

## Monte Carlo Confidence Intervals for the Indirect Effect with Missing Data

Ivan Jacob Agaloos Pesigan and Shu Fai Cheung

Department of Psychology

University of Macau

### Author Note

Ivan Jacob Agaloos Pesigan  <https://orcid.org/0000-0003-4818-8420>; Shu Fai Cheung  <https://orcid.org/0000-0002-9871-9448>

The simulation was performed in part at the High-Performance Computing Cluster (HPCC) which is supported by the Information and Communication Technology Office (ICTO) of the University of Macau.

Correspondence concerning this article should be addressed to Ivan Jacob Agaloos Pesigan, Department of Psychology, Faculty of Social Sciences, University of Macau, Avenida da Universidade, Taipa, Macao SAR, China, or by email (i.j.a.pesigan@connect.um.edu.mo).

**Abstract**

Missing data is a common occurrence in mediation analysis. As a result, the methods used to construct confidence intervals around the indirect effect should take missing data into consideration. Previous research has demonstrated that, for the indirect effect in data with complete cases, the Monte Carlo method performs as well as nonparametric bootstrap confidence intervals (see MacKinnon et al., 2004; Preacher & Selig, 2012; Tofighi & MacKinnon, 2015). In this manuscript, we propose a simple, fast, and accurate two-step approach for generating confidence intervals for the indirect effect, in the presence of missing data, based on the Monte Carlo method. In the first step, an appropriate method, for example, full information maximum likelihood or multiple imputation, is used to estimate the parameters and their corresponding sampling variance-covariance matrix in a mediation model. In the second step, the sampling distribution of the indirect effect is simulated using estimates from the first step. A confidence interval is constructed from the resulting sampling distribution. A simulation study with various conditions is presented. Implications of the results for applied research are discussed.

*Keywords:* Monte Carlo method, nonparametric bootstrap, indirect effect, mediation, missing completely at random, missing at random, full information maximum likelihood, multiple imputation

## Monte Carlo Confidence Intervals for the Indirect Effect with Missing Data

Mediation is an important model in psychology. A Google Scholar search using the search phrase ‘psychology AND mediation’ returns almost two million entries (as of May 8, 2022), and over forty thousand entries, restricted to entries in 2021. Baron and Kenny (1986) describe mediation as a mechanism through which an independent variable  $X$  influences a dependent variable  $Y$ , through a mediator variable  $M$ . The ability to decompose the indirect effect of  $X$  on  $Y$  via  $M$  allows researchers to provide a clear picture of the interplay of variables and to discover possible pathways for intervention.

Although the indirect effect in a mediation model is simply a product of regression coefficients, estimating it has its unique challenges. The sampling distribution of the mediation effect is asymmetric if it is not exactly zero in the population, and complicated (Aroian, 1947; Craig, 1936), making the computation of a  $p$ -value of the indirect effect difficult. Therefore, it is common to make inferences for the indirect effect by forming the confidence interval (CI). Early attempts, such as the normal-theory approach, are simple to apply but assume that the sampling distribution is symmetric (Aroian, 1947; Goodman, 1960; Sobel, 1982, 1986, 1987). Several other methods have been proposed to take into account the asymmetry of the sampling distribution, such as (1) the distribution of the product method, which examines the sampling distribution of the product of normally distributed coefficients analytically (MacKinnon et al., 2004, 2007); (2) the nonparametric bootstrapping, which involves resampling the original sample data and fitting the model many times to produce a sampling distribution of the indirect effect (Bollen & Stine, 1990; Preacher & Hayes, 2008; Shrout & Bolger, 2002); (3) the Monte Carlo method, which involves simulating regression coefficients to produce a sampling distribution of the indirect effect (MacKinnon et al., 2004; Preacher & Selig, 2012); and (4) the likelihood-based approach where CIs are estimated using the likelihood ratio test (M. W.-L. Cheung, 2009a, 2009b; Pawitan, 2013; Pesigan & Cheung, 2020; Venzon & Moolgavkar, 1988).

Several researchers have concluded that nonparametric bootstrapping is the

recommended method of generating CIs for the indirect effect (e.g., Biesanz et al., 2010; G. W. Cheung & Lau, 2007; M. W.-L. Cheung, 2009a; MacKinnon et al., 2004; Preacher & Hayes, 2008; Shrout & Bolger, 2002; Taylor et al., 2007). Koopman et al. (2015), however, showed that nonparametric bootstrapping's performance in small sample sizes might be overstated, and attention has to be paid to nonparametric bootstrapping's inflated Type I error rates found in some previous studies. See also Koopman et al. (2014) for a critical review of bootstrap confidence intervals for the indirect effect in the social and behavioral sciences.

Another issue that is common in practice but has not received enough attention, is forming CIs of an indirect effect in the presence of missing data. Methods have been proposed to allow nonparametric bootstrapping with missing data, such as combining nonparametric bootstrapping with maximum likelihood (ML) and multiple imputation (MI). However, as nonparametric bootstrapping requires multiple model fitting, these approaches are computationally intensive especially for complex models.

While not as widely studied as nonparametric bootstrapping, previous studies have shown that Monte Carlo CIs performed just as well as nonparametric bootstrap CIs (see MacKinnon et al., 2004; Preacher & Selig, 2012; Tofghi & MacKinnon, 2015). Therefore, in the present study, we propose the use of the Monte Carlo method coupled with full information maximum likelihood (FIML) or multiple imputation (MI) to handle missing data in mediation. We argue that this approach is simpler and quicker than combining nonparametric bootstrapping and MI or ML without sacrificing accuracy.

We first briefly review the simple mediation model and common missing data mechanisms. We then present the two selected methods, the nonparametric bootstrapping method and the Monte Carlo method, and highlight the advantages of the Monte Carlo method over existing methods when there are missing cases in a data set. A simulation study will then be reported to compare the performance of the Monte Carlo method and nonparametric bootstrapping in various situations.

## The Simple Mediation Model

The most basic mediation model involves three variables,  $X$ ,  $M$ , and  $Y$  whose associations are expressed in the following equations

$$Y = \delta_Y + \tau'X + \beta M + \varepsilon_Y, \quad (1)$$

$$M = \delta_M + \alpha X + \varepsilon_M, \quad (2)$$

where  $\alpha$  represents the effect of the independent variable  $X$  on the mediator variable  $M$ ;  $\beta$  represents the effect of the mediator variable  $M$  on the dependent variable  $Y$ ;  $\tau'$  represents the effect of the independent variable  $X$  on the dependent variable  $Y$ , adjusting for the mediator variable  $M$ ;  $\delta_M$  and  $\delta_Y$  are intercepts; and  $\varepsilon_M$  and  $\varepsilon_Y$  are stochastic error terms. The direct effect,  $\tau'$ , which is the change in the dependent variable corresponding to one unit change in the independent variable, while the mediator variable is held constant, can be contrasted with the indirect effect (the product of  $\alpha$  and  $\beta$  or  $\alpha\beta$ ), which represents the change in the dependent variable via the mediator if the independent variable increases by one unit. See Chapter 3 of MacKinnon (2008) for more details on the simple mediation model.

## Basic Missing Data Mechanisms

For a complete presentation of the missing data mechanisms, see Rubin (1976) and Little and Rubin (2019). We briefly review three basic missing data mechanisms relevant to missing data in mediation: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Missingness is MCAR when the probability of being missing is equal for all cases. This means that missingness is not related to any variable. Missingness in this context is ignorable and the consequence of dropping missing cases is a decrease in sample size and the associated loss of information and power. In real situations, MCAR is typically an unrealistic assumption to make.

Missingness is MAR when the probability of being missing is equal within the groups defined by the observed data. This means that missingness is related to the observed variables. Missingness in this context cannot be ignored. The advantage of this mechanism is that, since missingness can be accounted for by the observed data, the missing observations can be recovered using the observed data.

Missingness is MNAR when missingness is related to the missingness itself or other unobserved variables. A strategy to deal with MNAR is to collect more data that can explain the missingness. In this context, MNAR can become MAR by adding what is called auxiliary variables in the analysis.

## Multiple Imputation and Maximum Likelihood

Multiple imputation (MI) and maximum likelihood (ML) approaches have become the state of the art in addressing missing data in research (Schafer & Graham, 2002). For a comprehensive review of applied missing data analysis that covers both MI and ML, see Enders (2010). In MI, several copies of the data are generated and the missing values are filled with plausible values for the missing data using an iterative algorithm. MI can be divided into two broad approaches, namely joint specification (JS) and fully conditional specification (FCS). In JS, a multivariate joint distribution is assumed to have generated the data and missing values are drawn simultaneously from this distribution (Rubin, 1987; Schafer, 1997). In FCS, missing data are imputed one at a time from a series of univariate conditional distributions (Raghunathan et al., 2001; van Buuren et al., 2006). Generating several data sets accounts for the uncertainty in the imputed data given the observed data. The model is fitted to each of the data sets that now has complete data. The parameter estimates and standard errors are pooled to produce the final parameter estimates and standard errors. See Little and Rubin (2019) and van Buuren (2018) for more information on MI.

In ML, missing values are not replaced or imputed, but are handled within the model fitting procedure. Two methods have been proposed, namely, the full information

maximum likelihood (FIML) and the two-stage maximum likelihood (TS-ML) approaches. In FIML (Arbuckle, 1996), the population parameters are estimated by maximizing the log-likelihood function of the model given the observed data. In TS-ML (Yuan & Bentler, 2000), the expectation-maximization (EM) algorithm is used to estimate the means and the covariance matrix without specifying the hypothesized model in the first stage. In the second stage, the saturated model means and covariance matrix from the first stage are used to estimate the parameters of the hypothesized model. MAR and multivariate normality are assumed in ML approaches. The MAR assumption allows us to integrate out the missing values in the likelihood and the multivariate normality assumption assures us that the subset of observations with available data are also multivariate normal. In this manner, the likelihood is computed based on the available information, per case.

## Missing Data in Mediation

Many popular procedures in mediation analysis require complete data. For example, PROCESS (Hayes, 2022; Preacher & Hayes, 2004), a popular tool in SPSS, SAS, and recently in R for mediation analysis using nonparametric bootstrapping, uses listwise deletion to handle missing data. AMOS (Arbuckle, 2021), a popular software for structural equation modeling, can use nonparametric bootstrapping to form CIs for indirect effects. It can also use FIML to handle missing data. However, it cannot do both in the same analysis.

One method for using nonparametric bootstrapping with missing data is to use listwise deletion, which means deleting cases with missing data and then performing nonparametric bootstrapping. However, this can result in the loss of statistical power, the increase in sampling variances, and hence the widths of CIs. Deleting cases also assumes MCAR which can result in inaccuracies when this assumption is not met (Rubin, 1976). There is on-going research on how to use nonparametric bootstrapping, combined with MI or ML estimation, to compute CIs around an indirect effect when the data set has missing data (Wu & Jia, 2013; Zhang & Wang, 2012; Zhang et al., 2015). We describe these methods and summarize the findings from previous research in the next sections.

### ***Nonparametric Bootstrapping***

Nonparametric bootstrapping (NB) involves a resampling step and a model-fitting step. In step 1, an independent sample with replacement of size  $n$  from the original sample data is obtained. In the presence of missing data, cases with missing observations can be sampled from the original sample data. In step 2, the model is fitted to the bootstrap sample obtained in step 1. In the presence of missing data, the estimation procedure should be able to account for missing values. MI and ML approaches can be used in this step. In the case of the simple mediation model, estimates of  $\alpha$  and  $\beta$ , as well as their products, are obtained. Steps 1 and 2 are repeated  $B$  number of times (e.g., 5,000) resulting in a sampling distribution of the indirect effect  $\hat{\alpha}^* \hat{\beta}^*$ . The sampling distribution can be used to generate CIs. The percentile CI (NBPC) is generated by obtaining the percentiles corresponding to the confidence limits for a given interval. For example, 2.5% and 97.5% for the 95% CI. Other types of CIs were introduced to account for bias in the percentile approach (Efron, 1987, 1988). In the bias-corrected approach (NBBC), the percentile confidence limits are adjusted for bias which is defined as the proportion of bootstrap replications less than the original estimate. In the bias-corrected and accelerated approach (NBBCA), in addition to the bias, the CI is adjusted for the acceleration parameter, which is proportional to the skewness of the sampling distribution. An exposition of the bootstrap and related methods can be found in Efron and Tibshirani (1993) and Davison and Hinkley (1997).

### ***Performance of Some Approaches Under Various Missing Data Mechanisms***

Existing approaches in dealing with missing data when NB is applied to mediation can be summarized as follows: (1) NB followed by listwise or pairwise deletion; (2) NB followed by ML approaches like FIML or TS-ML on each of the bootstrap samples; (3) NB followed by MI; and (4) MI followed by NB. See Wu and Jia (2013), Zhang and Wang (2012), and Zhang et al. (2015) for the performance of the approaches mentioned above. These methods have specific assumptions about the missing data mechanism. [Methods](#)

that use deletion techniques require MCAR, methods that employ ML and MI (depending on the imputation technique employed) require MAR.

While listwise or pairwise deletion performed well in simulations with MCAR, dropping cases resulted in the loss of statistical power (Zhang & Wang, 2012). MI or ML methods combined with NB worked well when missingness is MCAR (Wu & Jia, 2013; Zhang & Wang, 2012; Zhang et al., 2015). When missingness is not MCAR, deletion methods such as listwise deletion, produced biased parameter estimates and coverage probabilities lower than the nominal value (Zhang & Wang, 2012). When missingness is not MCAR but is MAR, MI and ML approaches take advantage of information from the observed data to handle missing data in this context. Results have shown that MI (NB then MI or MI then NB) or ML methods combined with NB performed well for MAR (Wu & Jia, 2013; Zhang & Wang, 2012; Zhang et al., 2015). Last, results have shown that MI or ML methods combined with NB performed well in MNAR with the addition of auxiliary variables, making MNAR become MAR (Wu & Jia, 2013; Zhang & Wang, 2012; Zhang et al., 2015).

Although some of these methods that combined NB with MI or ML are promising, they are not yet popular, probably because they are available only in a limited number of software packages, such as the R package `bmem` (Zhang & Wang, 2012), and scripts that can be requested from the authors (Wu & Jia, 2013). The R package `lavaan` (Rosseel, 2012) can combine NB and FIML. But speed can be an issue with complex models and large samples. `Mplus` (Muthén & Muthén, 2017) can also combine NB and FIML and is much faster than `lavaan`, yet it can still suffer from issues associated with fitting an SEM model many times.

## The Monte Carlo Method

The Monte Carlo (MC) method provides sampling distributions and CIs of statistics that are functions of other statistics using point estimates of the component statistics, their associated sampling variance-covariance matrix, and the parametric assumption on the distribution of the component statistics. MC can be described as a two-step process. In the

first step, which we call the estimation step, the parameters and their corresponding sampling variance-covariance matrix are estimated. In the second step, which we call the simulation step, a distribution of parameter estimates  $\hat{\boldsymbol{\theta}}^*$  using the estimated parameter vector  $\hat{\boldsymbol{\theta}}$  and estimates of the sampling variance-covariance matrix  $\mathbb{V}(\hat{\boldsymbol{\theta}})$  is generated assuming a multivariate normal joint distribution. A large number of random samples  $R$  (e.g., 20,000) are generated from the parametric model given by

$$\hat{\boldsymbol{\theta}}^* \sim \mathcal{N}(\boldsymbol{\mu} = \hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma} = \mathbb{V}(\hat{\boldsymbol{\theta}})). \quad (3)$$

The composite statistic of interest  $\mathbf{g}(\hat{\boldsymbol{\theta}})$  which is a function of elements of  $\hat{\boldsymbol{\theta}}$  are computed for each of the independent samples. The  $100(1 - \alpha)\%$  CIs can be generated using the corresponding percentiles from the sampling distribution of  $\mathbf{g}(\hat{\boldsymbol{\theta}}^*)$ .

In the case of the mediation model with missing data, in the estimation step, an appropriate method for estimating parameters and the sampling variance-covariance matrix in the presence of missing data is applied. FIML or MI approaches can be applied in this step. We then use the estimates from step 1 to generate a sampling distribution of estimates in step 2 as previously described. While the covariance between  $\hat{\alpha}$  and  $\hat{\beta}$  in a simple mediation model with observed variables is zero, this is not the case for a simple mediation model where  $X$ ,  $M$ , and  $Y$  are latent variables with multiple indicators (MacKinnon, 2008). Therefore, as a general approach, we recommend using the entire sampling covariance matrix, and not just the diagonal elements or the square of the standard errors in step 2, to take into account the covariances between the component statistics. We extract and multiply the generated  $\hat{\alpha}$  and  $\hat{\beta}$  from each of the independent samples to form the sampling distribution of the indirect effect  $\hat{\alpha}^* \hat{\beta}^*$ . The 95% CI for the indirect effect can then be formed using the 2.5% and the 97.5% percentiles from the sampling distribution of  $\hat{\alpha}^* \hat{\beta}^*$ .

### ***Advantages of The Monte Carlo Method***

MC shares the advantage of NB that it makes no parametric assumptions about the distribution of the composite statistics, for example, the indirect effect. While parametric distributional assumptions are made on the joint distribution of the component statistics, for example,  $\hat{\alpha}$  and  $\hat{\beta}$  in a mediation model, no distributional assumptions are made about their product  $\hat{\alpha}\hat{\beta}$ . This makes MC also a useful tool for generating CIs when the shape of the distribution of the statistic of interest is complex such as in the case of the indirect effect.

MC also has four advantages over NB, one on computational cost and three in handling missing data. First, its two-step approach, separating estimation in step 1 and simulation in step 2, makes MC a general approach that can be applied to different models and different contexts without a significant increase in computational overhead. While NB is routinely used in mediation models, repeated model fitting of complex models can be slow or difficult in certain contexts with the possibility of non-convergence (e.g., Bauer et al., 2006). In NB, the computational cost increases rapidly as model complexity increases. Furthermore, combining MI or ML with NB for complex models can be difficult to set up and computationally intensive. In MC, a complex model needs to be fitted only once, reducing issues associated with fitting a complex model many times. Even if we choose a large  $R$  in MC, the overhead of generating a large sample and computing the component statistic of interest for each replication is still lower than fitting a regression or SEM model using a large  $B$  in NB. As a consequence, we can make  $R$  arbitrarily large in order to generate a smooth distribution, making the confidence limits relatively stable.

Second, MC is computationally more efficient than NB in the presence of missing data. If FIML is an option for dealing with missing data, pairing MC with FIML is the most convenient approach as missing data is handled once in step 1 as opposed to NB followed by FIML where the model is fitted  $B$  times. When FIML is not an option, pairing MC with MI is an alternative. While more complicated than MC with FIML, MC with MI

is still computationally efficient as the whole process of MI that involves many imputations is done once in step 1. With NB followed by MI or MI followed by NB, the model is fitted  $B$  times the number of imputations.

Third, as step 2 is not dependent on step 1 except for the summary statistics produced, that is, one can do step 1 in any software package available to handle the missing data, obtain the necessary summary statistics, and then do step 2 in any software package that can generate random data. Due to the need of repeating computation  $B$  times, combining NB with ML or MI requires software packages that facilitates doing both NB and ML or MI efficiently, which usually means doing both in the same package.

Last, MC can be applied using secondary data sources when the raw data is “missing,” for example, in a meta-analysis, as long as the necessary summary statistics are available or can be derived from the available data. NB cannot be used to test a mediation effect in a meta-analytic structural equation model (M. W.-L. Cheung, 2021).

## The Present Study

We conducted a simulation study to assess the finite sample performance of MC for a simple mediation model with (1) complete data, and (2) missing data (MCAR and MAR). We also used NBBC, NBPC, and the joint-significance test (SIG; MacKinnon et al., 2002) as methods to compare MC with. In the joint-significance test, the indirect effect,  $\hat{\alpha}\hat{\beta}$ , is statistically significant if the  $p$ -values of the component parts,  $\hat{\alpha}$  and  $\hat{\beta}$ , are both less than an apriori significance level. Although our focus is on interval estimation, we included SIG in these conditions as it has recently been recommended to examine the components of the indirect effect (Yzerbyt et al., 2018). See Biesanz et al. (2010) and Tofghi and Kelley (2019) in addition to Yzerbyt et al. (2018) for more information on issues related to hypothesis testing and confidence intervals for indirect effects. We used the 95% CI for MC, NBBC, NBPC, and  $\alpha = .05$  for SIG.

For the complete data conditions, the conventional MC, NBBC, NBPC, and SIG were used to generate confidence intervals for  $\hat{\alpha}\hat{\beta}$  and assess the statistical significance.

The labels MC.COMPLETE, NBBC.COMPLETE, NBPC.COMPLETE, and SIG.COMPLETE will be used to refer to these approaches.

For the missing data conditions, four CI methods were used (1) MC with FIML, (2) MC with MI, (3) NBBC followed by FIML, and (4) NBPC followed by FIML. SIG was based on  $p$ -values of the model fitted using FIML. The labels MC.FIML, MC.MI, NBBC.FIML, NBPC.FIML, and SIG.FIML will be used to refer to these methods, respectively.

We evaluated the Type I error rate and statistical power of MC, NBBC, NBPC, and SIG. We also evaluated the miss rate or the proportion of replications where the MC, NBBC, and NBPC CIs did not contain the population  $\alpha\beta$ .

## Methods

### Computer Simulation

The simulation was performed using the R statistical software environment version 4.2.1 (R Core Team, 2022). The R package MASS (Venables & Ripley, 2002) version 7.3-58.1 particularly the function `mvrnorm` was used to generate multivariate normal data. The R package mice (van Buuren & Groothuis-Oudshoorn, 2011) version 3.14.0 particularly the function `ampute` was used to generate data with missing values. Mplus 8.8 (Muthén & Muthén, 2017) demo was used to fit the simple mediation model using normal-theory ML with complete and missing data (FIML); to generate NB CIs for the indirect effect; and to generate completed data sets using MI.<sup>1</sup> Custom wrapper functions were written in R to perform specific tasks required in the simulation. All the scripts used and other supplementary materials were compiled into an R package `manMCMedMiss` which serves as a research compendium for this manuscript. The simulation was performed in part at the

---

<sup>1</sup> While we use and recommend free and open-source statistical software, Mplus was used in the simulations because it is currently the fastest SEM software available. The speed was crucial for this study because it allowed us to investigate a wide range of simulation conditions (2,832) with adequate replication size (5,000). It also allowed us to use a large number of bootstrap samples (5,000) and imputations (100).

High-Performance Computing Cluster (HPCC) of the University of Macau with clusters of computers having Intel(R) Xeon(R) E3-1280 v6 @ 3.90GHz CPUs and 64G of RAM. See <https://github.com/jeksterslab/manMCMedMiss> for the `manMCMedMiss` repository; <https://jeksterslab.github.io/manMCMedMiss/reference/index.html> for documentation of the functions in the package; and <https://jeksterslab.github.io/manMCMedMiss/articles/index.html> for supplementary materials.

## Data Generation

The vector of expected values of  $X$ ,  $M$ , and  $Y$  from Equations 1 and 2 is given by

$$\mathbb{E} \begin{bmatrix} X \\ M \\ Y \end{bmatrix} = \begin{bmatrix} \mu_X \\ \mu_M \\ \mu_Y \end{bmatrix} = \begin{bmatrix} \mu_X \\ \delta_M + \alpha\mu_X \\ \delta_Y + \tau'\mu_X + \beta\mu_M \end{bmatrix}. \quad (4)$$

The means of  $X$ ,  $M$ , and  $Y$  namely  $\mu_X$ ,  $\mu_M$ ,  $\mu_Y$  were fixed to zero. The lower triangular elements of the model-implied variance-covariance matrix from Equations 1 and 2 are given by

$$\Sigma = \begin{bmatrix} \sigma_X^2 & & \\ \alpha\sigma_X^2 & \sigma_M^2 = \alpha^2\sigma_X^2 + \sigma_{\varepsilon_M}^2 & \\ \beta\alpha\sigma_X^2 + \tau'\sigma_X^2 & \beta\alpha^2\sigma_X^2 + \tau'\alpha\sigma_X^2 + \beta\sigma_{\varepsilon_M}^2 & \sigma_Y^2 = \tau'^2\sigma_X^2 + \alpha^2\beta^2\sigma_X^2 + \beta^2\sigma_{\varepsilon_M}^2 + 2\tau'\beta\alpha\sigma_X^2 + \sigma_{\varepsilon_Y}^2 \end{bmatrix} \quad (5)$$

The variances of  $X$ ,  $M$ , and  $Y$  namely  $\sigma_X^2$ ,  $\sigma_M^2$ , and  $\sigma_Y^2$  were fixed to one. It follows that the error variances are given by

$$\begin{aligned} \sigma_{\varepsilon_Y}^2 &= \sigma_Y^2 - \tau'^2\sigma_X^2 - \alpha^2\beta^2\sigma_X^2 - \beta^2\sigma_{\varepsilon_M}^2 - 2\tau'\beta\alpha\sigma_X^2 & \text{and} \\ &= 1 - \tau'^2 - \alpha^2\beta^2 - \beta^2\sigma_{\varepsilon_M}^2 - 2\tau'\beta\alpha \end{aligned} \quad (6)$$

$$\begin{aligned}\sigma_{\varepsilon_M}^2 &= \sigma_M^2 - \alpha^2 \sigma_X^2 \\ &= 1 - \alpha^2.\end{aligned}\tag{7}$$

Note that the choice of values for the means and the variances is arbitrary. However, the choice of the variances being equal and specifying the error variances the way we did in Equations 6 and 7 is by design, as this made the unstandardized and standardized regression slopes equal in the population in this parameterization. As such, the magnitude of the regression coefficients can be treated as effect sizes (M. W.-L. Cheung, 2009a).

Regression slopes  $\tau'$ ,  $\beta$ , and  $\alpha$  were varied. Assuming that the direct effect  $\tau'$  is zero, M. W.-L. Cheung (2009a) proposed effect sizes, that is, standardized regression coefficients, of 0, .37, .60, and .71 for  $\alpha$  and  $\beta$  to represent zero, small, medium, and large effect sizes. This is based on the interpretation that 0%, 2%, 13%, and 26% of the variance in the dependent variable constitutes zero, small, medium, and large effect sizes proposed by Cohen (1988)<sup>1</sup>. Following this convention, we used 0, .376, .600, and .714 as effect sizes for  $\alpha$  and  $\beta$ . We also varied the direct effect  $\tau'$ . Following the effect size convention, we used 0, .141, .361, and .509 as effect sizes for  $\tau'$ .<sup>2</sup> Nine sample sizes ( $n$ ) were used, including 50, 75, 100, 150, 200, 250, 500, and 1,000. These sample sizes are typical in published research in psychology (Fritz & MacKinnon, 2007). In summary, we have the following effect size and sample size combinations: 4 ( $\tau'$ )  $\times$  4 ( $\beta$ )  $\times$  4 ( $\alpha$ )  $\times$  8 ( $n$ ). There were cases where the variance explained by the model in Equation 1 exceeded the variance of  $Y$ . We removed these cases and ended up with a total of 472 conditions.

Using the expected values as the population mean vector and the model-implied variance-covariance matrix as the population covariance matrix, complete data was generated from a multivariate normal distribution [using the GenData function which we wrote as a wrapper around](#) the `mvrnorm` function from the `MASS` package. The complete data generated using the population parameters described above was subjected to a multivariate amputation procedure implemented in the `ampute` function (Schouten et al.,

2018) within the `mice` package using the wrapper function `AmputeData`. This results in data sets with missing observations following a particular missing data mechanism with predefined missing data patterns and the proportion of missing cases. Two missing data mechanisms, namely, MCAR and MAR were used. All possible missing data patterns for a three-variable data set was used. The percentages of missing cases, that is, the percentage of the original sample size  $n$  which is dropped if listwise deletion is used, were set to 10%, 20%, and 30%. For example, for a sample size of 1,000 and 30% missing data, 300 cases will have been dropped had listwise deletion been employed.<sup>2</sup> These percentages are similar to previous studies (Zhang & Wang, 2012; Zhang et al., 2015). For more information on the multivariate amputation procedure see Schouten et al. (2018). Considering the missing data mechanisms (2), the percentage of missing data (3), and the effect and sample size combinations described above (472), the overall number of simulation conditions is 2,832.

## Confidence Intervals

### *The Monte Carlo Method*

In the case of MC.COMPLETE and MC.FIML in the first step, the simple mediation model was estimated using normal-theory maximum likelihood estimator with `Mplus`. Note that, by default, when the normal-theory maximum likelihood estimator is used and raw data with missing values is supplied, `Mplus` uses FIML to handle missing values. The wrapper function `FitModelML` was used to fit the model in `Mplus` and extract the parameter estimates and the sampling variance-covariance matrix.

In the case of MC.MI, `Mplus` was used to generate  $M = 100$  data sets with complete data using joint specification multiple imputation from the amputed data set (Asparouhov

---

<sup>2</sup> The `ampute` function defines the proportion of missing cases as the proportion of the original case which would be dropped had listwise deletion been employed. There are other ways of thinking about proportion of missing data, for example, the number of cells in the data matrix with missing values over the total number of cells. For a more thorough presentation of the proportion of missing data used in the study see <https://jeksterslab.github.io/manMCMedMiss/articles/proportion-missing.html>.

& Muthen, 2022). Graham et al. (2007) for considerations in choosing the number of imputations. An unrestricted model, where no structural model is imposed, is specified in the imputation stage. Missing values are imputed, assuming that the variables come from a multivariate normal joint distribution, using Markov-Chain Monte Carlo. The default variance-covariance model approach was used. The simple mediation model was estimated using the normal-theory maximum likelihood estimator for each of the 100 completed data sets. The estimated parameters and their corresponding sampling variance-covariance matrix were extracted for each fitted model and pooled using Rubin's (1987) rules. The pooled vector of parameter estimates is given by

$$\bar{\boldsymbol{\theta}} = M^{-1} \sum_{m=1}^M \hat{\boldsymbol{\theta}}_m \quad (8)$$

where  $M$  is the number of imputations,  $m = \{1, 2, \dots, M\}$ , and  $\hat{\boldsymbol{\theta}}_m$  is the vector of parameter estimates for the  $m^{\text{th}}$  imputation. The pooled or total sampling variance-covariance matrix<sup>3</sup> consists of combining between and within imputation variances given by

$$\mathbf{V}_{\text{within}} = M^{-1} \sum_{m=1}^M \text{Var}(\hat{\boldsymbol{\theta}}_m) \quad (9)$$

$$\mathbf{V}_{\text{between}} = (M - 1)^{-1} \sum_{m=1}^M (\hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}})' \quad (10)$$

$$\mathbf{V}_{\text{total}} = \mathbf{V}_{\text{within}} + \mathbf{V}_{\text{between}} + M^{-1}\mathbf{V}_{\text{between}}. \quad (11)$$

The wrapper function `ImputeData` was used to perform the imputation stage while the `FitModelMI` was used to fit the model for each of the data sets, extract the parameter

---

<sup>3</sup> An alternative total variance was introduced by Li et al. (1991) and was used in the simulation. While results for this approach are available in the supplementary materials, they will be omitted in the manuscript as this approach did not provide a significant improvement to the performance of Equation 11.

estimates and the sampling-covariance matrix, and pool the parameter estimates and the sampling covariance matrix.

The second step is identical for MC.COMPLETE, MC.FIML, and MC.MI where the vector of parameter estimates and sampling variance-covariance matrix were used as arguments `mu` and `Sigma` for the `mvrnorm` function to generate the sampling distribution of parameter estimates using  $R = 20,000$ . The indirect effect is calculated for each of the generated vectors of parameter estimates. The quantiles corresponding to  $100(1 - \alpha)\%$  of the sampling distribution of the indirect effect correspond to the confidence intervals. The wrapper functions `MCML` and `MCMI` were used to implement the simulation step.

### ***Nonparametric Bootstrapping***

The conventional NB approach was implemented for simulation conditions with complete data (NB.COMPLETE). For conditions with missing data, we used NB followed by FIML (NB.FIML). It must be noted that assuming that the data is sampled from a multivariate normal distribution with adequate sample size and a large number of imputations is used, NB followed by MI and NB followed by FIML will produce similar results. This is based on the asymptotic behavior of MI, that is, MI will approach FIML as the number of imputations goes to infinity (Graham et al., 2007). Also, we used NB followed by FIML since FIML is more commonly used than TS-ML in applied research. As our focus is on the performance of MC with FIML and MC with MI, we believe that comparing the MC approaches proposed with NB followed by FIML is enough to accomplish our objectives. We used `Mplus` to perform this method. We requested  $B = 5,000$  bootstrap samples in each replication. The `NBML` wrapper function was used to implement this method.

### **Evaluation**

For each combination of simulation factors described above, 5,000 samples were generated. Each sample is referred to as a simulation replication. The CI methods and SIG were used to make inferences for the indirect effect in each of the samples. Type I error

rate and statistical power were evaluated for the CI methods and SIG. The miss rate was evaluated for the CI methods.

For the CI methods, our definition of Type I and Type II errors was different from the typical null hypothesis significance test. The Type I error rate was defined as the proportion of replications where the CI did not contain zero when the population  $\alpha\beta$  is zero. The Type II error rate was defined as the proportion of replications where the CI contained zero when the population  $\alpha\beta$  is greater than zero. Statistical power was defined as 1 minus the Type II error rate.

We also examined the coverage of the CIs, that is, the proportion of replications where the CIs contained the population indirect effect. To be consistent with examining error rates, we decided to evaluate the miss rate, that is 1 minus the coverage proportion. We excluded conditions where the population indirect effect is zero in the miss rate evaluation since this is equivalent to the Type I error rate defined above.

Type I error and miss rates within .035 and .065 were considered robust for a 95% CI (Serlin, 2000). This is a compromise between criteria proposed by Cochran (1952) and Bradley (1978). Statistical power greater than or equal to .80 was considered robust (Cohen, 1988).

## Results

### Type I Error

Figures 1, 2, and 3 present Type I error rates for complete data, MCAR with 30% missing cases, and MAR with 30% missing cases respectively. Results for 10% missing cases for MCAR and MAR closely resembled results for complete data and results for 20% missing cases for MCAR and MAR closely resembled results for 30% missing cases. As such, figures for these cases were not presented. The figures show that all methods had deflated Type I error rates when both  $\alpha$  and  $\beta$  are zero. NBBC had inflated Type I error rates associated with small effect and sample sizes, which is consistent with previous research (e.g., Biesanz et al., 2010; Hayes & Scharkow, 2013; MacKinnon et al., 2004;

Tofighi & Kelley, 2020). An increase in effect size in  $\alpha$  or  $\beta$  while the other effect size is zero and an increase in sample size reduced the inflation of the Type I error rate. For example, for  $.00(\alpha = .00, \beta = .38)$  and  $.00(\alpha = .38, \beta = .00)$ , NBBC requires sample sizes greater than 200. Furthermore, for  $.00(\alpha = .00, \beta = .68)$ ,  $.00(\alpha = .68, \beta = .00)$ ,  $.00(\alpha = .00, \beta = .71)$ , and  $.00(\alpha = .71, \beta = .00)$ , NBBC requires sample sizes greater than 100. If we ignore the deflated Type I error rates when both  $\alpha$  and  $\beta$  are zero, MC, NBPC, and SIG Type I error rates were well within the robustness criterion in all simulation cases.

### Statistical Power

Figures 4, 5, and 6 present statistical power for complete data, MCAR with 30% missing cases, and MAR with 30% missing cases respectively. Note that the overall pattern of results for cases with missing data is similar to the results for complete data. There are no substantial differences between the methods except when sample sizes were small (e.g., 50 and 75) and the effect sizes were on the smaller end of the spectrum. When  $n \geq 150$ , all cases for all methods had statistical power greater than or equal to .80 regardless of type and percentage of missingness. When  $n < 150$ , an increase in the percentage of missing data for MCAR and MAR resulted in some decrease in the statistical power compared to the results for complete data. The figures also show that, with smaller sample sizes, statistical power can differ for indirect effects with the same effect size but different combinations of  $\alpha$  and  $\beta$ . For instance,  $.27(\alpha = .71, \beta = .38)$  performed relatively poorly compared to  $.27(\alpha = .38, \beta = .71)$  regardless of the method. The same pattern is true for  $.23(\alpha = .60, \beta = .38)$  compared to  $.23(\alpha = .38, \beta = .60)$ . It seems that given the same indirect effect, the statistical power is less when  $\alpha$  is relatively bigger than  $\beta$ .

### Miss Rate

Figures 7, 8, and 9 present miss rates for nonzero effect sizes for complete data, MCAR with 30% missing cases, and MAR with 30% missing cases respectively. The pattern of results were virtually identical for cases with complete data and with missing data. Almost all the simulation cases were within the robustness criterion regardless of the

type and percentage of missingness and the method. There were, however, some cases outside the robustness criteria when sample sizes were small (e.g., 50 and 75). However, these cases were still within the liberal criteria proposed by Bradley (1978).

## Discussion

The simulation study results show that MC (both MC.FIML and MC.MI) is a viable method for constructing CIs for the indirect effect when missing cases exist. Not only is MC accurate, but it also outperforms NBBC in terms of Type I error rate. Previous research has suggested that NBPC be used instead of NBBC due to NBBC's inflated Type I error rate. SIG's Type I error rate was as accurate as MC's. All methods had comparable miss rates and statistical power.

Given these findings, as well as the advantages of MC described in the introduction, MC can be used when other methods are difficult or impossible to implement. When it comes to resampling techniques, the overhead of multiple model fitting is frequently the bottleneck, especially for methods that require the use of optimization algorithms, such as SEM. In the case of combining MI and NB, the user fits the model 500,000 times assuming 100 MIs and 5,000  $B$  bootstrap resamples. In the case of combining ML and NB, the user fits the model 5,000 times with methods that account for missingness such as FIML and TS-ML, assuming that 5,000  $B$  bootstrap resamples are used. The computational demand increases as model complexity increases. In terms of implementation, MC is straightforward. When combining MC and FIML, the user must fit the model once in the estimation step and then use the parameter estimates and sampling variance-covariance matrix in the simulation step. The MI part is only performed once in the case of MC.MI. Because generating the sampling distribution of the indirect effect does not require repeated model fitting, it is quick and efficient. **Benchmarks comparing MC with NB are available in <https://jeksterslab.github.io/manMCMedMiss/articles/benchmark.html> and <https://jeksterslab.github.io/manMCMedMiss/articles/benchmark-semcci.html>.**

Separating the estimation and simulation steps has two additional benefits. First,

the user can perform the usual model fitting in their preferred statistical package in step 1. For example, popular SEM software packages like **AMOS**, **EQS**, **Mplus**, or **lavaan** can be used to estimate the parameters and the sampling variance-covariance matrix. No special software is required to perform MC because it only requires parameter estimates and the sampling variance-covariance matrix. Any statistical package or programming language capable of simulating data from a given distribution can be used for step 2. The majority of popular open-source data analysis software can generate multivariate normal data (e.g., **R**, **Python**, and **Julia**). Sample scripts for implementing MC in **R**, **Python**, and **Julia** are available in <https://jeksterslab.github.io/manMCMedMiss/articles/open-source.html>. Second, MC can be used even when raw data is unavailable. As a result, MC can be used in secondary analyses like meta-analysis. While there is a growing trend toward data sharing for reproducibility, concerns about confidentiality and anonymity can prevent organizations and research institutes from making raw data available. As long as summary statistics are available, MC can still be used in these situations.

For researchers who are already using **lavaan** to fit structural equation models, we provide the **R** package **semcci**, which is a simple and convenient tool for generating MC CIs from **lavaan** output. **Lavaan** is used to fit the model, and the output is passed as an argument to the **MC** function. The **fit lavaan** object is passed to the **MC** function as an argument. The **semcci** package can also generate confidence intervals for standardized parameters including standardized indirect effects. The **semcci** package is available in CRAN (<https://CRAN.R-project.org/package=semcci>). Documentation and examples are available in <https://jeksterslab.github.io/semcci/>. The **MonteCarloCI** function in the **semTools** (Jorgensen et al., 2022) package is another option. For researcher who wish to implement MC using MI estimates in **R**, a worked example is provided as a supplementary material in <https://jeksterslab.github.io/manMCMedMiss/articles/mc-mi.html>.

```

# Step 1: Model Fitting

library(lavaan)

model <- "Y ~ cp * X + b * M
          M ~ a * X
          ab := a * b"

fit <- sem(model = model, data = data, missing = "fiml")

# Step 2: Monte Carlo CIs

library(semmcci)

MC(fit)

#> Monte Carlo Confidence Intervals

#>           est      se      R  0.05%   0.5%   2.5%  97.5% 99.5% 99.95%
#> cp        0.2333 0.0263 20000 0.1490 0.1655 0.1813 0.2846 0.3002 0.3183
#> b         0.5082 0.0273 20000 0.4205 0.4397 0.4549 0.5616 0.5779 0.6004
#> a         0.4820 0.0264 20000 0.3959 0.4140 0.4299 0.5337 0.5495 0.5642
#> Y~~Y      0.5462 0.0244 20000 0.4660 0.4832 0.4979 0.5944 0.6095 0.6263
#> M~~M      0.7527 0.0339 20000 0.6471 0.6660 0.6859 0.8181 0.8402 0.8613
#> ab        0.2449 0.0187 20000 0.1878 0.1995 0.2091 0.2829 0.2958 0.3085

```

Finally, we acknowledge the study's limitations and make recommendations for future research. As with any other simulation study, caution has to be made regarding generalizing the results to broader research settings. We only studied a limited number of missing value percentages. While the highest missing value percentage that we studied is close to the maximum percentage of missing cases in cross-sectional studies (27.84%), it is far from the maximum in longitudinal designs (67%) (see Peugh & Enders, 2004). Future research can look into other missing data patterns and situations when percentages of missing data are larger than what was considered in this study. Second, the simulation study used data generated from the multivariate normal distribution. Psychological research data deviate from the normal distribution, according to studies (Blanca et al.,

2013; Micceri, 1989). Consequently, when the data is nonnormal, the performance of MC should be investigated. In the future, the performance of MC.FIML can be investigated when the sampling variance-covariance matrix of the parameters is estimated using robust maximum likelihood (MLR in `Mplus` and `lavaan`) (see Savalei & Rosseel, 2021, for SEM options for nonnormal data with missing cases). The robustness of MC.MI when dealing with nonnormal data can also be investigated. In theory, MC will work even if the data comes from a nonnormal distribution if the parameter estimates are unbiased, the sampling variance-covariance matrix is efficient, and the parametric assumption on the distribution of the estimates is asymptotically normal. Third, we only used a simple mediation model, which may not be realistic in real-world research situations. This was a deliberate choice to demonstrate the performance of MC versus NB under a wide range of simulation conditions, including different effect size combinations, sample sizes, missing data mechanisms, and missing data percentages. However, future research can look into how an increase in model complexity affects the results. While future research is needed to address the limitations of our work, evidence of finite sample performance of MC in the two state of the art missing data approaches (MC.FIML and MC.MI) presented in this manuscript gives researchers fast and efficient alternative approaches to handle their missing data concerns.

### Open Practices Statement

The data and materials for this study are available in GitHub (see  
<https://github.com/jeksterslab/manMCMedMiss> and  
<https://jeksterslab.github.io/manMCMedMiss/index.html>).

## References

- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling*. Psychology Press. <https://doi.org/10.4324/9781315827414>
- Arbuckle, J. L. (2021). *Amos 28.0 user's guide*. Chicago, IBM SPSS.
- Aroian, L. A. (1947). The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics*, 18(2), 265–271.  
<https://doi.org/10.1214/aoms/1177730442>
- Asparouhov, T., & Muthén, B. (2022). *Multiple imputation with Mplus* (tech. rep.).  
<http://www.statmodel.com/download/Imputations7.pdf>
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.  
<https://doi.org/10.1037/0022-3514.51.6.1173>
- Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11(2), 142–163.  
<https://doi.org/10.1037/1082-989x.11.2.142>
- Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*, 45(4), 661–701. <https://doi.org/10.1080/00273171.2010.498292>
- Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology*, 9(2), 78–84.  
<https://doi.org/10.1027/1614-2241/a000057>
- Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 20, 115.  
<https://doi.org/10.2307/271084>

- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2), 144–152. <https://doi.org/10.1111/j.2044-8317.1978.tb00581.x>
- Cheung, G. W., & Lau, R. S. (2007). Testing mediation and suppression effects of latent variables. *Organizational Research Methods*, 11(2), 296–325. <https://doi.org/10.1177/1094428107300343>
- Cheung, M. W.-L. (2009a). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*, 41(2), 425–438. <https://doi.org/10.3758/brm.41.2.425>
- Cheung, M. W.-L. (2009b). Constructing approximate confidence intervals for parameters with structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(2), 267–294. <https://doi.org/10.1080/10705510902751291>
- Cheung, M. W.-L. (2021). Synthesizing indirect effects in mediation models with meta-analytic methods. *Alcohol and Alcoholism*, 57(1), 5–15. <https://doi.org/10.1093/alcalc/agab044>
- Cochran, W. G. (1952). The  $\chi^2$  test of goodness of fit. *The Annals of Mathematical Statistics*, 23(3), 315–345. <https://doi.org/10.1214/aoms/1177729380>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Craig, C. C. (1936). On the frequency function of  $xy$ . *The Annals of Mathematical Statistics*, 7(1), 1–15. <https://doi.org/10.1214/aoms/1177732541>
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511802843>
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397), 171–185. <https://doi.org/10.1080/01621459.1987.10478410>
- Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, 104(2), 293–296. <https://doi.org/10.1037/0033-2909.104.2.293>

- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.  
<https://doi.org/10.1201/9780429246593>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18(3), 233–239.  
<https://doi.org/10.1111/j.1467-9280.2007.01882.x>
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55(292), 708–713.  
<https://doi.org/10.1080/01621459.1960.10483369>
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3), 206–213. <https://doi.org/10.1007/s11121-007-0070-9>
- Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). Guilford Publications.
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis. *Psychological Science*, 24(10), 1918–1927. <https://doi.org/10.1177/0956797613480187>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling*.  
<https://CRAN.R-project.org/package=semTools>
- Koopman, J., Howe, M., & Hollenbeck, J. R. (2014). Pulling the Sobel test up by its bootstraps. In *More statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences* (pp. 224–243). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780203775851>
- Koopman, J., Howe, M., Hollenbeck, J. R., & Sin, H.-P. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology*, 100(1), 194–202. <https://doi.org/10.1037/a0036635>

- 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/10.1080/01621459.1991.10475152>
- Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley. <https://doi.org/10.1002/9781119482260>
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 39(3), 384–389. <https://doi.org/10.3758/bf03193007>
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83–104. <https://doi.org/10.1037/1082-989x.7.1.83>
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99–128. [https://doi.org/10.1207/s15327906mbr3901\\_4](https://doi.org/10.1207/s15327906mbr3901_4)
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Erlbaum Psych Press. <https://doi.org/10.4324/9780203809556>
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156–166. <https://doi.org/10.1037/0033-2909.105.1.156>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide. Eighth edition*. Los Angeles, CA, Muthén & Muthén.
- Pawitan, Y. (2013). *In all likelihood: Statistical modelling and inference using likelihood*. Oxford University Press.
- Pesigan, I. J. A., & Cheung, S. F. (2020). SEM-based methods to form confidence intervals for indirect effect: Still applicable given nonnormality, under certain conditions. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.571928>

- 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(4), 525–556. <https://doi.org/10.3102/00346543074004525>
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4), 717–731. <https://doi.org/10.3758/bf03206553>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/brm.40.3.879>
- Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6(2), 77–98. <https://doi.org/10.1080/19312458.2012.679848>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Raghunathan, T. E., Lepkowski, J. M., Hoewyk, J. V., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27(1), 85–95.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- 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*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470316696>
- Savalei, V., & Rosseel, Y. (2021). Computational options for standard errors and test statistics with incomplete normal and nonnormal data in SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(2), 163–181. <https://doi.org/10.1080/10705511.2021.1877548>

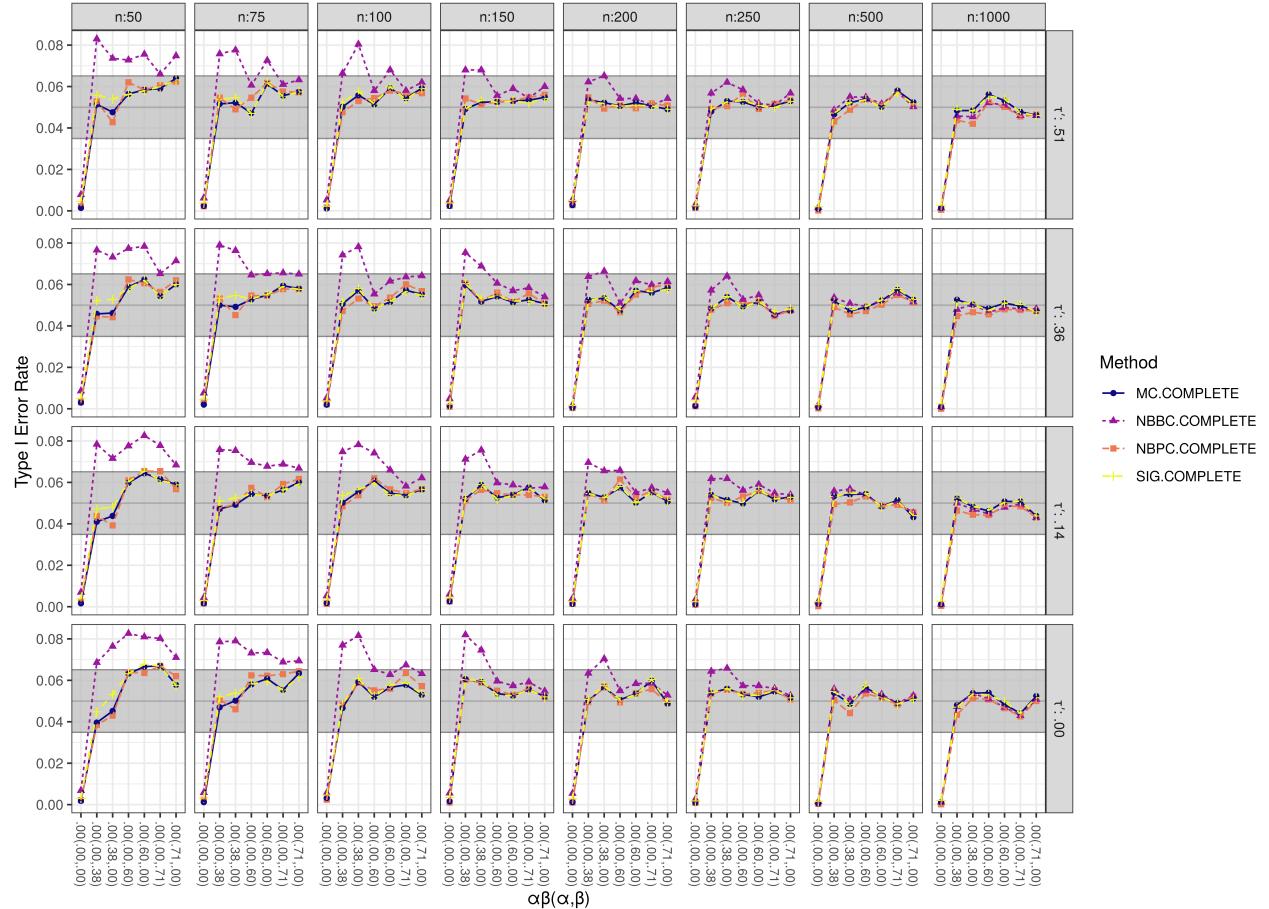
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman; Hall/CRC.  
<https://doi.org/10.1201/9780367803025>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989x.7.2.147>
- Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating missing values for simulation purposes: A multivariate amputation procedure. *Journal of Statistical Computation and Simulation*, 88(15), 2909–2930. <https://doi.org/10.1080/00949655.2018.1491577>
- Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods*, 5(2), 230–240. <https://doi.org/10.1037/1082-989x.5.2.230>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422–445. <https://doi.org/10.1037/1082-989x.7.4.422>
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290. <https://doi.org/10.2307/270723>
- Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological Methodology*, 16, 159.  
<https://doi.org/10.2307/270922>
- Sobel, M. E. (1987). Direct and indirect effects in linear structural equation models. *Sociological Methods & Research*, 16(1), 155–176.  
<https://doi.org/10.1177/0049124187016001006>
- Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2007). Tests of the three-path mediated effect. *Organizational Research Methods*, 11(2), 241–269.  
<https://doi.org/10.1177/1094428107300344>
- Tofghi, D., & Kelley, K. (2019). Indirect effects in sequential mediation models: Evaluating methods for hypothesis testing and confidence interval formation. *Multivariate Behavioral Research*, 55(2), 188–210.  
<https://doi.org/10.1080/00273171.2019.1618545>

- Tofghi, D., & Kelley, K. (2020). Improved inference in mediation analysis: Introducing the model-based constrained optimization procedure. *Psychological Methods*, 25, 496–515. <https://doi.org/10.1037/met0000259>
- Tofghi, D., & MacKinnon, D. P. (2015). Monte Carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(2), 194–205. <https://doi.org/10.1080/10705511.2015.1057284>
- van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman; Hall/CRC. <https://doi.org/10.1201/9780429492259>
- van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049–1064. <https://doi.org/10.1080/10629360600810434>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3). <https://doi.org/10.18637/jss.v045.i03>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer New York. <https://doi.org/10.1007/978-0-387-21706-2>
- Venzon, D. J., & Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics*, 37(1), 87. <https://doi.org/10.2307/2347496>
- Wu, W., & Jia, F. (2013). A new procedure to test mediation with missing data through nonparametric bootstrapping and multiple imputation. *Multivariate Behavioral Research*, 48(5), 663–691. <https://doi.org/10.1080/00273171.2013.816235>
- Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30(1), 165–200. <https://doi.org/10.1111/0081-1750.00078>

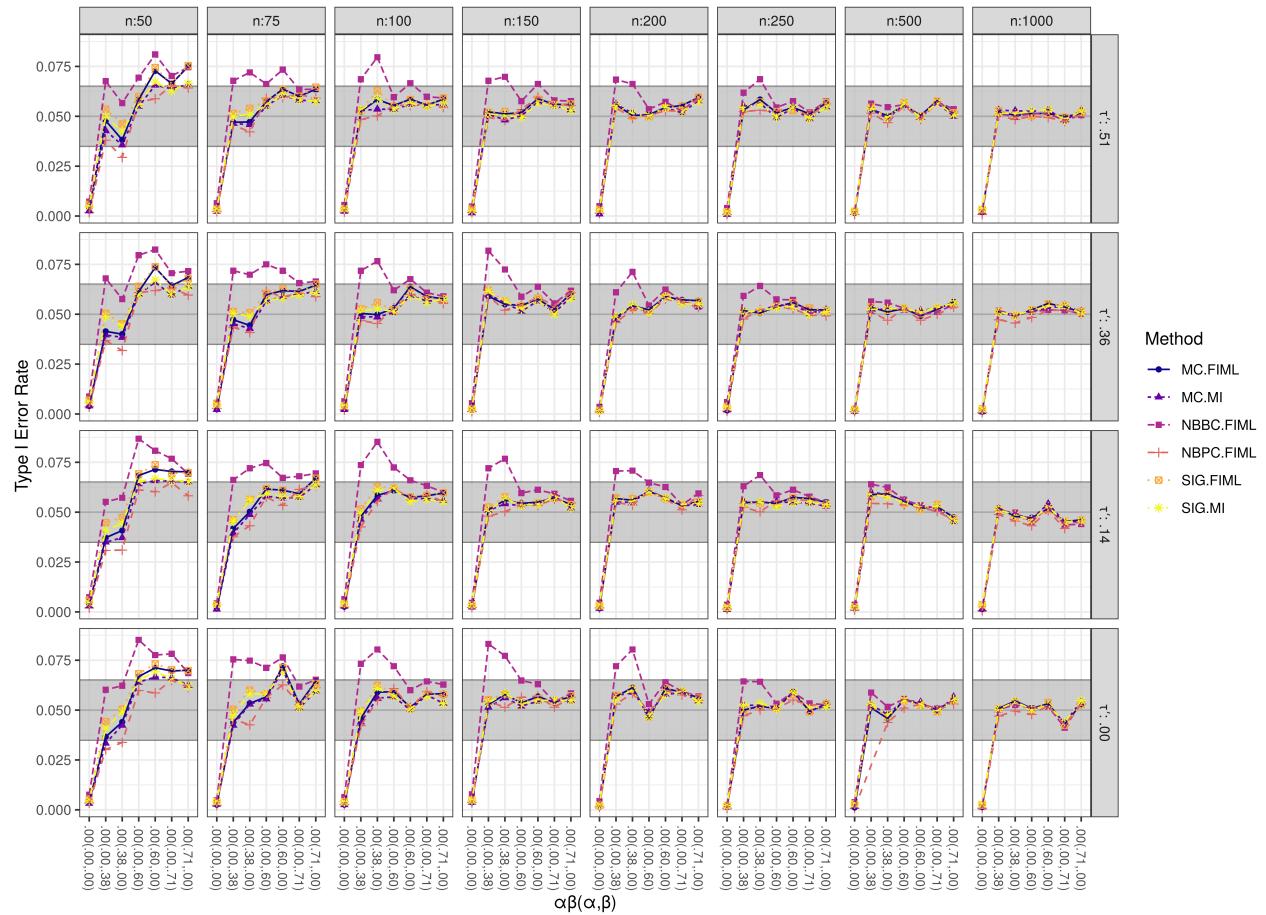
- Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology*, 115(6), 929–943.  
<https://doi.org/10.1037/pspa0000132>
- Zhang, Z., & Wang, L. (2012). Methods for mediation analysis with missing data. *Psychometrika*, 78(1), 154–184. <https://doi.org/10.1007/s11336-012-9301-5>
- Zhang, Z., Wang, L., & Tong, X. (2015). Mediation analysis with missing data through multiple imputation and bootstrap. In *Quantitative psychology research* (pp. 341–355). Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-19977-1\\_24](https://doi.org/10.1007/978-3-319-19977-1_24)

**Figure 1**

*Type I Error for Complete Data*



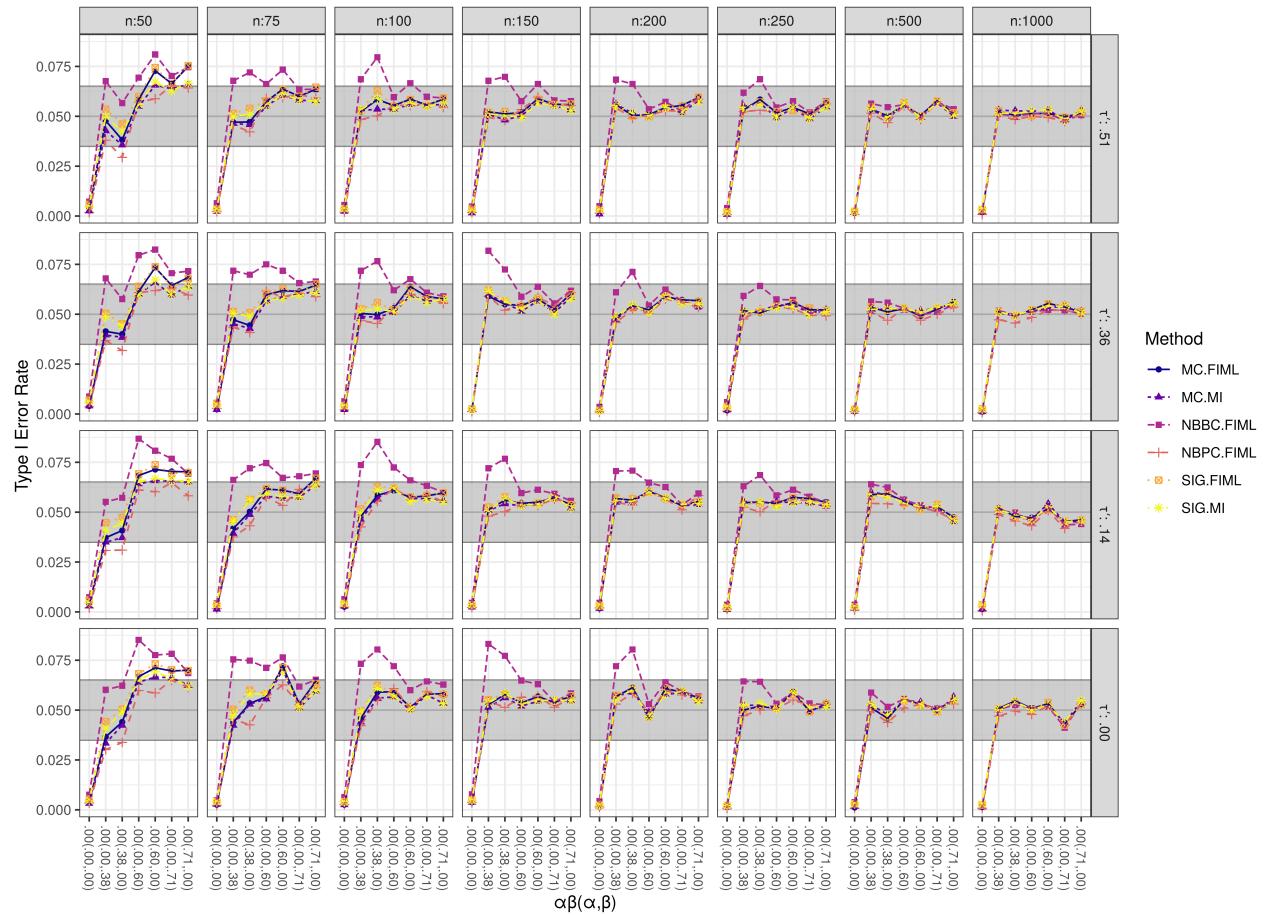
Note.  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.COMPLETE = Monte Carlo with complete data, NBBC.COMPLETE = Nonparametric bootstrap with bias-corrected confidence intervals with complete data, NBPC.COMPLETE = Nonparametric bootstrap with percentile confidence intervals with complete data, SIG.COMPLETE = Joint-significance test with complete data.

**Figure 2***Type I Error for MCAR (30% Missing)*

*Note.*  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.FIML = Monte Carlo using FIML estimates, MC.MI = Monte Carlo using MI estimates, NBBC.FIML = Nonparametric bootstrap with bias-corrected confidence intervals using FIML, NBPC.FIML = Nonparametric bootstrap with percentile confidence intervals using FIML, SIG.FIML = Joint-significance test using FIML estimates, SIG.MI = Joint-significance test using MI estimates.

**Figure 3**

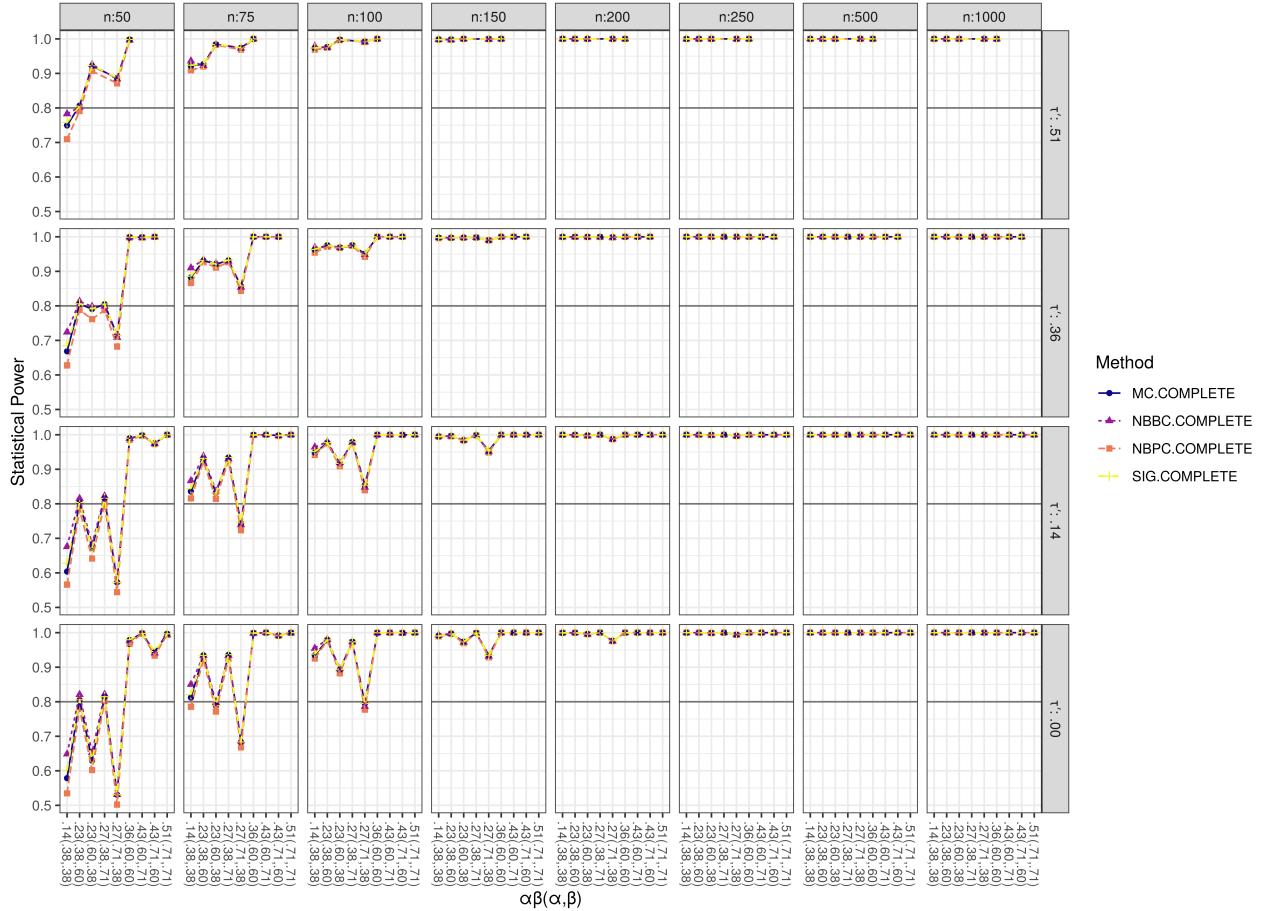
Type I Error for MAR (30% Missing)



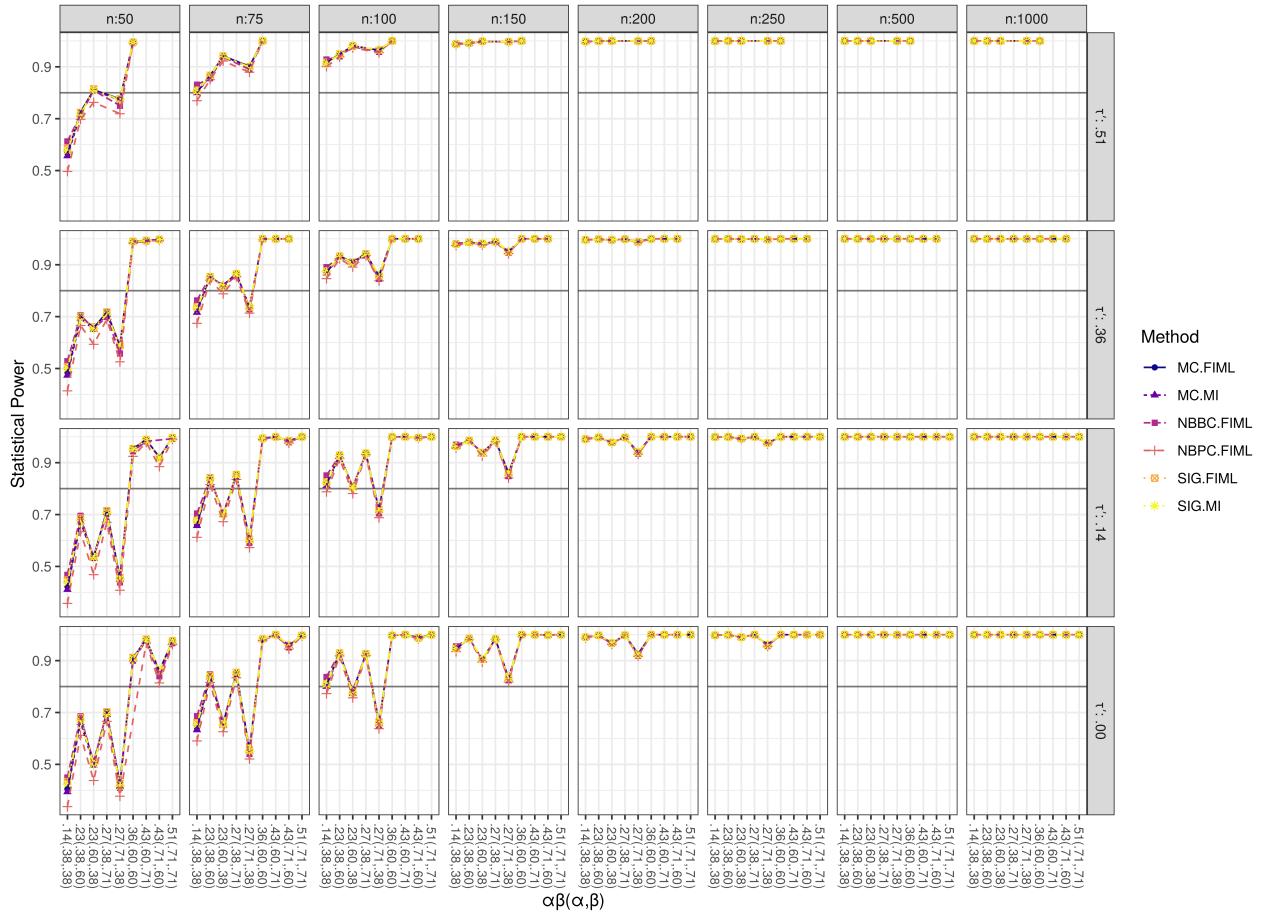
Note.  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.FIML = Monte Carlo using FIML estimates, MC.MI = Monte Carlo using MI estimates, NBBC.FIML = Nonparametric bootstrap with bias-corrected confidence intervals using FIML, NBPC.FIML = Nonparametric bootstrap with percentile confidence intervals using FIML, SIG.FIML = Joint-significance test using FIML estimates, SIG.MI = Joint-significance test using MI estimates.

**Figure 4**

*Statistical Power for Complete Data*



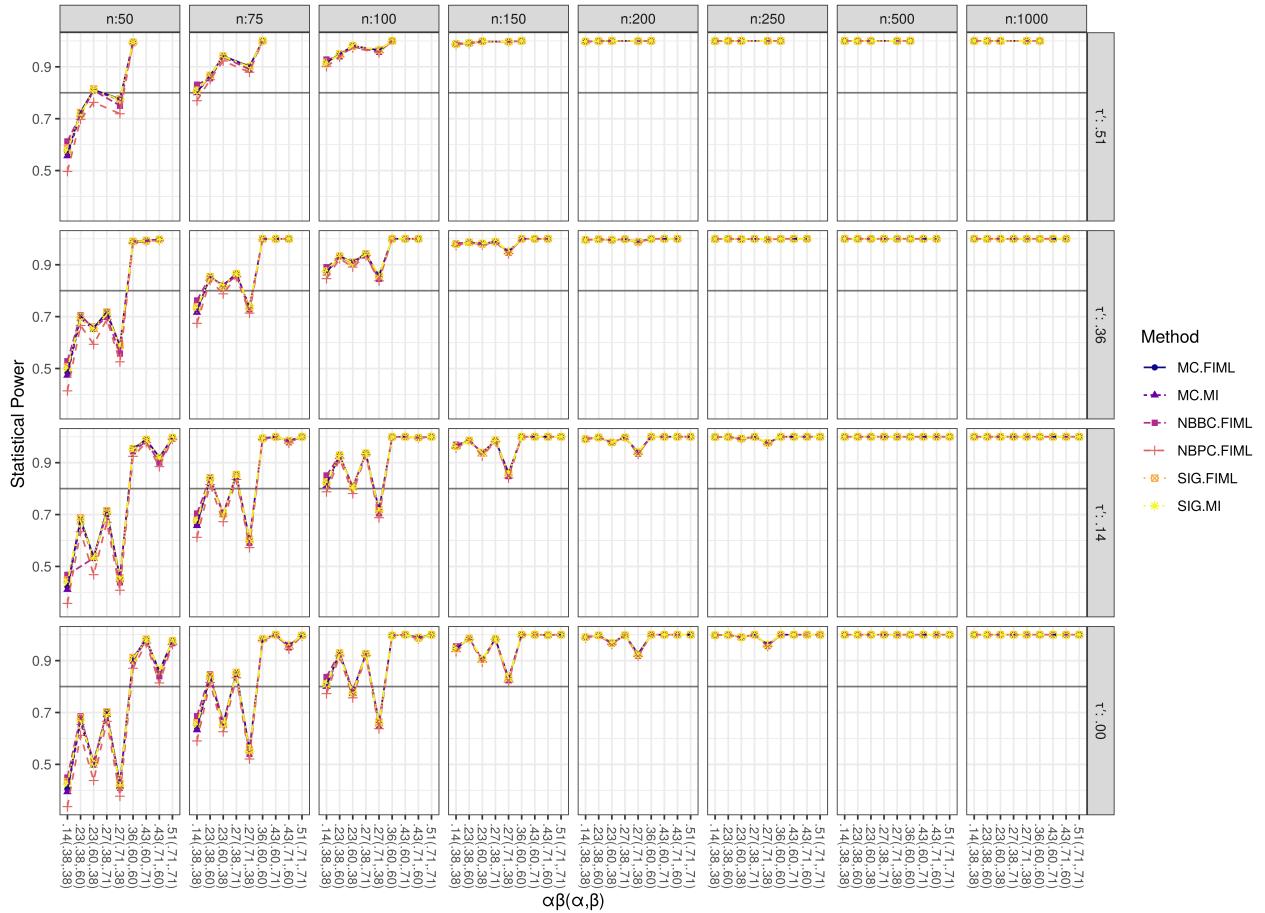
Note.  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.COMPLETE = Monte Carlo with complete data, NBBC.COMPLETE = Nonparametric bootstrap with bias-corrected confidence intervals with complete data, NBPC.COMPLETE = Nonparametric bootstrap with percentile confidence intervals with complete data, SIG.COMPLETE = Joint-significance test with complete data.

**Figure 5***Statistical Power for MCAR (30% Missing)*

*Note.*  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.FIML = Monte Carlo using FIML estimates, MC.MI = Monte Carlo using MI estimates, NBBC.FIML = Nonparametric bootstrap with bias-corrected confidence intervals using FIML, NBPC.FIML = Nonparametric bootstrap with percentile confidence intervals using FIML, SIG.FIML = Joint-significance test using FIML estimates, SIG.MI = Joint-significance test using MI estimates.

**Figure 6**

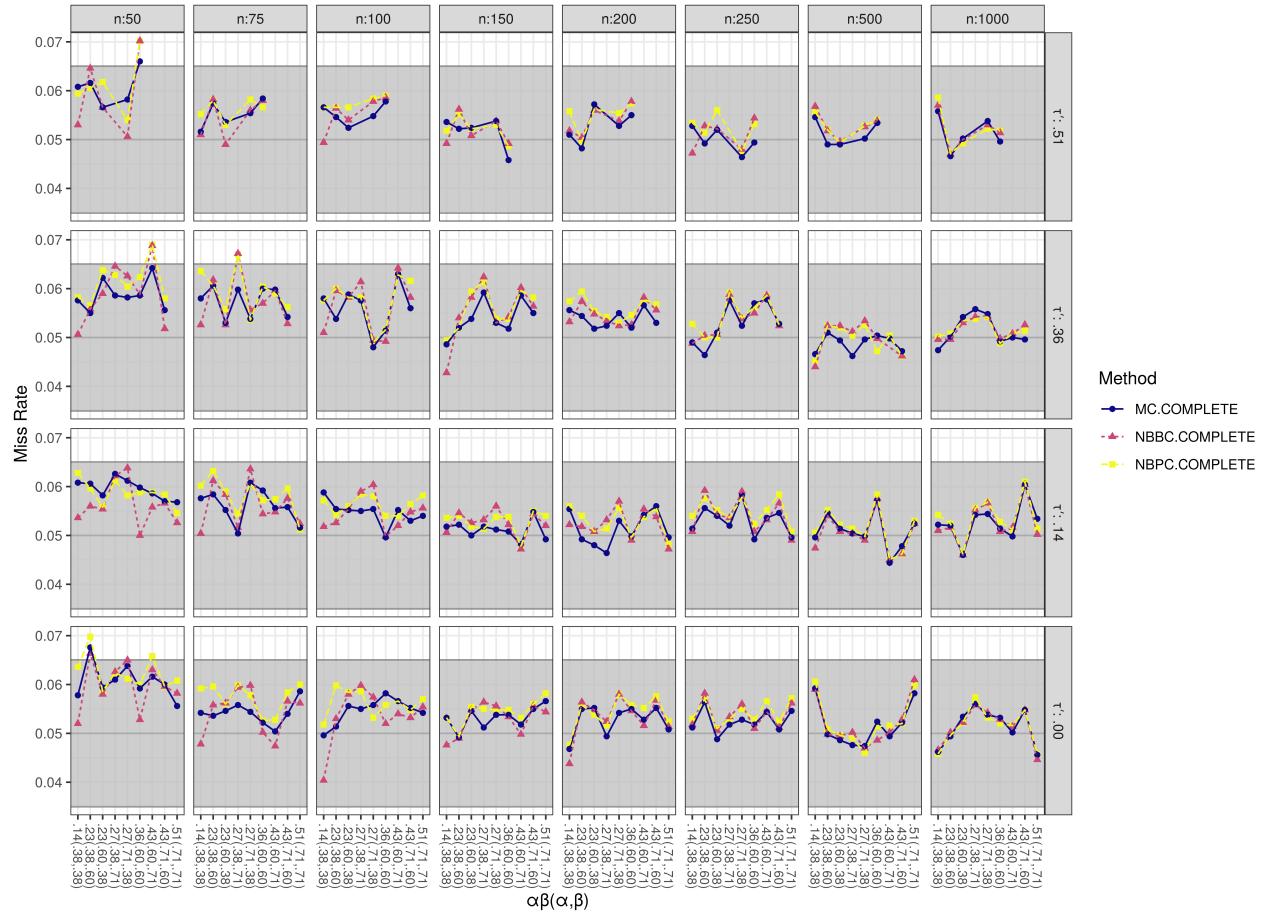
*Statistical Power for MAR (30% Missing)*



*Note.*  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.FIML = Monte Carlo using FIML estimates, MC.MI = Monte Carlo using MI estimates, NBBC.FIML = Nonparametric bootstrap with bias-corrected confidence intervals using FIML, NBPC.FIML = Nonparametric bootstrap with percentile confidence intervals using FIML, SIG.FIML = Joint-significance test using FIML estimates, SIG.MI = Joint-significance test using MI estimates.

**Figure 7**

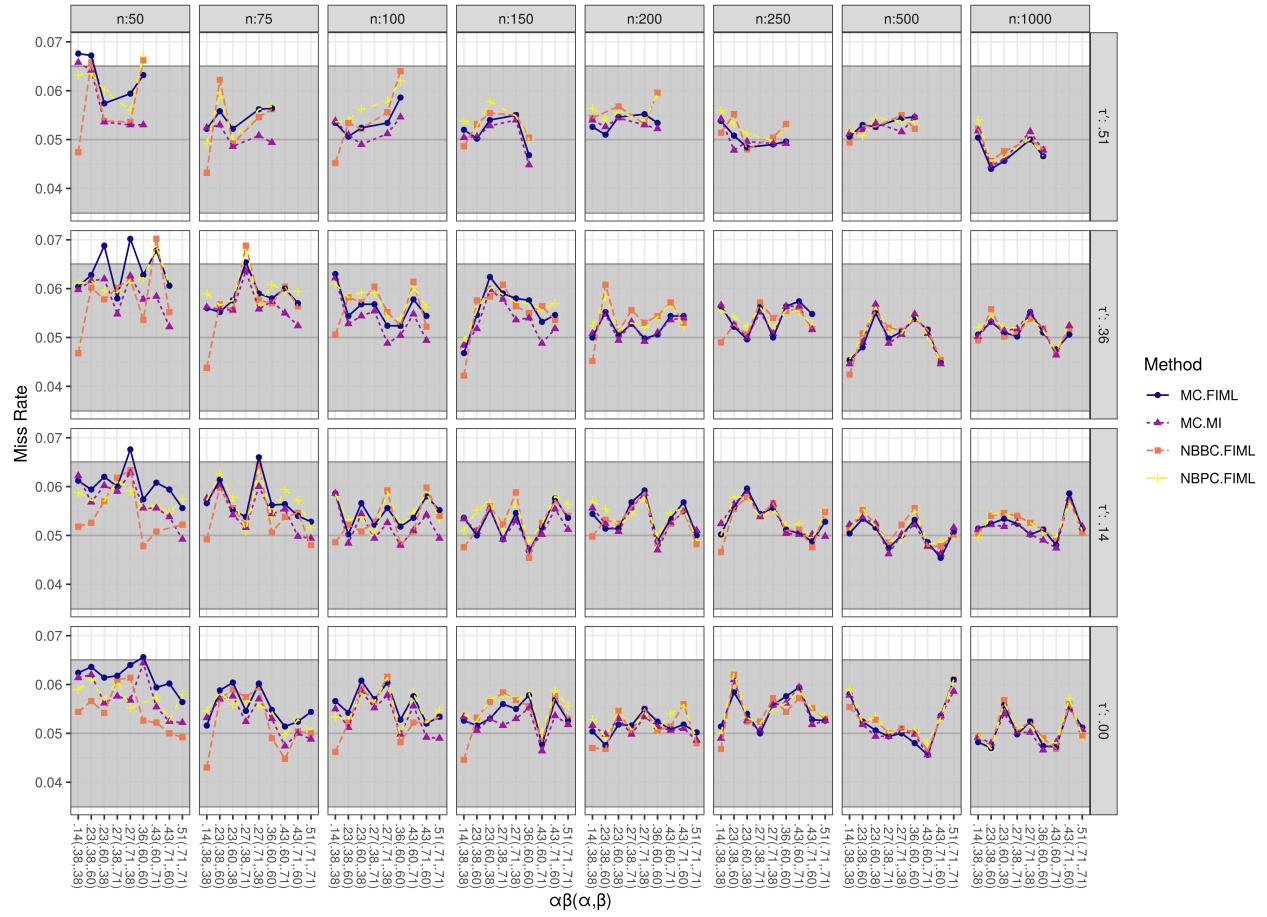
*Miss Rate for Complete Data*



Note.  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.COMPLETE = Monte Carlo with complete data, NBBC.COMPLETE = Nonparametric bootstrap with bias-corrected confidence intervals with complete data, NBPC.COMPLETE = Nonparametric bootstrap with percentile confidence intervals with complete data.

**Figure 8**

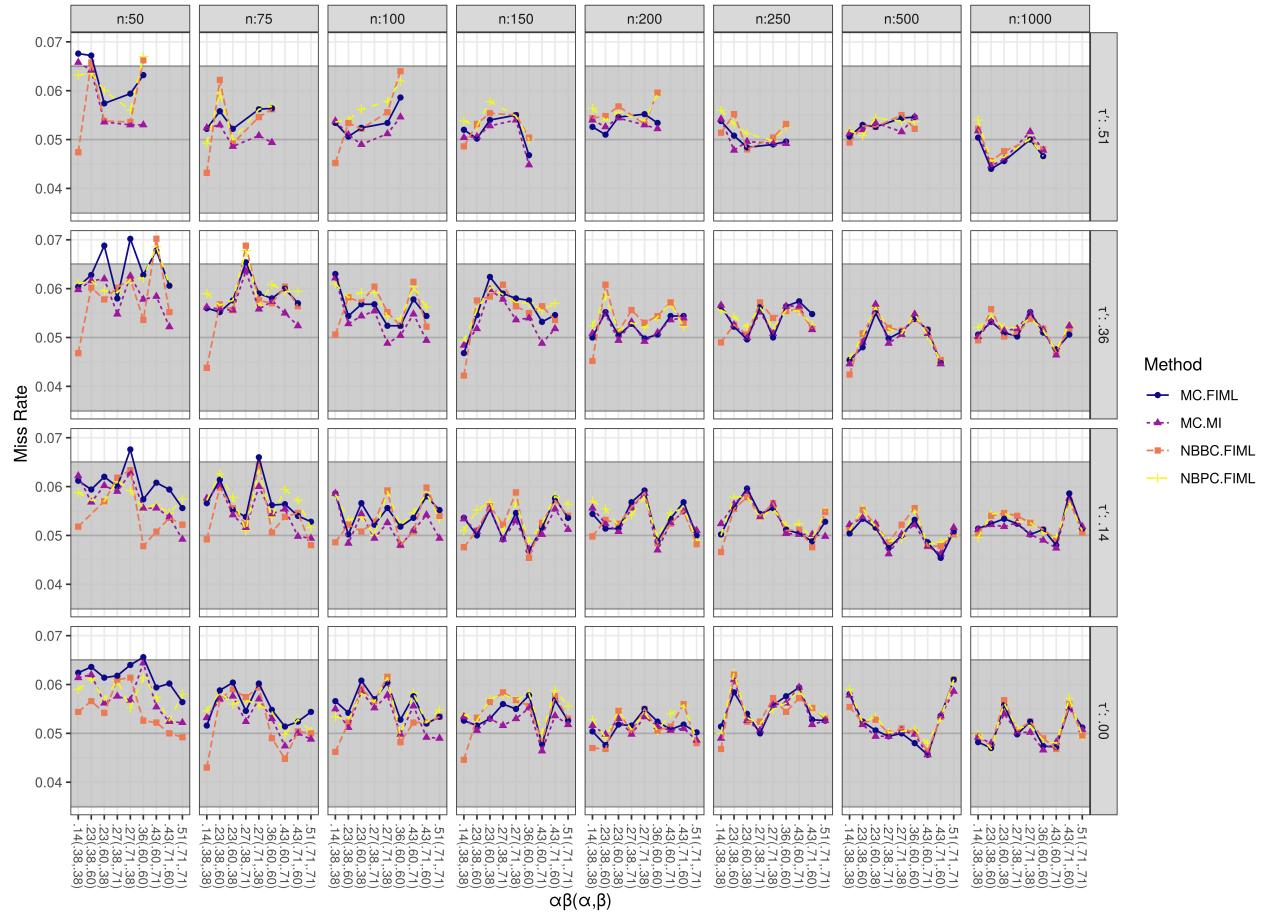
*Miss Rate for MCAR (30% Missing)*



*Note.*  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.FIML = Monte Carlo using FIML estimates, MC.MI = Monte Carlo using MI estimates, NBBC.FIML = Nonparametric bootstrap with bias-corrected confidence intervals using FIML, NBPC.FIML = Nonparametric bootstrap with percentile confidence intervals using FIML.

**Figure 9**

Miss Rate for MAR (30% Missing)



Note.  $n$  = sample size,  $\alpha$  = path from  $X$  to  $M$ ,  $\beta$  = path from  $M$  to  $Y$ ,  $\tau'$  = direct effect, MC.FIML = Monte Carlo using FIML estimates, MC.MI = Monte Carlo using MI estimates, NBBC.FIML = Nonparametric bootstrap with bias-corrected confidence intervals using FIML, NBPC.FIML = Nonparametric bootstrap with percentile confidence intervals using FIML.