Improved Inference in Mediation Analysis: Introducing the Model-Based Constrained Optimization Procedure

Davood Tofighi
University of New Mexico

Ken Kelley
University of Notre Dame

Abstract

Mediation analysis is an important approach for investigating causal pathways. One approach used in mediation analysis is the test of an indirect effect, which seeks to measure how the effect of an independent variable impacts an outcome variable through 1 or more mediators. However, in many situations the proposed tests of indirect effects, including popular confidence interval-based methods, tend to produce poor Type I error rates when mediation does not occur and, more generally, only allow dichotomous decisions of "not significant" or "significant" with regards to the statistical conclusion. To remedy these issues, we propose a new method, a likelihood ratio test (LRT), that uses nonlinear constraints in what we term the model-based constrained optimization (MBCO) procedure. The MBCO procedure (a) offers a more robust Type I error rate than existing methods; (b) provides a p value, which serves as a continuous measure of compatibility of data with the hypothesized null model (not just a dichotomous reject or fail-to-reject decision rule); (c) allows simple and complex hypotheses about mediation (i.e., 1 or more mediators; different mediational pathways); and (d) allows the mediation model to use observed or latent variables. The MBCO procedure is based on a structural equation modeling framework (even if latent variables are not specified) with specialized fitting routines, namely with the use of nonlinear constraints. We advocate using the MBCO procedure to test hypotheses about an indirect effect in addition to reporting a confidence interval to capture uncertainty about the indirect effect because this combination transcends existing methods.

Translational Abstract

Mediation analysis has become one of the most important approaches for investigating causal pathways. One instrument used in mediation analysis is a test of indirect effects that seeks to measure how the effect of an independent variable impacts an outcome variable through one or more mediators. However, in many situations the proposed tests of indirect effects, including popular confidence interval-based methods, tend to produce too few or too many false positives (Type I errors) and are commonly used to make *only* dichotomous decisions about acceptance ("not significant") or rejection ("significant") of a null hypothesis. To remedy these issues, we propose a new procedure to test an indirect effect. We call this new procedure the model-based constrained optimization (MBCO) procedure. The MBCO procedure (a) more accurately controls the false positive rate than existing methods; (b) provides a p value, which serves as a continuous measure of compatibility of data with the hypothesized null model (not just a dichotomous reject or fail-to-reject decision rule); (c) allows simple and complex hypotheses about mediation (i.e., one or more mediators; different mediational pathways); and (d) allows the mediation model to use observed or latent variables common in psychological research. We advocate using the MBCO procedure to test hypotheses about an indirect effect in addition to reporting a confidence interval to capture uncertainty about the indirect effect.

Keywords: indirect effect, mediation analysis, confidence interval, likelihood ratio, model-comparison test

Supplemental materials: http://dx.doi.org/10.1037/met0000259.supp

Mediation analysis has become one of the most important and widely used methods to study causal mechanisms in psychology, management, education, and other related fields (e.g., Ato García, Vallejo Seco, & Ato Lozano, 2014; Boies, Fiset, & Gill, 2015; Bulls et al., 2017; Carmeli, McKay, & Kaufman, 2014; Deković, Asscher, Manders, Prins, & van der Laan, 2012; Ernsting, Knoll, Schneider, & Schwarzer, 2015; Graça, Calheiros, & Oliveira, 2016; Haslam, Cruwys, Milne, Kan, & Haslam, 2016; Koning, Maric, MacKinnon, & Vollebergh, 2015; MacKinnon, 2008; Molina et al., 2014). Researchers have proposed several methods to test for the presence of mediation (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), including the widely recommended (a) Monte Carlo confidence interval (CI; MacKinnon, Lockwood, & Williams, 2004; Tofighi & MacKinnon, 2016), also known as parametric bootstrap method (Efron & Tibshirani, 1993); (b) per-

centile and bias-corrected (BC) bootstrap resampling CI (Bollen & Stine, 1990; Efron & Tibshirani, 1993; MacKinnon et al., 2004; Shrout & Bolger, 2002); (c) profile-likelihood CI method (Folmer, 1981; Neale & Miller, 1997; Pawitan, 2001; Pek & Wu, 2015); and (d) joint significance test (Kenny, Kashy, & Bolger, 1998; MacKinnon et al., 2002). Henceforth, we will use the term *CI-based methods* to emphasize a dual role that CIs have played in the mediation analysis literature. In addition to providing a range of plausible values for a population indirect effect, CIs have also controversially been used to test the null hypothesis of no mediation (i.e., a zero indirect effect) at the α level with a $(1-\alpha)$ 100% CI. However, testing an indirect effect is complicated because the indirect effect is the product of the coefficients along the mediation chain (MacKinnon, 2008). Further, each of the methods mentioned earlier has key limitations, many of which researchers commonly encounter. Thus, even though these procedures are currently recommended, that recommendation in research is questionable.

Many popular testing methods in mediation models are limited because they lack robustness of the observed Type I error rate. That is, the empirical probability of falsely rejecting the null hypothesis about an indirect effect does not equal, and can frequently fall outside of, a robustness interval for the Type I error rate, such as Bradley's (1978) liberal interval $[0.5\alpha, 1.5\alpha]$, where α is the nominal significance level (Biesanz, Falk, & Savalei, 2010; Tofighi & Kelley, 2019). The idea of using .025–.075 as an acceptable departure from the idealized properties of the statistical procedures, when $\alpha = .05$, acknowledges that a null hypothesis procedure may not work perfectly but is still useful. Bradley's (1978) liberal criterion accepts a procedure as robust when violations of the model's assumptions are small enough not to fundamentally change the statistical conclusion. The importance of the conclusion in the context of mediation, for example, is when sample size is less than 100 and $\alpha = .05$. Several methods, including widely recommended CI-based methods, can result in Type I error rates that fall outside of Bradley's liberal interval [0.025 .075] (Tofighi & Kelley, 2019) and are, therefore, unsatisfactory. We propose a method that is more robust than other procedures because it better satisfies Bradley's (1978) liberal criterion and has desirable statistical properties that allow its wide use, including in situations beyond which existing methods are applicable.

Currently used methods of testing mediation are limited because these tests are commonly used to make only dichotomous decisions about acceptance ("not significant") or rejection ("significant") of a null hypothesis. These dichotomous decisions provide a false sense of certainty about a model and data and are not generally recommended (Amrhein, Trafimow, & Greenland, 2019; Wasserstein, Schirm, & Lazar, 2019). Further, these methods do not offer either a *p* value, a continuous measure of compatibility between data and the null hypothesis (Greenland et al., 2016), or an additional means of measuring statistical evidence, such as a likelihood ratio (Blume, 2002). As recommended by Wilkinson and the Task Force on Statistical Inference, American Psychological Association, Science Directorate (1999), reporting exact *p* values, CIs, and effect sizes, and not just noting "reject" or "fail-to-reject" is the ideal for science to progress. None of the CI-based tests, by their very nature, offers additional ways of evaluating the strength of evidence against a null hypothesis of no mediation. A likelihood ratio, however, is a continuous measure that can compare the goodness of fit of two competing models

under standard assumptions, as measured by the ratio of the models' maximized likelihoods. One model, a null model, represents and satisfies the null hypothesis of no mediation (e.g., zero indirect effect), and the other model, a full model, shows that mediation does occur, as posited by a researcher (e.g., the indirect effect is nonzero). Further, a likelihood value can be used to compute information fit indices, such as Akaike information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC; Schwarz, 1978); such indices compare the fit of the full and null models while penalizing each model for a lack of parsimony (i.e., the number of model parameters).

To address these limitations, we propose a model-based constrained optimization (MBCO) method as a new procedure that tests any function representing an indirect effect[1] and that provides a formal way to recast testing simple and complex hypotheses about an indirect effect into a model-comparison framework. The MBCO procedure offers two formal mechanisms to evaluate hypotheses about indirect effects in the model-comparison framework. First, the MBCO procedure offers a likelihood ratio test (LRT) statistic, which we term LRT_{MBCO}, that has a large sample chi-squared distribution and provides a *p* value to test hypotheses about an indirect effect. Second, the MBCO procedure compares the fit of two models in terms of a likelihood ratio and information fit indices, such as the AIC and BIC, as well as their generalizations.

In the model-comparison framework, we estimate the null and full models representing different hypotheses about an indirect effect under a null and an alternative hypothesis, respectively (see Maxwell, Delaney, & Kelley, 2018). A null model represents the null hypothesis about the indirect effect. Unlike existing approaches to model comparisons, the MBCO procedure can estimate the null mediation model by using nonlinear constraints for testing a null hypothesis in mediation models. This innovative application of nonlinear constraints can be used to test both simple and complex hypotheses in mediation models. The MBCO procedure also allows the indirect effect of a full model, posited by a researcher, to be freely estimated. For example, a full model is a mediation model shown in Figure 1. Although model comparisons using the likelihood ratio test are common in techniques such as multilevel modeling (Raudenbush & Bryk, 2002) and structural equation modeling (SEM; Kline, 2016), in almost all such situations the parameters are subject to linear constraints. Testing indirect effects is problematic because constraining the indirect effect, H_0: $\beta_1\beta_2 = 0$, is inherently a nonlinear constraint optimization problem as the constraint is the product of two parameters (Snyman, 2005).

The MBCO procedure offers several advantages over the existing, recommended methods. First, the MBCO procedure offers the test statistic LRT_{MBCO} that has better statistical properties than the widely recommended approaches in the literature. This test statistic, widely used in statistics, has desirable properties for testing complex null hypotheses (Cox & Hinkley, 2000). The LRT_{MBCO} yields more robust Type I error rates than the currently recommended methods when the null hypothesis of no mediation occurs. Second, the MBCO procedure offers researchers multiple ways of evaluating a null hypothesis of an indirect effect rather than only

---

[1] More formally, a function of model parameters should be a *smooth* function, which is a function that has continuous derivatives up to a desired order (Weisstein, 2018).
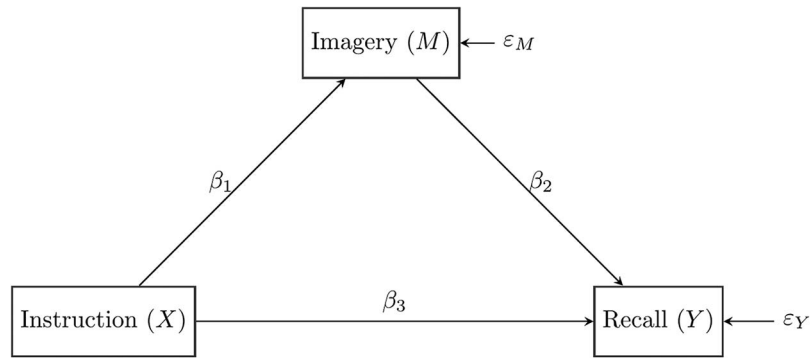
Figure 1. An example of a randomized single-mediator model. Instruction (X) is a random assignment with two conditions: instruction to use mental imagery rehearsal ($X = 1$) versus instruction to use repetition ($X = 0$). Imagery is a self-reported score of using mental imagery to memorize words, and recall is the number of words out of a total of 20 words that each student correctly remembered at the end of the experiment. A solid arrow between two variables indicates a direct effect of the variable on the left on the other variable. Greek letters denote population values, where βs denote the regression (path) coefficients, with β_1 being the treatment effect on imagery, β_2 being the partial effect of imagery on recall controlling for treatment effect, and β_3 is the partial effect of instruction on recall controlling for imagery. Terms ε_Y and ε_M denote residuals that are assumed to be normally distributed.

allowing a dichotomous reject or fail conclusion that has been criticized by researchers (e.g., Cohen, 1994; Harlow, Mulaik, & Steiger, 1997; Wasserstein et al., 2019). The MBCO procedure also allows researchers to test simple and complex hypotheses for a variety of mediation models involving multiple mediators (observed or latent) in a flexible manner. This flexibility aids researchers interested in complex causal models that are often posited by theories and can be implemented in an SEM framework. Such flexibility and desirable statistical performance make the MBCO procedure a desirable advancement for mediation studies. To make our proposed method immediately available to researchers, we provide detailed instructions along with the necessary R (R Core Team, 2019) code (see the online supplemental materials). Nonlinear constraint optimization algorithms to implement the MBCO procedure and to compute the LRT_{MBCO} are already implemented in at least one R software package, and we expect other software packages to soon make such allowances as well.

The Model-Based Constrained Optimization Procedure

We begin by explaining the MBCO procedure conceptually and then discussing the required steps required to perform this procedure. Without loss of generality, we elaborate on the MBCO procedure to test the zero indirect effect null hypothesis, H_0: $\beta_1\beta_2 = 0$, for the single-mediator model example (see Figure 1). In the Empirical Example of Using the MBCO Procedure on a Complex Mediation Model section, we discuss extensions of the MBCO procedure to the more elaborate, parallel two-mediator model (see Figure 2). In the Simulation Study section, we briefly extend the MBCO procedure to a sequential two-mediator model and then compare the performance of the MBCO procedure with the most commonly used tests of indirect effects for both a single-mediator model and a sequential two-mediator model.

The MBCO procedure recasts hypotheses about indirect effects into a model-comparison framework. To illustrate, consider the single-mediator model (see Figure 1), which can be represented by the following regression equations:

$$M = \beta_{0M} + \beta_1 X + \varepsilon_M \tag{1}$$

$$Y = \beta_{0Y} + \beta_2 M + \beta_3 X + \varepsilon_Y, \tag{2}$$

where β_1 is the effect of instruction ($X = 1$ corresponds to instruction to use imagery rehearsal; $X = 0$ corresponds to instruction to use repetition rehearsal) on the use of imagery (imagery); β_2 is the effect of imagery on the number of words recalled (recall) controlling for instruction, and β_3 is the direct effect of the instruction on recall controlling for imagery. β_{0M} and β_{0Y} denote the intercepts while ε_M and ε_Y are residual terms. The indirect effect of instruction on recall through imagery is defined as whether instruction to use imagery increases the use of imagery and then, in turn, increases the number of words participants recalled. Under the no-omitted-confounder assumption, which assumes that a variable should not be omitted from the model that would influence both the mediator and the outcome variables, the indirect effect equals the product of two coefficients, $\beta_1\beta_2$ (Imai, Keele, & Tingley, 2010; Judd & Kenny, 1981; Valeri & Vander-Weele, 2013; VanderWeele, 2010). Suppose that we are interested in testing whether the indirect effect of instruction on recall through imagery is zero. We can use the following null hypothesis to answer this question:

$$H_0: \beta_1\beta_2 = 0$$

$$H_1: \beta_1\beta_2 \neq 0$$

To recast the hypotheses about the indirect effect into a model-comparison framework, we estimate two mediation models: a full (alternative) model and a null (restricted) model (Maxwell et al., 2018). Which model to estimate first is arbitrary. For our example, the full model is the single-mediator model in (1)–(2). Note that the full model does not constrain the value of the indirect effect, $\beta_1\beta_2$. Next, for the null model, we estimate a mediation model in which the indirect effect $\beta_1\beta_2$ is constrained to equal 0. Both the full and the null model can be represented by the set of equations in (1) and (2) with the limitation that $\beta_1\beta_2$ is constrained to be 0 in the null model. The null model is nested within the full model;
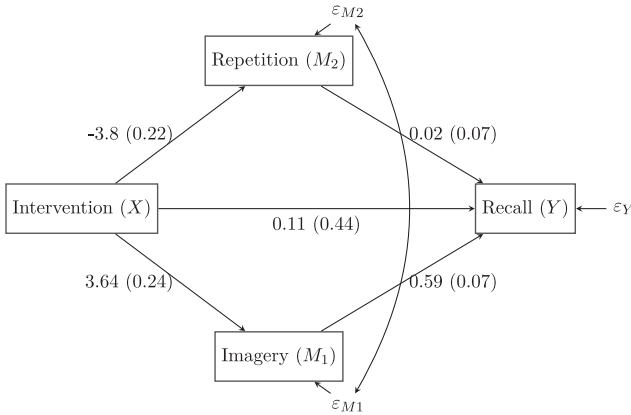
Figure 2. A randomized parallel (covarying) two-mediator model. Instruction (*X*) is a random assignment with two conditions: instruction to use mental imagery rehearsal ($X = 1$) versus instruction to use repetition ($X = 0$). Repetition is a self-reported score of using a repetition rehearsal to memorize words. Imagery is a self-reported score of using mental imagery to memorize words. Recall is the number of words out of a total of 20 words that each student correctly remembered at the end of the experiment and is the outcome variable of interest. A solid arrow from one variable to another indicates a direct effect of the variable on the other variable. The curved double-headed arrow shows covariance between two residual terms associated with the mediators, ε_{M1} and ε_{M2}. The numbers next to each arrow denote regression (path) coefficient estimates while the numbers in parentheses denote the respective standard errors. The ε terms denote residuals, which we assume are normally distributed.

that is, by setting the indirect effect to zero in the full model, we would obtain the null model.

The final step of the MBCO procedure is to formally compare the two models. One such approach compares the fit of the two models by computing LRT_{MBCO}. To compute the LRT_{MBCO}, we first calculate the deviance, denoted by $\mathcal{D}$, for each model:

$$\mathcal{D} = -2LL, \qquad (3)$$

where (*LL*) denotes the log-likelihood. The equation proposes that deviance equals twice the negative of the maximum value of the log-likelihood (*LL*) of the corresponding model. Deviance is a positive number quantifying the "badness" of fit or the "misfit" of a model. The larger the value of deviance for the same data set, the worse the fit of the model to that sample data.

Next, we compute the LRT_{MBCO} as the difference between the deviance for the null and full models. Under the null hypothesis, LRT_{MBCO} has a large sample chi-squared distribution (Wilks, 1938) as follows:

$$LRT_{MBCO} = \mathcal{D}_{null} - \mathcal{D}_{full} \sim \chi^2(df), \qquad (4)$$

where the degrees of freedom, *df*, equal the difference in the number of the free parameters between the two models. The larger values of the LRT_{MBCO} indicate that the null model, which estimates no indirect effect, fits the data worse than does the full model in which the indirect effect is freely estimated. To obtain a *p*-value for the LRT_{MBCO}, we calculate the upper tail probability from the chi-squared distribution (i.e., the area larger than the observed chi-squared statistic) with the specified degrees of freedom. While using a likelihood ratio test is standard in SEM, the LRT_{MBCO} uses the nonlinear

constraint of $\beta_1\beta_2 = 0$ that has not been previously explored for testing the null hypothesis of no indirect effect. The nonlinear constraint of $\beta_1\beta_2 = 0$ is, in fact, much more difficult to implement than is simply fixing a single parameter to a specified value (such as zero). Because of this nonlinear constraint of $\beta_1\beta_2 = 0$, specialized algorithms must be used (discussed below), which at present are not implemented in all programs.

Additionally, the MBCO procedure can compare the fit of the null and full models in terms of information fit indices such as the AIC (Akaike, 1974) or BIC (Schwarz, 1978). Such indices consider the fit of the model (as measured by the maximized values of likelihood for the model) to the data while penalizing the model for estimating more parameters. To clarify, consider the formulas for AIC and BIC:

$$AIC = \mathcal{D} + 2k \qquad (5)$$

$$BIC = \mathcal{D} + 2k \ln(n), \qquad (6)$$

where *k* is the number of free parameters (e.g., degrees of freedom) and *n* is the sample size in a mediation model. Both the AIC and the BIC penalize estimating more parameters in a model by adding the term $2k$ (for AIC) and $2k\ln(n)$ (for BIC) to deviance. If two models have the same deviance, the AIC favors the model with fewer free parameters (i.e., the more parsimonious model). For the BIC, if two models have the same deviance and sample size, the BIC favors the more parsimonious model. A lower AIC or BIC value indicates a better fit.

The MBCO procedure and LRT_{MBCO} have not previously been explored for testing the null hypothesis of no mediation for a few reasons. First, the nonlinear constraints require a sophisticated optimization algorithm (e.g., Zahery, Maes, & Neale, 2017). To our knowledge, the only structural equation modeling programs that are equipped to implement the nonlinear constraint are OpenMx (Boker et al., 2011; Neale et al., 2016) and SAS PROC CALIS (SAS Institute Inc., 2016). Second, researchers commonly rely on the CI-based tests because other researchers recommend them (e.g., MacKinnon et al., 2004; Preacher & Hayes, 2008; Shrout & Bolger, 2002). Third, few researchers have realized the infinite number of ways to realize and test the null hypothesis of no indirect effect, such as $H_0: \beta_1\beta_2 = 0$. For example, β_1 can be zero but β_2 may take on any value; β_2 can be zero but β_1 may take on any value; or β_1 and β_2 are both zero. Thus, in an infinite number of ways, mediation does not happen. The MBCO procedure offers flexibility in testing and improvement in evaluating simple and complex hypotheses about indirect effects, which we demonstrate next using an empirical example.

Empirical Example of Using the MBCO Procedure on a Complex Mediation Model

The MBCO procedure can be applied to a mediation model with two parallel mediators; in fact, the MBCO procedure can test a variety of simple and complex hypotheses about indirect effects and can offer flexibility in testing and improvement in evaluating these hypotheses. Using R (R Core Team, 2019) and the OpenMx (Boker et al., 2011; Neale et al., 2016) package, we apply and illustrate parts of the code and output to a sample problem. Full descriptions of the OpenMx code and output are given in the online supplemental materials.

Table 1
Means, Standard Deviations, and Correlations With Confidence Intervals

Variable	M	SD	1	2	3
1. Repetition	6.08	2.84			
2. Recall	12.07	3.4	−.28 [−.37, −.19]		
3. Imagery	5.66	2.96	−.56 [−.63, −.49]	.51 [.43, .58]	
4. Instruction	0.51	0.5	−.67 [−.72, −.61]	.32 [.22, .41]	.62 [.55, .68]

Note. M and SD are used to represent mean and standard deviation, respectively. Instruction (X) is a random assignment with two conditions: instruction to use mental imagery rehearsal ($X = 1$) versus instruction to use repetition ($X = 0$). Repetition is a self-reported score of using a repetition rehearsal to memorize words. Imagery is a self-reported score of using mental imagery to memorize words. Recall is the number of words out of a total of 20 words that each student correctly remembered at the end of the experiment and is the outcome variable of interest. Values in square brackets indicate the 95% confidence interval for each correlation.

MacKinnon, Valente, and Wurpts (2018) discussed a study on using mental imagery to improve the memorization of words. Table 1 shows descriptive statistics for this empirical example. Oliver, Bays, and Zabrucky (2016) showed individuals who are instructed to create mental images of words appear to recall more of them than do individuals who were simply instructed to remember the words. MacKinnon et al. (2018) conducted another study to test the mediation hypothesis that instructions to participants to create mental images of words would increase their using mental imagery and, in turn, would increase the number of words participants recalled (see Figure 2). The study was replicated eight times, each with a different group of undergraduate students enrolled in introductory psychology courses; the sample sizes were 77, 43, 24, 79, 22, 45, 35, and 44. At the beginning of each assessment, the instructor informed the students that they would hear a number of words and that they were to memorize the words using techniques described in the instructional sheets that they would receive. The students then were randomly assigned to receive instructions on one of two different memorization techniques: repetition or imagery rehearsal. The students who received repetition instructions were asked to memorize each word by repeating it to themselves. The students who received imagery instructions were asked to memorize each word by forming a mental image of the word along with other words they had heard during the experiment. For example, upon hearing the word *camel* and after hearing the word *women*, they might imagine a woman riding a camel. The students listened to 20 words with a 10-s interval to rehearse each word. Ten seconds after hearing the last word, the students were asked to write down as many words as they remembered. Next, the students were all asked to (a) rate the degree to which they formed images of the words in order to memorize the words using a scale of 1 = *not at all* to 9 = *absolutely*; and (b) rate the degree to which they memorized each word by repetition on a scale of 1 = *not at all* to 9 = *absolutely*.

For this mediation model, we were interested in testing the following hypotheses:

Research Question 1: Does the instruction to create mental images of words increase use of mental imagery that, in turn, increases the number of words recalled?

Research Question 2: Does the instruction to repeat words increase use of repetition to memorize the words that, in turn, increases the number of words recalled over and above using mental imagery?

Research Question 3: Is the indirect effect of instruction to create mental images on the number of words recalled through use of mental imagery greater than the indirect effect of instruction to repeat words on the number of words recalled through use of repetition?

To answer each research question, we implemented the MBCO procedure in three steps:

Step 1: Specify and estimate the full mediation model.
Step 2: Specify and estimate the null mediation model.
Step 3: Compare fit of the full and null models using the MBCO procedure.

For Research Question 1, we briefly explain OpenMx (Boker et al., 2011; Neale et al., 2016) syntax for conducting the MBCO procedure and provide the relevant part of OpenMx code and output. For Research Questions 2 and 3, we summarize the MBCO procedure statistical results.[2]

Research Question 1

Because we were interested in testing whether the indirect effect of instruction of recall through imagery is different from zero, we answered this question by using the single-mediator model (see Figure 1). We briefly introduced the key parts of OpenMx code (Boker et al., 2011; Neale et al., 2016) and R (R Core Team, 2019) and then followed the three steps outlined in the article to implement the MBCO procedure.

Before reporting the formal modeling process, we show a glimpse of the data set and its structure using the `glimpse` function from the dplyr package (Wickham, François, Henry, & Müller, 2019):

---

[2] All OpenMx scripts for the model along with a detailed explanation are provided in the online supplemental materials, which can also be found at https://github.com/quantPsych/mbco. For a comprehensive list of resources about the OpenMx software, we encourage readers to visit the project website https://openmx.ssri.psu.edu. For a detailed document on each function within R, use the `help()` function. For example, type `help(mxModel)` within the R console.

```
glimpse(memory_df)

## Observations: 369
## Variables: 4
## $ X          <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
## $ repetition <int> 1, 1, 1, 3, 7, 2, 1, 7, 3, 6, 6, 9, 3, 1, 4, 4, 3, 7,…
## $ recall     <int> 9, 14, 17, 8, 10, 10, 8, 12, 12, 15, 7, 10, 8, 9, 15,…
## $ imagery    <int> 6, 8, 9, 7, 7, 6, 3, 6, 7, 7, 8, 8, 5, 5, 8, 9, 4, 7,…
```

Step 1

The MBCO procedure begins by specifying the full single-mediator model (see Figure 1). We fit the full single-mediator model using the code shown below:

```
endVar <- c('imagery', 'recall') # name of the endogenous variables
maniVar <- c('X', 'imagery', 'recall') # name of the observed variables
# Specify the full single-mediator model using mxModel()
single_med_full <- mxModel(
  "single_med_full",
  type = "RAM",
  manifestVars = maniVar, #specify the manifest (observed) variables
  # specify the path from X to the M (imagery) and Y (recall)
  mxPath(
    from = "X",
    to = endVar,
    arrows = 1,
    free = TRUE,
    values = .2,
    labels = c("b1", "b3")
  ),
  # specify the path from M to Y
  mxPath(
    from = 'imagery',
    to = 'recall',
    arrows = 1,
    free = TRUE,
    values = .2,
    labels = "b2"
  ),
  # specify (residual) variances for the observed variables
  mxPath(
    from = maniVar,
    arrows = 2,
    free = TRUE,
    values = .8,
    labels = c("s2x", "s2em", "s2ey")
  ),
  # specify the intercepts for the endogenous variables
  mxPath(
    from = "one",
    to = endVar,
    arrows = 1,
    free = TRUE,
    values = .1,
    labels = c("int1","int2")
  ),
  # define the indirect effect and call it ind
  mxAlgebra(b1 * b2, name = "ind"),
  # specify the data set for analysis
  mxData(observed = memory_df, type = "raw")
)
# run the single mediator null model
fit_single_med_full <- mxRun(single_med_full)
# runs multiple times until convergence achieved
fit_single_med_full <- mxTryHard(single_med_full)
# save the results
stat_single_med_full <- summary(fit_single_med_full)
# print results of the full model
stat_single_med_full
```

In the first two lines of the script above, before specifying the model parts with `mxModel`, we can simplify the process by grouping the names of variables as a vector that will be used in model specification. For example, we specified a vector of the names of observed variables and saved them as `maniVar`. Then, we specified a vector of the names of endogenous variables and saved them as `endVar`. Next, we specified the full single-mediator model.

The main command to specify an SEM is `mxModel`. The arguments provided to the mxModel function specified all the elements of the mediation model. For the single-mediator example, we specified the manifest (observed) variables, the paths (regression coefficients), the indirect effect to be tested (or any function of model parameters), the constraint distinguishing the full and null models, and the variances and covariances among the variables. The first argument to `mxModel` is `single_med_full`, which is a name we chose for the single-mediator model. The next argument, `type="RAM"`, specifies that OpenMx uses the reticular action model (RAM; McArdle & McDonald, 1984), a symbolic algebraic notation to specify an SEM. In the argument `manifestVars`, we introduced the vector of the names of the observed (manifest) variables. The variable names must match the names in the data set `memory_df`, as shown earlier in the output from the `glimpse` function.

Next, we specified the paths between the variables using `mxPath`, a function that corresponds to the graphical representation of paths in an SEM such as the model in Figure 1. For example, we used `mxPath` to indicate a path (coefficient) corresponding to an arrow between the two variables specified in the arguments `from` (predictors) and `to` (response variables) in Figure 1. The argument `arrows=1` indicates a unidirectional arrow that starts from the variable in the argument `from` and ends at the variable specified in the argument as `to`. The argument `arrows=2` indicates a bidirectional arrow representing a covariance between the two variables. The argument `free=TRUE` indicates that the parameter is freely estimated; otherwise,

`free=FALSE` indicates that the parameter is fixed at the values set by the argument `values`. If the parameter is freely estimated, the argument `values` would provide starting values. The argument `labels` provides labels for the coefficients. Because we specified more than one coefficient in the arguments `to` and `from`, we could provide a vector of labels corresponding to the stated order of the coefficients. For our example, `b1` is the coefficient for $X \rightarrow$ imagery and `b3` is the coefficient for $X \rightarrow$ recall.

We used the function `mxAlgebra` to define the indirect effect that, in general, is a function of model parameters. A function may include combinations of mathematical operations, such as $+$, $-$, $*$, and $/$ (e.g., `b1*b2/(b1*b2+b3)`), exponentials (e.g., `exp`), and logarithms (e.g., `log`). The first argument to `mxAlgebra` was the product of two coefficients, `b1*b2`, in which `b1` and `b2` had been defined in `mxPath`. The argument `name="ind"` named the indirect effect. Next, we specified the data set for the model. The `mxData` identified the data set to be analyzed. The argument `observed=memory_df` specified the name of the data set in R. The second argument, `type="raw,"` indicated that the data set was in the raw format, which meant that the data set included observations on the participants as opposed to being a summary statistic such as a covariance matrix.

Finally, we ran the model using `mxRun`, where the first argument was the name of the `mxModel` that was then saved as `fit_single_med_full`. Because we received a warning (not shown here) from the optimizer that the convergence criterion was not satisfied, we used the function `mxTryHard`. This function would either run the model multiple times until an acceptable solution, according to the convergence criterion set by the estimation algorithm, was found or it would run a preset maximum number of times until an acceptable solution was reached. In the subsequent analysis, we first used `mxRun` and, if the model had convergence issues, we then used `mxTryHard`. We used the function `summary` to save or print the summary of the results. We saved the summary of the results as `stat_single_med_full` and then printed the summary. Below, is the relevant part of the summary results.

```
## Model Statistics:
##                | Parameters | Degrees of Freedom | Fit (-2lnL units)
##        Model:        8             1099                 4305.727
##    Saturated:        9             1098                 NA
## Independence:        6             1101                 NA
## Information Criteria:
##       | df Penalty | Parameters Penalty | Sample-Size Adjusted
## AIC:     2107.727        4321.727              4322.127
## BIC:    -2190.239        4353.013              4327.632
```

Below the title `Model Statistics`, the row that starts with `Model` gives the pertinent information for the full model. The output showed that the full single-mediator model had eight free parameters, with $df_{Full} = 1099$ and $\mathcal{D}_{Full} = 4305.727$. Under `Information Criteria`, we present the estimates of the information indices. The information indices that match the formulas in (5) and (6) are under the `Parameters Penalty` column. For the full model, the information fit indices were $AIC_{Full} = 4321.727$ and $BIC_{Full} = 4353.013$.

Before proceeding, after fitting a mediation model, we checked the degree to which the statistical assumptions about normality of the residuals and the presence of outliers to ensure that the esti-

mates were unbiased and the inference about the parameters was valid (Cohen, Cohen, West, & Aiken, 2003). We checked for normality of the residuals using QQ plots, which indicated that the normality assumption for the residuals was reasonable. Using the methods described by Fox (2016), we did not find any outliers that would change the results.

Step 2

We ran the null model (code below), which is the single-mediator model in which the indirect effect of instruction on recall through imagery is constrained to zero, $\beta_1\beta_2 = 0$.

```
single_med_null <-

  mxModel(model = single_med_full, name = "single_med_null",

          mxConstraint(ind == 0, name = "b1b2_equals_0"))

fit_single_med_null <- mxRun(single_med_null)

stat_single_med_null <- summary(fit_single_med_null)

stat_single_med_null
```

Instead of specifying all parts of the null mediation model using `MxModel` in the above code, we modified the full model `single_med_full`, as specified in the argument `model`. Next, we specified the nonlinear constraint for the indirect effect through `mxConstraint`. The first argument, `ind == 0`, constrains the indirect effect defined in the `mxAlgebra` statement to zero. The argument `name` assigns a name to the constraint. Finally, we saved the null model to `single_med_null`. Next, we ran the null model and saved the results to `fit_single_med_null`. Relevant parts of the summary of the model results are shown below:

```
## Model Statistics:
##               | Parameters  |  Degrees of Freedom  |  Fit (-2lnL units)
##      Model:          8                1100                  4481.493
##    Saturated:        9                1099                     NA
## Independence:        6                1102                     NA
## Number of observations/statistics: 369/1108
## Constraint 'b1b2_equals_0' contributes 1 observed statistic.
## Information Criteria:
##        | df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:      2281.493            4497.493                4497.893
## BIC:     -2020.384            4528.779                4503.398
```

For the null model, $df_{\text{Null}} = 1100$ and $\mathcal{D}_{\text{Null}} = 4481.493$, and the information fit indices were $\text{AIC}_{\text{Null}} = 4497.493$ and $\text{BIC}_{\text{Null}} = 4528.779$.

Step 3

We compared the full and null model both in terms of the LRT_{MBCO} and the information fit indices to evaluate H_0: $\beta_1\beta_2 = 0$. The LRT_{MBCO} equals the difference between the deviance of the two models: $\text{LRT}_{\text{MBCO}} = \mathcal{D}_{\text{Null}} - \mathcal{D}_{\text{Full}} = 4481.493 - 4305.727 = 175.766$. The degrees of freedom for the LRT_{MBCO} equal the difference in the degrees of freedom associated with each model; that is, $df_{\text{LRT}} = df_{\text{Null}} - df_{\text{Full}} = 1100 - 1099 = 1$. Given that the LRT_{MBCO} has a large sample χ^2 distribution with $df_{\text{LRT}} = 1$, we have the p-value $= \Pr(\chi^2(1) > 175.766) = 4.073738E - 40$. More conveniently, we computed the LRT_{MBCO} using the `mxCompare` or `anova` function where the arguments are the names of the full and null models:

```
mxCompare(fit_single_med,fit_single_med_null)

##           base        comparison    ep minus2LL   df     AIC      diffLL   diffdf
## 1 single_med_full                    8 4305.727  1099   2107.727    NA       NA
## 2 single_med_full single_med_null    8 4481.493  1100   2281.493  175.7657    1
##             p
## 1          NA
## 2 4.073738e-40
```

The first row of the above output shows the results for the full model, which is the single-mediator model in Figure 1. The columns `ep`, `minus2LL`, `df`, and `AIC` show the number of estimated parameters, deviance, degrees of freedom, and AIC, respectively. The columns `diffLL`, `diffdf`, and `p` represent the difference in deviance (*not* the log-likelihoods), the difference in degrees of freedom, and *p*-value for the two models being compared. The second row shows the results for the null model under the columns `ep`, `minus2LL`, `df`, and `AIC`. The results of comparing the null and full model are shown under the columns `diffLL`, `diffdf`, and `p`. The value for the test statistic $\text{LRT}_{\text{MBCO}} = 175.766$ is located under the column `diffLL`. The degrees of freedom are $df_{\text{LRT}} = 1$ and the *p*-value $= 4.073738E - 40$; these amounts are located under the columns `diffdf` and `p`, respectively.

We used the following commands to compute the Monte Carlo CI for the indirect effect. First, we used the functions `coef` and `vcov` to extract the path coefficients and covariance matrix of the coefficients, respectively. Next, we used the `ci` function in the RMediation package (Tofighi & MacKinnon, 2011, 2016) to compute the Monte Carlo CI. The first argument `mu` to this function is

a vector of the coefficient estimates, and the second argument `Sigma` is a covariance matrix of the coefficient estimates. The

```
Mu <- coef(fit_single_med) # coefficient estimates
Sigma <- vcov(fit_single_med) #covariance matrix of the coefficient estimates
ci(mu = Mu, Sigma = Sigma, quant = ~ b1 * b2)[1:3] #95% Monte Carlo CI

## [[1]]
##     2.5 %    97.5 %
## 1.599516 2.682124
##
## $Estimate
## [1] 2.120964
##
## $SE
## [1] 0.2761099
```

The numbers below 2.5% and 97.5% show the lower and upper limits of the 95% CI, [1.6, 2.682]; the numbers below `Estimate` and `SE` show the estimates of indirect effect and its standard error, $\hat{\beta}_1\hat{\beta}_2 = 2.121$ and $SE = 0.276$, respectively.

```
rsq(model = fit_single_med_full, name = c("imagery","recall") )

##   imagery    recall
## 0.3806254 0.2642803

rsq(model = fit_single_med_null, name = c("imagery","recall") )
##     imagery        recall
## 0.002711604 0.264281019
```

The first argument `model` to `rsq` specifies an OpenMx model, and the second argument `name` specifies names of endogenous variables (i.e., variables in which a single-headed arrow enters; namely those that are a function of another variable).

To summarize, the LRT_{MBCO} result showed that the indirect effect of instruction on recall through imagery appeared to be greater than zero, $\hat{\beta}_1\hat{\beta}_2 = 2.121$ ($SE = 0.276$), 95% Monte Carlo CI [1.6, 2.682], $LRT_{MBCO} = 175.766$, $df_{LRT} = 1$, $p = 4.073738E - 40$. We recommend researchers compute the difference in R^2s between the full and null models and then examine the change in the effect sizes that occurs as a result of the indirect effect through imagery. For recall, R^2 remained unchanged to four decimal places while for imagery $\Delta R^2 = .3779$. Further, the information fit indices supported the assertion that the indirect effect was greater than zero because the AIC and the BIC for the full single-mediator model were smaller than those of the null single-mediator model. This result indicates that constraining the indirect effect to zero worsens the fit of the full single-mediator model. On average, the instruction to students to use mental imagery increased use of mental imagery that, in turn, increased the number of words they recalled by approximately two words. We have inferential support that mediation occurred. Note these results are valid if the no-omitted confounder assumption is met. That is, an omitted variable may not exist that influences the relations between instruction, imagery, and recall.

argument quant accepts a formula for the indirect effect that starts with the symbol "~".

We next obtained R^2 for the endogenous variables of the full and null models using the `rsq` function available in the online supplemental materials.

Research Question 2

Here, we were interested in whether the instruction to use repetition increased the use of repetition to memorize the words that, in turn, improved participants' memory over and above the indirect effect of instruction on recall through imagery. In other words, we wanted to test H_0: $\beta_4\beta_5 = 0$ for the parallel two-mediator model in Figure 2. We again applied the three-step MBCO procedure.

Step 1

We estimated the model with two parallel mediators in Figure 2, which is the full model for this research question. For this model, the two specific indirect effects associated with imagery and repetition were freely estimated. Below are the regression equations for the full parallel two-mediator model:

$$M_1 = \beta_{0,M1} + \beta_1 X + \varepsilon_{M1} \quad (7)$$

$$M_2 = \beta_{0,M2} + \beta_4 X + \varepsilon_{M2} \quad (8)$$

$$Y = \beta_{0,Y} + \beta_3 X + \beta_2 M_1 + \beta_5 M_2 + \varepsilon_Y, \quad (9)$$

where β_1 is the effect of the instruction (X) on imagery (M_1), β_4 is the effect of the instruction on repetition (M_2), β_3 is the direct effect of the instruction on recall (Y), β_2 is the effect of imagery on recall controlling for the instruction and repetition, and β_5 is the effect of repetition on recall controlling for instruction and imag-

ery; $\beta_{0,M1}$, $\beta_{0,M2}$, $\beta_{0,Y}$ are the intercepts; ε_{M1}, ε_{M2}, and ε_Y are the residuals.

We fitted the full two-mediator model within OpenMx. The results showed that the full two-mediator model had 13 free parameters, with $df_{\text{Full}} = 1463$ and $\mathcal{D}_{\text{Full}} = 5874.685$. For the full model, the results were that $AIC_{\text{Full}} = 5900.685$ and $BIC_{\text{Full}} = 5951.526$; the effect sizes were $R^2_{\text{Imagery}} = .38$, $R^2_{\text{Repetition}} = .45$, and $R^2_{\text{Recall}} = .26$. We computed the specific indirect effects through imagery and repetition and the 95% CI for each specific indirect effect using the `ci` function in the RMediation package. The results showed that the specific indirect effect through imagery was $\hat{\beta}_1 \hat{\beta}_2 = 2.139$ ($SE = 0.284$), 95% Monte Carlo CI [1.602, 2.716], and the specific indirect effect through repetition was $\hat{\beta}_4 \hat{\beta}_5 = -0.082$ ($SE = 0.285$), 95% Monte Carlo CI [-0.643, 0.477].

Step 2

Recall that we were interested in whether instruction to use repetition increased the use of repetition to improve memory over and above the indirect effect of instruction on recall through imagery. Thus, in the null parallel two-mediator model, we constrained the specific indirect effect through repetition to zero while letting the specific indirect effect through imagery be freely estimated. The null parallel two-mediator model was estimated by fitting the equations in (7)–(9) subject to the null hypothesis constraint H_0: $\beta_4\beta_5 = 0$, which fixed the specific indirect effect through repetition to zero. In OpenMx, we specified the null model by adding the nonlinear constraint $\beta_4 \beta_5 = 0$ to the full model in Step 1. The results showed that the null parallel two-mediator model had $df_{\text{Null}} = 1464$ and $\mathcal{D}_{\text{Null}} = 5874.768$. The information fit indices for the null two-mediator model were $AIC_{\text{Null}} = 5900.768$ and $BIC_{\text{Null}} = 5951.609$. The effect sizes were $R^2_{\text{Imagery}} = .38$, $R^2_{\text{Repetition}} = .45$, and $R^2_{\text{Recall}} = .26$.

Step 3

We compared the two models and computed LRT_{MBCO}. The results showed that $LRT_{\text{MBCO}} = 0.083$, $df_{\text{LRT}} = 1$, and $p = .773$. The specific indirect effect through Repetition was, therefore, not different from zero, $\hat{\beta}_4\hat{\beta}_5 = -0.08$ ($SE = 0.29$), 95% Monte Carlo CI [-0.64, 0.48]. Further, the R^2 for imagery, repetition, and recall remained unchanged to three decimal places, and the information fit indices between the two models were roughly the same. These results indicate that the specific indirect effect through repetition above and beyond the specific indirect effect through Imagery does not appear to be different from zero.

Research Question 3

For this question, we were interested in comparing the sizes of the two specific indirect effects: the indirect effect of instruction on recall through imagery (i.e., $\beta_1\beta_2$) and the indirect effect of instruction on recall through repetition (i.e., $\beta_4\beta_5$). The null hypothesis for this research question is:

$$H_0: \beta_1\beta_2 = \beta_4\beta_5 \tag{10}$$

To test this null hypothesis, we again employed the three-step MBCO procedure.

Step 1

The full model for this research question was the same as the full model in Research Question 2; that is, they matched the parallel two-mediator model in (7)–(9). Thus, we used the results (i.e., the vector of the coefficient estimates, the covariance matrix of the coefficient estimate, R^2, indirect effects estimates, AIC, and BIC) of the full parallel two-mediator model from Research Question 2.

Step 2

The null model was the parallel two-mediator model in (7)–(9) subject to the null hypothesis constraint in (10). That is, we estimated the two-mediator model while we constrained the two specific indirect effects to be equal. Alternatively, we could specify the contrast of the two specific indirect effects to zero. We fitted the new null model in OpenMx. The results showed that the null model had $df_{\text{Null}} = 1464$ and $\mathcal{D}_{\text{Null}} = 5900.514$. The information fit indices for the null model were $AIC_{\text{Null}} = 5926.514$ and $BIC_{\text{Null}} = 5977.354$. The effect sizes were $R^2_{\text{Imagery}} = .38$, $R^2_{\text{Repetition}} = .45$, and $R^2_{\text{Recall}} = .21$.

Step 3

We compared the full parallel two-mediator model in Step 1 and the null parallel two-mediator model in Step 2. We computed the LRT_{MBCO} as well as the 95% Monte Carlo CI for the contrast of the two indirect effects using the `ci` function in the RMediation package. The results of the MBCO procedure showed that $LRT_{\text{MBCO}} = 25.828$, $df_{\text{LRT}} = 1$, and $p = 3.731857E - 07$. These outcomes indicate that the indirect effect through imagery appeared to be larger than the indirect effect through repetition by $\hat{\beta}_1\hat{\beta}_2 - \hat{\beta}_4\hat{\beta}_5 = 2.222$ ($SE = 0.445$) words, 95% Monte Carlo CI [1.364, 3.11]. In comparing the R^2s for the endogenous variables obtained from Step 1 and 2 for imagery and repetition, R^2s remained unchanged to three decimal places while for recall $\Delta R^2 = .05$. Comparing the information fit indices of the null model with the full model also supported the conclusion that the full model fitted the data better than did the null model.

Comparing Methods of Testing Indirect Effect Via Monte Carlo Simulations

In this section, we describe simulation studies assessing the Type I error rates and the statistical power of the LRT_{MBCO} of our proposed MBCO procedure; we then compare the Type I error rates and power of LRT_{MBCO} to those of the currently recommended methods of testing indirect effects. Monte Carlo simulation studies are necessary because there are no known ways of assessing the performance via mathematical proofs or derivations. In our simulations, we generated data from known population models, fitted the same models to the data, repeated this process many times, and assessed the properties of the different methods of testing indirect effects.

The simulation studies were designed to answer the following questions:

1. Does the LRT_{MBCO} provide more robust Type I error rates than the recommended methods? In particular,

would the LRT_{MBCO} remain robust across the combination of the parameter values for smaller sample sizes?

2. Is the LRT_{MBCO} as powerful as, if not more powerful than, the currently recommended methods across different effect sizes, sample sizes, and other types of mediation models?

To answer these questions, we compared the Type I error rate and power of the LRT_{MBCO} to five currently recommended methods of testing indirect effects: (a) the percentile bootstrap CI (Bollen & Stine, 1990; Efron & Tibshirani, 1993; MacKinnon et al., 2004); (b) the bias-corrected (BC) bootstrap CI (Bollen & Stine, 1990; Efron & Tibshirani, 1993; MacKinnon et al., 2004); (c) the Monte Carlo CI (MacKinnon et al., 2004; Tofighi & MacKinnon, 2016); (d) the profile-likelihood CI (Folmer, 1981; Neale & Miller, 1997; Pawitan, 2001; Pek & Wu, 2015); and (e) the joint significance test (Kenny et al., 1998; MacKinnon et al., 2002). We describe these methods in the next section.[3]

For a two-path indirect effect, we did not separately consider distribution of the product CI because its performance is similar to that of the Monte Carlo CI (MacKinnon et al., 2004). We also did not study the Wald (1943) (multivariate delta) z or the causal steps test because research has shown that these two tests are overly conservative and do not provide adequate power in smaller sample sizes and effect sizes (MacKinnon et al., 2002). For completeness of the simulation study and because of its popularity and endorsements (MacKinnon et al., 2004; Preacher & Hayes, 2008; Shrout & Bolger, 2002), we studied the bias-corrected (BC) bootstrap CI. However, we do not recommend using the BC bootstrap CI to test an indirect effect because it shows inflated Type I error rates for both small and large samples (Falk & Biesanz, 2015; Koopman, Howe, Hollenbeck, & Sin, 2015).

In the simulation studies, we considered two types of indirect effects commonly found in empirical research as well as in simulation studies (Tofighi & Kelley, 2019): (a) a two-path indirect effect (e.g., $\beta_1\beta_2$), which is the product of two coefficients for a single-mediator chain, $X \xrightarrow{\beta_1} M \xrightarrow{\beta_2} Y$; and (b) a three-path indirect effect (e.g., $\beta_1\beta_2\beta_3$), which is the product of three coefficients for a sequential two-mediator chain, $X \xrightarrow{\beta_1} M_1 \xrightarrow{\beta_2} M_2 \xrightarrow{\beta_3} Y$. We chose these two population models because they have been extensively discussed in both substantive and methodological literature (Tofighi & Kelley, 2019).

Percentile and Bias-Corrected Bootstrap CI

We now discuss two methods of computing a CI for nonparametric bootstrap (Bollen & Stine, 1990; Efron & Tibshirani, 1993). In this technique, many samples (e.g., $R = 1,000$) with replacement were drawn from the sample data. The hypothesized mediation model was fit to each resampled data set, and indirect effects were computed for each model. This process resulted in R estimates of indirect effects, which is called the *bootstrap sampling distribution* of the indirect effects. To compute a CI, we obtained quantiles corresponding to $1 - \alpha/2$ and $\alpha/2$ percentiles of the bootstrap sampling distribution, resulting in what is called a percentile CI. As a modified version of the percentile bootstrap, the BC bootstrap uses adjusted percentiles $1 - \alpha'/2$ and $\alpha'/2$ to obtain the upper and lower limits of the CI, where α' is computed to

correct for potential bias due to skewness (Efron, 1987). Although not in the context of mediation analysis, Efron (1987) argued that, for smaller sample sizes, the BC bootstrap CI yields a more accurate coverage (i.e., the proportion of times a CI contains true values of the parameter is closer to the nominal coverage of $1 - \alpha$) than does the percentile CI.

Monte Carlo CI

The Monte Carlo method, also known as the parametric bootstrap (Efron & Tibshirani, 1993), is another sampling-based technique used to compute a CI for indirect effects (MacKinnon et al., 2004; Tofighi & MacKinnon, 2016). In this method, the sampling distribution for each parameter is estimated by drawing R random samples from a multivariate normal distribution where the mean of the distribution is the vector of the coefficient estimates and the covariance matrix of the distribution is the covariance matrix of the coefficient estimates. The Monte Carlo method is based on the theory that the maximum likelihood (ML) estimates of the coefficients in an SEM asymptotically have a multivariate normal distribution (Bollen, 1989). The population parameters (i.e., mean vector and covariance matrix) for this multivariate normal distribution are the parameter estimates from the estimated model. The number of Monte Carlo samples, R, should be large, typically 1,000 or more. The indirect effect is estimated within each sample, resulting in a total of R estimates. The mean and standard deviation of the R indirect effect estimates are then used to estimate the indirect effect and its standard error, respectively. To compute a $(1 - \alpha)100\%$ CI, we obtained the $\alpha/2$ and $1 - \alpha/2$ quantiles of the Monte Carlo sample of the indirect effects.

Profile Likelihood CI

In its simplest form, the profile-likelihood approach produces a CI for a single parameter using a profile-likelihood function (Folmer, 1981; Neale & Miller, 1997; Pawitan, 2001). In mediation analysis, the profile-likelihood method has been extended to produce a CI for the indirect effect in a single-mediator chain (Cheung, 2007; Pek & Wu, 2015). Let θ be a vector of all the free parameters in a single-mediator model. The profile-likelihood function for a single-mediator model, $L(\theta \mid \beta_1\beta_2)$, is computed by assuming that the indirect effect, $\beta_1\beta_2$, is a known value. Next, the profile-likelihood function is maximized. In practice, the indirect effect takes on different values, and the profile-likelihood function is maximized while the indirect effect is fixed at specific values.

Next, we compared the deviance of the maximized profile-likelihood function and the deviance of the maximized likelihood function, where deviance equals negative twice the maximized log-likelihood function as shown in (3). Asymptotically, the difference in the deviance of the two functions had a chi-squared distribution with degrees of freedom equaling the difference in the number of free parameters between the profile likelihood and likelihood function for the model (Pawitan, 2001). For the indirect effect, in general, this difference asymptotically had a chi-squared distribution with one degree of freedom. The lower and upper

---

[3] We do not name all the existing methods for single-mediator models because they are considered elsewhere (e.g., Falk & Biesanz, 2015; MacKinnon et al., 2002).

limits for the $100(1 - \alpha)\%$ profile-likelihood CI corresponded to the minimum and maximum of all values of the indirect effect that satisfied the following equality:

$$\mathcal{D}(\hat{\boldsymbol{\theta}}_{\text{prof}} | \beta_1\beta_2) - \mathcal{D}(\hat{\boldsymbol{\theta}}) = \chi^2_\alpha(1) \qquad (11)$$

where $\hat{\boldsymbol{\theta}}_{\text{prof}}$ is the ML estimate of the model parameters given that the indirect effect is fixed and $\chi^2_\alpha(1)$ denotes upper-tail α critical value of the chi-squared distribution with one degree of freedom. For a specific mediation model, the values of $\mathcal{D}(\hat{\boldsymbol{\theta}})$ (computed by estimating the mediation model) and the critical value of the chi-squared distribution (e.g., $\chi^2_{.05}(1) = 3.84$) were known. A 95% profile-likelihood CI was computed by solving the expression in (11) for two values of $\beta_1\beta_2$ corresponding to the upper and lower limits.

Joint Significance Test

For a single-mediator model, MacKinnon, Lockwood, Hoffman, West, and Sheets (2002) proposed a joint significance test to describe a variation of the causal steps test (Kenny et al., 1998). This method declares an indirect effect to be significantly different from zero if every coefficient in the indirect effect is significantly different from zero. The joint significance test does not produce a robust Type I error rate for smaller sample sizes, a CI, or a p value for the indirect effect (MacKinnon et al., 2002). While the joint significance test can be extended to test indirect effects that are the products of two or more coefficients in a mediational chain, the method cannot be used to test a complex function of indirect effects such as a contrast of two indirect effects (e.g., H_0: $\beta_1\beta_2 - \beta_3\beta_4 = 0$). The joint significance test has been recommended to test indirect effects in single-mediator and two-mediator sequential chains (Falk & Biesanz, 2015; Taylor, MacKinnon, & Tein, 2008; Yzerbyt, Muller, Batailler, & Judd, 2018).

Simulation Design and Population Values

In the simulation studies, we manipulated three factors: (a) effect size R^2, which quantifies how well the predictors account for the variance in the endogenous variables; (b) sample size; and (c) the method of testing the indirect effect(s). Previous work showed that sample size and the effect size R^2 influence the Type I error and the power of the existing tests of the indirect effect for single-mediator and sequential two-mediator chains (MacKinnon et al., 2002, 2004; Williams & MacKinnon, 2008). Based on Cohen's (1988) guidelines, we specified the population effect sizes in the simulation to be $R^2 = .02$, .13, and .26, which leads to the corresponding nonzero population regression coefficients of 0.14, 0.39, and 0.59, respectively (see the method described by Thoemmes, MacKinnon, & Reiser, 2010, which we used).

Sample size took on the following values: $N = 50$, 75, 100, and 200. We chose these values to cover a range of sample sizes reported in the applied literature of psychology, in related disciplines, and in the simulation studies of mediation effects (e.g., MacKinnon et al., 2002; Tofighi & Kelley, 2019). The smallest sample size we considered, 50, might not be viewed as a best practice in the SEM literature, and a sample size of 200 is roughly equal to the median sample size in an SEM (Jaccard & Wan, 1995; MacCallum & Austin, 2000). However, the

studies with a sample size of 50 do appear in the applied literature (Tofighi & Kelley, 2019). We do not report results from a sample size greater than 200 because our preliminary simulation study for the LRT_{MBCO} as well as the previous simulation studies for the existing methods (e.g., MacKinnon et al., 2002, 2004; Taylor et al., 2008; Williams & MacKinnon, 2008) found that the performance of the tests did not differ between methods at larger sample sizes. In these large samples, all the methods (which are based on large sample theory) worked effectively.

The method conditions included six commonly used tests of indirect effects for both the Type I error and power study. These conditions resulted in 168 conditions for the Type I error and 216 conditions for the power study for the single-mediator model; and 240 conditions for the Type I error and 648 conditions for the power study for the sequential two-mediator model. We used a mixed full factorial design for each simulation study where the factors effect size and sample size were between-subjects factors and the method was a within-subjects factor.

Data Generation

We generated 5,000 independent data sets for each combination of the between-subjects factors using a known population model, either $X \xrightarrow{\beta_1} M \xrightarrow{\beta_2} Y$ for the single-mediator model or $X \xrightarrow{\beta_1} M_1 \xrightarrow{\beta_2} M_2 \xrightarrow{\beta_3} Y$ for the sequential two-mediator model. The variables X, M_1, M_2, and Y were observed. Without loss of generality, the population values for the intercepts were fixed at zero but were estimated in the simulation study. Values for X were sampled from a binomial distribution with .5 probability of two categories (i.e., treatment vs. control). Data for each residual term were generated from the standard normal distribution.

We chose $\alpha = .05$ to test the indirect effects, as is commonly done in psychology and related areas. We used OpenMx Version 2.9.6 (Boker et al., 2011; Neale et al., 2016) to conduct the LRT_{MBCO} as well as the profile-likelihood CI.[4] For the Monte Carlo, percentile and BC bootstrap CI, and for the joint significance test, we estimated the model using lavaan package Version 0.5–23 (Rosseel, 2012). We then used the model estimates and ci() function in the RMediation package Version 1.1.4 (Tofighi & MacKinnon, 2011, 2016) to compute a Monte Carlo CI; both

---

[4] We studied the software manual and references for the optimization techniques implemented in OpenMx (Boker et al., 2011; Neale et al., 2016; Zahery et al., 2017), communicated with the authors/programmers, and conducted an extensive simulation study to examine whether the LRT_{MBCO} can be implemented using different optimization methods in OpenMx. In addition, we examined whether the software packages lavaan Version 0.5-23.1097, Mplus Version 8.0, and PROC CALIS (SAS Institute Inc., 2016) allowed nonlinear constraints to estimate the LRT_{MBCO}. We also communicated with the authors/programmers of Mplus and lavaan. Based on our communications, at the time of writing, Mplus and lavaan do not guarantee obtaining optimal results for the mediation models with nonlinear constraints in the form of null hypothesis about the indirect effect. We chose the NPSOL (Gill, Murray, Saunders, & Wright, 1986) optimizer instead of the default CSOLNP (Zahery et al., 2017) optimizer in OpenMx because the CSOLNP showed convergence problems in our preliminary simulation studies. Based on our personal communication with OpenMx developers, they suggested that the SLSQP (Snyman, 2005) optimizer may be successful as well. However, we did not formally evaluate this option because it is beyond the scope of the current article.

percentile and BC bootstrap CI were estimated using the facilities provided in the lavaan package. For the joint significance test, we used the z test to examine the joint significance of the coefficients of the indirect effects.[5]

Results

For the simulation studies assessing the robustness of the Type I error, the empirical Type I error rate is the proportion of times out of 5,000 replications that a test incorrectly rejects the null hypothesis of zero indirect effect. For the statistical power studies, the empirical power is the proportion of times out of 5,000 replications that a test correctly rejects the null hypothesis of zero indirect effect. The large sample size of 5,000 within each condition precluded the use of repeated measures ANOVA for the simulation results. To determine if a test was robust, as discussed, we used Bradley's (1978) liberal interval of .025 and .075 to determine a Type I error rate that was "good enough" for practical purposes. We considered a test robust (i.e., good enough) according to Bradley's (1978) criteria if its empirical Type I error rate fell within the robustness interval. Otherwise, a test was considered conservative or liberal depending on whether the Type I error rate was smaller or greater than Bradley's lower or upper bound, respectively.

Type I error. To facilitate comparison of the Type I error rates across different combinations of the sample sizes and effect sizes, we created graphs showing the empirical Type I error rate as a function of sample size and size of nonzero coefficients. To save space, we only present a few graphs (Figures 3–6). More graphs can be found in the online supplemental materials. For the two-path indirect effect where both coefficients were zero (see Figure 3), the empirical Type I error rate for the LRT$_{MBCO}$ was robust as it fell within Bradley's (1978) robustness interval. The other five methods were not robust across all sample sizes; that is, their empirical Type I error rates were consistently below .01. The LRT$_{MBCO}$ produced more robust Type I error rates because it estimated the null mediation model and the sampling distribution of the indirect effect under the null hypothesis.

As can be seen in Figure 4, when one of the coefficients was zero, the LRT$_{MBCO}$ was the most robust method across all combinations of the nonzero coefficients and sample sizes; its Type I error rate remained close to the nominal .05 value and stayed within the limits of Bradley's (1978) robustness interval. The BC bootstrap showed conservative Type I error rates or inflated Type I error rates in certain conditions. For example, when $\beta_2 = 0.59$ and $N = 50$, and $\beta_2 = 0.39$ and $N = 100$, the BC bootstrap showed an inflated Type I error rate; when $\beta_2 = 0.14$ and $N = 75$, the BC bootstrap showed a conservative Type I error rate. In general, the Type I error rates for the percentile bootstrap, Monte Carlo, profile-likelihood, and joint significance methods were all robust except when one of the nonzero coefficients was small (i.e., 0.14). More specifically, when the magnitude of the coefficient was small and the sample size was ≤100, all four methods showed a conservative Type I error rate. The results for CI-based and joint significance tests, when one of the coefficients was zero, essentially matched the results from previous studies (Biesanz et al., 2010; Falk & Biesanz, 2015), as would be expected. Finally, which coefficient was zero did not appear to change the overall conclusion about the Type I error rate robustness of the methods.

For the sequential two-mediator model, the LRT$_{MBCO}$ was the most robust test across the combinations of the sample sizes and effect sizes. When all three coefficients were zero (see Figure 5) or when two of the coefficients were zero (see Figure 6), the LRT$_{MBCO}$'s Type I error rate remained robust within the Bradley's (1978) limits. The five other methods showed conservative Type I error rates. Finally, which coefficient was zero did not appear to change the Type I error rates.

Power. The simulation study showed that the LRT$_{MBCO}$ was as powerful as the other methods, except for the BC bootstrap method, for both two-path and three-path indirect effects. The BC bootstrap method showed slightly more power although that power was at the expense of yielding inflated Type I error rates. This result is consistent with previous research (Biesanz et al., 2010; Falk & Biesanz, 2015; Taylor et al., 2008). The results of power simulation studies were similar between the single-mediator model and the sequential two-mediator model. We explain the similar power between the CI-based methods and the LRT$_{MBCO}$ in the next section.

Summary

One explanation for why the simulation studies showed that the LRT$_{MBCO}$ produces more robust empirical Type I error rates than other CI-based methods is that the sampling distribution of the indirect effect tends to differ from the null sampling distribution of the indirect effect estimate. In the null sampling distribution, the population value of the indirect effect is fixed to zero. The sampling distribution used by the CI-based methods is likely to reflect an alternative hypothesis of a nonzero population indirect effect rather than the null hypothesis of zero population indirect effect. As a result, the CI-based methods tend to produce nonrobust empirical Type I error rates, especially for smaller sample sizes. However, the LRT$_{MBCO}$ uses both the null and the full models to test an indirect effect and, thus, more appropriately maps the statistical method onto the question of interest when seeking to determine if a hypothesized mediator mediates the relationship between X and Y. Hence, the LRT$_{MBCO}$ appears more appropriate than other methods that have been recommended in the literature in the conditions we studied.

The simulation results also showed that the LRT$_{MBCO}$ is as powerful as the other recommended tests except for the BC bootstrap method; however, we do not recommend the BC bootstrap because of its inflated Type I error rate. On the other hand, the CI-based tests have adequate power when the null hypothesis is false (i.e., nonzero indirect effect) and the sampling distribution is computed under an estimated alternative hypothesis in which the indirect effect is nonzero. In other words, the estimated alternative sampling distribution of CI-based methods is more accurate when the null hypothesis is false as opposed to when the null hypothesis is true. In these cases, the CI-based tests and LRT$_{MBCO}$ appear to exhibit similar power.

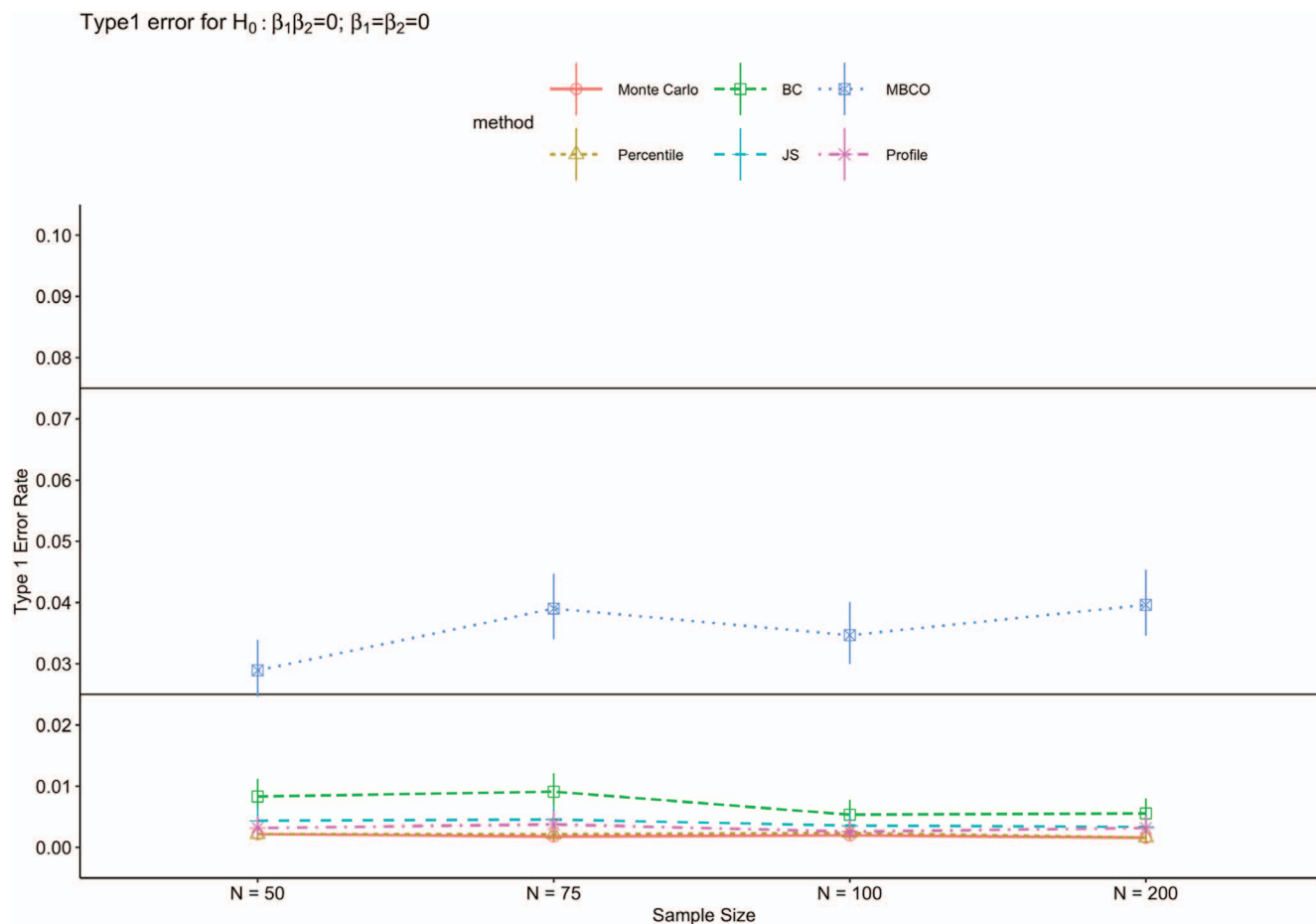[5] The simulation study code scripts can be found at https://github.com/quantPsych/mbco.

Figure 3. Point and 95% CI estimate of the Type I error rate for six methods of testing a two-path indirect effect, $\beta_1\beta_2$, where both parameters were fixed at zero. Horizontal solid lines show the limits of the Bradley's (1978) liberal interval of .025 and .075 for $\alpha = .05$. A test is robust according to a Bradley's (1978) criterion if its Type I error rate falls within the criterion's interval. MBCO = model-based constrained optimization (LRT_{MBCO}); Percentile = percentile bootstrap; BC = bias-corrected bootstrap; JS = joint significance test; Profile = profile-likelihood CI. See the online article for the color version of this figure.

Limitations

In our simulations, we made a few common assumptions for estimating a mediation model. The data for the simulation studies were generated assuming that residuals associated with the mediators and outcome variable had a multivariate normal distribution and that the correct model was fit (e.g., the variables were not nonlinearly related and there were no omitted confounders). Though assessing the effectiveness of the LRT_{MBCO} under the multivariate normality of the simulated data is a limitation, using the multivariate normal data is necessary to compare performance of the LRT_{MBCO} to other commonly used methods of testing mediation in known conditions. Therefore, the results of the simulation studies apply to situations in which these assumptions are reasonably met. Assessing the performance of the LRT_{MBCO} for non-normally distributed residuals remains a topic for future research. We also assumed the simulation study included correct functional forms of the relationships between the posited variables.

For the empirical study, we evaluated if there were outliers (and we did not find any in the sample data), and we assumed that no

common method biasing effect of measuring the endogenous variables existed (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Moreover, we assumed that the observed mediator and outcome variables were measured without error (Dwyer, 1983; Fritz, Kenny, & MacKinnon, 2016); otherwise, the results could be misleading (Cole & Preacher, 2014). In addition, one of the more stringent of these assumptions, which is untestable, is the no-omitted-confounder assumption (Imai et al., 2010; Judd & Kenny, 1981; Valeri & VanderWeele, 2013; VanderWeele, 2010). That is, no variable should be omitted from the model that would affect both the mediator and the outcome variables, given the independent variable (X) and covariates (if they exist and are included in the model). Even when participants are randomly assigned to a treatment or control group, making causal claims about an indirect effect requires the no-omitted-confounder assumption. If researchers believe that not all confounders are included in the model, then the claims about the magnitude and existence of an indirect effect need to be relaxed. We recommend that researchers conduct a sensitivity analysis to investigate the biasing impact of a potential
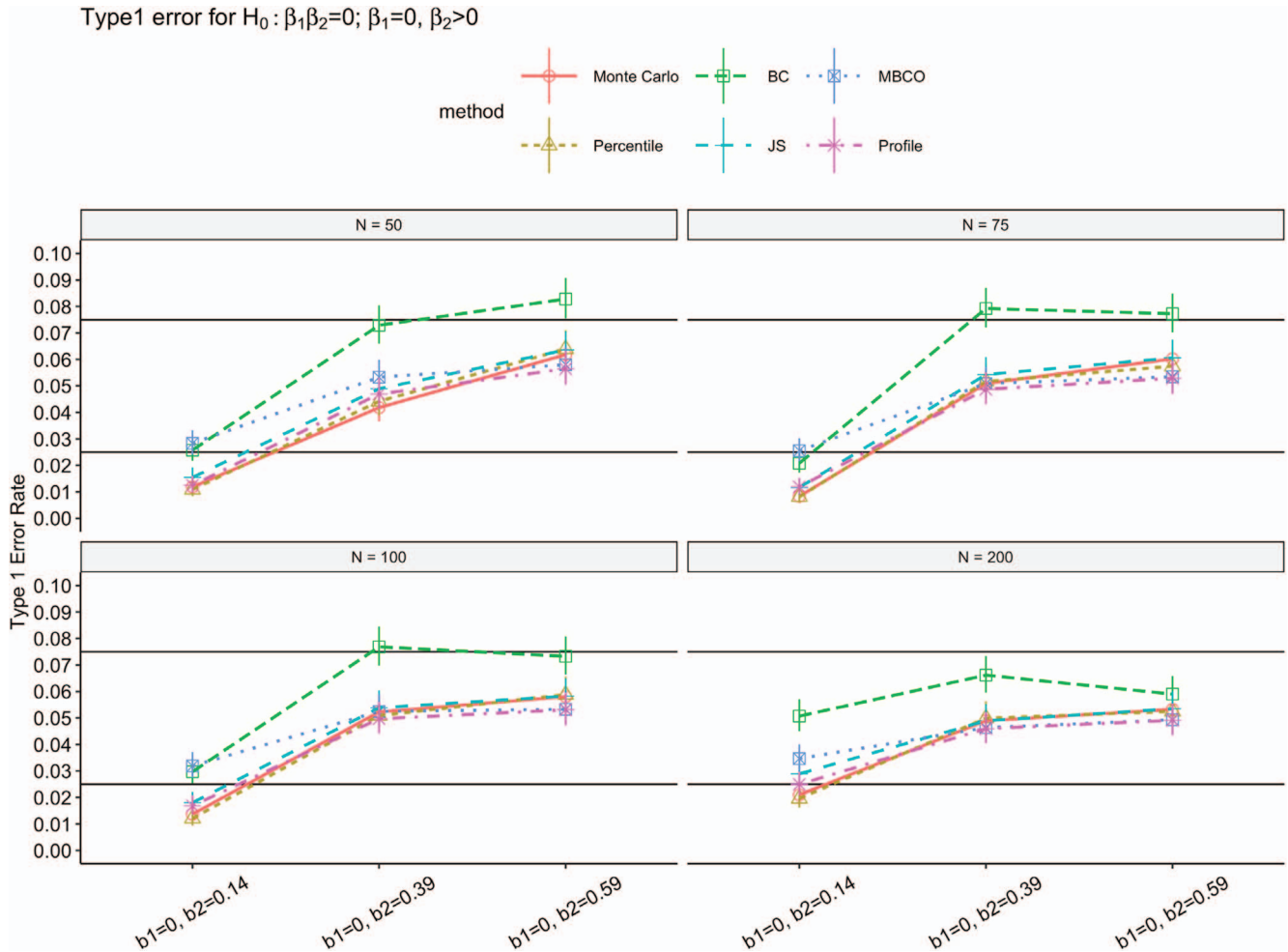
Type1 error for $H_0 : \beta_1\beta_2=0$; $\beta_1=0$, $\beta_2>0$

Figure 4.   Point and 95% CI estimate of the Type I error rate for six methods of testing a two-path indirect effect, $\beta_1\beta_2$, where only β_1 was fixed at zero. The nonzero parameters take on the values: 0.14, 0.39, and 0.59. The *x*-axis shows the parameter values. Horizontal solid lines show the limits of the Bradley's (1978) liberal interval of .025 and .075 for $\alpha = .05$. A test is robust according to a Bradley's (1978) criterion if its Type I error rate falls within the criterion's interval. MBCO = model-based constrained optimization (LRT_{MBCO}); Percentile = percentile bootstrap; BC = bias-corrected bootstrap; JS = joint significance test; Profile = profile-likelihood CI. See the online article for the color version of this figure.

confounder on the indirect effect point and interval estimate (Cox, Kisbu-Sakarya, Miočević, & MacKinnon, 2013; Tofighi et al., 2019; Tofighi & Kelley, 2016). MacKinnon and Pirlott (2015) provide an excellent discussion of different approaches to sensitivity analyses in mediation analysis. Sensitivity analyses for certain mediation models can also be conducted in an SEM framework in Mplus software, which may include observed and latent variables (Muthén & Asparouhov, 2015; Muthén & Muthén, 2018). Conducting sensitivity analyses using the LRT_{MBCO} also remains a topic for future study.

One requirement of our proposed LRT_{MBCO} and the profile-likelihood method is that the mediation model be estimated within the SEM (or other appropriate multivariate) framework. That is, the equations for the mediation model must be simultaneously estimated. These methods do not work when using OLS regression to separately estimate the equations representing a mediation

model, which is the classical way of testing mediation (Baron & Kenny, 1986).

We discussed applying the LRT_{MBCO} to the models when the independent variable and mediator do not interact and when the mediator and outcome are continuous. For such models, both the SEM and causal inference methods provide the same estimate of the indirect effect (Muthén & Asparouhov, 2015); thus, the LRT_{MBCO} can be used to test an indirect effect. However, when the independent variable and a mediator interact or when either a mediator or the outcome variable is a categorical variable, researchers need to use the potential outcome framework to correctly estimate the indirect effect (VanderWeele, 2015). Extension of the MBCO procedure and LRT_{MBCO} to the causal inference framework for these scenarios remains a topic for future study. Finally, we did not discuss application of the MBCO procedure to multilevel mediation
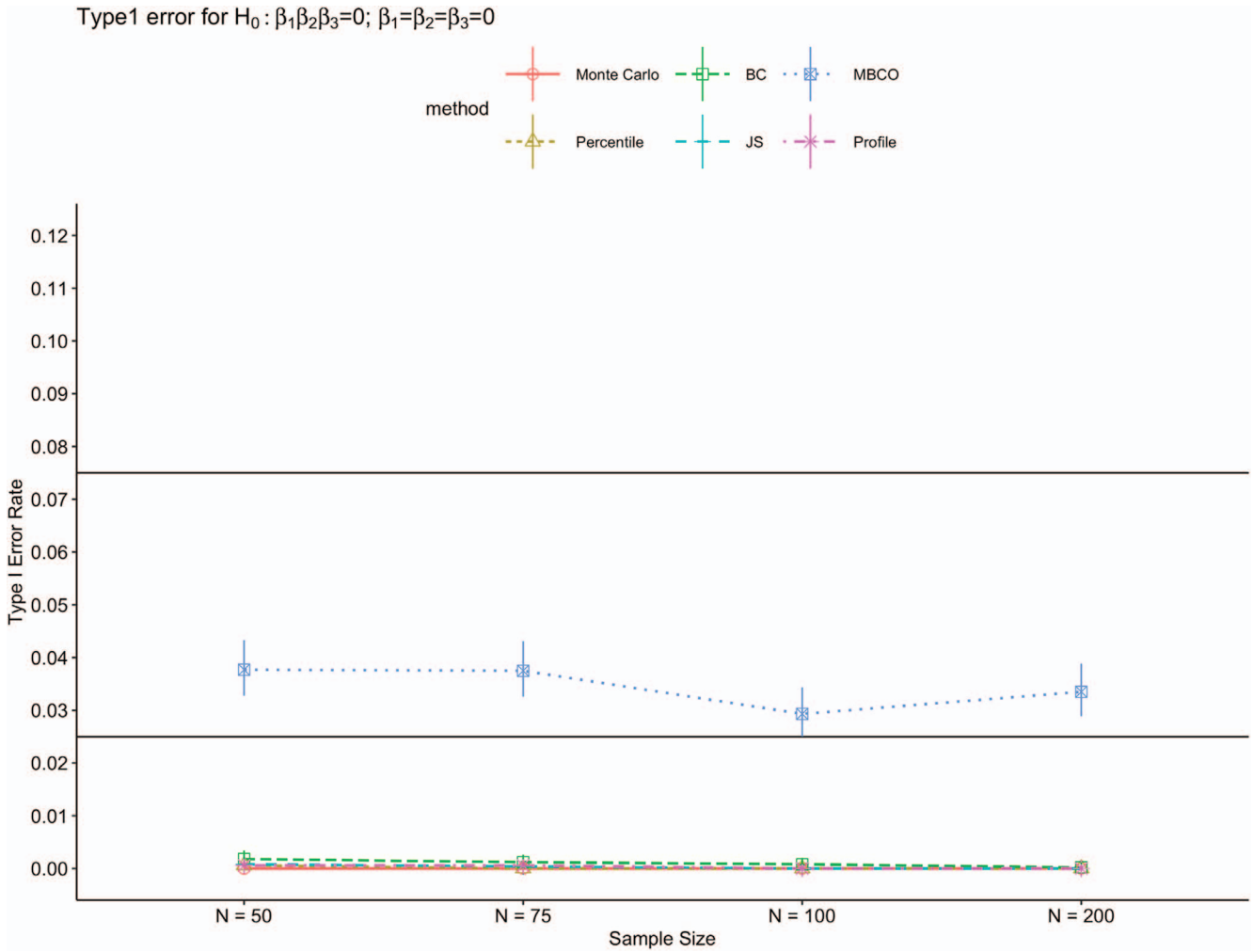
Figure 5. Point and 95% CI estimates of the Type I error rate for six methods of testing a three-path indirect effect, $\beta_1\beta_2\beta_3$, where all three parameters were zero. Horizontal solid lines show the limits of the Bradley's (1978) liberal interval of .025 and .075 for $\alpha = .05$. A test is robust according to a Bradley's criterion if its Type I error rate falls within the criterion's interval. MBCO = model-based constrained optimization (LRT$_{MBCO}$); Percentile = percentile bootstrap; BC = bias-corrected bootstrap; JS = joint significance test; Profile = profile-likelihood CI. See the online article for the color version of this figure.

analysis (Krull & MacKinnon, 2001; Tofighi, West, & MacKinnon, 2013), where the data have a multilevel (i.e., clustered, repeated measures) structure (Snijders & Bosker, 2012). Extending the MBCO procedure to multilevel data also remains a topic for future study.

Discussion

This article proposed an MBCO procedure that uses a model-comparison approach to make inferences about any function representing an indirect effect in mediation analysis. An innovation of our proposed MBCO procedure is using a nonlinear constraint to test a variety of simple and complex hypotheses about indirect effects in a model-comparison framework. The MBCO procedure offers the following advantages compared with the existing methods. First, the MBCO procedure produces a likelihood ratio test,

termed LRT$_{MBCO}$, to formally evaluate simple and complex hypotheses about indirect effects and produces a *p* value, a continuous measure of compatibility between data and the null hypotheses. Second, through the model-comparison framework, the MBCO procedure computes a likelihood ratio, a continuous measure of comparing goodness of fit of the null and full models. It also computes information fit indices such as AIC and BIC, used to compare the null and full models based on both goodness fit and parsimony.

To assess robustness of the LRT$_{MBCO}$ and the five most recommended methods of testing indirect effects in a single-mediator or in a sequential two-mediator model, we conducted a Monte Carlo simulation study. The results showed that the LRT$_{MBCO}$ is more robust in terms of the empirical Type I error rate (i.e., it was within Bradley's, 1978 liberal interval of .025 and .075 for $\alpha =$
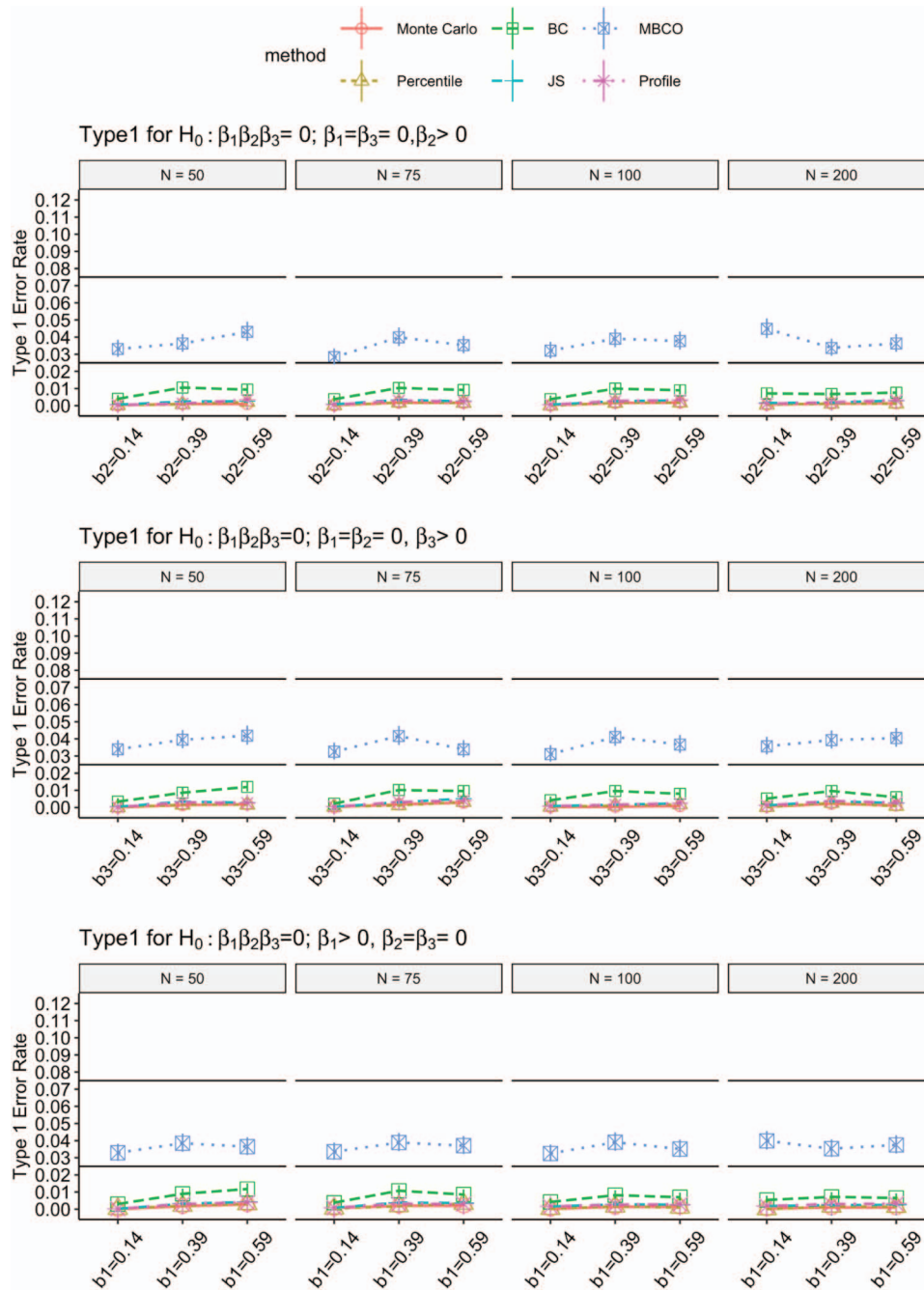
Figure 6. Point and 95% CI estimate of the Type I error rate for six methods of testing a three-path indirect effect, $\beta_1\beta_2\beta_3$, where two out of the three parameters were fixed at zero. The nonzero parameter took on the values: 0.14, 0.39, and 0.59. The *x*-axis shows the values of the nonzero parameter. Horizontal solid lines show the limits of the Bradley's (1978) liberal interval of .025 and .075 for $\alpha = .05$. A test is robust according to a Bradley's (1978) criterion if its Type I error rate falls within the criterion's interval. MBCO = model-based constrained optimization (LRT$_{MBCO}$); Percentile = percentile bootstrap; BC = bias-corrected bootstrap; JS = joint significance test; Profile = profile-likelihood CI. See the online article for the color version of this figure.

.05) than the other methods when the residuals were generated under an ideal condition of the multivariate normality. In addition, the LRT$_{MBCO}$ is as powerful as the currently used tests (except for the BC bootstrap) and offers more robust empirical Type I error rates in situations typical of psychology and related fields.

In addition, the MBCO procedure provides a model-comparison framework to compare one or more alternative mediation models. This important new feature relieves researchers of the restriction to test a null hypothesis with a single hypothesized model. Researchers can test multiple competing models in addition to fitting a

single hypothesized model and can focus on testing and comparing one or more indirect effects. More than a significance testing framework, a model-comparison framework allows researchers to compare one null model to one or more alternative models. These comparisons also can address the significance of indirect effects, the effect sizes associated with the mediators and outcome variables, and the overall fit and complexity of the models as measured by various fit indices such as the AIC (Akaike, 1974) and the BIC (Schwarz, 1978).

We applied the MBCO procedure to an empirical example. The accompanying computer code in OpenMx (Boker et al., 2011; Neale et al., 2016), the data set, and detailed analysis results are in the online supplemental materials. As the empirical example shows, after conducting the MBCO procedure, researchers should report the resulting LRT_{MBCO}, exact p value as well as the CIs for the indirect effects calculated using the Monte Carlo, profile-likelihood, or percentile bootstrap method. A p value should be interpreted as a measure of compatibility of a null model with the sample data and should not be used to make dichotomous decisions about the significance of a null hypothesis. Further, researchers should report R^2 effect sizes associated with mediators and the outcome variable and compute differences in the respective R^2s between the competing models. Computing differences in R^2s allows researchers to gauge potential changes in the effect sizes between the competing models, changes that could be a result of nonzero indirect effects. In addition, researchers should report the information fit indices AIC or BIC. Although using CI-based methods to test the existence of an indirect effect or to make dichotomous decisions about significance of the indirect effect is not recommended, researchers should report a CI to convey a range of plausible values for an indirect effect estimate. These recommendations are consistent with and enhance the APA recommendations (APA Publication Manual, 2010; Appelbaum et al., 2018; Wilkinson & Task Force on Statistical Inference, American Psychological Association, Science Directorate, 1999) for reporting statistical analysis results.

In conclusion, we believe that our proposed MBCO procedure provides multiple ways to evaluate hypotheses about mediation effects beyond the methods widely recommended in the literature (e.g., CI-based approaches). We believe that work using the MBCO procedure will advance the rich literature on testing and interpreting mediation models.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723. http://dx.doi.org/10.1109/TAC.1974.1100705

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician, 73,* 262–270. http://dx.doi.org/10.1080/00031305.2018.1543137

APA Publication Manual. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73,* 3–25. http://dx.doi.org/10.1037/amp0000191

Ato García, M., Vallejo Seco, G., & Ato Lozano, E. (2014). Classical and causal inference approaches to statistical mediation analysis. *Psicothema, 26,* 252–259.

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,* 1173–1182. http://dx.doi.org/10.1037/0022-3514.51.6.1173

Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research, 45,* 661–701. http://dx.doi.org/10.1080/00273171.2010.498292

Blume, J. D. (2002). Likelihood methods for measuring statistical evidence. *Statistics in Medicine, 21,* 2563–2599. http://dx.doi.org/10.1002/sim.1216

Boies, K., Fiset, J., & Gill, H. (2015). Communication and trust are key: Unlocking the relationship between leadership and team performance and creativity. *The Leadership Quarterly, 26,* 1080–1094. http://dx.doi.org/10.1016/j.leaqua.2015.07.007

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76,* 306–317. http://dx.doi.org/10.1007/s11336-010-9200-6

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley. http://dx.doi.org/10.1002/9781118619179

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20,* 115–140. http://dx.doi.org/10.2307/271084

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical & Statistical Psychology, 31,* 144–152. http://dx.doi.org/10.1111/j.2044-8317.1978.tb00581.x

Bulls, H. W., Lynch, M. K., Petrov, M. E., Gossett, E. W., Owens, M. A., Terry, S. C., . . . Goodin, B. R. (2017). Depressive symptoms and sleep efficiency sequentially mediate racial differences in temporal summation of mechanical pain. *Annals of Behavioral Medicine, 51,* 673–682. http://dx.doi.org/10.1007/s12160-017-9889-x

Carmeli, A., McKay, A. S., & Kaufman, J. C. (2014). Emotional intelligence and creativity: The mediating role of generosity and vigor. *The Journal of Creative Behavior, 48,* 290–309. http://dx.doi.org/10.1002/jocb.53

Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling, 14,* 227–246. http://dx.doi.org/10.1080/10705510709336745

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003. http://dx.doi.org/10.1037/0003-066X.49.12.997

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19,* 300–315. http://dx.doi.org/10.1037/a0033805

Cox, D. R., & Hinkley, D. V. (2000). *Theoretical statistics.* Boca Raton, FL: Chapman & Hall/CRC.

Cox, M. G., Kisbu-Sakarya, Y., Miočević, M., & MacKinnon, D. P. (2013). Sensitivity plots for confounder bias in the single mediator model. *Evaluation Review, 37,* 405–431. http://dx.doi.org/10.1177/0193841X14524576

Deković, M., Asscher, J. J., Manders, W. A., Prins, P. J. M., & van der Laan, P. (2012). Within-intervention change: Mediators of intervention effects during multisystemic therapy. *Journal of Consulting and Clinical Psychology, 80,* 574–587. http://dx.doi.org/10.1037/a0028482

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. New York, NY: Oxford University Press.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82,* 171–185. http://dx.doi.org/10.1080/01621459.1987.10478410

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.

Ernsting, A., Knoll, N., Schneider, M., & Schwarzer, R. (2015). The enabling effect of social support on vaccination uptake via self-efficacy and planning. *Psychology Health and Medicine, 20,* 239–246. http://dx.doi.org/10.1080/13548506.2014.920957

Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal, 22,* 24–38. http://dx.doi.org/10.1080/10705511.2014.935266

Folmer, H. (1981). Measurement of the effects of regional policy instruments by means of linear structural equation models and panel data. *Environment & Planning A, 13,* 1435–1448. http://dx.doi.org/10.1068/a131435

Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). Los Angeles, CA: SAGE.

Fritz, M. S., Kenny, D. A., & MacKinnon, D. P. (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behavioral Research, 51,* 681–697. http://dx.doi.org/10.1080/00273171.2016.1224154

Gill, P. E., Murray, W., Saunders, M. A., & Wright, M. H. (1986). *Fortran package for nonlinear programming. User's guide for NPSOL* (Version 4. 0). Stanford, CA: Stanford University. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a169115.pdf

Graça, J., Calheiros, M. M., & Oliveira, A. (2016). Situating moral disengagement: Motivated reasoning in meat consumption and substitution. *Personality and Individual Differences, 90,* 353–364. http://dx.doi.org/10.1016/j.paid.2015.11.042

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, *P* values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31,* 337–350. http://dx.doi.org/10.1007/s10654-016-0149-3

Harlow, L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Haslam, C., Cruwys, T., Milne, M., Kan, C.-H., & Haslam, S. A. (2016). Group ties protect cognitive health by promoting social identification and social support. *Journal of Aging and Health, 28,* 244–266. http://dx.doi.org/10.1177/0898264315589578

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15,* 309–334. http://dx.doi.org/10.1037/a0020761

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin, 117,* 348–357. http://dx.doi.org/10.1037/0033-2909.117.2.348

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602–619. http://dx.doi.org/10.1177/0193841X8100500502

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). Boston, MA: McGraw-Hill.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Koning, I. M., Maric, M., MacKinnon, D., & Vollebergh, W. A. M. (2015). Effects of a combined parent-student alcohol prevention program on intermediate factors and adolescents' drinking behavior: A sequential

mediation model. *Journal of Consulting and Clinical Psychology, 83,* 719–727. http://dx.doi.org/10.1037/a0039197

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,* 194–202. http://dx.doi.org/10.1037/a0036635

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36,* 249–277. http://dx.doi.org/10.1207/S15327906MBR3602_06

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226. http://dx.doi.org/10.1146/annurev.psych.51.1.201

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Erlbaum.

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,* 83–104. http://dx.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,* 99–128. http://dx.doi.org/10.1207/s15327906mbr3901_4

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review, 19,* 30–43. http://dx.doi.org/10.1177/1088868314542878

MacKinnon, D. P., Valente, M. J., & Wurpts, I. C. (2018). Benchmark validation of statistical models: Application to mediation analysis of imagery and memory. *Psychological Methods, 23,* 654–671. http://dx.doi.org/10.1037/met0000174

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2018). *Designing experiments and analyzing data: A model comparison perspective* (3rd ed.). New York, NY: Routledge.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical & Statistical Psychology, 37,* 234–251. http://dx.doi.org/10.1111/j.2044-8317.1984.tb00802.x

Molina, B. S. G., Walther, C. A. P., Cheong, J., Pedersen, S. L., Gnagy, E. M., & Pelham, W. E. J. (2014). Heavy alcohol use in early adulthood as a function of childhood ADHD: Developmentally specific mediation by social impairment and delinquency. *Experimental and Clinical Psychopharmacology, 22,* 110–121. http://dx.doi.org/10.1037/a0035656

Muthén, B. O., & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling, 22,* 12–23. http://dx.doi.org/10.1080/10705511.2014.935843

Muthén, L. K., & Muthén, B. O. (2018). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., . . . Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika, 81,* 535–549. http://dx.doi.org/10.1007/s11336-014-9435-8

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27,* 113–120. http://dx.doi.org/10.1023/A:1025681223921

Oliver, M. C., Bays, R. B., & Zabrucky, K. M. (2016). False memories and the DRM paradigm: Effects of imagery, list, and test type. *The Journal of General Psychology, 143,* 33–48. http://dx.doi.org/10.1080/00221309.2015.1110558

Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood*. New York, NY: Oxford University Press.

Pek, J., & Wu, H. (2015). Profile likelihood-based confidence intervals and regions for structural equation models. *Psychometrika, 80,* 1123–1145. http://dx.doi.org/10.1007/s11336-015-9461-1

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879–903. http://dx.doi.org/10.1037/0021-9010.88.5.879

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,* 879–891. http://dx.doi.org/10.3758/BRM.40.3.879

R Core Team. (2019). R: A Language and Environment for Statistical Computing (Version 3.6.0) [Computer software]. Retrieved from http://www.R-project.org/

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

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

SAS Institute Inc. (2016). *SAS/STAT 14.2 user's guide.* Cary, NC: Author. Retrieved from http://support.sas.com/documentation/cdl/en/stathpug/67524/PDF/default/stathpug.pdf

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. http://dx.doi.org/10.1214/aos/1176344136

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and non-experimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445. http://dx.doi.org/10.1037/1082-989X.7.4.422

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: SAGE.

Snyman, J. A. (2005). *Practical mathematical optimization: An introduction to basic optimization theory and classical and new gradient-based algorithms.* New York, NY: Springer.

Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods, 11,* 241–269. http://dx.doi.org/10.1177/1094428107300344

Thoemmes, F., Mackinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling, 17,* 510–534. http://dx.doi.org/10.1080/10705511.2010.489379

Tofighi, D., Hsiao, Y.-Y., Kruger, E. S., MacKinnon, D. P., Van Horn, M. L., & Witkiewitz, K. A. (2019). Sensitivity analysis of the no-omitted confounder assumption in latent growth curve mediation models. *Structural Equation Modeling, 26,* 94–109. http://dx.doi.org/10.1080/10705511.2018.1506925

Tofighi, D., & Kelley, K. (2016). Assessing omitted confounder bias in multilevel mediation models. *Multivariate Behavioral Research, 51,* 86–105. http://dx.doi.org/10.1080/00273171.2015.1105736

Tofighi, D., & Kelley, K. (2019). Indirect effects in sequential mediation models: Evaluating methods for hypothesis testing and confidence interval formation. *Multivariate Behavioral Research.* Advance online publication. http://dx.doi.org/10.1080/00273171.2019.1618545

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43,* 692–700. http://dx.doi.org/10.3758/s13428-011-0076-x

Tofighi, D., & MacKinnon, D. P. (2016). Monte Carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling, 23,* 194–205. http://dx.doi.org/10.1080/10705511.2015.1057284

Tofighi, D., West, S. G., & MacKinnon, D. P. (2013). Multilevel mediation analysis: The effects of omitted variables in the 1–1–1 model. *British Journal of Mathematical & Statistical Psychology, 66,* 290–307. http://dx.doi.org/10.1111/j.2044-8317.2012.02051.x

Valeri, L., & Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods, 18,* 137–150. http://dx.doi.org/10.1037/a0031034

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology, 21,* 540–551. http://dx.doi.org/10.1097/EDE.0b013e3181df191c

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction.* New York, NY: Oxford University Press.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54,* 426–482. http://dx.doi.org/10.1090/S0002-9947-1943-0012401-3

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "$p < 0.05$". *The American Statistician, 73,* 1–19. http://dx.doi.org/10.1080/00031305.2019.1583913

Weisstein, E. W. (2018). *Smooth function.* Retrieved from http://mathworld.wolfram.com/SmoothFunction.html

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604. http://dx.doi.org/10.1037/0003-066X.54.8.594

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9,* 60–62. http://dx.doi.org/10.1214/aoms/1177732360

Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling: A Multidisciplinary Journal, 15,* 23–51.

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,* 929–943. http://dx.doi.org/10.1037/pspa0000132

Zahery, M., Maes, H. H., & Neale, M. C. (2017). CSOLNP: Numerical optimization engine for solving non-linearly constrained problems. *Twin Research and Human Genetics, 20,* 290–297. http://dx.doi.org/10.1017/thg.2017.28