lee, T. (2009). Covariance structure model fit testing under missing data: An application of the supplemented EM algorithm. *Multivariate Behavioral Research*, *44*(2), 281–304. https://doi.org/10.1080/00273170902794255

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), 330–351. https://doi.org/10.1037/1082-989X.6.4.330

Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika*, *57*(3), 357–369.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. *Psychological Bulletin*, *109*(3), 512–519. https://doi.org/10.1037/0033-2909.109.3.512

Enders, C. K. (2001). The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352–370. https://doi.org/10.1037/1082-989X.6.4.352

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

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*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Enders, C. K., & Mansolf, M. (2018). Assessing the fit of structural equation models with multiply imputed data. *Psychological Methods*, *23*(1), 76–93. https://doi.org/10.1037/met0000102

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*(4), 409–420. https://doi.org/10.1007/BF02296204

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, *10*(1), 80–100. https://doi.org/10.1207/S15328007SEM1001_4

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Graham, J. W. (2012). *Missing data: Analysis and design*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-4018-5

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*(2), 197–218. https://doi.org/10.1207/s15327906mbr3102_3

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). Sage.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. https://doi.org/10.1037/1082-989X.11.4.323

Huebner, E. S. (1991). Initial development of the student's life satisfaction scale. *School Psychology International*, *12*(3), 231–240.

Jamshidian, M., & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics*, *24*(1), 21–24.

Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods*, *51*(5), 1–19. https://doi.org/10.3758/s13428-018-1187-4

Jiang, X., Shi, D., Topps, A. K., & Archer, C. M. (2020). From family support to goal-directed behaviors: examining the mediating role of cognitive well-being factors. *Journal of Happiness Studies*, *21*(3), 1015–1035.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7. A guide to the program and applications* (2nd ed.). International Education Services.

Kaplan, D. (1995). The impact of BIB spiraling-induced missing data patterns on goodness-of-fit tests in factor analysis. *Journal of Educational and Behavioral Statistics*, *20*(1), 69–82.

Kenward, M. G., & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, *13*(3), 236–247. https://doi.org/10.1214/ss/1028905886

Lee, T., & Cai, L. (2012). Alternative multiple imputation inference for mean and covariance structure modeling. *Journal of Educational and Behavioral Statistics*, *37*(6), 675–702. https://doi.org/10.3102/1076998612458320

Li, K. H., Raghunathan, T. E., & Rubin, D. B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association*, *86*(416), 1065–1073. https://doi.org/10.2307/2290525

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley. https://doi.org/10.1002/9781119013563

Longford, N. T. (2006). *Missing data and small-area estimation: Modern analytical equipment for the survey statistician*. Springer Science & Business Media.

MacCallum, R. C. (2003). 2001 Presidential Address: Working with imperfect models. *Multivariate Behavioral Research*, *38*(1), 113–139. https://doi.org/10.1207/S15327906MBR3801_5

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*(3), 502–511. https://doi.org/10.1037/0033-2909.109.3.502

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611–637. https://doi.org/10.1207/S15327906MBR3604_06

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, *82*(3), 533–558. https://doi.org/10.1007/s11336-016-9552-7

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Meng, X. L., & Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, *86*(416), 899–909. https://doi.org/10.1080/01621459.1991.10475130

Meng, X. L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, *79*(1), 103–111. https://doi.org/10.1093/biomet/79.1.103

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (Sixth edition).

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*(1), 190–204. https://doi.org/10.1006/jmps.1999.1283

Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and application. In L. M. Lecam, J. Neyman, & E. L. Scott (Eds.), *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability* (pp. 697–715). University of California Press.

Peng, C. Y. J., Harwell, M., Liou, S. M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. Sawilowsky (Ed.), *Real data analysis* (pp. 31–78). Information Age Publishing, Inc.

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489. https://doi.org/10.1080/01621459.1996.10476908

Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, *15*(4), 352–367. https://doi.org/10.1037/a0020143

Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling*, *16*(3), 477–497. https://doi.org/10.1080/10705510903008238

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press. https://doi.org/10.1201/9781439821862

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Shi, D., Lee, T., Fairchild, A. J., & Maydeu-Olivares, A. (2020). Fitting ordinal factor analysis models with missing data: A comparison between pairwise deletion and multiple imputation. *Educational and Psychological Measurement*, *80*(1), 41–66.

Shi, D., Maydeu-Olivares, A., & DiStefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multivariate Behavioral Research*, *53*(5), 676–694.

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Systat, Inc.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540. https://doi.org/10.1080/01621459.1987.10478458

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*(4), 421–459. https://doi.org/10.1007/BF02290601

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics*, *3*(3), 163–195. https://doi.org/10.1214/aoms/1177732885

Wu, H., & Browne, M. W. (2015). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *80*(3), 571–600. https://doi.org/10.1007/s11336-015-9451-3

Yuan, K. H., Yang-Wallentin, F., & Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. *Sociological Methods & Research*, *41*(4), 598–629. https://doi.org/10.1177/0049124112460373

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, *44*(1), 41–61. https://doi.org/10.1006/jmps.1999.1276

Appendix

Matrix Definition of the Relative Increase in Variance

A somewhat technical derivation of the matrix expression of the increase in variance due to missingness can be found in Rubin (1987, Chapter 3), Meng and Rubin (1991), Li et al. (1991), and Cai and Lee (2009). Here, for the convenience of the reader, we re-derive the matrix definition of the relative increase in variance (RIV) due to nonresponse using the notations employed in the present article. To this end, the general expression of the *observed-data* log-likelihood can be written as

$$\log L(\vartheta \mid Y_{obs}) = \log L(\vartheta \mid Y_{com}) - \log P(Y_{mis} \mid Y_{obs}, \vartheta) \tag{A.1}$$

where $\log L(\vartheta \mid Y_{com})$ is the *complete data* log-likelihood and the second term on the right-hand side in Equation A.1 is the logarithm of the predictive distribution of the missing data given ϑ, which conveys the evidence about ϑ available in Y_{mis} over and above that in Y_{obs} (Schafer, 1997). From Equation A.1, one can arrive at the *missing information principle* due to Orchard and Woodbury (1972):

$$J(\vartheta \mid Y_{obs}) = J(\vartheta \mid Y_{com}) - J(\vartheta \mid Y_{mis}) = (I_p - \Lambda) J(\vartheta \mid Y_{com}) \tag{A.2}$$

where

$$\begin{aligned} J(\vartheta \mid Y_{obs}) &= -\frac{\partial \log L(\vartheta \mid Y_{obs})}{\partial \vartheta \, \partial \vartheta'}, \\ J(\vartheta \mid Y_{com}) &= -\frac{\partial \log L(\vartheta \mid Y_{com})}{\partial \vartheta \, \partial \vartheta'}, \\ J(\vartheta \mid Y_{mis}) &= -\frac{\partial \log P(Y_{mis} \mid Y_{obs}, \vartheta)}{\partial \vartheta \, \partial \vartheta'}, \\ \Lambda &= J(\vartheta \mid Y_{mis}) J(\vartheta \mid Y_{com})^{-1}, \end{aligned}$$

and $\boldsymbol{I}_p$ is a $p \times p$ identify matrix. In the missing data literature, the matrix $\boldsymbol{\Lambda}$ is a matrix representation of the *fraction of missing information* (Little & Rubin, 2002; Rubin, 1987; Schafer, 1997). From Equation A.2, it can be shown that

$$\hat{V}_{obs} = \hat{V}_{com}(I_p - \Lambda)^{-1} = \hat{V}_{com}(I_p + \nabla)$$

where $\nabla = \Lambda(\mathrm{I} - \Lambda)^{-1}$. It follows that

$$\nabla = \hat{V}_{com}^{-1}\hat{V}_{obs} - I_p = \hat{V}_{com}^{-1}(\hat{V}_{obs} - \hat{V}_{com}) \tag{A.3}$$

That is, ∇ quantifies the RIV due to nonresponse compared with the sampling variability that would have resulted had there been no missing data. It naturally follows that the average relative increase in variance (ARIV) across q parameters is defined as

$$\bar{\nabla} = \mathrm{tr}\left[V_{com}^{-1}(V_{obs} - V_{com})\right]/q. \tag{A.4}$$

The scalar definition of Equation A.3 or A.4 for a single parameter estimate is given and well explained in Enders (2010, p. 226).

An estimate of Equation A.4 due to Rubin (1987, Chapter 3) is given as

$$r_1 = (1 + M^{-1}) tr\left(B_M \bar{V}_*^{-1}\right)/q \tag{A.5}$$

with M, B_M, and $\bar{V}_*$ defined as before. An alternative estimate of Equation A.4 that is asymptotically equivalent to Equation A.5 was proposed by Meng and Rubin (1992)

$$r_3 = \frac{M+1}{d(M-1)}(\bar{T} - \tilde{T}) \tag{A.6}$$

with M, d, $\bar{T}$, and $\tilde{T}$ defined as before. The two likelihood ratio statistics $\tilde{T}$ and $\bar{T}$ represent the average of the likelihood ratio statistics evaluated at the pooled estimates and the average of the likelihood ratio statistics evaluated with the imputation-specific estimates, respectively.

Received May 26, 2019
Revision received September 29, 2020
Accepted October 14, 2020 ■