



# An evaluation of methods to handle missing data in the context of latent variable interaction analysis: multiple imputation, maximum likelihood, and random forest algorithm

Tacksoo Shin<sup>1</sup> · Jeffrey D. Long<sup>2</sup> · Mark L. Davison<sup>3</sup>

Received: 3 January 2022 / Revised: 14 July 2022 / Accepted: 25 July 2022 /  
Published online: 11 August 2022

© The Author(s) under exclusive licence to Japanese Federation of Statistical Science Associations 2022

## Abstract

The objective of this study was to examine the performance of the two most popular missing data methods (i.e., multiple imputation and maximum likelihood), as well as newly developed machine learning framework based on random forest algorithm for missing data under various reserach conditions. The design of the simulation study included random and non-random missingness (i.e., MCAR, MAR, and MNAR), small samples, and different levels of missing rates. All statistical inferences were investigated using latent variable interaction modeling. Consistent with the missing data literature, the combined effects of small sample sizes, higher missing rates, and non-ignorable missingness along with complicated modeling structure adversely affected the accuracy of statistical inferences. Although there is a possibility for overparameterization, it is a good way to select MI when convergence is concerned. If the primary goal of research is to investigate the relationship between variables as in many studies, ML would be attractive. MF presented similar performance compared to MI and ML across all research conditions and outperformed when estimating the variability of parameter estimates. Other practical issues pertaining to the missing data methods were also discussed.

**Keywords** Missing data · Multiple imputation · Maximum likelihood estimation · Random forest algorithm · Latent interaction model

---

✉ Tacksoo Shin  
shin16@mju.ac.kr

<sup>1</sup> Department of Youth Education and Leadership, Myongji University, Seoul, South Korea

<sup>2</sup> Department of Psychiatry, University of Iowa, Iowa City, USA

<sup>3</sup> Department of Educational Psychology, University of Minnesota at Twin-Cities, Minneapolis, USA

## 1 Introduction

One of the most common problems that researchers face in data analysis is dealing with missingness. If missing values are not properly managed, inaccurate inferences about the data are obtained. To date, a wide variety of missing data studies (Allison, 2006; Arbuckle, 1996; Baraldi & Enders, 2013; Collins et al., 2001; Enders, 2001a, 2006, 2010; Ho et al., 2001; Ji et al., 2018; Larsen, 2011; Little & Rubin, 2002; Savalei et al., 2005; Savalei & Rhemtulla, 2012; Schafer & Graham, 2002; Shin et al., 2009; Yuan et al., 2012a, 2012b; Wothke, 2000) have proposed several missing data methods which can reduce bias as well as the efficiency and sensitivity of the statistical analysis and examined their performance under various research conditions (e.g., longitudinal design, quasi- or true-experimental designs, and survey nonresponses). According to the missing data literature, maximum likelihood (ML) estimation and multiple imputation (MI) are superior to deletion methods (i.e., listwise and pairwise deletion). ML and MI are more efficient and yield less biased statistical inferences under missing completely at random (MCAR) and missing at random (MAR)<sup>1</sup> (Arbuckle, 1996; Enders, 2001a, 2001b; Gold & Bentler, 2000; Graham et al., 1996; Rubin, 1996; Savalei et al., 2005; Schafer, 1997, 1999; Schafer & Graham, 2002; Schafer & Olsen, 1998; Sinharay et al., 2001; Wothke, 2000). However, analyzing data with both ML and MI under the missing not at random (MNAR)<sup>2</sup> assumption can lead to relatively extreme bias when the missing mechanism is truly MNAR (Shin et al., 2017).

Although MI and ML have been considered promising approaches, Tang and Ishwaran (2017) pointed out that these methods can lead to overparameterization (Rubin, 1996) and computational issues (Loh & Wainwright, 2011; e.g., the occurrence of non-convexity due to missing data). Also, these techniques can be inefficient to implement in settings where mixed data (i.e., data having both nominal and categorical variables) and the nonlinearity of variables are analyzed (Aittokallio, 2009; Doove et al., 2014; Liao et al., 2014). In recent years, a machine learning method for missing data imputation [i.e., random forest (RF) algorithm for missing data; Breiman, 2001] has been proposed. RF originated from classification and regression trees (CART), which are commonly used in data mining. CART constructs trees by conducting binary splits of certain predicting variables of the data, aiming to produce homogeneous subsets of the data concerning the outcome (Tang, 2017). However, a single CART has a weakness in terms of instability in prediction; thus, Breiman (1996) used bagging (bootstrap aggregation) to enhance the tree methodology. In bagging, several trees are fit to bootstrapped or subsampled data. Averaged values or majority votes of the predictions produced by each tree are used as predictions (Hapfelmeier, 2012). As an extension of bagging, each split is searched for in a subset of variables in the RF (Breiman, 2001; Breiman and Cutler, 2002). A popular choice is to randomly select the square root

<sup>1</sup> Denoting complete data as  $Y_{\text{com}}$  and it partitioned as  $Y_{\text{obs}}$  and  $Y_{\text{mis}}$ . In missing completely at random (MCAR), the probability of an observation being missing ( $R$ ) does not depend on observed ( $Y_{\text{obs}}$ ) and unobserved ( $Y_{\text{mis}}$ ) measurements ( $P(R|Y_{\text{com}}) = P(R)$ ), whereas missing at random (MAR) indicates that missingness depends on observed characteristics of the individuals but not on the missing values ( $P(R|Y_{\text{com}}) = P(R|Y_{\text{obs}})$ ) (Enders, 2010; Rubin, 1996; Shin et al., 2017).

<sup>2</sup> Data are missing not at random (MNAR) when the probability of missing data on a variable  $Y$  can depend on other variables (i.e.,  $Y_{\text{obs}}$ ) as well as on the underlying values of  $Y$  itself (i.e.,  $Y_{\text{mis}}$ )  $p(R|Y_{\text{obs}}, Y_{\text{mis}})$ .

of the number of available predictors as a candidate for the split (Díaz-Uriarte & de Andrés, 2006). This enables a more diverse set of variables to contribute to the joint prediction of a RF, which results in improved prediction accuracy (Hapfelmeier, 2012). The advantages of RF include handling mixed types of missing data and addressing nonlinearity including interactions, as well as scaling to high dimensions (Ishioka, 2013; Tang & Ishwaran, 2017).

Because RF is a newly developed missing data method, no studies have examined the performance of MI, ML and RF for treating missing data. In addition, there is no information about MI and ML comparing them with RF with combined research conditions of missingness, small samples, and nonlinear relationships between variables. As mentioned above, incomplete data are a common feature in applied research circumstance, and the treatment of missing data is one of the prominent topics. Another potential challenge is an inadequate sample size leading to overestimation of the potential non-centrality parameters (Herzog & Boomsma, 2009) and fit tests (Lee & Song, 2004). Although robust techniques (e.g., robust ML) yield more precise estimates of standard errors (SEs) and make corrections for fit statistics (Arminger & Sobel, 1990; Gold et al., 2003; Yuan & Bentler, 1998, 2000), some problems (e.g., convergence failure, inaccurate parameter estimates, etc.) especially in covariance structure methodologies may not be fully resolved. Unfortunately, the sample sizes in experimental and longitudinal studies, even in observational studies driven by individual researchers are usually small.

Along with missingness and small samples, nonlinear relationships such as the interaction effect between variables would threaten the statistical inference. For instance, even if observed indicator and latent exogenous variables are normally distributed, the multivariate distribution of indicators of latent endogenous variables deviates substantially from normality (Moosbrugger et al., 1997). Klein and Moosbrugger (2000, p.458) described,

Ignoring the nonnormal distribution of indicator variables, the application of an estimation procedure can lead to different statistical problem: either, if an estimation procedure is used under the assumption of multivariate normality, it must be robust against the type of nonnormality implied by latent interaction. Or, if an asymptotical distribution-free estimation method is used, it does not exploit the specific distributional characteristics of latent interaction models, which might lower the method's efficiency and power, especially when sample size is not very high.

Numerous substantive theories within the social and behavioral sciences hypothesize nonlinear effects including interaction, quadratic effects or both between multiple independent and dependent variables (Ajzen, 1987; Cronbach & Snow, 1977; Karasek, 1979; Lusch & Brown, 1996; Snyder & Tanke, 1976). For example, Ganzach (1997) hypothesized a complex interactive and quadratic relationship between parents' educational level and children's educational expectations. Likewise, estimating the interactive effect of two independent variables on an outcome variable is an important concern in the social sciences, which has resulted in a plethora of statistical approaches (Lin et al., 2010). Various approaches have been proposed to manage interaction effects

(e.g., unconstrained approach, mean-centered [un]constrained approach, orthogonalizing approach, etc.; Algina & Moulder, 2001; Jöreskog & Yang, 1996; Little et al., 2006; Kenny & Judd, 1984; Marsh et al., 2004, 2006). For latent variable interaction analysis, newer distribution analytic approaches (Klein & Moosbrugger, 2000; Klein & Muthén, 2007) are easy to use and provide parameter estimates that can be more efficient, yielding greater statistical power, particularly with more complex models (Kelava et al., 2011).

The objective of this study was to examine the performance of MI, ML and RF under various research conditions. The design of the simulation study included random and nonrandom missingness (i.e., MCAR, MAR, and MNAR), small samples, and different levels of missing rates. All statistical inferences were investigated using latent variable interaction modeling. The convergence rates and degrees of biases in the parameter estimates and corrected ML standard errors were based on the evaluation of which missing data method is preferred to others.

## 2 Methods for missing data

### 2.1 Multiple imputation (MI)

There are two simple approaches to handling missing values: (1) delete the missing cases (i.e., listwise and pairwise deletion), or (2) fill in the missing value with a plausible one and then, continues with the statistical method as if the data are completely observed. MI falls in the latter case. Replacing missing values with a single value cannot adequately reflect the uncertainty of missingness, and thus, yield inaccurate estimates (Little, 1992). Rather, MI generates several possible values for missing observations in the data to obtain a set of parallel completed datasets (Rubin, 1987, 1996; Schafer, 1999). MI analysis consists of three phases: the imputation-posterior (I–P) phase, the analysis phase, and the pooling phase (Enders, 2010). For the I–P phase, every aspect of MI is rooted in the Bayesian paradigm, which is an MCMC algorithm (Allison, 2010; Baraldi & Enders, 2010; Croy & Novins, 2005; Kenward & Carpenter, 2007; Lee & Song, 2004; Peugh & Enders, 2004; Taylor & Zhou, 2009). In the first step of the I–P phase, missing scores are imputed by certain statistical models (e.g., the stochastic regression procedure). From this Bayesian perspective, these imputed values are randomly drawn from a distribution of plausible replacement values, given the observed mean vector and covariance matrix, as well as the preceding mean vector and covariance matrix from the previous P-step. The P-step is essentially a standalone Bayesian analysis that describes the posterior distributions of the mean vector and covariance matrix (Enders, 2010, Chapter 7). From these imputed data, the P-step is used to simulate the posterior population mean and covariance, and then generate alternate estimates of the mean vector and covariance matrix, which are drawn from the respective posterior distributions. After drawing new values, the subsequent I-step uses updated estimates to construct a different set of imputations (Enders, 2010; Rubin, 2004; Schafer, 1997; Shin et al., 2017). These two steps are repeated to obtain multiple sets of data including unique estimates of missing values.

After obtaining multiple sets of imputed data ( $m$  sets) through thousands of iterations, the completed data are then analyzed in the analysis phase. Schafer (1997) noted that  $m = 5$  imputed datasets typically result in unbiased estimates. If there are five imputed datasets from the I-P phase, researchers perform the targeted statistical analysis on each data set separately and then obtain five sets of parameters. Several estimation techniques (e.g., ML, Bayesian, and weighted least squares estimator) are available in this phase. The last step of MI is the pooling phase, which combines all estimates of parameters from  $m$  sets of data based on Rubin (1987, 2004) and Schafer's (1997) rule. Schafer and Graham (2002) noted that MI is an alternative to ML estimation and is the other state-of-the-art missing data technique that methodologists currently recommend. Both MI and ML are more efficient than deletion methods and yield less biased statistical inferences under MCAR and MAR (Arbuckle, 1996; Enders, 2001a, 2001b; Savalei & Bentler, 2005; Schafer & Graham, 2002; Wothke, 2000). Several studies (e.g., Gold & Bentler, 2000; Graham et al., 1996; Rubin, 1996; Schafer, 1997; Schafer & Olsen, 1998; Sinharay et al., 2001) have reported that MI has performed well in handling missing data in the structural equation modeling context. The MI approach is well documented and has exhibited good qualities when the imputation model approximates the true missing data mechanism (Duncan & Duncan, 1994; Rubin, 1996, 2004).

## 2.2 Maximum likelihood estimation (ML)

With the wise acceptance of the principle of drawing inferences from a likelihood function, ML algorithms for use with missing data are becoming commonplace. In the likelihood formulas, candidate parameter estimates are repeatedly replaced, and for each replacement, the likelihood (or log-likelihood) value is computed using the sample data (Anderson, 1957; Dempster et al., 1977; Finkbeiner, 1979; Hartley & Hocking, 1971; Wilks, 1932). Then, it identifies the set of parameter estimates that produce the highest likelihood. Although the likelihood function can be maximized directly without iteration especially under monotone pattern missing data (Little & Rubin, 2002, Chapter 7), missing data analyses generally require iterative optimization algorithms, with the well-known ML approach being the expectation and maximization (EM) algorithm (Enders, 2010; Shin et al., 2017). The EM algorithm uses a two-step iterative procedure in which missing observations are filled in or imputed, and unknown parameters are subsequently estimated (Dempster et al., 1977). In the first step (expectation [E]-step), missing values are replaced with the conditional expectation of the missing data given the observed data and an initial estimate of the covariance matrix. The purpose of the E-step is to fill in the missing values in a manner that resembles stochastic regression imputation (Enders, 2010). The second step (maximization [M]-step) uses the expected value of the sum of variables to estimate the population mean and covariance. The process cycles back and forth until the estimates do not change substantially (Pigott, 2001).

Traditional missing data methods, such as listwise and pairwise deletion, require the strict assumption that the missing data are MCAR for valid inferences, whereas more accurate results are obtained by both MI and ML with the weaker assumption

(i.e., MAR). Even under MCAR, ML yields estimates with lower sample variability than others (Enders, 2001a). At the same time, both MI and ML share two statistical assumptions: the multivariate normality of the joint distribution of the data and ignorable missing data mechanisms (i.e., MCAR and MAR). Multivariate normality plays an integral role in every aspect of ML analysis, while distributions of real data tend to be skewed and have heterogeneous marginal kurtosis (Micceri, 1989). Thus, several extensions and variations of the EM algorithm to deal with the issue of non-normal missing data have been developed. King et al. (2001) investigated a method for improving the convergence of these EM-type algorithms. In addition, EM can be applied to non-normal missing data in a way that adds estimated standard errors and model fit statistics to the ML parameter estimates (Arminger & Sobel, 1990; Enders, 2001b; Gold et al., 2003; Savalei & Bentler, 2009; Yuan & Bentler, 1998, 2000). However, Shin et al. (2009) noted that the combined effects of non-normality, small samples and non-ignorable missingness would still threaten the proper performance of the ML method. Compared to MI, some studies (e.g., Allison, 2006; Enders, 2006; Larsen, 2011; Yuan et al., 2012a, 2012b) insist that ML is better than MI, and ML estimation is more straightforward, whereas others (e.g., Cham et al., 2013; Collins et al., 2001; Ho et al., 2001) suggest that MI is comparable with ML. Recently, Shin et al. (2017) examined ML versus MI for the effect of specifying informative priors in the I–P phase along with Bayesian estimation in the analysis phase. They concluded that ML appears to be preferable to MI in research conditions with small missing samples and multivariate non-normality, regardless of whether strong prior information for the I–P phase analysis is available.

### 2.3 Random forest algorithm (RF) for missing data

According to Tang and Ishwaran (2017, p. 3), there are three general strategies used for RF missing imputation:

- (A) Preimpute the data; grow the forest; update the original missing values using proximity of the data. Iterate for improved results.
- (B) Simultaneously impute data while growing the forest; iterate for improved results.
- (C) Preimpute the data; grow a forest using in turn each variable that has missing values; predict the missing values using the grown forest. Iterate for improved results.

Proximity imputation and on-the-fly-imputation use strategies (A) and (B) respectively. *missForest* (MF), which is considered in this study, uses strategy (C). MF is a nonparametric imputation method applicable to any type of data, as well as nonlinear relations, complex interactions and high dimensionality (Stekhoven, 2016). Data were fitted to the observed part by RF and then predicting the missing data for the dependent variable using a fitted forest (Tang & Ishwaran, 2017). The proximity approach is used to define the closeness between pairs of cases. Breiman (2003) defined the data proximity as follows: the  $(i, j)$  element of the proximity matrix produced by a RF is the fraction of trees in which elements  $i$  and  $j$  fall in the same terminal node. The intuition is that similar observations should be in the same terminal nodes more

often than dissimilar ones (Ishioka, 2013; Liaw & Wiener, 2002). For continuous predictors, the imputed value is the weighted average of the non-missing observations, where the weights are the proximities. For categorical predictors, the imputed value is the category with the largest average proximity and this process is iterated a few times (Ishioka, 2013; Liaw & Wiener, 2002).

After each iteration the difference between the newly imputed data matrix and the previous one is assessed and the stopping criterion is defined such that the imputation process is stopped as soon as both differences become larger once (Stekhoven, 2016). For the continuous variables, the normalized root mean square error is used to evaluate the performance (Oba et al., 2003). Stekhoven and Bühlmann (2012) used the proportion of falsely classified entries over categorical missing values. In both cases, 0 indicates the good performance and bad performance leads to a value of approximately 1. When fitting RF to the observed part of the data for each time, the out-of-bag (OOB) error estimate of RF is given. The performance of this estimation is obtained by averaging the absolute difference between the true imputation error and the OOB imputation error estimate in all simulation runs (Stekhoven & Bühlmann, 2012).

The new algorithm RF has been presented to allow for missing value imputation (Breiman, 2001; Ishioka, 2013; Pantanowitz & Marwala, 2008). A few studies (e.g., Liaw & Wiener, 2002; Shah et al., 2014; Stekhoven & Bühlmann, 2012; Waljee et al., 2013; Tang & Ishwaran, 2017) have suggested that MF is a better method for moderate to high missingness and for some missing mechanisms, including MNAR, than other RF based imputation methods. However, there are almost no studies that compare performance of the most popular and traditional methods (i.e., MI and ML) and MF. Therefore, this simulation study investigated which method among MI, ML and MF is preferred under different missing rates, three types of missing mechanisms, and small to medium sample sizes.

## 2.4 Latent variable interaction modeling

In behavioral research analysis, theory suggests that the effect of a latent exogenous variable on a latent endogenous variable is moderated by a second exogenous variable (Klein & Moosbrugger, 2000). Several researchers have called for methods to estimate latent interaction effects in structural equation models (Aiken & West, 1991; Cohen & Cohen, 1975; Jaccard et al., 1990; Schmitt, 1990). Many approaches for the estimation of interaction models with continuous variables have been developed. One of the most basic techniques is the product indicator (PI) approach developed by Kenny and Judd (1984). It simply identified the latent interaction term as being measured by the products of each latent variable's indicators. Unfortunately, this approach has rarely been used by researchers. One reason is that the PI approach involves the specification of nonlinear parameter constraints that are difficult for researchers to implement (Kelava et al., 2011). To overcome the problem of PI, several other models (e.g., unconstrained approach, extended unconstrained approach), which freely estimate factor loadings, measurement error variances, as well as the variance of the latent nonlinear effect, have been considered (Kelava, 2009; Kelava & Brandt, 2009; Marsh et al., 2004, 2006; Moosbrugger et al., 2009). However, these traditional product indicator approaches

still suffer from the violated assumption of multivariate normally distributed variables. Thus, distribution-analytic approaches have been developed to address the non-normal distributions. Klein and Moosbrugger (2000) developed a latent moderated structural equations (LMS) approach that employs a unique model specification that does not involve PIs. LMS produces asymptotically correct standard errors for nonlinear effects (Kelava et al., 2011). Because this approach becomes computationally (numerically) intensive as the number of nonlinear effects increases, Klein and Muthén (2007) subsequently developed a quasi-maximum likelihood (QML) approach. QML permits the estimation of multiple nonlinear effects with a smaller increase of computational burden by taking a small loss of precision because a “quasi” likelihood (described later) is maximized. It approximates the likelihood of a multivariate non-normally distributed indicator vector using a normal and a conditionally normal distribution. The parameter estimates are obtained using the Newton–Raphson algorithm (Kelava & Brandt, 2009).

In this study, we focus on a LMS approach to estimate parameters with missing scores for small samples. Klien and Moosbrugger (2000) described that in LMS, the density function of the joint indicator vector  $(x, y)$  is represented as a finite mixture of normal densities, and LMS utilizes the model-implied mean vectors and covariance matrices of the mixture components for an iterative estimation of the model parameters with the EM algorithm. They added that LMS takes the non-normality of distribution explicitly into account based on the distribution analysis of the joint indicator vector (p. 467). Several studies (Coenders et al., 2008; Klein & Muthén, 2007; Klein et al. 2009; Little et al., 2006; Wall & Amemiya, 2000) have noted that LMS is superior to other approaches including constrained PI, unconstrained PI, orthogonalizing PI (OPI) developed by Little et al. (2006), and double-mean-centering strategies by Lin et al. (2010).

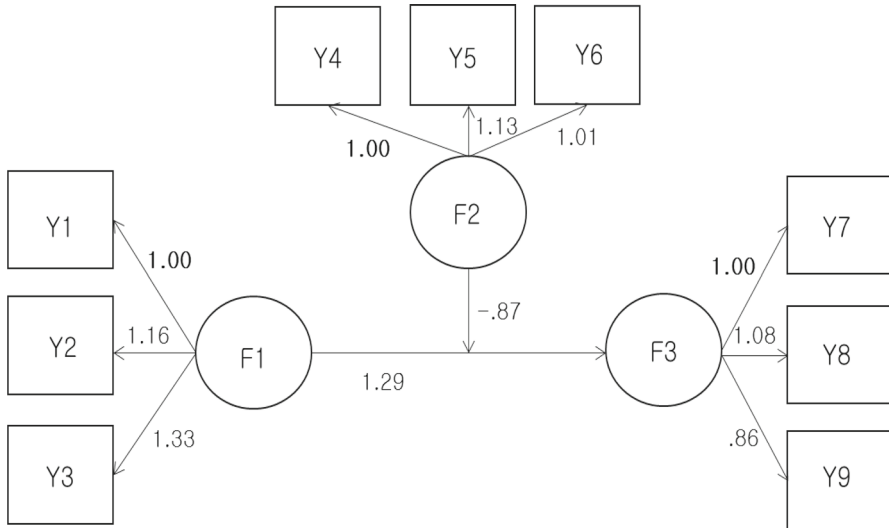
### 3 Methods

In this simulation study, we compared the performance of three missing data methods, MI, ML and MF under various research conditions. We investigated the combined effects of three independent variables (sample size, missing rate, and missing mechanisms) on three dependent variables (convergence rate, bias of parameter and standard error estimates).

#### 3.1 Population model

Data were obtained from Kang and Shin (2015). They explored the direct, indirect and moderation effects among adolescents’ stress, depression, social support, and suicidal ideation. A total of 372 students were randomly selected from four high schools located in the largest urban school district of Korea. The population model for this simulation study reflected some of the properties of the hypothesized model. As seen in Fig. 1, depression (F3) was influenced by stress (F1), and social support (F2) was considered as the moderated variable in the relationship between stress and depression.





**Fig. 1** Population latent variable interaction modeling

For depression, Children’s Depression Scale (CDS) by Kovacs (1983) was utilized. Depression was further subcategorized as loneliness (seven items), helplessness (11 items), and no value (seven items). The daily stress questionnaire, developed by the Korean Youth Policy Institute (2007) was used to measure students’ stress. The instrument consisted of 18 items mapped onto three subordinate constructs: school life stress (six items), stress among peers (six items) and parents (six items). Lastly, the revised Dubow and Ullman’s (1989) Social Support Evaluation Scale for youth was assessed to measure social support. It consisted of peer (five items), family (five items) and teacher (five items) supports. Responses to all scales were selected on a five-point Likert scale to indicate agreement.

According to a previous study (Kang & Shin, 2015), these theory-driven factor structures in all instruments fit well with given set of data. Along with the high-reliability coefficients, these CFA results demonstrated unidimensionality of sub-factors, and thus, scores for each subscale can be calculated by averaging the items from that particular subscale. This indicates a single score represented by the characteristics of each construct. Then, the calculated scores (Y1 through Y9) of depression, stress and social support scales were used for this simulation study. The parameter estimates for factor loadings and regression coefficients including the interaction term (F1 by F2) obtained by the hypothesized model, were adopted as the population parameters. In addition, descriptive information (i.e., mean and variance) of measured variables (Y1–Y9) was set to the starting values when simulating the data. The population information is presented in Table 1, and Fig. 1 displays the path diagram for the population model.

**Table 1** Descriptive information and population parameters

	Descriptive information		Population parameters		
	Mean	Standard deviation		Estimates	Standard deviation
Y1	2.27	0.96	F1–Y1 <sup>a</sup>	1.00	–
Y2	2.46	1.04	F1–Y2	1.16	2.72
Y3	2.59	1.21	F1–Y3	1.33	2.76
Y4	3.57	1.10	F2–Y4 <sup>a</sup>	1.00	–
Y5	3.50	1.00	F2–Y5	1.13	1.08
Y6	3.62	0.94	F2–Y6	1.01	1.01
Y7	1.96	1.42	F3–Y7 <sup>a</sup>	1.00	–
Y8	1.74	1.32	F3–Y8	1.08	1.34
Y9	1.50	1.11	F3–Y9	0.86	1.49
			F1–F3	1.29	4.47
			Interaction–F3	–0.87	4.32

<sup>a</sup>Indicates the fixed parameters

### 3.2 Research design and data generation

For the simulation, the sample sizes were selected to represent both experimental and non-experimental research: 75, 100, 200, 300, and 500. Based on Fig. 1, *Mplus* 8.2 (a *Monte Carlo* option; Muthén & Muthén, 2018) was used to generate the complete data along with the descriptive information (i.e., means and variances of variables; Table 1). Then, three types of missing data mechanisms (MCAR, MAR, and MNAR) were simulated using the multivariate amputation procedure of the *R* programming (*ampute* of *R* function; Schouten et al., 2018; van Burren, 2012). The steps of *ampute* missing generation are followed (Brand, 1999; Brand et al., 2003; van Buuren et al., 2006). First, the amputation procedure starts with the researchers deciding what kind of missing data patterns on desires to establish. The study considered nine combinations of variables with missing values and variables remaining complete, as shown in Table 2. Second, the complete data set is randomly divided into  $k$  subsets based on the number of missing data patterns  $k$ . Then, the missing candidate is defined when calculating the weighted sum score. These weight sum scores are used to determine whether a data cell is missing, and based on its weighted sum score, each candidate receives a probability of being missing for a given variable (Schouten et al., 2018, pp. 2914–2915). For MCAR, the candidates have an equal probability to have missing values. Rather, for MAR and MNAR, the allocation of the probabilities is applied by four logistic distribution functions on the weighted sum scores (van Burren, 2012). For example, for a “right” like mechanism, scoring in one of the higher quantiles should have high missing odds, whereas higher probability values are given to the candidates with low, average, or extreme weighted sum scores, respectively with a left-tailed, centered, and both-tailed (Schouten et al., 2018; van Burren, 2012). A right-tailed type of missing

**Table 2** Missing patterns

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9
Pattern 1	M	M	O	M	O	M	M	M	O
Pattern 2	M	O	O	M	O	O	M	O	O
Pattern 3	O	M	M	O	M	M	O	M	M
Pattern 4	O	M	O	O	M	O	O	M	O
Pattern 5	O	O	M	O	O	M	O	O	M
Pattern 6	M	O	M	M	O	M	M	O	M

*O* observed values, *M* missing values

data was used in this study. Lastly, two levels of missing rates (i.e., 10% and 20%) were considered within R-function “*ampute*”.

### 3.3 Model estimation and analysis

For the simulation analysis of MI and ML, *Mplus* and *MplusAutomation* (Hallquist & Wiley, 2017) within the R programming language (*RStudio*, version 1.1.463) were implemented. Following suggestions from previous MI studies (e.g., Graham et al., 2007; Rubin, 1987; Schafer, 1997; von Hippel, 2007), five imputed data sets per single MI analysis were used. In the ML missing data analysis, factor loadings and regression parameters were analyzed using the EM algorithm and the corrected method (SML; Satorra & Bentler, 1988; Yuan & Bentler, 2000). Similar to ML, the parameters from the MI analysis are also obtained from ML and SML. Using *missForest* package (Stekhoven, 2016; version 1.4), missing values were imputed for MF analysis. The maximum number of iterations to be performed given the stopping criterion and the number of trees to grow in each forest was set to 10 and 100, respectively. Then, the MF parameters were again estimated using ML and SML. The latent interaction regression coefficients were obtained from the LMS technique.

### 3.4 Evaluation criteria

Based on the default criteria (iteration = 1000; convergence criteria = 0.00001) of *Mplus*, the performances of MI, ML and MF were first evaluated by the convergence rate. In the case of MI, the main criterion used in *Mplus* to determine the convergence of the MCMC sequence is based on the potential scale reduction (Asparouhov & Muthén, 2010; Gelman & Rubin, 1992). *Mplus* runs 100 MCMC iterations by default. Then, the convergence criterion is checked every 100th iteration using the default value of 0.05. Second, the average bias in the parameter and standard error (SE) estimates was assessed to measure the degree of bias. The bias statistic was computed as  $\frac{\sum_{i=1}^n (\hat{\theta}_i - \theta)}{n}$  where  $\hat{\theta}_i$  is the corresponding parameter and standard error estimates for each replication  $i$ .  $\theta$  is the true population parameter, and  $n$  is the number of all converged cases for each cell condition. Based on the population distribution of growth

parameters, the standard deviation of the parameters refers to the extent to which the estimates vary from sample to sample. Since our missing samples were random samples from the population, the statistics of these samples should have distributions determined by the population parameters. At this point, the standard errors of the parameter estimates of each sample were also determined by population parameters. With the population standard deviations given in Table 1, population standard errors were separately computed for each sample design (e.g.,  $SE_{N=75} = \frac{SD_{pop}}{\sqrt{75}}$  where  $SD_{pop}$  indicates population standard deviation). Then, biases of corrected SE estimates by three missing data methods were investigated. In addition, the root mean squared relative difference (RMSRD, Alkasawneh et al., 2007; Chiarella et al., 2014; Kroll & Stedinger, 1996) was examined to evaluate the recovery of parameters and SE estimates (i.e., parameter estimation error). This was calculated as  $RMSRD \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(\hat{\theta}_i - \theta)^2}{\theta^2}}$ .

## 4 Results

### 4.1 Proportion of convergence successes

The percentages of convergence by sample sizes, missing mechanisms, and missing rates are given in the Table 3. As expected, the convergence failures increased as the sample size decreased. Even when the completed data sets were generated, sample sizes of 75 failed to obtain the proper solutions more than 30 percent of the times. Boomsma (1985) noted that the seriousness of the non-convergence problem depends heavily on the sample size. Anderson and Gerbing (1984) and Enders and Bandalos (2001) added that sample sizes of 100 or less resulted in higher rates of non-convergence. Missing rates and nonignorable missingness also negatively impacted the convergence (i.e., the effect of the missing mechanism increased as the missing rates increased). Thus, the proportion of convergence was the worst in the condition of sample size = 75, missing rate = 20%, and MNAR. A large amount of missing data can lead to a covariance or correlation matrix that is not positive definite (Arbuckle, 1996; Shin et al., 2009; Wothke, 2000). Moreover, the singularity properties of the information matrix can be frequently found in the presence of MNAR (Copas & Li, 1997; Jansen et al., 2006; Lee, 1993; Rotnitzky et al., 2000). Singularity problems causing negative variance and linear dependency among factors are more likely to occur with MNAR and small sample sizes (Shin et al., 2017). Another reason for the higher rates of non-convergence may originate from the population model and parameters. The computational process for the nonlinear relationship model is more complex and the variances of the latent factor are relatively small. Thus, the structure of the latent interaction model and small effects cause improper results for some replications in both the complete and incomplete simulations.

Among the missing data methods, MI resulted in the highest percentage of admissible solutions across all cell designs. When comparing ML and MF, ML tended to show higher convergence rates especially in unfavorable research conditions (i.e., small samples, MAR and MNAR with higher missing rate percentages). However, the

**Table 3** Convergence rate

Sample size	Complete data	Missing rate = 10%												Missing rate = 20%											
		MCAR				MAR				MNAR				MCAR				MAR				MNAR			
		MI	ML	RF	MI	ML	RF	MI	ML	RF	MI	ML	RF	MI	ML	RF	MI	ML	RF	MI	ML	RF	MI	ML	RF
$N = 75$	68.4	88.2	65.6	68.0	89.3	67.2	64.4	64.4	87.0	64.8	64.8	64.8	84.0	62.4	66.4	66.4	78.0	66.4	66.4	64.4	81.6	65.2	61.2		
$N = 100$	81.6	91.6	81.6	82.8	94.8	75.8	76.4	76.4	90.0	74.4	73.6	73.6	92.4	74.0	72.4	72.4	92.8	72.0	70.8	70.8	93.2	69.6	71.6		
$N = 200$	96.8	100	96.8	96.6	100	95.2	94.8	94.8	100	94.4	96.4	96.4	100	93.6	95.6	95.6	100	94.0	93.6	93.6	100	93.6	93.2		
$N = 300$	99.2	100	99.2	100	100	98.8	99.2	99.2	100	99.2	98.0	98.0	100	98.0	99.2	99.2	100	98.8	97.6	97.6	100	99.6	99.2		
$N = 500$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	

differences between the two missing data methods were relatively small. An interesting result in the analysis of convergence is that MI presented higher convergence rates than even in the analysis of complete data cases. The lowest MI convergence rate was approximately 10% higher than that of the complete data. Rather, ML and MF showed similar proportions of convergence success compared to the complete cases. The superiority of MI over other missing data methods in proportion of convergence success is unclear. A theoretical fundamental form of MI is repeated imputation (Rubin, 1987, pp. 75–76). As noted in the introduction, these repeated imputations could be “superefficient” from the perspective of the data analyst because the imputation use extra information (Rubin, 1996). Although these imputations would effectively provide additional data values (then, contribute to a better estimate), the procedure also leads to reparameterization or overparameterization. In other words, this MI strategy has some advantages, which are much less susceptible to becoming stuck near zero variance parameter values than other algorithms, while it may cause “overdoing when unnecessary or undesirable”.

### 4.2 Bias and recovery of parameter estimates

Only models that converged within iterations and had admissible parameter estimates were included in the evaluation of the parameter estimates. For this reason, screening mechanisms were implemented to identify unreasonable solutions. The indication of non-convergence and unacceptable results for each replication was evident from the error messages of the *Mplus* program. Figures 2, 3, 4 and 5 show the biases of free factor loadings and regression coefficient parameter estimates by sample size, missing rates, missing data mechanisms, and missing data methods. The results of parameter

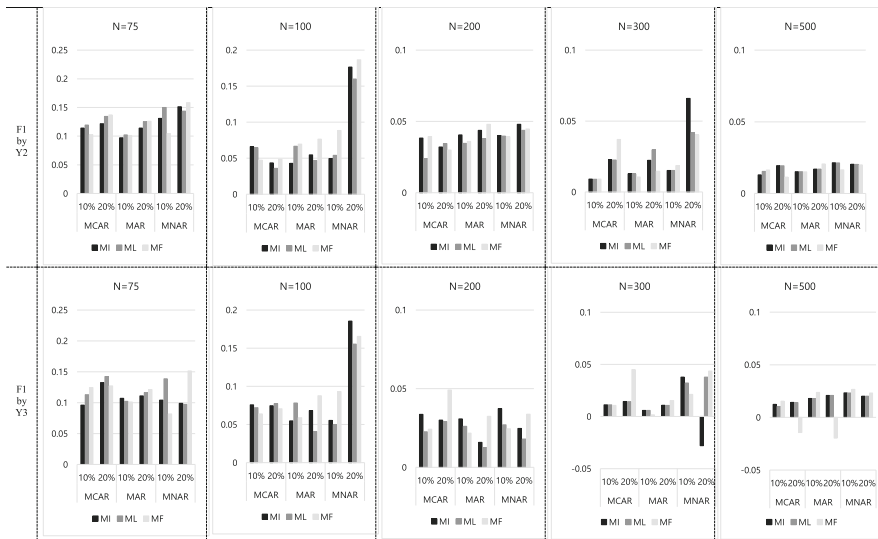


Fig. 2 Average bias of loadings of variables Y2 and Y3 to F1

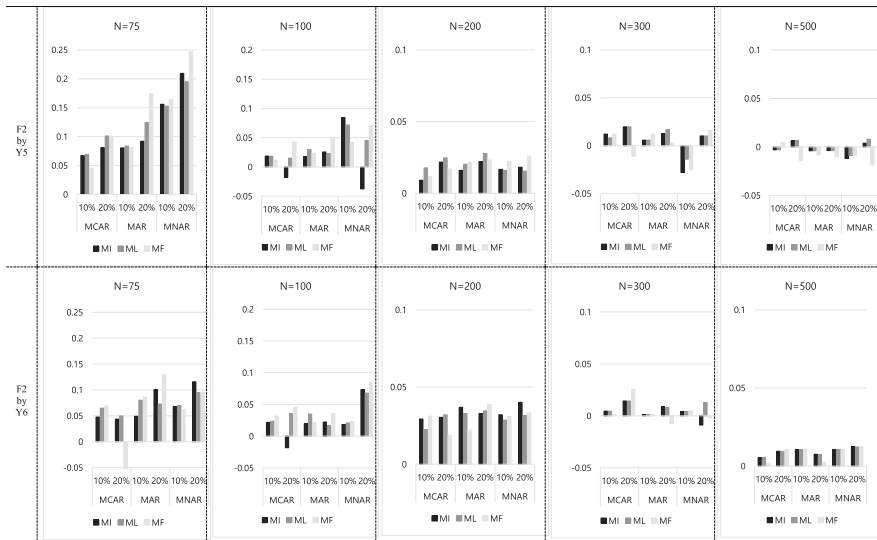


Fig. 3 Average bias of loadings of variables Y5 and Y6 to F2

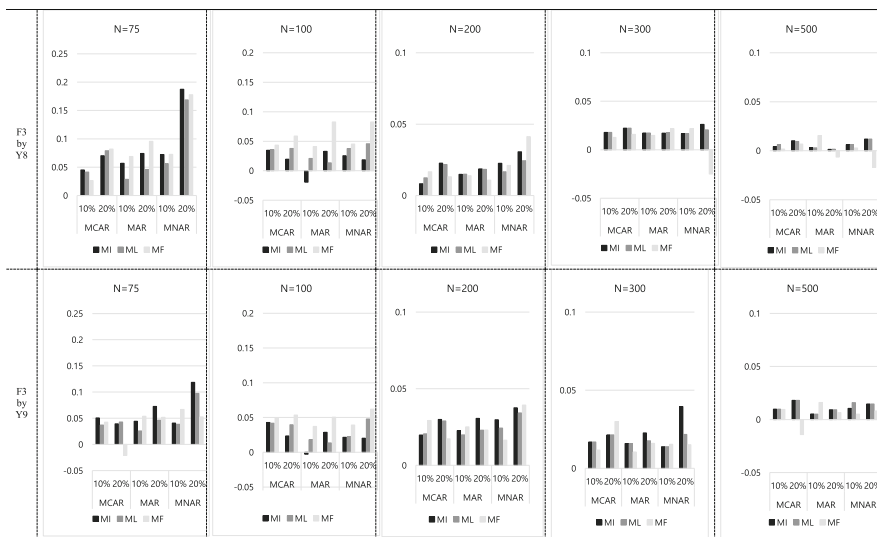


Fig. 4 Average bias of loadings of variables Y8 and Y9 to F3

estimation errors (i.e., recovery of parameter estimates) are presented in Figs. 6, 7, 8 and 9.

As expected, the bias of parameter estimates of all missing data methods was close to the population parameters with sufficient sample sizes. However, the presence of large missing rates and MNAR along with small samples had negative effects on the

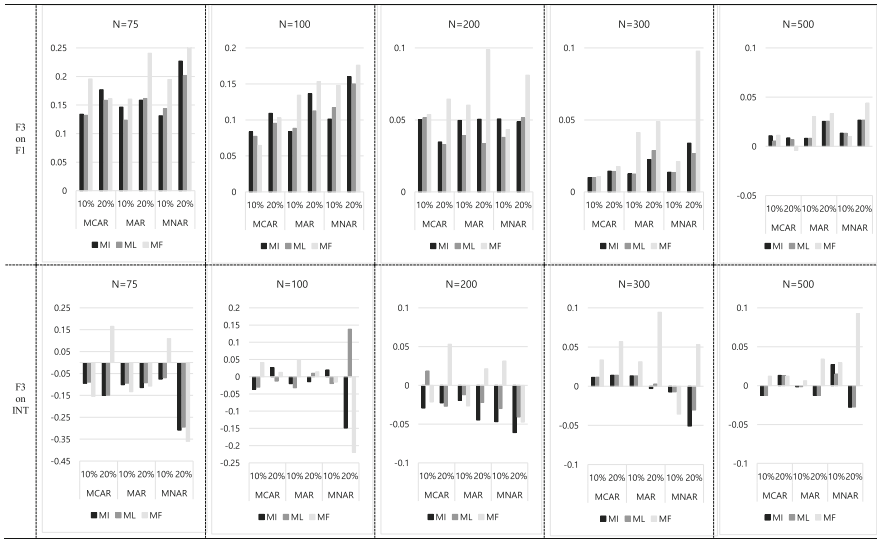


Fig. 5 Average bias of regression coefficients of F1 and interaction (F1 × F2) on F3

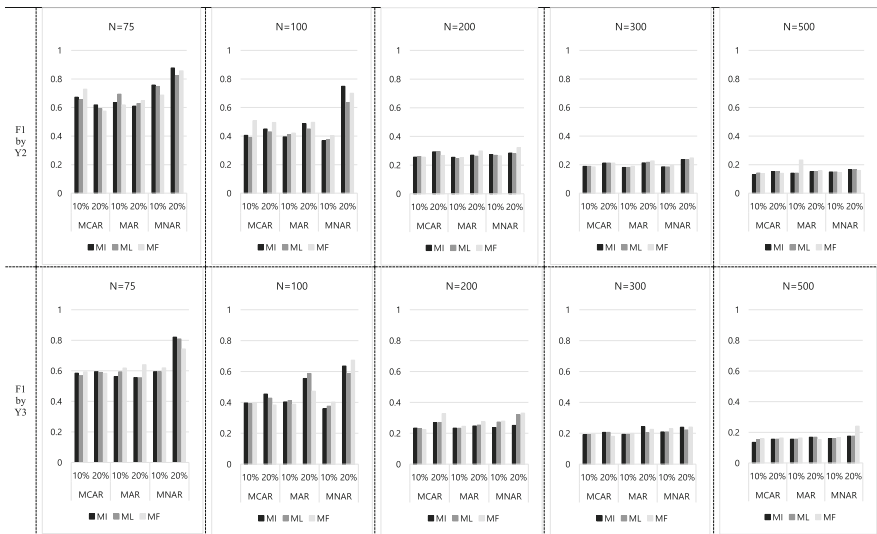


Fig. 6 RMSRD of loadings of variables Y2 and Y3 to F1

degree of bias in the parameter values. No missing data methods significantly outperformed the others, and the differences among them were relatively small. ML was more likely to yield a smaller bias especially in regression coefficients including the interaction effect. When examining the number of cells, which indicates the smallest bias across research designs, ML had the most, followed by MI and MF. Again, the



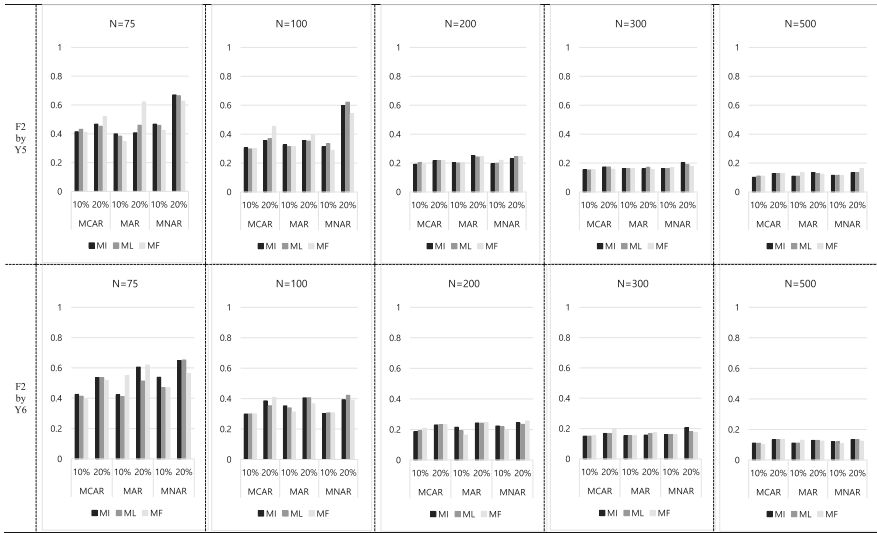


Fig. 7 RMSRD of loadings of variables Y5 and Y6 to F2

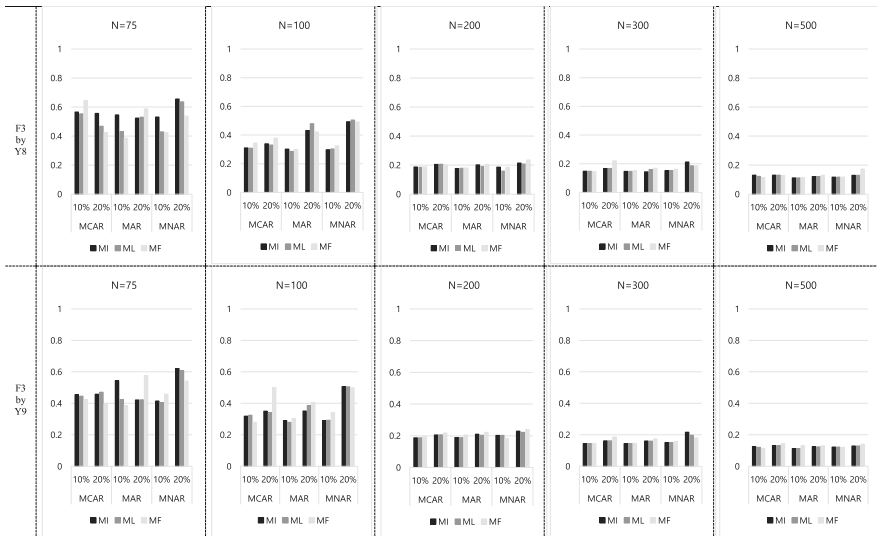


Fig. 8 RMSRD of loadings of variables Y8 and Y9 to F3

differences were sometimes based on three decimal places. The amount of parameter estimation error (RMSRD) had similar patterns to the bias of estimates. Small samples and large missing rates with MNAR led to larger RMSRD values. For the analysis of the RMSRD, all missing data techniques yielded similar performance. MI tended to have smaller errors in MCAR, whereas ML was more likely to present smaller RMSRDs

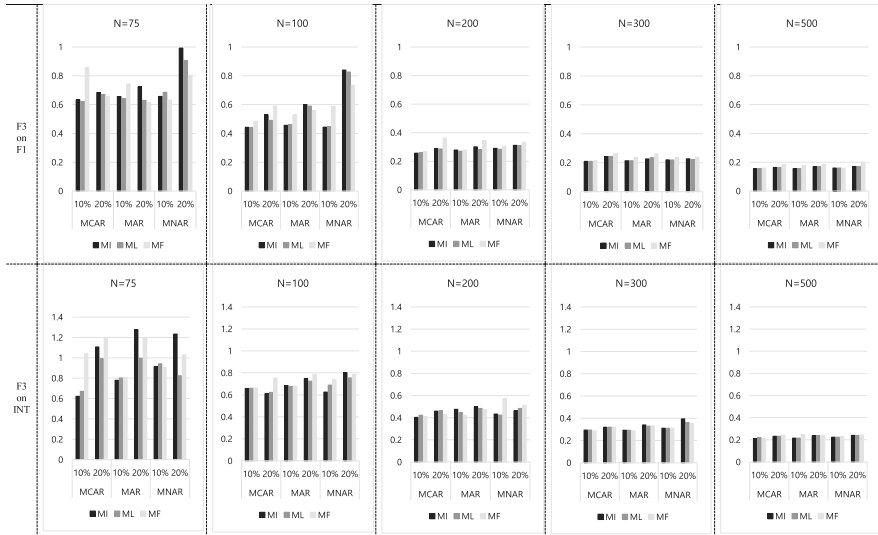


Fig. 9 RMSRD of regression coefficients of F1 and interaction ( $F1 \times F2$ ) on F3

in MAR. MF showed slightly better results with smaller sample sizes and MNAR. In general, ML showed slightly fewer parameter estimation errors in the regression coefficients. However, the differences were negligible and both MI and ML yielded very similar RMSRDs especially for larger sample sizes. In the analysis of MF, it tended to show somewhat inconsistent results compared to MI and ML indicating that it yielded relatively less bias and RMSRD values in some conditions, while larger bias and RMSRD values were sometimes found. In other words, the differences between MI and ML were very small across all conditions, whereas MF sometimes presented somewhat different results regardless of whether we had less or more biased outcomes.

### 4.3 Bias and recovery of standard error estimates

The exclusion rules stated above were used in the analysis of the standard error estimates. Figures 10, 11, 12, 13, 14, 15, 16 and 17 show the biases and RMSRDs in the standard error estimates for each missing data method.

As the sample size increased, the bias and RMSRDs for standard error estimates decreased. In addition, when the missing rates became larger and the missing mechanism was MNAR, all methods tended to increase bias and estimation errors. Significantly, MF yielded more accurate standard error estimates with smaller errors than other missing data methods. However, ML presented relatively similar (or sometimes better) results with MF especially with respect to the analysis of regression and interaction coefficients. For larger sample sizes, the results obtained from all missing data methods tended to be close to each other. ML was more likely to show better performance than MI specifically under larger missing rates and for estimates of regression and interaction coefficient estimates, whereas the differences again tended

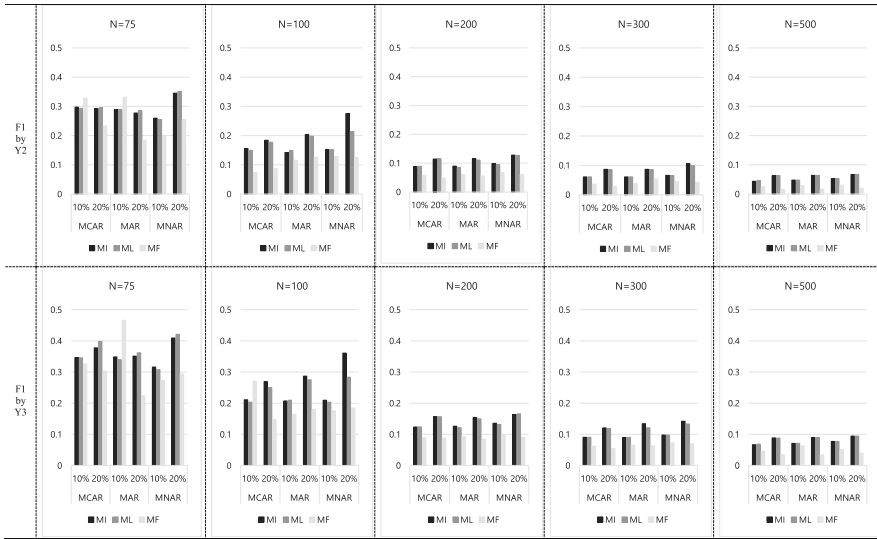


Fig. 10 Average bias of standard error for loadings of variables Y2 and Y3 to F1

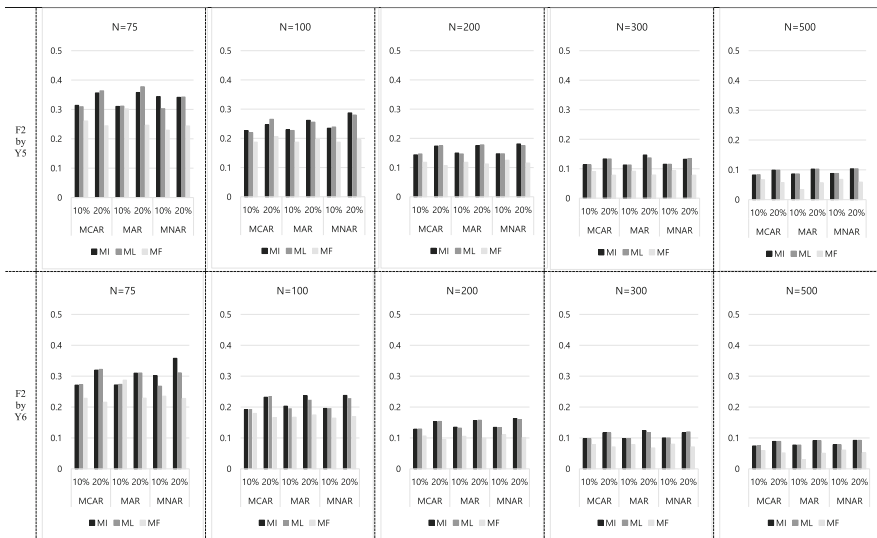


Fig. 11 Average bias of standard error for loadings of variables Y5 and Y6 to F2

to be small consistent with the analysis of parameter estimates. Smaller biased standard error estimates of MF originate from the nearest neighbor algorithms. In RF, the basic idea is to compute a distance measure between each pair of observations based on non-missing variables. Then the  $k$ -nearest observations that have non-missing values for that particular variable are used to impute a missing value through a weighted

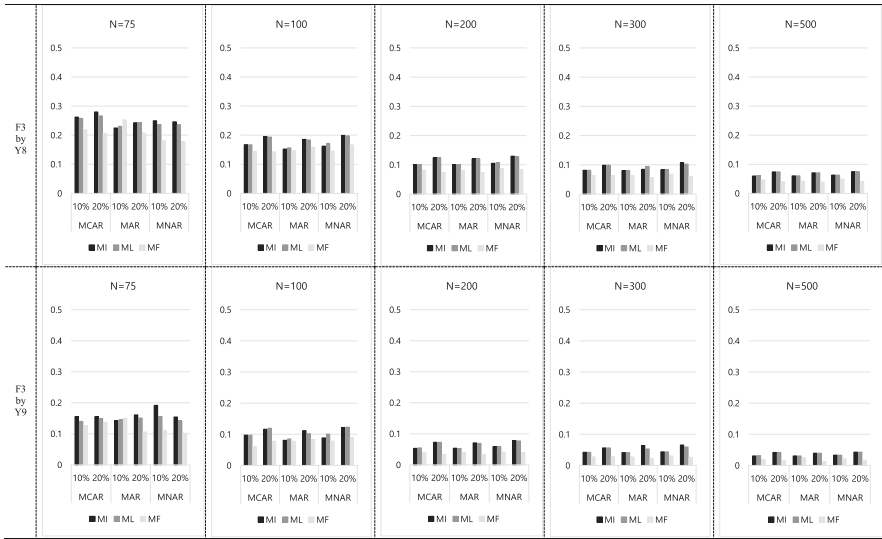


Fig. 12 Average bias of standard error for loadings of variables Y8 and Y9 to F3

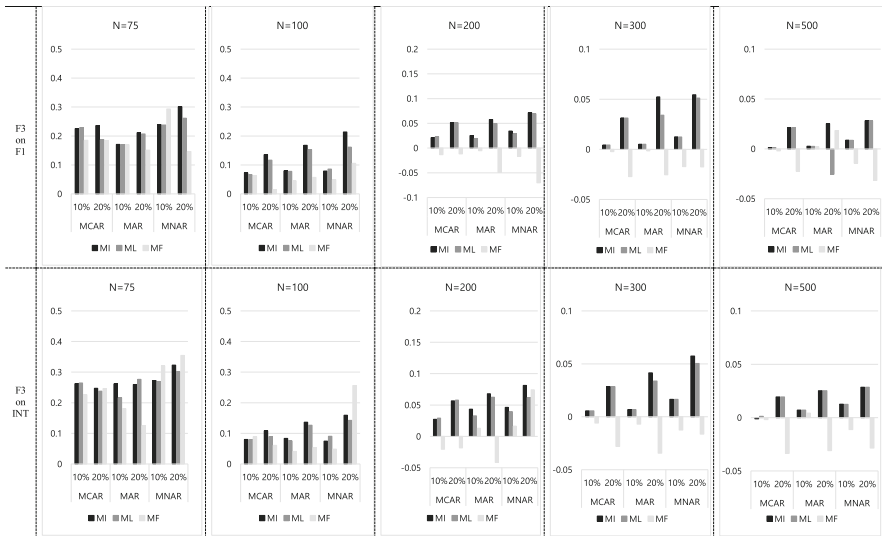


Fig. 13 Average bias of standard error for regression coefficients of F1 and interaction ( $F1 \times F2$ ) on F3

mean of the neighboring values (i.e., for continuous variables, the values are imputed using the proximity weighted average non-missing values; Waljee et al., 2013). This bootstrapped aggregation of multiple trees may lead to accurate prediction, while the variability between imputed missing values and complete ones would be relatively small.

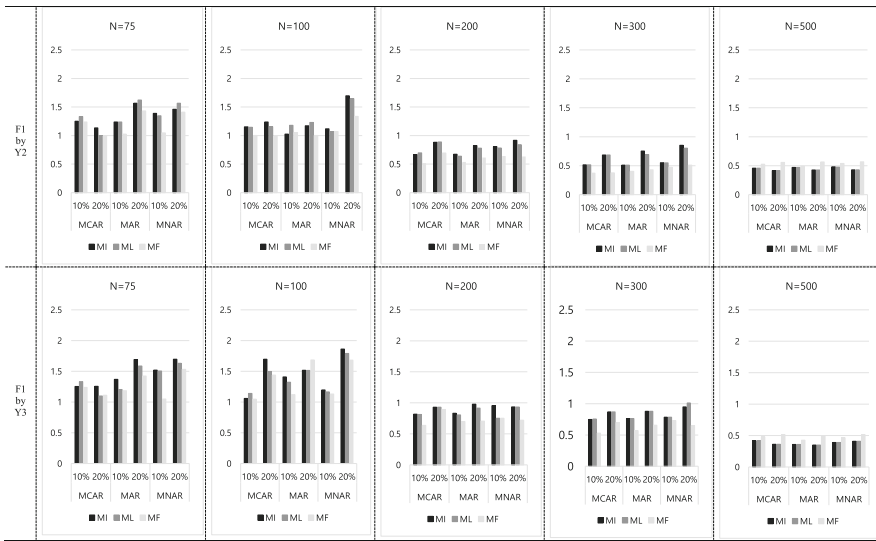


Fig. 14 RMSRD of standard error for loadings of variables Y2 and Y3 to F1

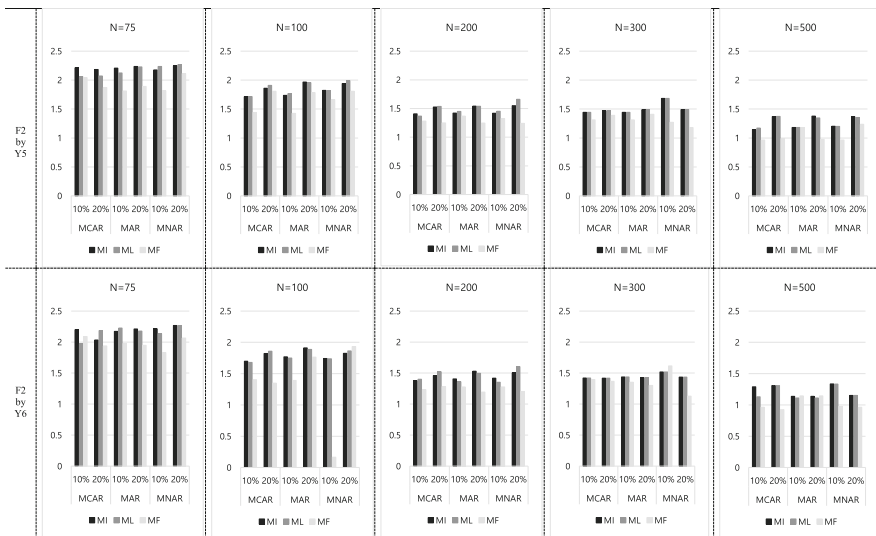


Fig. 15 RMSRD of standard error for loadings of variables Y5 and Y6 to F2

### 5 Discussion

Since missing values in most social and behavioral studies are more often the norm than the exception, dealing with missing data effectively is of great importance to applied researchers. To date, MI and ML are very useful tools for handling missingness and

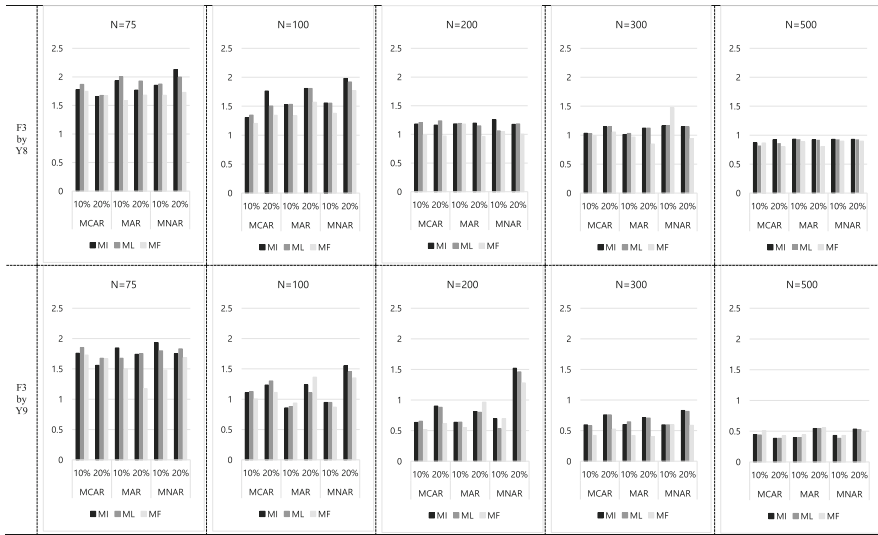


Fig. 16 RMSRD of standard error for loadings of variables Y8 and Y9 to F3

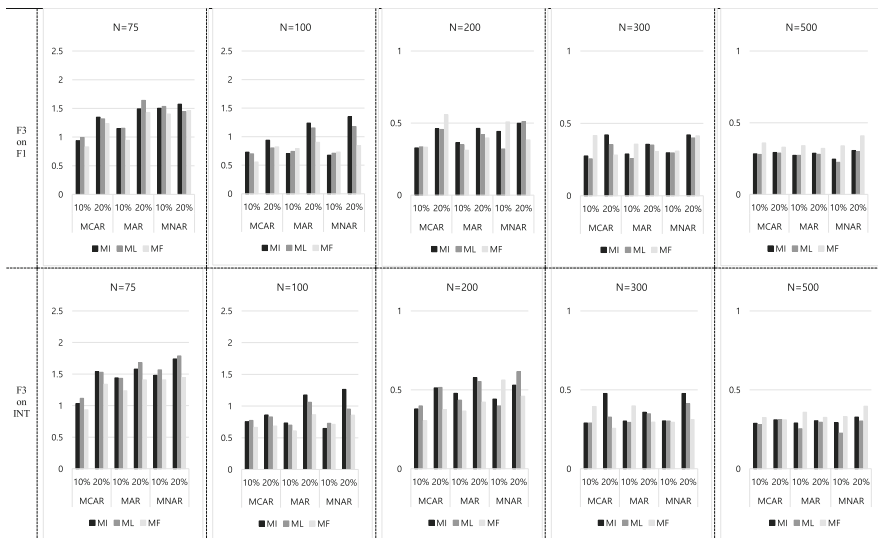


Fig. 17 RMSRD of standard error for regression coefficients of F1 and interaction (F1 × F2) on F3

several studies (e.g., Allison, 2003; Arbuckle, 1996; Croy & Novins, 2005; Enders, 2001a, 2001b, 2010; Gold et al., 2003; Larsen, 2011; Little & Rubin, 2002; Kenward & Carpenter, 2007; Rubin, 2004; Savalei & Bentler, 2005; Savalei & Rhemtulla, 2012; Schafer & Graham, 2002; Shin et al., 2009, 2017; Ho et al., 2001; Yuan & Bentler, 2000; Yuan et al., 2012a, b) demonstrated that these two missing data methods provide

less bias and greater efficiency in many circumstances. Recently, a RF based algorithm was proposed and a series of studies (e.g., Shah et al., 2014; Stekhoven & Bühlmann, 2012; Tang & Ishwaran, 2017; Waljee et al, 2013) suggested that a machine learning method represents a potentially attractive solution to missing data problem. However, no systematic study of RF procedures, which compares MI and ML, has been attempted in various missing data settings. Thus, the primary goal of this simulation study was to investigate the performance of MI, ML and MF under as close to real research conditions as possible. These assumptions included small to moderate sample sizes, two missing rates and three missing data mechanisms. These conditions were examined for nonlinear modeling (i.e., latent interaction variable modeling) and the effects of these factors on convergence rate and bias of parameter estimation.

The proportion of convergence success showed that the degree of sample size and missing rates with nonignorable missing data mechanism had major impacts on the convergence rate of the methods. When the sample size is over 300, the convergence does not make a significant impact even with complicated modeling structures (e.g., nonlinear relationships) and smaller variances (e.g., close to near 0) for latent factors. However, as expected, the combination of small sample sizes, larger missing rates and MNAR prevented valid statistical inferences. Across all the cell designs, MI yielded higher convergence rates. ML and MF showed lower rates, while the results regarding convergence success were very similar to those when generating the complete data. Because failure of data augmentation is one of the prominent problems in missing data analysis, being able to yield higher and stable convergence rates must be very important when selecting a missing data method. Thus, MI may be superior to other methods when accounting for the uncertainty brought about by the presence of missing values. Black et al. (2011) also indicated that analysts who use ML may face problems with convergence more often than those who use MI. The MI data augmentation algorithm belongs to a family of MCMC procedures (Jackman, 2000). MCMC simulates random draws from non-informative prior distribution for the covariance matrix. Then, to avoid convergence failure under MI analysis, the ridge prior, which is a semi-informative distribution, contributes additional information about the covariance matrix and effectively reduces the convergence problem (Enders, 2010). Moreover, MI is repeated imputation indicating that it has additional information about the data. Although Rubin (1996) noted that adding random noise to data is only being used to handle missing information, there is no doubt that the imputations use an extra source of information. MI with additional data values could lead to valid inferences for a variety of estimates. In particular, when the variance estimate is near zero (or varies around zero), the pooling phase (i.e., averaging over estimates of multiple sets of data) may be able to avoid negative variances with smoothing information in the distribution of data. Thus, strong superefficiency is one of the great advantages of MI over others. However, one would be concerned with whether MI estimates too much (overparameterization). As shown in Table 3, some convergence failures were found even when the complete data were generated. Thus, is it supposed to have similar (or same) convergence rates in the analysis of the missing data method? In other words, if a negative variance is estimated in one certain complete data, is it fair that the missing data method with incomplete data, in which missing scores are generated based on the complete data, yield a negative variance? We might need to be more careful about

this advantage of MI. Within the framework of ML and MF, using auxiliary variables chosen based on theory and previous experience may be helpful in gaining more knowledge about missingness as well as model estimation information (Asparouhov & Muthén, 2008; Yuan et al., 2015). Another consideration is that the rate of convergence depends on the parameters used, such as the stop criterion and the number of growing trees. Since the default values of iteration and convergence criteria were set, the convergence results obtained in this simulation study would be only a guide.

In the analysis of parameter estimates, abnormal research conditions (i.e., small sample sizes, higher missing rate, and MNAR) had negative influence on the performance of missing data methods. The factor loading estimates obtained from both MI and ML were very similar to each other, while ML was more likely to yield less biased estimates for regression and interaction coefficients than MI. In addition, ML tended to show the smallest estimation errors compared to others. According to previous studies (Allison, 2003, 2010; Dong & Peng, 2013; von Hippel, 2016; Yuan et al., 2012a, 2012b), MI does not generate consistent results because of its lack of a mechanism for handling mis-specified distributions and ML outperformed MI in small samples. Larsen (2011) also suggested that ML is superior to MI as it correctly estimates standard errors in the analysis of hierarchical modeling (i.e., second-level dependency model). Shin et al. (2017) noted that ML showed better performance and provided more diverse corrected model fit statistics under non-normality. However, we confirmed that the differences between MI and ML became negligible as sample sizes increased, in the line with the study of Collins et al., (2001, p. 336). The simulation study presented MF as a good competitor in the analysis of missingness. Across all research designs, MF provided a similar degree of bias and efficiently estimated parameters in MCAR and MAR as both MI and ML did. However, one concern is that MF tended to have unstable performance indicating that it showed the largest degree of bias especially for regression and interaction coefficients in the most severe research conditions (i.e., small sample sizes, higher missing rates, and MNAR). Shah et al. (2014) noted that the MF algorithm aims to predict individual missing values accurately rather than take random draws from a distribution (i.e., non-parametric approach), so the imputed values may lead to biased parameter estimates in statistical models. In summary, although the differences among missing data methods tended to be small and all methods showed similar recovery of estimates, ML tended to yield slightly less biased parameter estimates (especially regression and interaction coefficients) and smaller parameter estimation errors than others in MAR and MNAR.

The biggest advantage of MF over other methods is the accurate estimation of the standard errors. Across almost all conditions, the MF yielded less biased SE estimates with small errors. As explained in the results section, this feature originates from the RF based algorithm. According to Breiman (2003), Ishwaran et al. (2008) and Liaw and Wiener (2002), RF first roughly imputes the data; missing values for continuous variables are replaced with the median of complete values or data are imputed using the most frequent complete data for categorical variables). RF then analyzed the roughly imputed data using proximity measures. Finally, imputed data for continuous variables is selected based on the proximity-weighted average of the complete data (the integer value having the largest average proximity over no missing data for categorical variables). With a small prediction error, this approach may lead to a smaller



variability in the data. However, this benefit should be examined under the conditions of multivariate non-normality. Theoretically, non-normality leads underestimation of standard errors of parameter estimates and overestimation of the chi-squared statistic leading to inflated statistics and hence possibly erroneous attributions of the significance of specific relationships in the model (Muthén & Kaplan, 1985). Thus, there may be a higher chance that the RF-based missing data algorithm (MF) underestimates the standard error especially under the violation of the normal distribution assumption. In this study, the MF only underestimated the standard error estimates for the regression and interaction coefficients compared to MI and MF. Further MF studies are required to examine the impact of non-normality.

## 6 Conclusion

The findings of this simulation study verified the past applied statistical literature in that the combined effects of small sample sizes, higher missing rates, and non-ignorable missingness along with complicated modeling structure adversely affected the accuracy of statistical inferences. Although there is a possibility for overparameterization, it is a good way to select MI when convergence is concerned. If the primary goal of the research is to investigate the relationship between variables as in many studies, ML would be attractive. ML produces a deterministic result and yields smaller sampling variance than MI estimates. MF presented similar performance compared to MI and ML across all research conditions and outperformed when estimating the variability of parameter estimates. However, it requires an additional check for this accuracy of standard error estimation especially under non-normality research conditions. Although our results confirm much theoretical and simulation work, few caveats that should be considered. First, the performance of missing data methods should be investigated under complicated modeling structures and categorical missing variables, as well as mixed-type variables. Liao et al. (2014) claimed that one of the most beneficial effects of MF over MI is the effective analysis of the interaction effect with various types of variables. To handle the interaction between observed variables properly in MI, Enders (2010) suggested that the imputation model include the product of the two variables if both are continuous. For categorical variables, Enders suggested performing MI separately for each subgroup defined by the combination of the levels of categorical variables. However, this procedure needs to be verified not only in terms of computational bothersome but also in terms of accuracy. Rather, likelihood-based methods do not require these additional steps for analysis of interaction. One concern of ML is that it may not fit as log-linear models with categorical variables due to distributional assumption (Didelez, 2002; Raghunathan, 2004). Peng and Zhu (2008) described that MI has a clear advantage over ML-based methods in the analysis of logistic regression (i.e., categorical variables), and Asparouhov and Muthén (2010) also suggested that MI is preferable for categorical responses. Recently, some pioneers (Cho & Rabe-Hesketh, 2011; Dong & Yin, 2017; Edwards et al., 2018; Jeon & Rijmen, 2014) have developed ML estimators with new algorithms (e.g., Fuch's model, Plackett–Luce model, variational maximization-maximization, alternating imputation posterior) for categorical variable analysis. However, these devised ML approaches

must be a refreshing topic for applied researchers, and it is worthwhile to conduct further studies under research conditions with nonlinearity and mixed-type variables.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.
- Aittokallio, T. (2009). Dealing with missing values in large-scale studies: Microarray data imputation and beyond. *Briefings in Bioinformatics*, 2(2), 253–264.
- Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 1–63). Academic Press.
- Algina, J., & Moulder, B. C. (2001). A note on estimating the Jöreskog-Yang model for latent variable interaction using LISREL 8.3. *Structural Equation Modeling*, 8, 40–52.
- Alkasawneh, Pan, & Green. (2007) Multiple imputation for missing data. A caution tale. *Sociological Methods and Research*, 28(3), 301–309.
- Allison, P. D. (2003). Missing data techniques for structural equation models. *Journal of Abnormal Psychology*, 112, 545–557.
- Allison, P. D. (2006). Multiple imputation of categorical variables under the multivariate normal model. In *Paper presented at the annual meeting of the American Sociological Association*, Montreal Convention Center, Montreal, Quebec, Canada, Aug. 11, 2006.
- Allison, P. D. (2010). Missing data. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of survey research* (pp. 631–657). Emerald Group Publishing Ltd.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155–173.
- Anderson, T. W. (1957). Maximum likelihood estimates for the multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, 52, 200–203.
- Arbuckle, J. (1996). *AMOS-Analysis of moment structures*. Small Waters Corporation.
- Arminger, G., & Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association*, 85, 195–203.
- Asparouhov, T. & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. <http://statmodel.com/download/Bayes3.pdf>
- Asparouhov, T., & Muthén, B. (2008). *Auxiliary variables predicting missing data. Technical appendix*. Muthén & Muthén.
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48, 5–37.
- Baraldi, A. N., & Enders, C. K. (2013). Missing data methods. In T. D. Little (Ed.), *Oxford library of psychology. The Oxford handbook of quantitative methods: Statistical analysis* (pp. 635–664). Oxford University Press.
- Black, A. C., Harel, O., & McCoach, D. B. (2011). Missing data techniques for multilevel data: Implications of model misspecification. *Journal of Applied Statistics*, 38(9), 1845–1865.
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, 50, 229–242.
- Brand, J. (1999). *Development, implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets*. University of Medical Center, Rotterdam.
- Breiman, L. (2003). *Manual for setting up, using, and understanding random forest V4.0*. [https://www.stat.berkeley.edu/~breiman/Using\\_random\\_forests\\_v4.0.pdf](https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf)
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

- Breiman, L., & Cutler, A. (2002). *Manual on setting up, using, and understanding random forests V3.1*. Berkeley: University of California, Berkeley. [http://oz.berkeley.edu/users/breiman/Using\\_random\\_forests\\_V3.1.pdf](http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf)
- Cham, H., Baraldi, A. N., & Enders, C. K. (2013). Applying maximum likelihood estimation and multiple imputation to moderated regression models with incomplete predictor variables. *Multivariate Behavioral Research*, 45, 153–154.
- Chiarella, C., Kang, B., Meyer, G., & Ziogas, A. (2014). Computational methods for derivatives with early exercise features. In K. Schmedders & K. L. Judd (Eds.), *Handbook of computational economics* (3rd ed., chap. 5). Elsevier.
- Cho, S. J., & Rabe-Hesketh, S. (2011). Alternating imputation posteriors estimation of models with crossed random effects. *Computational Statistics & Data Analysis*, 55, 12–25.
- Coenders, G., Batista-Foguet, J. M., & Saris, W. E. (2008). Simple, efficient and distribution-free approach to interaction effects in complex structural equation models. *Quality & Quantity*, 42, 369–396.
- Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analyses for the behavioral sciences*. Erlbaum.
- Collins, L. M., Schafer, J. L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4), 330–351.
- Copas, J. B., & Li, H. G. (1997). Inference for non-random samples (with discussion). *Journal of Royal Statistical Society (series b)*, 59, 55–96.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.
- Croy, C. D., & Novins, D. K. (2005). Methods for addressing missing data in psychiatric and developmental research. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 1230–1240.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistics Society (series b)*, 39, 1–38.
- Díaz-Uriarte, R & de Andrés, A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1), 3. Retrieved from <http://www.biomedcentral.com/1471-2105/7/3>
- Didelez, V. (2002). ML- and semiparametric estimation in logistic models with incomplete covariate data. *Statistica Neerlandica*, 56, 330–345.
- Dong, F., & Yin, G. (2017). Maximum likelihood estimation for incomplete multinomial data via the weaver algorithm. *Statistics and Computing (published on-line)*.
- Dong, Y., & Peng, C.-Y.J. (2013). Principled missing data methods for researchers. *Springer plus*, 2, 222. <https://doi.org/10.1186/2193-1801-2-222>
- Doove, L., Van Buuren, S., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis*, 72, 92–104.
- Dubow, E. F., & Ullman, D. G. (1989). Assessing social support in elementary school children: The survey of children's social support. *Journal of Clinical Child Psychology*, 18(1), 52–64.
- Duncan, S., & Duncan, T. (1994). Modeling incomplete longitudinal substance use data using latent variable growth curve methodology. *Multivariate Behavioral Research*, 29, 313–338.
- Edwards, S. L., Berzofsky, M. E., & Biemer, P. P. (2018). *Addressing nonresponse for categorical data items using full information maximum likelihood with latent GOLD 5.0*. RTI Press Publication No. MR-0038-1809. RTI Press.
- Enders, C. K. (2001a). A Primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, 8, 128–141.
- Enders, C. K. (2001b). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, 6, 352–370.
- Enders, C. K. (2006). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 313–342). Information Age Publishing.
- Enders, C. K. (2010). *Applied missing data analysis*. The 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, 430–457.
- Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, 44, 409–420.
- Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, 2, 235–247.

- Gelman, A., & Rubin, D. (1992). A single series from the Gibbs sampler provides a false sense of security. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 625–631). Oxford University Press.
- Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*, 7, 319–355.
- Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete nonnormal data. *Structural Equation Modeling*, 10(1), 47–79.
- 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, 197–218.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarification of multiple imputation theory. *Prevention Science*, 8, 206–213.
- Hallquist, M. N., & Wiley, J. F. (2017). *MplusAutomation*: an R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, 25(4), 621–638.
- Hapfelmeier, A. (2012). Analysis of missing data with random forests (Doctoral dissertation. Ludwig Maximilian University of Munich, Munich, Germany). Retrieved from <https://edoc.ub.uni-muenchen.de/15058/>
- Hartley, H. O., & Hocking, R. (1971). The analysis of incomplete data. *Biometrics*, 27, 783–808.
- Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling*, 16(1), 1–27.
- Ho, P., Silva, M., & Hogg, T. (2001). Multiple imputation and maximum likelihood principal component analysis of incomplete multivariate data from a study of the ageing of port. *Chemometrics and Intelligent Laboratory Systems*, 55(2), 1–11.
- Ishioka, T. (2013). Imputation of missing values for unsupervised data using the proximity in random forests. In *The fifth international conference on mobile, hybrid, and on-line learning* (pp. 30–36). The National Center for University Entrance Examinations.
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *Annals of Applied Statistics*, 2, 841–860.
- Jaccard, J., Turrissi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression* (Sage university papers series. Quantitative applications in the social sciences; Vol. no. 07-072). Sage Publications.
- Jackman, S. (2000). Estimation and inference via Bayesian simulation: an introduction to Markov Chain Monte Carlo. *American Journal of Political Science*, 44(2), 375–404.
- Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G., & Kenward, M. G. (2006). The nature of sensitivity in monotone missing not at random models. *Computational Statistics and Data Analysis*, 50, 830–858.
- Jeon, M., & Rijmen, F. (2014). Recent developments in maximum likelihood estimation of MTMM models for categorical data. *Frontier in Psychology*, 5(1), 1–7.
- Ji, L., Chow, S.-M., Schermerhorn, A. C., Jacobson, N. C., & Cummings, E. M. (2018). Handling missing data in the modeling of intensive longitudinal data. *Structural Equation Modeling*, 25(5), 715–736.
- Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57–87). Lawrence Erlbaum Associates.
- Kang, J., & Shin, T. (2015). The effects of adolescents' stress on suicidal ideation: Focusing on the moderating and mediating effects of depression and social support. *Korean Journal of Youth Studies*, 22(5), 27–51.
- Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implication for job redesign. *Administrative Science Quarterly*, 24, 285–308.
- Kelava, A. (2009). *Multicollinearity in nonlinear structural equation models*. (Doctoral dissertation, Goethe University, Frankfurt, Germany). Retrieved from <http://publikationen.ub.uni-frankfurt.de/volltexte/2009/6336/>
- Kelava, A., & Brandt, H. (2009). Estimation of nonlinear latent structural equation models using the extended unconstrained approach. *Review of Psychology*, 16, 123–131.
- Kelava, A., Werner, C. S., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., Cham, H., Aiken, L. S., & West, S. G. (2011). Advanced nonlinear latent variable modeling: distribution analytic LMS and QML estimators of interaction and quadratic effects. *Structural Equation Modeling*, 18(3), 465–491.

- Kenny, D., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96, 201–210.
- Kenward, M. G., & Carpenter, J. (2007). Multiple imputation: Current perspectives. *Statistical Methods in Medical Research*, 16, 199–218.
- King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1), 49–69.
- Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65, 457–474.
- Klein, A. G., & Muthén, B. O. (2007). Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, 42, 647–674.
- Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Kelava, A. (2009). Assessing spurious interaction effects. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 13–28). Sense.
- Korean Youth Policy Institute. (2007). *Korean children and youth panel survey*, Sejong-si.
- Kovacs, M. (1983). *The children's depression inventory: A self-rated depression scale for school-aged youngsters*. University of Pittsburgh school of medicine, Department of Psychiatry, Western Psychiatric Institute and Clinic.
- Kroll, C. N., & Stedinger, J. R. (1996). Estimation of moments and quantiles using censored data. *Water Resource Research*, 32(4), 1005–1012.
- Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling*, 18, 649–662.
- Lee, L. E. (1993). Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. *Econometric Theory*, 9, 413–430.
- Lee, S. Y., & Song, X. Y. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal data. *British Journal of Mathematical and Statistical Psychology*, 57, 131–150.
- Liao, S. G., Lin, Y., Kang, D., Chandra, D., Bon, J., Kaminski, N., Sciurba, F. C., & Tseng, G. C. (2014). Missing value imputation in high-dimensional phenomic data: Imputable or not, and how? *BMC Bioinformatics*, 5(15), 346. <https://doi.org/10.1186/s12859-014-0346-6>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Lin, G.-C., Wen, Z., Marsh, H., & Lin, H.-S. (2010). Structural equation models of latent interactions: Clarification of orthogonalizing and double-mean-centering strategies. *Structural Equation Modeling*, 17(3), 374–391.
- Little, R. J. A. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association*, 87, 1227–1237.
- Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Wiley.
- Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling*, 13(4), 497–519.
- Loh, P. L., & Wainwright, M. J. (2011). High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Advances in Neural Information Processing Systems*, 24, 2726–2734.
- Lusch, R. F., & Brown, J. R. (1996). Interdependency, contracting, and relational behavior in marketing channels. *Journal of Marketing*, 60, 19–38.
- Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9, 275–300.
- Marsh, H. W., Wen, Z., & Hau, K. T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 225–265). Information Age.
- Miccéni, T. (1989). The unicorn, the normal curve, and other improbably creatures. *Psychological Bulletin*, 105, 156–166.
- Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation approaches. In T. Teo & M. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 103–136). Sense Publishers.
- Moosbrugger, H., Schermelleh-Engel, K., & Klein, A. G. (1997). Methodological problems of estimating latent interaction effects. *Methods of Psychological Research Online*, 2, 95–111.

- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171–189.
- Muthén, L. K., & Muthén, B. O. (2018). *Mplus version 8.2 [Computer software]*. Muthén & Muthén.
- Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., & Ishii, S. (2003). A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, 19(16), 2088–2096.
- Pantanowitz, A., & Marwala, T. (2008). *Evaluating the impact of missing data imputation through the use of the random forest algorithm*. School of Electrical and Information Engineering. University of the Witwatersrand Private Bag x3. Wits. 2050. Republic of South Africa. Retrieved from <http://arxiv.org/ftp/arxiv/papers/0812/0812.2412.pdf>
- Peng, C.-Y.J., & Zhu, J. (2008). Comparison of two approaches for handling missing covariates in logistic regression. *Educational and Psychological Measurement*, 68(1), 58–77.
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: a review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525–556.
- Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation*, 7(4), 353–383.
- Raghunathan, T. E. (2004). What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health*, 25, 99–117.
- Rotnitzky, A., Cox, D. R., Botta, M., & Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6, 243–284.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Rubin, D. B. (1996). Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association*, 91, 473–489.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. Wiley.
- Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *ASA proceedings of the business and economic section* (pp. 308–313).
- Savalei, V., & Bentler, P. M. (2005). A statistically justified pairwise ML method for incomplete nonnormal data: A comparison with direct ML and pairwise ADF. *Structural Equation Modeling*, 12, 183–214.
- 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.
- Savalei, V., & Rhemtulla, M. (2012). On obtaining estimates of the fraction of missing information from full information maximum likelihood. *Structural Equation Modeling*, 19(3), 37–62.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall.
- Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8, 3–15.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, 7, 147–177.
- Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33, 545–571.
- Schmitt, M. (1990). *Konsistenz als Persönlichkeitseigenschaft? Moderatorvariablen in der Persönlichkeits- und Einstellungsforschung [Consistency as a personality trait? Moderator variables in personality and attitude research]*. Springer.
- Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating missing values for simulation purpose: A multivariate amputation procedure. *Journal of Statistical Computation and Simulation*, 88(15), 2909–2930.
- Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A caliber study. *American Journal of Epidemiology*, 179(6), 764–774.
- Shin, T., Davison, M. L., & Long, J. D. (2009). Effects of missing data methods in structural equation modeling with nonnormal longitudinal data. *Structural Equation Modeling*, 16, 70–98.
- Shin, T., Davison, M. L., & Long, J. D. (2017). Maximum likelihood versus multiple imputation for missing data in small longitudinal samples with nonnormality. *Psychological Methods*, 22(3), 426–449.
- Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, 6, 317–329.
- Snyder, M., & Tanke, E. D. (1976). Behavior and attitude: some people are more consistent than others. *Journal of Personality*, 44, 501–517.
- Stekhoven, J. D. (2016). *missForest: Nonparametric missing value imputation using random forest*. R package version 1.4.
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.

- Tang, F. (2017). Random forest missing data approaches. *Open Access Dissertations*. Retrieved from [https://scholarlyrepository.miami.edu/oa\\_dissertations/1852](https://scholarlyrepository.miami.edu/oa_dissertations/1852)
- Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining*, 10, 363–377.
- Taylor, L., & Zhou, X. H. (2009). Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics*, 65(1), 88–95.
- van Brand, J., Buuren, S., & Groothuis-Oudshoorn, C. (2003). A toolkit in SAS for the evaluation of multiple imputation methods. *Statist Neerlandica*, 57(1), 36–45.
- van Burren, S. (2012). *Flexible imputation of missing data*. Chapman & Hall/CRC.
- van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, K., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049–1064.
- von Hippel, P. (2007). Regression with missing y's: An improved method for analyzing multiple imputed data. *Sociological Methodology*, 37, 83–117.
- Von Hippel, P. (2016). New confidence intervals and bias comparisons show that maximum likelihood can be multiple imputation in small samples. *Structural Equation Modeling*, 23(3), 422–437.
- Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., & Higgins, P. D. R. (2013). Comparison of imputation methods for missing laboratory data in medicine. *British Medical Journal Open*, 3, 1–7.
- Wall, M. M., & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, 95, 929–940.
- Wilks, S. S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics*, 3, 163–195.
- Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches, and specific examples* (pp. 219–240). Erlbaum.
- Yuan, K. H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51, 289–309.
- Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30, 165–200.
- Yuan, K. H., Fan, Y., & Bentler, P. M. (2012a). ML versus MI for missing data with violation of distribution conditions. *Sociological Methods & Research*, 41(4), 598–629.
- Yuan, K. H., Yang-Wallentin, F., & Bentler, P. M. (2012b). ML versus MI for missing data with violation of distribution conditions. *Social Methods Research*, 4(4), 598–629.
- Yuan, K. H., Tong, X., & Zhang, Z. (2015). Bias and efficiency for SEM with missing data and auxiliary variables: Two-stage robust method versus two-stage ML. *Structural Equation Modeling*, 22(2), 178–192.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.