A Simulation Study Assessing Mediator to Outcome Confounding Bias in Mediation

Analysis

by

Diana Alvarez Bartolo

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved March 2022 by the
Graduate Supervisory Committee:

David P. MacKinnon, Chair
Kevin J. Grimm
Daniel McNeish

ARIZONA STATE UNIVERSITY

May 2022

ABSTRACT

This project studied a four-variable single mediator model, a single mediator model: X (independent variable) to M (mediator) to Y (dependent variable), and a confounder (U) that influences M and Y. Confounding represents a threat to the causal interpretation in mediation analysis. For instance, if X represents random assignment to control and treatment conditions, the effect of X on M and the effect of X on Y have a causal interpretation under certain reasonable assumptions. However, the randomization of X does not allow for a causal interpretation of the M to Y effect unless certain confounding assumptions are satisfied. The aim of this project was to develop a significance test and an effect size comparison for two sensitivity to confounding analyses methods: Left Out Variables Error (L.O.V.E.) and the correlated residuals method. Further, the project assessed the accuracy of the methods for identifying confounding bias by simulating data with and without confounding bias.

i

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The simplest cause-effect relation involves two variables (e.g., X and Y), and in terms of effects, there are four possible combinations: X and Y are unrelated, X causes Y ($X{\rightarrow}Y$), Y causes X ($Y{\rightarrow}X$), or X and Y cause each other ($Y{\leftrightarrow}X$) (MacKinnon, 2008). More complex combinations arise when a third variable is included. Some of these relations include mediators, confounders, and colliders.

A mediator represents the causal mechanism through which an independent variable influences a dependent variable (Baron & Kenny, 1986, p. 1173). In other words, a mediator (M) is an intermediate variable in the causal path from the independent variable (X) to the dependent variable (Y). Where X causes M, which in turn causes Y ($X{\rightarrow}M{\rightarrow}Y$) (MacKinnon, 2008).

A confounder has been defined in different ways; for instance, "… a variable that changes the relationship between an independent variable and dependent variable because it is related to both the independent and the dependent variable." (MacKinnon, 2008, p. 7), "a variable related to two factors of interest that falsely obscures or accentuates the relationship between the factors." (Meinert & Tonascia, 1986, p. 285). In a diagram, the arrows point from the confounder (U), to the independent variable (X) and to the dependent variable (Y) as follows: ($X{\leftarrow}U{\rightarrow}Y$).

In a single confounder case, statistical adjustment for the confounder provides an unbiased estimate of the relation between the independent and the dependent variable

1

(Hafeman, 2011; MacKinnon, Krull, & Lockwood, 2000). Also, for the single confounder case, the sign of the arrows can predict the direction of the bias. For instance, if both arrows have the same sign (both positive or negative), the confounding bias is called *positive*, and the estimate will be biased upwards without adjusting for U. However, if the arrows have opposite signs, the confounding bias is called *negative*, and the estimate will be biased downwards without adjustment for U (Hernán & Robins, 2020).

A collider variable is caused by two other variables (Elwert & Winship, 2014). In a diagram, the arrows are from X and Y to the collider variable ($X \rightarrow Collider \leftarrow Y$) rather than arrows from the third variable to X and to Y, for a confounder. Opposite to the confounder case, researchers should not control for a collider because conditioning on a collider induces a spurious association between the two variables (in this case, X and Y). Collider bias can also arise from restricting sampling based on the collider variable, known as endogenous selection bias (Elwert & Winship, 2014).

Mediators, confounders, and colliders involve three variables. The current project focuses on a model that combines mediation and confounding in a four-variable model. First, this document presents the single mediator model by including the path diagrams, equations, and assumptions. Second, it incorporates confounding within mediation analysis, focusing on how confounding represents a threat to causal interpretation. The focus is on the single mediator model, where a confounder U influences both M and Y. Third, the document includes the topic of sensitivity analyses, a group of methods that are used to help deal with unmeasured confounders. Then, it describes the proposed research questions,

the methods to answer these questions, and the results. Finally, it presents the discussion, limitations of the study, and future research directions.

Single mediator model

Mediation analysis is a method that helps to evaluate the underlying mechanisms in which an independent variable causes a dependent variable. The following three equations represent the single mediator model:

$$Y = i_1 + cX + e_1 \qquad (1)$$

$$M = i_2 + aX + e_2 \qquad (2)$$

$$Y = i_3 + c'X + bM + e_3 \qquad (3)$$

Figure 1 represents the single mediation model path diagrams, where X represents the independent variable, M represents the mediator, and Y represents the dependent variable. The coefficient c represents the effect of X on Y; a represents the direct effect of X on M; c' is the coefficient of X to Y adjusted for the effects of M; the coefficient b represents the effect of M on Y adjusted for the effects of X; e_1, e_2, and e_3 represent the error variances; and i_1, i_2, and i_3 represent the intercepts.

Figure 1

Path Diagrams of the Single Mediator Model
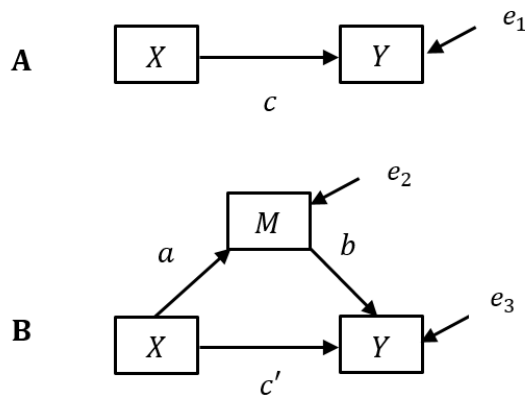
The conclusions from the single mediator model analysis rely on the following assumptions: a) each mediation regression equation assumes a linear relationship among variables, b) the regression equations represent the correct model; thus, it does not omit any important variable (e.g., confounders), c) the scores of X, M, and Y are valid and reliable, d) the residuals are assumed to be uncorrelated with the predictors, to be independent of each other, to have constant variance, and to be uncorrelated between equations, e) the single mediator model assumes correct temporal order, where X precedes M, which in turn precedes Y, f) it is assumed that the timing of measurement of the mediator and the outcome matches the right changing time in these variables (MacKinnon, 2008).

There are four additional no-confounding assumptions to enhance a causal interpretation of the direct and indirect effects in the single mediator model: (1) no unmeasured X–Y confounding, (2) no unmeasured M–Y confounding, (3) no unmeasured X–M confounding, and (4) no measured or unmeasured confounders of M and Y that have been affected by the exposure X (Pearl, 2001; Valeri & Vanderweele, 2013). Assumptions

4

(1) and (3) hold if X represents random assignment to control and treatment conditions. However, assumptions (2) and (4) are not guaranteed with random assignment of X.

Single mediator model and confounding bias

The model of interest in this study was a single mediator model with a confounder (U) influencing, M and Y, but not X (Figure 2). The model is widely used for mediation analysis of studies with randomization to conditions represented by X. The model is described by equations 4 and 5, where X represents the independent variable, M the mediator, Y the dependent variable, and U the confounder. The coefficient a represents the effect of X on M adjusted for the effects of U; d represents the effect of U on M adjusted for the effects of X; c' represents the effect of X on Y adjusted for the effects of M and U; the coefficient b represents the effect of M on Y adjusted for the effects of X and U; and e represents the effect of U on Y adjusted for the effects of M and X; r_1, and r_2 represent the error variances; i_1, i_2 represent the intercepts. The true covariance matrix for this model is included in Appendix A, and the analytical program for this model is included in Appendix B.

$$M = i_1 + aX + dU + r_1 \qquad (4)$$

$$Y = i_2 + c'X + bM + eU + r_2 \qquad (5)$$

Figure 2

Path Diagram: Single Mediator Model With a Confounder (U) That is Influencing the

Mediator (M) and the Outcome (Y)
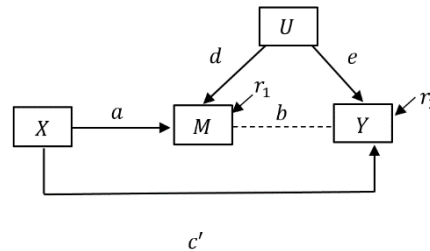
Several techniques help strengthen the causal interpretation of the mediation

analysis results, which are divided into design-based and analysis-based techniques. The

design-based techniques include randomization of X, randomization of M, and

randomization of both X and M (e.g., sequential double randomization design, concurrent

double randomization design, and parallel randomization). The analysis-based techniques

include inverse probability weighting, sequential G-estimation, the inclusion of

instrumental variables in the analysis, principal stratification, the usage of comprehensive

models, and sensitivity analyses (see MacKinnon & Pirlott, 2015; Pirlott & MacKinnon,

2016; Valente et al., 2017). This research focuses on sensitivity to confounding analyses to

strengthen the causal interpretation of the M to Y effect.

Sensitivity analyses is a group of methods that help deal with confounding bias in

mediation analysis when there is no measure of the potential confounders. In the following

section, I describe four sensitivity analyses methods for unmeasured confounders: the Left

Out Variables Error (L.O.V.E.) method (Cox, Kisbu-Sakarya, Miocevic, & MacKinnon,

2013; Mauro, 1990), the phantom variable method (Harring, McNeish, & Hancock, 2017;

MacKinnon, 2022; Rindskopf, 1984), the correlated residuals method (Imai, Keele, & Yamamoto, 2010), and the binary confounder method (Vanderweele, 2010).

Left Out Variables Error (L.O.V.E.; Cox et al., 2013; Mauro, 1990)

Cox et al. (2013) adapted Mauro's (1990) L.O.V.E. method for mediation analysis. This method assesses confounding in the M to Y relation in mediation analysis by using correlation methods, focusing on 1) the correlation between the omitted variable[1] and the mediator (r_{um}) and 2) the correlation between the omitted variable and the dependent variable (r_{uy}).

This method specifies the model equations when the confounder (U) is included in the single mediator model, with X as the treatment, M the mediator, and Y as the outcome (Equations 6 and 7).

$$M = a^*M + d^*U + r_1 \qquad (6)$$

$$Y = b^*M + c'^*X + e^*U + r_2 \qquad (7)$$

Based on these equations, the standardized regression coefficients (b^*, c'^*, and a^*) can be estimated using Equations 8, 9, 10 (Cox et al., 2013).

$$b^* = \frac{r_{ym}(1-r_{ux}^2)+r_{yx}\,(r_{um}r_{xu}-r_{mx})+\,r_{uy}(r_{mx}r_{ux}-\,r_{um})}{1+2r_{mx}r_{xu}r_{um}-r_{um}^2-r_{ux}^2-r_{mx}^2} \qquad (8)$$

$$c'^* = \frac{r_{yx}(1-r_{um}^2)+r_{ym}\,(r_{ux}r_{um}-\,r_{mx})+\,r_{uy}(r_{um}r_{ux}-\,r_{xm})}{1+2r_{mx}r_{um}r_{ux}-r_{ux}^2-r_{um}^2-r_{mx}^2} \qquad (9)$$

---

1 Throughout the document, the term "omitted variable" is used interchangeably with "unmeasured confounder".

$$a^* = \frac{(r_{mx} - r_{um} * r_{ux})}{1 - r_{ux}^2} \qquad (10)$$

If there is a correlation between the omitted variable U and the other predictors in the model (M and Y), the estimates b^* and c'^* will be biased if the regression model does not include U (Cox et al., 2013). The L.O.V.E. method assesses how large the correlation between the omitted variable and the mediator (r_{um}) and the correlation between the omitted variable and the dependent variable (r_{uy}) should be to make the standardized indirect effect zero (Cox et al., 2013). The indirect effect can be considered robust if the sensitivity analysis shows that implausibly large values of r_{um} and r_{uy} render the standardized indirect effect equal to zero (Valente et al., 2017).

Appendix C shows the code using the SAS programming language that conducts sensitivity analysis using the L.O.V.E. method as in Cox et al. (2013). The code generates a range of mathematically possible values for two correlations: r_{um} and r_{uy}. Also, consistent with a randomized X variable, the correlation between the treatment and the omitted variable is fixed to zero ($r_{xu} = 0$). The program requires the value of three correlations values as input: 1) the correlation between the dependent variable and the treatment (r_{yx}), 2) the correlation between the mediator and the treatment (r_{mx}), and 3) the correlation between the dependent variable and the mediator (r_{ym}).

The L.O.V.E. method compares the correctly specified model coefficients (which includes the confounder U) with the misspecified model coefficients (which ignores the

confounder U). The standardized regression coefficients for the misspecified model are estimated with Equations 11, 12, and 13:

$$b^* biased = \frac{r_{ym} - r_{yx} * r_{mx}}{1 - r_{mx}^2} \quad (11)$$

$$c'^* biased = \frac{r_{yx} - r_{ym} * r_{mx}}{1 - r_{mx}^2} \quad (12)$$

$$a^* biased = r_{mx} \quad (13)$$

The bias of the correctly specified and misspecified model can be calculated based on the regression coefficients:

$$biasb^* = b^* biased - b^* \quad (14)$$

$$biasc'^* = c'^* biased - c'^* \quad (15)$$

$$biasa^* = a^* biased - a^* \quad (16)$$

The method estimates the standardized indirect effect $(a^* * b^*)$ and creates a plot for all the values of r_{um} and r_{uy} that make the standardized indirect effect $(a^* * b^*)$ zero. Figure 3 shows a plot of the L.O.V.E method, where the values of r_{uy} are shown on x-axis, and the values of r_{um} are shown on the y-axis.

Figure 3

Plot of the L.O.V.E. Sensitivity to Confounding Method
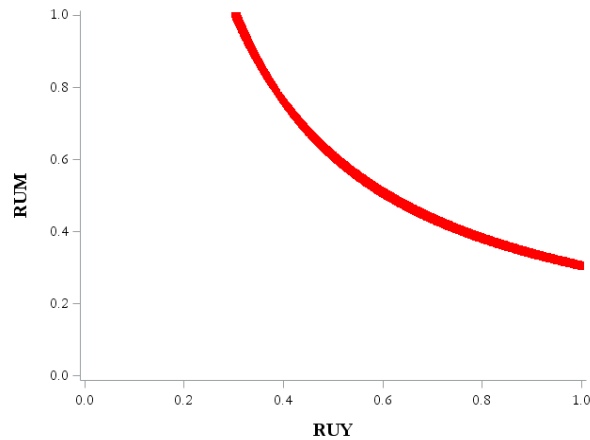
The phantom variable method

A phantom variable is defined as "a latent variable with no manifest indicators" (Harring et al., 2017, p. 619). Rindskopf (1984) first used phantom variables to constrain parameters in LISREL. Harring, et al. (2017) applied this method to assess a model sensitivity to potential external misspecifications using fixed and random parameter approaches.

The phantom variable method can be used as a sensitivity to confounding method in mediation analysis by specifying a phantom variable representing the hypothetical confounder. This method fixes the effect of the phantom variable to the mediator (*phantom variable* → *M*) and the effect of the phantom variable to the dependent variable (*phantom variable* → *Y*) to different values and then estimates the indirect effect. The goal is to find the value that makes the indirect effect zero. Note that if *X*

represents random assignment to conditions, the relation between the phantom variable and

X, would be fixed at zero because the variables are expected to be uncorrelated.

Appendix D shows the code in Mplus that performs a sensitivity analysis to

confounding using the phantom variable method, following MacKinnon (2022). The

program specifies the variance of the phantom variable, which is fixed to one

(phantom@1;). Also, since *X* represent random assignment to conditions, the effect

between the phantom variable and the independent variable *X* is fixed at zero (X ON

phantom@0;). The phantom variable is specified within the two mediation equations. The

effect of the phantom variable on the mediator (M ON X phantom@1.744;) and the effect

of the phantom variable on the outcome were fixed to 1.744 (Y ON M X

phantom@1.744;), which is the value that renders the indirect effect to zero in this dataset.

Appendix E also includes the code to perform the phantom variable method, but in the SAS

programming language using the PROC CALIS statement. Table 1 shows the estimated

coefficients for both programs.

Table 1

Phantom Variable Method Results: SAS and Mplus

		Estimates	Standard Error	t-value	p-value
SAS	$\hat{a}$	3.558	0.681	5.222	<.0001
phantom@1.744	$\widehat{c'}$	2.468	1.403	1.758	0.078
	$\hat{b}$	0.013	0.255	0.052	0.958
	$\hat{a}\hat{b}$	0.046			
Mplus	$\hat{a}$	3.558	0.632		0.000
phantom@1.744	$\widehat{c'}$	2.518	1.488		0.091
	$\hat{b}$	0.000	0.244		0.998
	$\hat{a}\hat{b}$	0			

In practice, a researcher can assess different values for the effect of the phantom variable and create a plot similar to the L.O.V.E. method plot (e.g., see program Confoundit by Lovis-McMahon & MacKinnon, 2014). For instance, Figure 4 shows a plot with the effects of the phantom variable on the mediator and the outcome on the x-axis and the indirect effect (ab) on the y-axis. The graph shows that values -1.76 and 1.76 of the phantom variable on M and Y make the indirect effect ~ zero (0.00850). These values differ from the value in the Mplus program, which was 1.744.

Figure 4

Plot of the Phantom Variable Sensitivity to Confounding Method
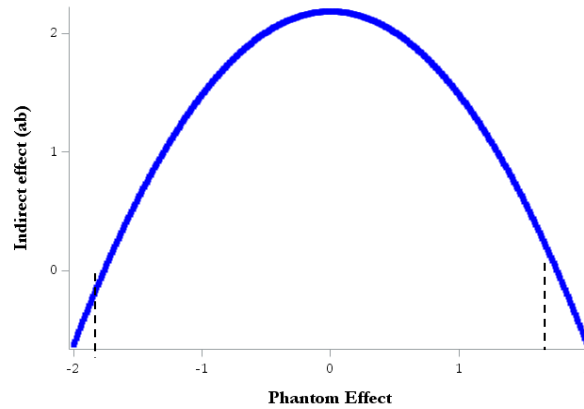
Note. The dashed lines indicate the values of phantom variable effect on M and Y needed to make the indirect effect zero.

Correlated residuals method (Imai et al., 2010)

Imai et al. (2010) developed a sensitivity analysis derived from the potential outcomes framework of mediation analysis, which evaluates the average causal indirect effect (indirect effect) for violations of the sequential ignorability assumption. This assumption involves the following two conditions. First, there are no unobserved confounders on the X to M relation or on the X to Y relation. Second, there are no unobserved (baseline or post-exposure) M to Y confounders. The first condition could be justified when the treatment is randomly assigned or when researchers have measured enough baseline covariates that differences between groups are ignorable (Imai et al., 2010). The correlated residuals method assumes randomization of X; hence it focuses on testing the latter condition.

13

This sensitivity analysis assesses confounding bias by analyzing how the indirect effect changes as a function of the correlation between the residual variances e_2 and e_3 (ρ) of the single mediator model equations (Equations 2 and 3). If the sequential ignorability assumption holds, the residuals (e_2, e_3) should be uncorrelated, hence $\rho = 0$. That means that there is no confounding on the M to Y relation. On the contrary, if the sequential ignorability assumption does not hold, the residuals (e_2, e_3) are correlated, hence $\rho \neq 0$, which could be interpreted as the existence of an omitted variable related to the mediator and to the dependent variable (Imai et al., 2010; James, 1980).

This sensitivity analysis is performed by plotting the average causal indirect effect as a function of the correlation between the residuals e_2 and e_3 (ρ). Imai et al. (2010) provide the following equation to estimate the average causal indirect effect given the parameter ρ.

$$\bar{\delta}(0) = \bar{\delta}(1) = \frac{(\beta_2 \sigma_1)}{\sigma_2} \left\{ \tilde{\rho} - \rho \sqrt{(1 - \tilde{\rho}^2/(1 - \rho^2))} \right\} \qquad (17)$$

Where $\sigma_j^2 \equiv var\ (\varepsilon_{ij})$ for $j = 1, 2$ and $\tilde{\rho} \equiv Corr\ (\varepsilon_{i2}, \varepsilon_{i3})$. If the sequential ignorability assumption is violated, the bounds are (-1 and 1) for the sensitivity parameter ρ are $(-\infty, \infty)$ for the indirect effect.

The package *mediation* in the R software package (Imai et al., 2010) performs this sensitivity analysis (included in Appendix F). The program estimates the causal mediation analysis and then runs the sensitivity analysis providing plots of the estimated average

causal indirect effect as a function of *ρ* (see Figure 5). This method can also be performed

in the Mplus programming language (Muthén & Muthén, 2012).

Figure 5

Plot of the Correlated Residuals Sensitivity to Confounding Method
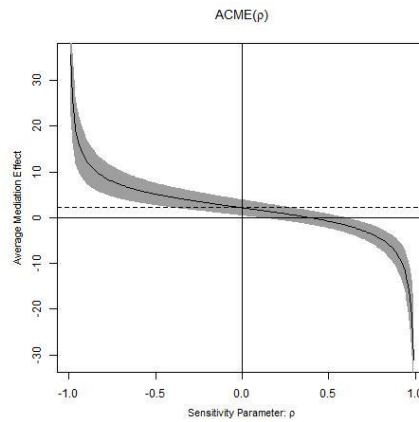
Binary confounder method (Vanderweele, 2010)

VanderWeele (2010) proposed a sensitivity analysis that focuses on binary

confounders. Like the correlated residuals method (Imai et al., 2010), this method is based

on the mediation analysis potential outcomes framework. The sensitivity analysis plot

displays the Natural Indirect Effect (NIE) as a function of two parameters (γ and δ). The

first parameter, γ, represents the effect of the binary confounder on Y, for individuals with

the same value of the mediator. The second parameter, δ, "represents the difference in the

prevalence of the confounder variable U, for people with the same value of M in both the

treatment and control groups" (Cox et al., 2013, p. 412).

15

VanderWeele (2010) specified the NIE's bias as a function of both parameters, as shown below:

$$Bias\ NIE\ (x) = -\delta\gamma \qquad (18)$$

There have been additional contributions on sensitivity analyses methods. For example, the development of sensitivity analyses that allows the assessment of the no-omitted-confounders assumption and the no-measurement error assumption jointly (Fritz, Kenny, & MacKinnon, 2016; Liu & Wang, 2020), a sensitivity analysis that allows assessing confounding bias for growth curve mediation models (Tofighi et al., 2019), and a sensitivity analysis that assesses confounding bias in a Bayesian mediation analysis (McCandless & Somers, 2019). Also, VanderWeele and Chiba (2014) developed a non-parametric sensitivity analysis applicable even if the M to Y confounder is affected by X. Albert and Wang (2014) developed two sensitivity analyses. The first allows latent versions of the mediator and the outcome, and second considers a hybrid causal-observational model, which considers two effects of X on Y: a causal effect and cohort effect (selection bias).

To sum up, sensitivity analyses is a group of methods that address confounding bias when there is no measure of the confounders. In the last section, multiple sensitivity analyses methods were described, but only two of them were selected for this project, the L.O.V.E. and correlated residuals methods. In particular, the interest was to select methods that help dealing with continuous unmeasured confounders. For instance, the sensitivity analysis proposed by VanderWeele (2010) is focused on binary confounders.

This project aims to develop a significance test and an effect size comparison for two sensitivity analyses methods: L.O.V.E. and correlated residuals (Cox et al., 2013; Imai et al., 2010; Mauro, 1990). Further, the project aims to assess the accuracy of these methods to identify confounding bias by simulating data with and without confounding bias.

CHAPTER 2

METHOD (STUDY 1)

Simulation Study

A Monte Carlo Simulation study was conducted in the SAS 9.4 programming language. The data were generated based on a single mediator model with a confounder (U) influencing the mediator (M) and the outcome (Y). In this model M, Y, and U were generated as continuous variables, and X was generated as a binary variable, representing random assignment to control and treatment groups.

In the simulation, sample size (N = 100, 200, 500), and parameter values (a, b, $c' =$ 0, 0.59; d and $e =$ -0.59, 0, 0.59) varied for a total of 3 X 2 X 2 X 2 X 3 X 3 equals 216 combinations. There were 1,000 replications for each condition. There were seven outcomes for the indirect effect (ab) in Study 1: Bias, Relative Bias, Standardized Bias, Empirical Type I Error Rates, Empirical Power, Confidence Intervals Coverage, and Imbalance. These outcomes were calculated for two scenarios: 1) when the data were analyzed with a single mediator model with a confounder (U) influencing the mediator (M) and the outcome (Y), and 2) when the data were analyzed with a misspecified model, where the confounder was omitted from the model. The simulation program is shown in Appendix G.

Bias of the indirect effect (a*b)

Three type of Biases of the indirect effect were calculated in the study: Bias, Relative bias, and Standardized bias. Bias was calculated as the difference between the estimate of the indirect effect minus the true indirect effect (Equation 19). Relative bias was

18

calculated by dividing bias (shown in Equation 19) by the true indirect effect (Equation 20).

Standardized bias was calculated as the difference between the indirect effect estimate

minus the true indirect effect, divided by the standard deviation of the indirect effect

estimate (Equation 21). Relative bias values between -0.10 and 0.10 were considered

acceptable.

$$Bias = \hat{\omega} - \omega \qquad (19)$$

$$Relative\ Bias = \frac{\hat{\omega} - \omega}{\omega} \qquad (20)$$

$$Standardized\ Bias = \frac{\hat{\omega} - \omega}{SD_{\hat{\omega}}} \qquad (21)$$

Confidence Limit Estimation

Confidence intervals were estimated using normal theory approximations and the

percentile bootstrap resampling method. Normal theory confidence limits were constructed

based on the estimate of the indirect effect and its standard error. The normal theory 95%

confidence intervals (Equation 22) were obtained by multiplying the indirect effect

standard error by its critical value (1.96). The resulting value was added and subtracted

from the indirect effect estimate.

$$95\%\ CI = \hat{a}\hat{b} \pm 1.96 * SE_{\hat{a}\hat{b}} \qquad (22)$$

Asymmetric confidence limits were obtained using the percentile bootstrap

confidence intervals. The bootstrap method generated 1000 random samples with

replacement for each dataset, and the indirect effect was estimated in each of those 1000

samples. Confidence intervals were calculated using the percentile bootstrap. This method

orders the estimates from the lowest to the highest. It then creates the confidence interval

with the values of the 2.5% and 97.5% percentile, which correspond to the Lower Bound and Upper Bound of the confidence interval, respectively.

Confidence Interval Coverage

Confidence Interval Coverage was assessed by the proportion of times the true population parameter (ab) was included in the 95% percentile bootstrap confidence interval. Poor coverage was considered when values fell outside the acceptable range [0.925, 0.975].

Imbalance

The accuracy of the Percentile Bootstrap Confidence Limits was evaluated with the proportion of the times the true value (ab) was above the upper confidence limit and the proportion of the times the true value (ab) was below the lower confidence limit. These proportions were then compared to the predicted proportion. For a 95% confidence interval, the most accurate methods should have a proportion of true values above the upper confidence limit of 0.025 and a proportion of true values below the lower confidence limit of 0.025. This study used the liberal robustness criterion (Bradley, 1978) to assess if proportions deviated significantly from the expected proportion. For instance, proportions were considered robust if they fell within the range $.5\alpha/2$ to $1.5\alpha/2. = (0.0125, 0.0375)$.

20

Empirical Type I Error Rates and Empirical Power

The statistical significance test for the indirect effect using the multivariate delta standard error divides the indirect effect estimate ($\hat{a}\hat{b}$) by its standard error

($\sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}$). Then this value was compared to the values in the normal distribution. If

($|\dfrac{\hat{a}\hat{b}}{\sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2}}|$) exceeded 1.96, the hypothesis of no indirect effect was rejected at a two-

tailed alpha level of 0.05 (MacKinnon, 2008).

The statistical significance for confidence intervals (normal theory confidence intervals and percentile bootstrap confidence intervals) was assessed by evaluating whether zero was included in the confidence interval.

The Empirical Type I Error Rate and the Empirical Power were calculated in both cases. Using Bradley's criterion (1978), acceptable values for Type I Error Rate fell in the range of $\alpha = 0.025 - 0.075$. Values above 0.80 for Empirical Power were considered acceptable using the conventional standard of 0.80 power.

CHAPTER 3

RESULTS (STUDY 1)

The results for the seven outcomes (Bias, Relative bias, Standardized bias,

Empirical Type I Error Rates, Empirical Power, Confidence Intervals Coverage, and

Imbalance) are discussed in the section below. Figures 6, 7, 8, and 9 show the sample size

(N = 100, 200, and 500) on the x-axis, and the outcome (e.g., Bias) on the y-axis. The

confounding effect is represented with the following symbols: negative confounding (-), no

confounding (0), and positive confounding (+).

Scenario 1. The data were analyzed with a single mediator model with a confounder

(U) influencing the mediator (M) and the outcome (Y).

Bias of the indirect effect

The results for Bias, Relative bias, and Standardized bias followed a consistent

pattern. As expected, when the data were analyzed with a single mediator model and the

confounder (U) was included, the values for Bias, Relative Bias, and Standardized bias fell

within the acceptable range (e.g., values between -.10 and .10 for Relative Bias). Figure 6

shows the Standardized Bias results. Figures with the Bias and Relative bias results are

shown in Appendix H.

Empirical Type I Error Rate for the indirect effect (ab)

The next outcome was the Empirical Type I Error Rate for the indirect effect (ab).

Empirical Type I Error Rates were computed using the Z test, Normal Confidence

Intervals, and Percentile Bootstrap Confidence Intervals. The results from these methods

show a consistent pattern, so this section describes the Empirical Type I Error Rates

calculated with the Percentile Bootstrap Confidence Intervals (Figure 7). Figures for

Empirical Type I Error Rates computed using the Z test, Normal Confidence Intervals are

shown in Appendix H.

Using Bradley's criterion (1978), Empirical Type I Error Rate values that fell in the

range of $\alpha = 0.025 - 0.075$ were considered acceptable. As expected, Empirical Type I Error

Rates did not have values above 0.075; however, some values were below 0.025.

Empirical Power for the indirect effect (ab)

The next outcome was the Empirical Power for the indirect effect (ab). This

outcome was estimated using the Z test, the Normal Confidence Intervals, and the

Percentile Bootstrap Confidence Intervals.

Overall, Empirical Power values were above the acceptable value of 0.80, for $N =$

200 and $N = 500$, for each method used to calculate the outcome (Z test, Normal

Confidence Intervals, and Percentile Bootstrap Confidence Intervals). However, differences

in Empirical Power were found in sample size $N = 100$. For instance, there were some

values between 0.75 and 0.80 when Empirical Power was estimated with a Z test and the

Normal Confidence Intervals (see Appendix H), but all values were higher than 0.80 when

Empirical Power was calculated using the Percentile Bootstrap Confidence Intervals

(Figure 8).

Confidence Interval Coverage and Imbalance

Confidence Interval Coverage was calculated using the Percentile Bootstrap

Confidence Intervals. Using Bradley's (1978) criterion, poor coverage was considered for

values that fell outside the range [0.925, 0.975]. When the data were analyzed with the true

model, no values were below 0.925 (see Figure 9). The imbalance results are shown in

Appendix H.

Figure 6

*Standardized Bias of the Indirect Effect (ab). The Data Were Analyzed With a Single
Mediator Model With a Confounder (U) That Influences the Mediator (M) and the Outcome
(Y)*
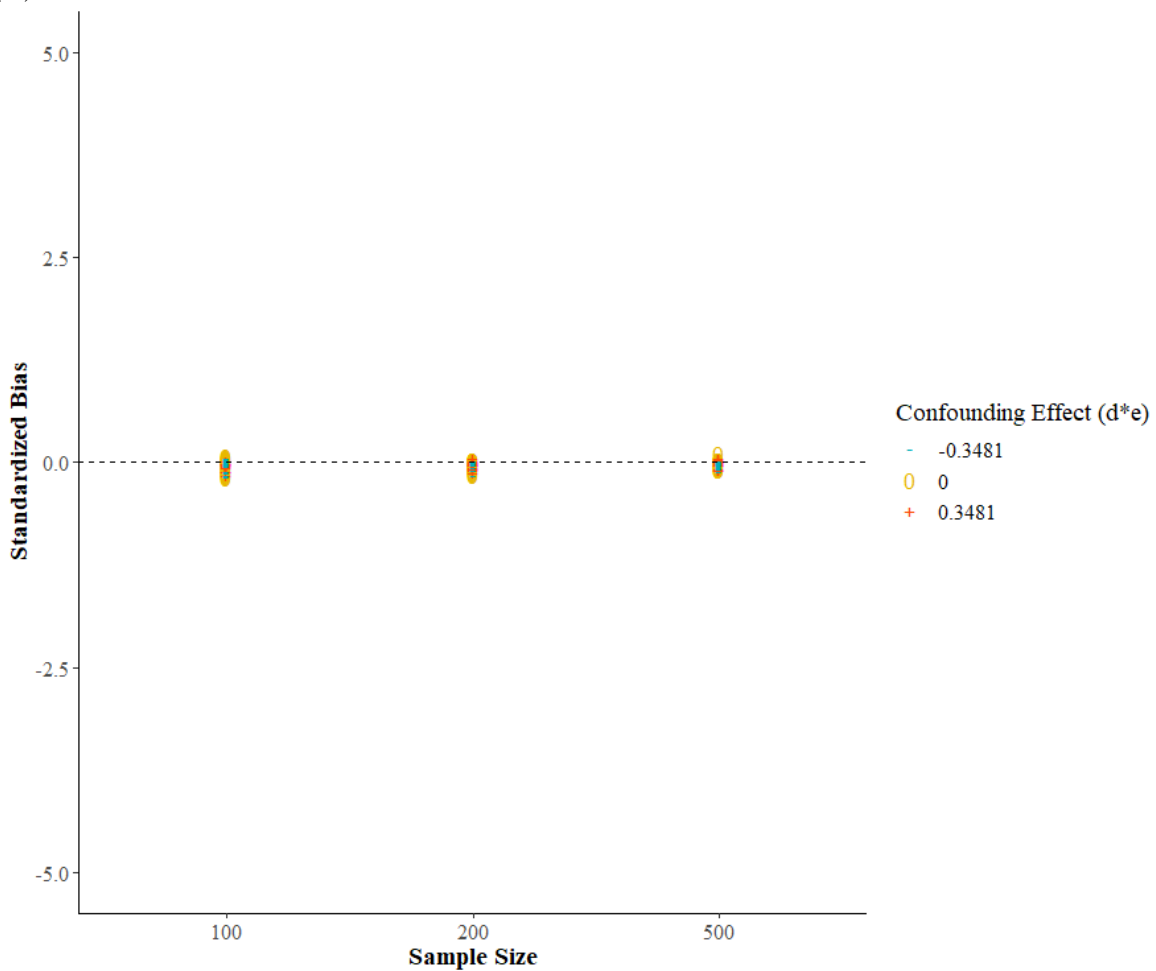
Note. The dashed line indicates a Standardized bias of zero.

Figure 7

Empirical Type I Error Rate for the Indirect Effect (ab). Statistical Significance Was Assessed Using Percentile Bootstrap Confidence Intervals. The Data Were Analyzed With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
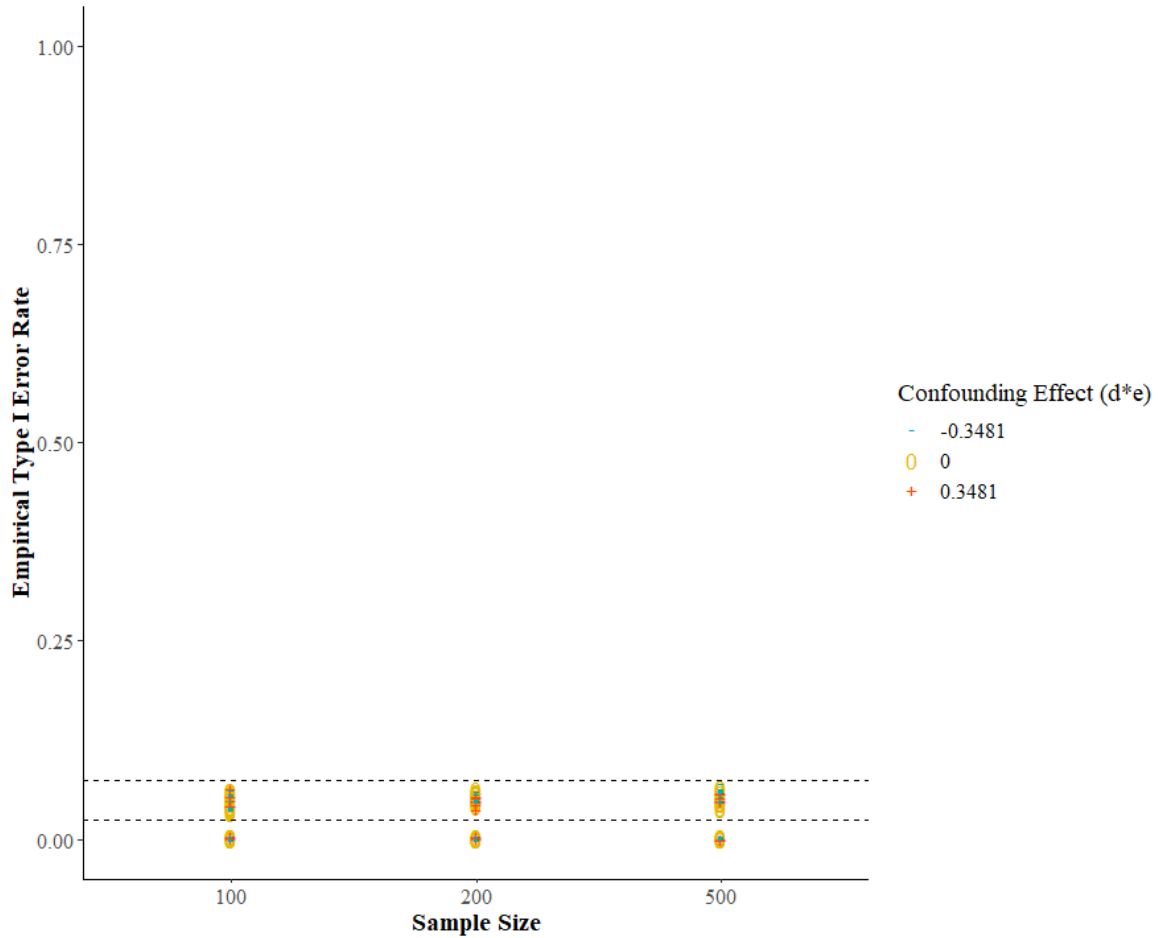
Statistical significance was assessed using Percentile Bootstrap Confidence Intervals

Note. The dashed lines indicate the adequate Empirical Type I Error Rate range [.025, .075].

Figure 8

Empirical Power for the Indirect Effect (ab). Statistical Significance was Assessed Using Percentile Bootstrap Confidence Intervals. The Data Were Analyzed With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
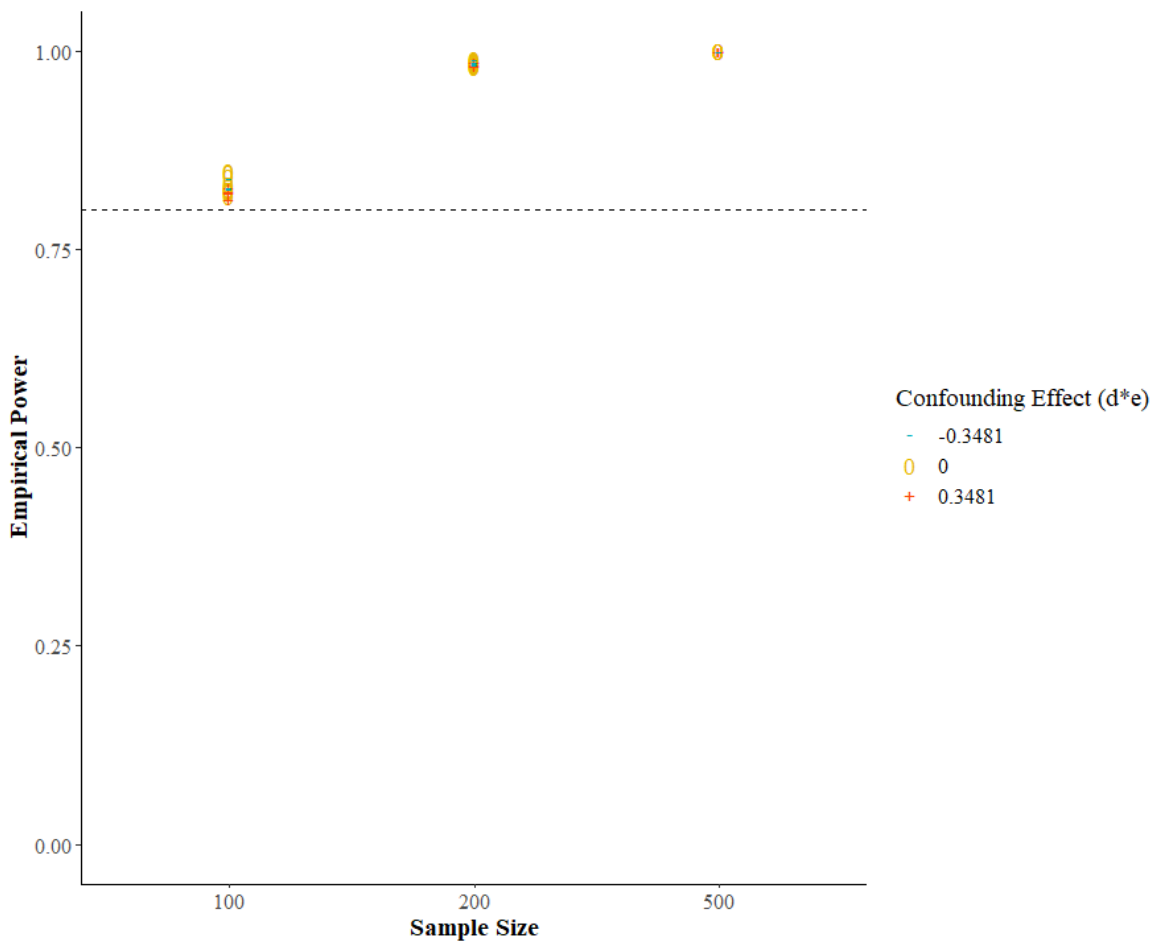
Statistical significance was assessed using Percentile Bootstrap Confidence Intervals

Note. The dashed line represents the conventional standard of 0.80 power.

Figure 9

Coverage for the Indirect Effect (ab). The Data Were Analyzed With a Single Mediator Model With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
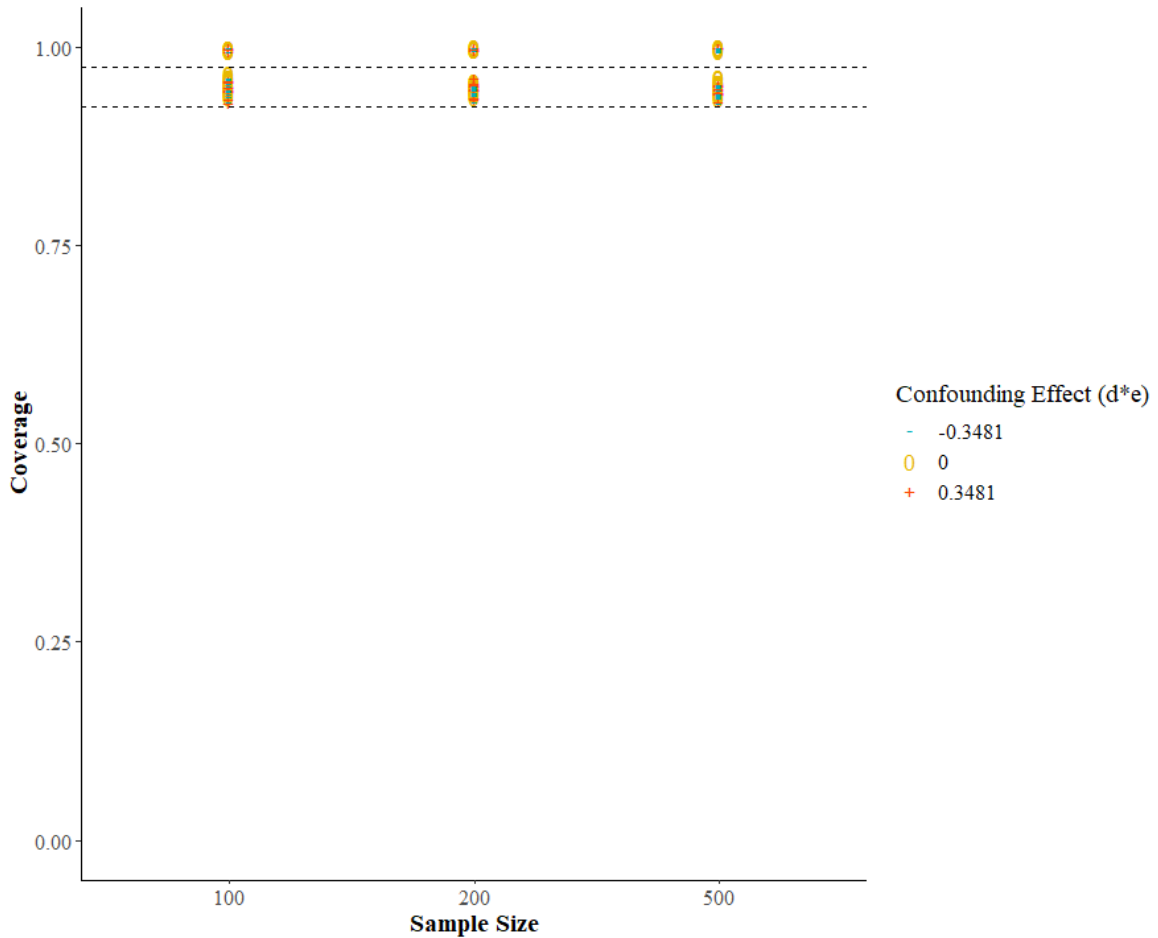
Coverage was assessed using Percentile Bootstrap Confidence Intervals

Note. The dashed lines indicate the adequate Confidence Interval Coverage range [0.925, 0.975]

Scenario 2. The data were analyzed with a misspecified model, where the confounder was omitted from the model.

Bias of the indirect effect (ab)

Because the results for Bias, Relative bias, and Standardized bias of the indirect effect (ab) followed a consistent pattern, this section focuses on reporting the results of Standardized bias (Figure 10).

When the data were analyzed with the misspecified model, Standardized bias had large values, but only for conditions where the true a parameter was large (0.59), not for cases where the true a parameter was zero. Negative confounding (de = -0.3481) resulted in negative standardized bias values, and positive confounding (de = 0.3481) resulted in positive standardized bias values. Figures with the results for Bias and Relative bias are shown in Appendix I.

Empirical Type I Error Rate for the indirect effect (ab)

The next outcome was the Empirical Type I Error Rate for the indirect effect (ab). Empirical Type I Error Rates were computed using the Z test, Normal Confidence Intervals, and Percentile Bootstrap Confidence Intervals. The results from these methods show a consistent pattern. For that reason this section describes the Empirical Type I Error Rates calculated with the Percentile Bootstrap Confidence Intervals (Figure 11). Figures for Empirical Type I Error Rates computed using the Z test, Normal Confidence Intervals are shown in Appendix I.

When the data were analyzed with a misspecified model, there were Empirical Type I Error Rates values above the acceptable criterion of 0.075 when large confounding was present, both negative (de = -0.3481) and positive (de = 0.3481); but only when the true a parameter was large (0.59). For these conditions, the Empirical Type I Error Rates were around 0.50 for sample size N = 100, increased to values around 0.80 for N = 200, and were around 1 for sample size N = 500 (See Figure 11).

Empirical Power for the indirect effect

The next outcome was the Empirical Power for the indirect effect (ab). This outcome was estimated using the Z test, the Normal Confidence Intervals, and the Percentile Bootstrap Confidence Intervals. Overall, Empirical Power values were above the acceptable value of 0.80, for N = 200 and N = 500, with no difference in the method used to calculate the outcome (Z test, Normal Confidence Intervals, and Percentile Bootstrap Confidence Intervals).

However, the true confounding effect influenced Empirical Power when N = 100. Negative confounding (de = -0.3481) was the instance where Empirical Power was affected the most, with values around 0.50 when Empirical Power was calculated using the Z test and the Normal Theory Confidence Intervals, and values around 0.60 using the Percentile Confidence Intervals. Figure 12 show the Empirical Power estimated with the Percentile Bootstrap Confidence Intervals. Results for Empirical Power estimated with the Z test, the Normal Confidence Intervals are shown in Appendix I.

Confidence Interval Coverage and Imbalance

The Confidence Interval Coverage was calculated using the Percentile Bootstrap Confidence Intervals. In general, large confounding ($de = 0.3481$ and $de = 0.3481$) affected Confidence Interval Coverage, but only when the true a path was large ($a = 0.59$) (see Figure 13). The imbalance results are shown in Appendix I.

Figure 10

Standardized Bias of the Indirect Effect (ab). The Data Were Analyzed With a Misspecified Model (Single Mediator Model)
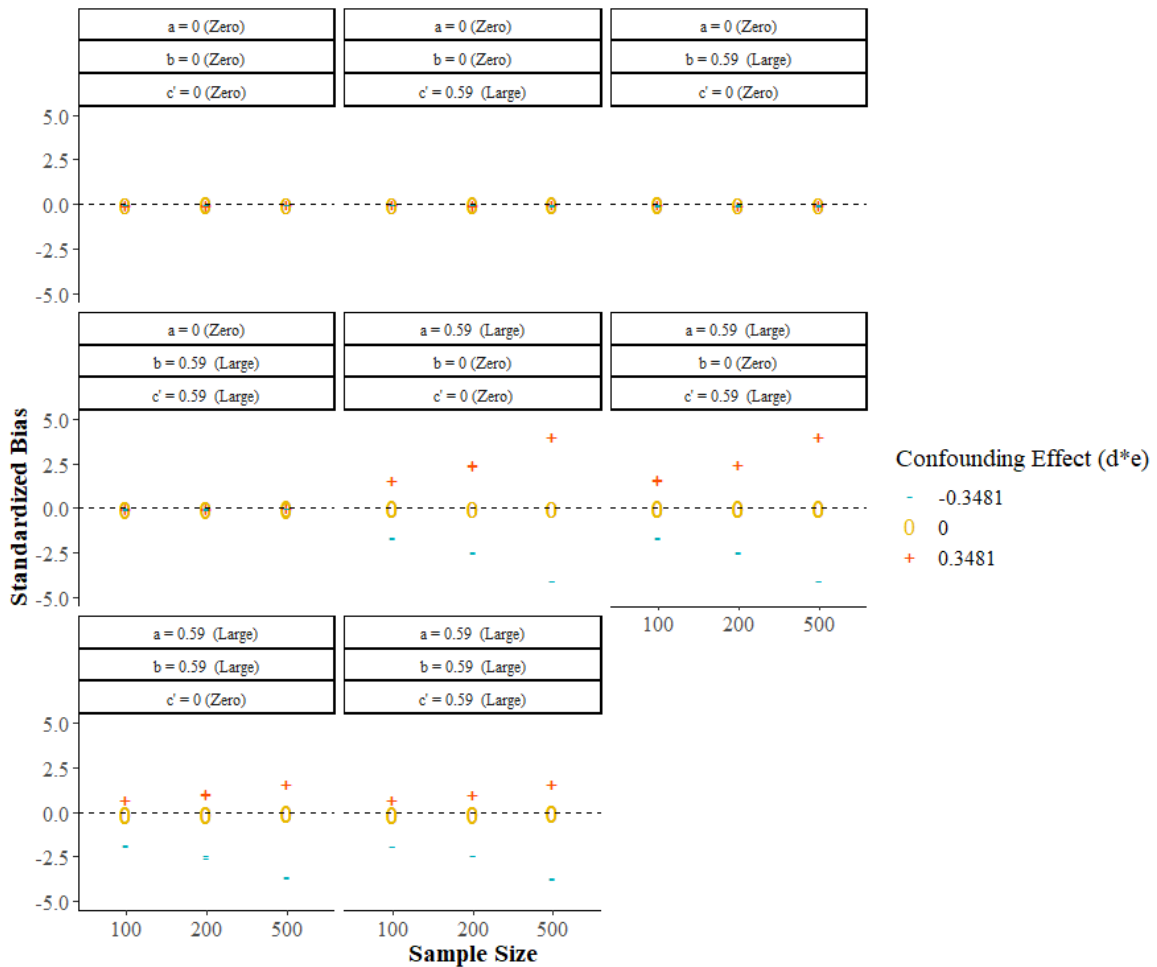
Note. The dashed line indicates a Standardized bias of zero.

Figure 11

Empirical Type I Error Rate for the Indirect Effect (ab). Statistical Significance Was Assessed Using Percentile Bootstrap Confidence Intervals). The Data Were Analyzed With a Misspecified Model (Single Mediator Model)
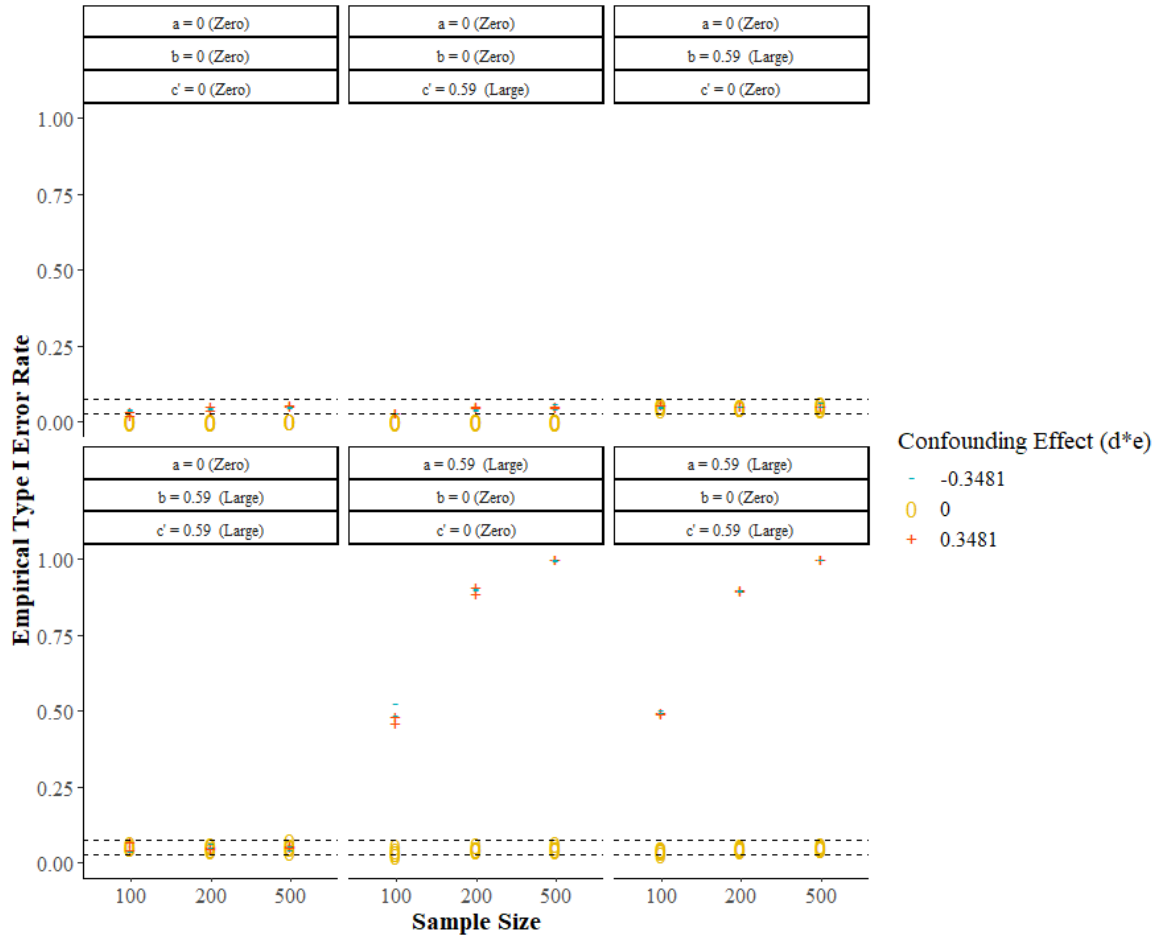
Note. The dashed lines indicate the adequate Empirical Type I Error Rate range [0.025, 0.075].

Figure 12

Empirical Power for the Indirect Effect (ab). Statistical Significance Was Assessed Using Percentile Bootstrap Confidence Intervals). The Data Were Analyzed With a Misspecified Model (Single Mediator Model)
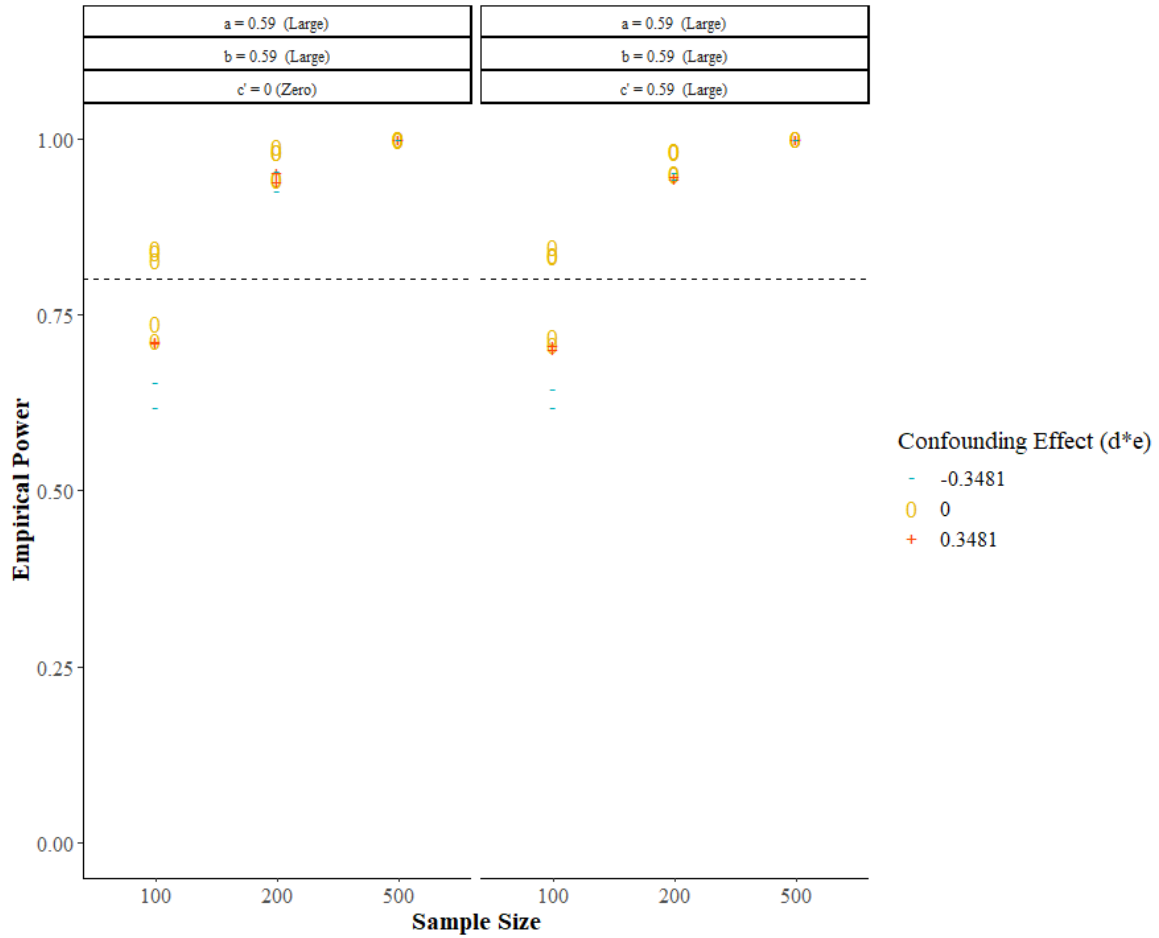
Note. The dashed line represents the conventional standard of 0.80 power.

Figure 13

Coverage for the Indirect Effect (ab). The Data Were Analyzed with a Misspecified Model (Single Mediator Model)
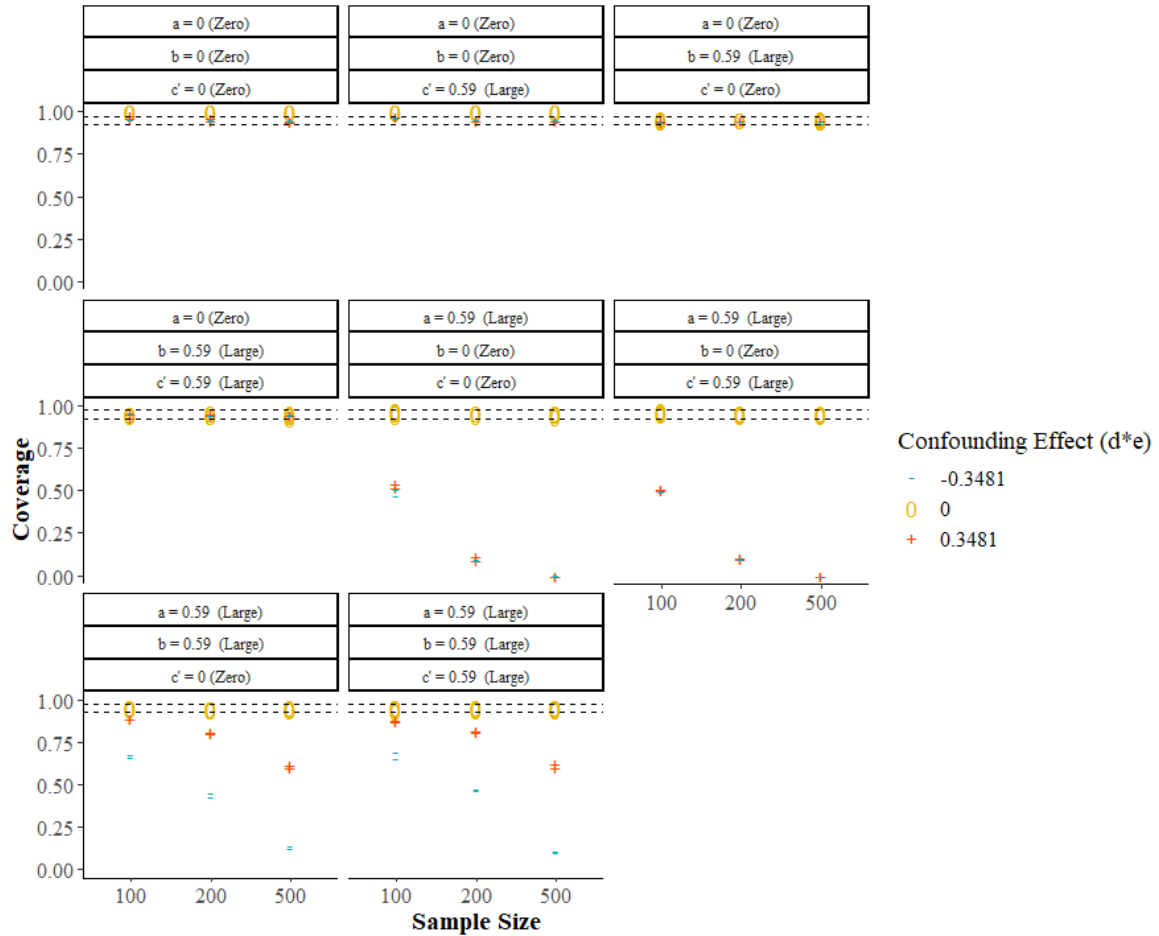
Coverage was assessed using Percentile Bootstrap Confidence Intervals

Note. The dashed lines indicate the adequate Confidence Interval Coverage range [0.925, 0.975].

CHAPTER 4

METHOD (STUDY 2)

Simulation Study

A Monte Carlo Simulation study was conducted using the SAS 9.4 programming language. The data were generated based on a single mediator model with a confounder (U) influencing the mediator (M) and the outcome (Y). In this model M, Y, and U were generated as continuous variables. X was generated as a binary variable, representing random assignment to treatment and control groups. Residuals for each variable were generated as normal variates.

The simulation had two sample sizes (N = 100 and 500) and 16 sets of parameter values (a, b, c', d and e) for a total of 32 combinations. There were 1,000 replications for each condition. The simulated data were analyzed with two sensitivity analyses (L.O.V.E. and correlated residuals). Each sensitivity analysis method had a significance test and an effect size comparison. The simulation programs for the sensitivity to confounding methods are included in the Appendices; Appendix J shows the program of the correlated residuals method, and Appendix K shows the program of the L.O.V.E. method.

There were two factors of interest in the simulation study. The first factor was the magnitude of the true confounding effect, which is denoted by the product of d times e (de = 0, 0.019, 0.1521, 0.3481). The second factor was the magnitude of the true indirect effect, which is denoted by the product of a times b (ab = 0, 0.019, 0.1521, 0.3481).

Significance tests

Significance test for the Left Out Variables Error (L.O.V.E.) method

For the L.O.V.E. method, the significance test evaluated the product of the correlation between the confounder and the mediator times the correlation between the confounder and the outcome $(r_{um} * r_{uy})$.

The first step was to estimate the standardized regression coefficients (b^*, c'^*, and a^*) for a model where the confounder U is included using equations 8, 9, and 10. Then, all the values of r_{um} and r_{uy} that make the standardized indirect effect zero $(a^* * b^* = 0)$ were selected. The significance test was performed for the pair of correlations where the values were equal $(r_{um} = r_{uy})$. If multiple pairs of correlations met the criterion, the smallest value of the product of the correlations $(r_{um} * r_{uy})$ was selected for the significance test.

The variance and covariances of the correlations were estimated with the following equations (Olkin & Finn, 1995):

$$var(r_{um}) = \frac{(1 - r_{um}^2)^2}{N} \tag{23}$$

$$var(r_{my}) = \frac{(1 - r_{my}^2)^2}{N} \tag{24}$$

$$var(r_{uy}) = \frac{(1 - r_{uy}^2)^2}{N} \tag{25}$$

$$cov(r_{um}, r_{my}) = \frac{\frac{1}{2}(2r_{my} - r_{um}r_{my})(1 - r_{uy}^2 - r_{um}^2 - r_{my}^2) + r_{uy}^3}{N} \tag{26}$$

$$cov(r_{um}, r_{uy}) = \frac{\frac{1}{2}(2r_{my} - r_{um}r_{uy})(1 - r_{um}^2 - r_{my}^2 - r_{uy}^2) + r_{my}^3}{N} \tag{27}$$

$$cov(r_{uy}, r_{my}) = \frac{\frac{1}{2}(2r_{um} - r_{my}r_{uy})(1 - r_{um}^2 - r_{my}^2 - r_{uy}^2) + r_{um}^3}{N} \tag{28}$$

The standard error for the product of correlations $(r_{um} * r_{uy})$ (equation 30) was estimated using the Multivariate delta method, where the variance–covariance matrix of the correlation coefficients was pre- and postmultiplied by the vector of partial derivatives.

Variance–covariance matrix:

$$\begin{bmatrix} & r_{um} & r_{my} & r_{uy} \\ r_{um} & var(r_{um}) & & \\ r_{my} & cov(r_{um}, r_{my}) & var(r_{my}) & \\ r_{uy} & cov(r_{um}, r_{uy}) & cov(r_{uy}, r_{my}) & var(r_{uy}) \end{bmatrix}$$

$$var_{r_{um}*r_{uy}} = r_{um}^2 * var(r_{uy}) + 2 * cov(r_{um}, r_{uy}) * r_{um} * r_{uy} + r_{uy}^2 * var(r_{um}) \tag{29}$$

$$SE = \sqrt{var_{r_{um}*r_{uy}}} \tag{30}$$

Finally, a Z value was calculated dividing $(r_{um} * r_{uy})$ by its standard error (equation 31).

$$Z = \frac{r_{um} * r_{uy}}{SE} \tag{31}$$

Significance test for the correlated residuals method

The correlated residuals method was performed in the Structural Equation Modeling (SEM) framework using the PROC CALIS statement in the SAS language programming. First, equations 2 and 3 from the single mediator model were specified with regression coefficients a, c', and b as free parameters. The COV statement was used to fix the covariances between residuals (e_2, e_3) to a range of values from -3 to 3.

The covariances between residuals were converted into correlations by dividing each covariance between residuals by the square root of residual of M ($\sqrt{D1}$) times the square root of the residual of Y ($\sqrt{D2}$).

The exact correlation between residuals (r_{e2e3}) needed to change the indirect effect to zero was identified. This value was used to perform the significance test using two-tailed hypotheses (H_0: $\rho = 0$, and H_1: $\rho \neq 0$). The t statistic was estimated using equation 32 and compared to the $t_{critical}$ with $n - 2$ degrees of freedom.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{32}$$

The 95% Confidence Intervals were estimated with the Fisher's z-transformation test. This method first transformed the correlation value to z′(equation 33). The lower and upper confidence limits for the Fisher's z test for the correlation were estimated using equation 35. Moreover, the confidence intervals were converted to r again using equation 36 (Cohen, Cohen, West, & Aiken, 2003). Finally, it was assessed whether zero was part of the 95% Confidence Interval.

$$z' = \tfrac{1}{2}[ln(1+r) - ln(1-r)] = 0.5\,ln\left(\tfrac{1+r}{1-r}\right) = arctanh(r) \tag{33}$$

$$SE_{z'} = \tfrac{1}{\sqrt{n-3}} \tag{34}$$

$$95\%\ CI = \widehat{z'} \pm 1.96 * SE_{z'} \tag{35}$$

$$r = \frac{e^{2z}-1}{e^{2z}-1} = tanh(z) \tag{36}$$

Effect size

        The amount of confounding was assessed by the effect size of a) the product of the correlation between the confounder and the mediator times the correlation between the confounder and the outcome ($r_{um} * r_{uy}$) for the L.O.V.E. method, and b) the correlation between residuals (r_{e2e3}) for the correlated residuals method.

        Cox et al. (2013) mentioned that there is no scale for evaluating whether the confounding effect is small, medium, or large. However, in the present study each estimate was evaluated at four effect sizes. Whether the estimate was greater than 0.05 (half small effect), greater than 0.1 (small effect), greater than 0.3 (medium effect), and greater than 0.5 (large effect).

Interpretation

Left Out Variables Error (L.O.V.E.) method

        In terms of the interpretation, Cox et al. (2013) described that if r_{um} and r_{uy} are large (e.g., 0.5), confounding bias is not likely because a confounder with such large effects is unlikely to exist. On the contrary, if the hypothetical correlations with the confounder are small, confounding of the indirect effect would be more plausible since there could be many possible confounders with smaller effects.

Correlated residuals method

        The correlated residuals method uses the correlation between residuals (r_{e2e3}) needed to change the indirect effect to zero. The interpretation follows the same pattern as the L.O.V.E method. If the correlation between residuals (r_{e2e3}) that is needed to

change the indirect effect to zero is large; confounding is less likely to explain the indirect effect results. And if the correlation between residuals (r_{e2e3}) needed to make the indirect effect zero is small, confounding is more likely to explain the indirect effect results.

Joint test

The accuracy of these methods was evaluated with a joint test, which multiplies the probability of a statistically significant indirect effect by the probability that confounding does not explain the results. The following section describes how each of these probabilities were estimated.

$$Joint\ test = Probability\ of\ a\ statistically\ significant\ indirect\ effect$$
$$* Probability\ that\ confounding\ does\ not\ explain\ the\ results$$

Probability of a statistically significant indirect effect

The probability of a statistically significant indirect effect was estimated using the 1000 replications for each condition. The Sobel test was used as the significance test of the indirect effect, which corresponds to the Type I Error Rate of the indirect effect when $ab = 0$, and to the Empirical Power of the indirect effect when $ab \neq 0$.

Probability that confounding does not explain the results

The probability that confounding does not explain the results was estimated by applying the tests for sensitivity to confounding to the datasets where the indirect effect estimate was statistically significant in the sample. This procedure resembles what a researcher would typically do, first testing if the indirect effect is statistically significant

39

and then assessing whether that statistically significant indirect effect can be explained by confounding bias by looking at the results of the sensitivity analyses. The probability that confounding does not explain the results of a statistically indirect effect was estimated by calculating the proportion of the times the Null Hypothesis (H$_0$: $r_{um} * r_{uy} = 0$ and H$_0$: $r_{e2e3} = 0$) of the sensitivity to confounding methods was rejected, and the proportion of times the sensitivity values were higher than a certain cutoff (0.05, 0.1, 0.3, 0.5).

CHAPTER 5

RESULTS (STUDY 2)

This section describes the simulation results of the sensitivity analyses methods.
Tables 2, 3, 4, and 5 show the results of the L.O.V.E method, and Tables 6, 7, 8, and 9
show the results of the correlated residuals method.

Type I Error Rates

For conditions when sample size N = 500 and the true confounding effect was
medium and large ($de = 0.1521 \ and \ 0.3481$), the Type I Error Rates for the test of the
indirect effect fell outside the acceptable range (0.117 and 0.226, respectively). The
Empirical Type I Error Rates for the indirect effect did not improve when multiplied by
the significance tests for the L.O.V.E method or the correlated residuals method (see
Tables 2 and 6). However, the Empirical Type I Error Rates for the test of the indirect
effect fell into the acceptable range [0.025, 0.075] when multiplied by the proportion of
times the sensitivity values were higher than 0.3 and 0.5 for both the L.O.V.E. and
correlated residuals methods (see Tables 3 and 7).

Empirical Power

The Empirical Power of the test of the indirect effect (ab) fell into the acceptable
range (above 0.80) when sample size N = 500 and the true indirect effect (ab) was
medium and large ($de = 0.1521 \ and \ 0.3481$). The Empirical Power of the indirect
effect did not change when multiplied by the significance tests or the proportion of times
the sensitivity values were higher than 0.05, 0.1, and 0.3. However, when the Empirical

Power values were multiplied by the proportion of times the sensitivity values were

higher than 0.5, decreased Empirical Power values below the acceptable value (0.80).

This finding implies that using 0.5 as a cutoff for either $r_{um} * r_{uy}$ or r_{e2e3} (see Tables 5

and 9), it would be too strict of a cutoff for sensitivity to confounding.

Table 2

Left Out Variables Error (L.O.V. E.) Simulation Results. Joint Test: Probability of a
Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain
the Results. Conditions When the True Indirect Effect Is Zero (ab = 0)

	True parameter values								Empirical Type I Error Rate of the indirect effect (ab) ■	Probability that confounding does not explain the results ●	Joint test ■ * ●
N	a	b	c'	d	e	de	ab	Mean $r_{um} * r_{uy}$		Z-test $H_0: r_{um} * r_{uy} = 0$	
100	0.14	0	0	0	0	0	0	-	0	-	-
100	0.14	0	0	0.14	0.14	0.0196	0	-	0	-	-
100	0.14	0	0	0.39	0.39	0.1521	0	0.3206	**0.003**	1	**0.003**
100	0.14	0	0	0.59	0.59	0.3481	0	0.3208	**0.016**	1	**0.016**
500	0.14	0	0	0	0	0	0	0.0961	**0.003**	1	**0.003**
500	0.14	0	0	0.14	0.14	0.0196	0	0.0900	**0.002**	1	**0.002**
500	0.14	0	0	0.39	0.39	0.1521	0	0.1312	0.117	1	0.117
500	0.14	0	0	0.59	0.59	0.3481	0	0.2238	0.226	1	0.226

Note. (-) Indicates cases when the indirect effect estimate was not statistically significant
in any of the datasets, hence no data was considered. Joint test = The joint test was
calculated by multiplying the probability of a statistically significant indirect effect
(Empirical Type I Error Rate of the indirect effect) times the probability that confounding
does not explain the results (Proportion of times the H₀: $r_{um} * r_{uy} = 0$ was rejected using
the Z-test).

Table 3

Left Out Variables Error (L.O.V. E.) Simulation Results. Joint Test: Probability of a Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain the Results. Conditions When the True Indirect Effect Is Zero (ab = 0)

True parameter values									Empirical Type I Error Rate of the indirect effect (ab)	Probability that confounding does not explain the results				Joint tests			
										Proportion of times $r_{um} * r_{uy}$ was higher than							
N	a	b	c'	d	e	de	ab	Mean $r_{um} * r_{uy}$	■	0.05 ▲	0.1 ▶	0.3 ▼	0.5 ◀	0.05 ■*▲	0.1 ■*▶	0.3 ■*▼	0.5 ■*◀
100	0.14	0	0	0	0	0	0	-	0	-	-	-	-	-	-	-	-
100	0.14	0	0	0.14	0.14	0.0196	0	-	0	-	-	-	-	-	-	-	-
100	0.14	0	0	0.39	0.39	0.1521	0	0.3206	**0.003**	1	1	0.6666	0	**0.003**	**0.003**	**0.0019**	**0**
100	0.14	0	0	0.59	0.59	0.3481	0	0.3208	**0.016**	1	1	0.5622	0	**0.016**	**0.016**	**0.0089**	**0**
500	0.14	0	0	0	0	0	0	0.0961	**0.003**	1	0	0	0	**0.003**	**0**	**0**	**0**
500	0.14	0	0	0.14	0.14	0.0196	0	0.0900	**0.002**	1	0	0	0	**0.002**	**0**	**0**	**0**
500	0.14	0	0	0.39	0.39	0.1521	0	0.1312	0.117	1	0.8290	0	0	0.117	0.0969	**0**	**0**
500	0.14	0	0	0.59	0.59	0.3481	0	0.2238	0.226	1	1	0.0575	0	0.226	0.226	**0.0129**	**0**

Note. (-) Indicates cases when the estimate of the indirect effect was not statistically significant in any of the datasets, hence no data was considered. Joint tests = The joint test was calculated by multiplying the probability of a statistically significant indirect effect (Empirical Type I Error Rate of the indirect effect) times the probability that confounding does not explain the results (Proportion of times the sensitivity values $r_{um} * r_{uy}$ were higher than 0.05, 0.1, 0.3, 0.5).

Table 4

Left Out Variables Error (L.O.V.E) Simulation Results. Joint Test (Probability of a Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain the Results). Conditions When the True Indirect Effect Is Different from Zero (ab ≠ 0)

	True parameter values								Empirical Power of the indirect effect (ab) ■	Probability that confounding does not explain the results ●	Joint test ■ * ●
N	a	b	c'	d	e	de	ab	Mean $r_{um} * r_{uy}$		Z-test $H_0: r_{um} * r_{uy} = 0$	
100	0.14	0.14	0	0	0	0	0.019	0.3078	0.005	1	0.005
100	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.2724	0.007	1	0.007
100	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.3302	0.027	1	0.027
100	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3582	0.038	1	0.038
100	0.39	0.39	0	0	0	0	0.1521	0.3685	0.316	1	0.316
100	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.3783	0.325	1	0.325
100	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4401	0.39	1	0.39
100	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5228	0.341	1	0.341
100	0.59	0.59	0	0	0	0	0.3481	0.4702	0.796	1	0.796
100	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.4762	**0.808**	1	**0.808**
100	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5426	0.74	1	0.74
100	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6088	0.682	1	0.682
500	0.14	0.14	0	0	0	0	0.019	0.1341	0.136	1	0.136
500	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.1475	0.174	1	0.174
500	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.2401	0.252	1	0.252
500	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3561	0.257	1	0.257
500	0.39	0.39	0	0	0	0	0.1521	0.3435	**0.991**	1	**0.991**
500	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.3598	**0.991**	1	**0.991**
500	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4476	**0.984**	1	**0.984**
500	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5394	**0.953**	1	**0.953**
500	0.59	0.59	0	0	0	0	0.3481	0.4784	1	1	1
500	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.4903	1	1	1
500	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5618	1	1	1
500	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6342	1	1	1

Note. Joint test = The joint test was calculated by multiplying the probability of a statistically significant indirect effect (Empirical Power of the indirect effect) times the probability that confounding does not explain the results (Proportion of times the H_0: $r_{um} * r_{uy} = 0$ was rejected using the Z-test).

Table 5

Left Out Variables Error (L.O.V.E) Simulation Results. Joint Test (Probability of a Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain the Results). Conditions When the True Indirect Effect Is Different from Zero (ab ≠ 0)

	True parameter values								Empirical Power of the indirect effect (ab)	Probability that confounding does not explain the results				Joint tests			
										Proportion of times $r_{um} * r_{uy}$ was higher than							
N	a	b	c'	d	e	de	ab	Mean $r_{um}*r_{uy}$	■	0.05 ▲	0.1 ►	0.3 ▼	0.5 ◄	0.05 ■*▲	0.1 ■*►	0.3 ■*▼	0.5 ■*◄
100	0.14	0.14	0	0	0	0	0.019	0.3078	0.005	1	1	0.80	0	0.005	0.005	0.004	0
100	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.2724	0.007	1	1	0.1428	0	0.007	0.007	0.0009	0
100	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.3302	0.027	1	1	0.6666	0	0.027	0.027	0.0179	0
100	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3582	0.038	1	1	0.7105	0.0526	0.038	0.038	0.0269	0.0019
100	0.39	0.39	0	0	0	0	0.1521	0.3685	0.316	1	1	0.8539	0.0317	0.316	0.316	0.2698	0.0100
100	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.3783	0.325	1	1	0.8641	0.0493	0.325	0.325	0.2808	0.0160
100	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4401	0.39	1	1	0.9731	0.2144	0.39	0.39	0.3795	0.0836
100	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5228	0.341	1	1	1	0.6419	0.341	0.341	0.341	0.2188
100	0.59	0.59	0	0	0	0	0.3481	0.4702	0.796	1	1	0.9816	0.3757	0.796	0.796	0.7813	0.2990
100	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.4762	**0.808**	1	1	0.9927	0.3976	**0.808**	**0.808**	**0.8021**	0.3212
100	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5426	0.74	1	1	1	0.7418	0.74	0.74	0.74	0.5489
100	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6088	0.682	1	1	1	0.9701	0.682	0.682	0.682	0.6616
500	0.14	0.14	0	0	0	0	0.019	0.1341	0.136	1	0.8308	0	0	0.136	0.1129	0	0
500	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.1475	0.174	1	0.9080	0	0	0.174	0.1579	0	0
500	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.2401	0.252	1	1	0.0952	0	0.252	0.252	0.0239	0
500	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3561	0.257	1	1	0.9027	0	0.257	0.257	0.2319	0
500	0.39	0.39	0	0	0	0	0.1521	0.3435	**0.991**	1	1	0.8799	0	**0.991**	**0.991**	**0.8719**	0
500	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.3598	**0.991**	1	1	0.9505	0.0010	**0.991**	**0.991**	**0.9419**	0.0009
500	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4476	**0.984**	1	1	1	0.0823	**0.984**	**0.984**	**0.984**	0.0809
500	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5394	**0.953**	1	1	1	0.9002	**0.953**	**0.953**	**0.953**	0.8578
500	0.59	0.59	0	0	0	0	0.3481	0.4784	**1**	1	1	1	0.3024	**1**	**1**	**1**	0.3024
500	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.4903	**1**	1	1	1	0.4300	**1**	**1**	**1**	0.43
500	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5618	**1**	1	1	1	0.9773	**1**	**1**	**1**	**0.9773**
500	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6342	**1**	1	1	1	1	**1**	**1**	**1**	**1**

Note. Joint test = The joint test was calculated by multiplying the probability of a statistically significant indirect effect (Empirical Power of the indirect effect) times the probability that confounding does not explain the results (Proportion of times the sensitivity values $r_{um} * r_{uy}$ were higher than 0.05, 0.1, 0.3, 0.5).

Table 6

Correlated Residuals Simulation Results. Joint Test (Probability of a Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain the Results. Conditions When the True Indirect Effect Is Zero (ab=0)

	True parameter values								Empirical Type I Error Rate Indirect effect (ab) ■	Probability that confounding does not explain the results		Joint tests	
N	a	b	c'	d	e	de	ab	Mean r_{e2e3}		t-test $H_0: r_{e2e3} = 0$ ●	95% CI ◊	■*●	■*◊
100	0.14	0	0	0	0	0	0	-	0	-	-	-	-
100	0.14	0	0	0.14	0.14	0.0196	0	-	0	-	-	-	-
100	0.14	0	0	0.39	0.39	0.1521	0	0.3572	**0.003**	1	1	**0.003**	**0.003**
100	0.14	0	0	0.59	0.59	0.3481	0	0.3424	**0.016**	1	1	**0.016**	**0.016**
500	0.14	0	0	0	0	0	0	-0.0284	**0.003**	1	1	**0.003**	**0.003**
500	0.14	0	0	0.14	0.14	0.0196	0	0.1010	**0.002**	1	1	**0.002**	**0.002**
500	0.14	0	0	0.39	0.39	0.1521	0	0.1617	0.117	0.9487	0.9487	0.1109	0.1109
500	0.14	0	0	0.59	0.59	0.3481	0	0.2542	0.226	1	1	0.226	0.226

Note. (-) Indicates cases when the indirect effect estimate was not statistically significant in any of the datasets, hence no data was considered. Joint test = The joint test was calculated by multiplying the probability of a statistically significant indirect effect (Empirical Type I Error Rate of the indirect effect) times the probability that confounding does not explain the results (Proportion of times the H_0: $r_{e2e3} = 0$ was rejected using the t-test, and the proportion zero was not inside the 95% CI).

Table 7

Correlated Residuals Simulation Results. Joint Test (Probability of a Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain the Results). Conditions When the True Indirect Effect Is Zero (ab=0)

	True parameter values								Empirical Type I Error Rate Indirect effect (ab)		Probability that confounding does not explain the results				Joint tests			
											Proportion of times r_{e2e3} was higher than							
N	a	b	c'	d	e	de	ab	Mean r_{e2e3}	■	0.05 ▲	0.1 ▶	0.3 ▼	0.5 ◀	0.05 ■*▲	0.1 ■*▶	0.3 ■*▼	0.5 ■*◀
100	0.14	0	0	0	0	0	0	-	0	-	-	-	-	-	-	-	-
100	0.14	0	0	0.14	0.14	0.0196	0	-	0	-	-	-	-	-	-	-	-
100	0.14	0	0	0.39	0.39	0.1521	0	0.3572	**0.003**	1	1	0.6666	0	**0.003**	**0.003**	**0.0019**	**0**
100	0.14	0	0	0.59	0.59	0.3481	0	0.3424	**0.016**	1	1	0.6875	0	**0.016**	**0.016**	**0.011**	**0**
500	0.14	0	0	0	0	0	0	-0.0284	**0.003**	1	0.3333	0	0	**0.003**	**0.0009**	**0**	**0**
500	0.14	0	0	0.14	0.14	0.0196	0	0.1010	**0.002**	1	1	0	0	**0.002**	**0.002**	**0**	**0**
500	0.14	0	0	0.39	0.39	0.1521	0	0.1617	0.117	1	0.8205	0	0	0.117	0.0959	**0**	**0**
500	0.14	0	0	0.59	0.59	0.3481	0	0.2542	0.226	1	1	0.1548	0	0.226	0.226	**0.0349**	**0**

Note. (-) Indicates cases when the indirect effect estimate was not statistically significant in any of the datasets, hence no data was considered. Joint tests = The joint test was calculated by multiplying the probability of a statistically significant indirect effect (Empirical Type I Error Rate of the indirect effect) times the probability that confounding does not explain the results (Proportion of times the sensitivity values r_{e2e3} were higher than 0.05, 0.1, 0.3, 0.5).

Table 8

Correlated Residuals Simulation Results. Joint Test (Probability of a Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain the Results). Conditions When the Indirect Effect Is Different from Zero (ab ≠ 0)

	True parameter values								Empirical Power Indirect effect (ab) ■	Probability that confounding does not explain the results		Joint tests	
N	a	b	c'	d	e	de	ab	Mean r_{e2e3}		t-test (H_0: $r_{e2e3} = 0$) ●	95% CI ◊	■*●	■*◊
100	0.14	0.14	0	0	0	0	0.019	0.3236	0.005	1	1	0.005	0.005
100	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.2914	0.007	1	1	0.007	0.007
100	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.3543	0.027	1	1	0.027	0.027
100	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3770	0.038	1	1	0.038	0.038
100	0.39	0.39	0	0	0	0	0.1521	0.3948	0.316	1	1	0.316	0.316
100	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.4021	0.325	1	1	0.325	0.325
100	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4677	0.39	1	1	0.39	0.39
100	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5553	0.341	1	1	0.341	0.341
100	0.59	0.59	0	0	0	0	0.3481	0.5090	0.796	1	1	0.796	0.796
100	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.5179	**0.808**	1	1	**0.808**	**0.808**
100	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5861	0.74	1	1	0.74	0.74
100	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6549	0.682	1	1	0.682	0.682
500	0.14	0.14	0	0	0	0	0.019	0.1614	0.136	1	1	0.136	0.136
500	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.1803	0.174	1	1	0.174	0.174
500	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.2698	0.252	1	1	0.252	0.252
500	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3804	0.257	1	1	0.257	0.257
500	0.39	0.39	0	0	0	0	0.1521	0.3625	**0.991**	1	1	**0.991**	**0.991**
500	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.3772	**0.991**	1	1	**0.991**	**0.991**
500	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4660	**0.984**	1	1	**0.984**	**0.984**
500	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5550	**0.953**	1	1	**0.953**	**0.953**
500	0.59	0.59	0	0	0	0	0.3481	0.5067	**1**	1	1	**1**	**1**
500	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.5192	**1**	1	1	**1**	**1**
500	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5900	**1**	1	1	**1**	**1**
500	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6599	**1**	1	1	**1**	**1**

Note. (-) Indicates cases when the estimate of the indirect effect was not statistically significant in any of the datasets, hence no data was considered. Joint test = The joint test was calculated by multiplying the probability of a statistically significant indirect effect (Empirical Power of the indirect effect) times the probability that confounding does not explain the results (Proportion of times the H_0: $r_{e2e3} = 0$ was rejected using the t-test, and the proportion zero was not inside the 95% CI).

Table 9

Correlated Residuals Simulation Results. Joint Test (Probability of a Statistically Significant Indirect Effect*Probability That Confounding Does Not Explain the Results). Conditions When the Indirect Effect Is Different from Zero (ab ≠ 0)

True parameter values									Empirical Power Indirect effect (ab)	Probability that confounding does not explain the results				Joint tests			
										Proportion of times r_{e2e3} is higher than							
N	a	b	c'	d	e	de	ab	Mean r_{e2e3}	■	0.05 ▲	0.1 ▶	0.3 ▼	0.5 ◀	0.05 ■*▲	0.1 ■*▶	0.3 ■*▼	0.5 ■*◀
100	0.14	0.14	0	0	0	0	0.019	0.3236	0.005	1	1	0.8000	0	0.005	0.005	0.004	0
100	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.2914	0.007	1	1	0.4285	0	0.007	0.007	0.0029	0
100	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.3543	0.027	1	1	0.7777	0	0.027	0.027	0.0209	0
100	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3770	0.038	1	1	0.8115	0.052	0.038	0.038	0.0308	0.0019
100	0.39	0.39	0	0	0	0	0.1521	0.3948	0.316	1	1	0.9018	0.0917	0.316	0.316	0.2849	0.0289
100	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.4021	0.325	1	1	0.9046	0.0800	0.325	0.325	0.2939	0.026
100	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4677	0.39	1	1	0.9769	0.3333	0.39	0.39	0.3809	0.1299
100	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5553	0.341	1	1	1	0.7947	0.341	0.341	0.341	0.2709
100	0.59	0.59	0	0	0	0	0.3481	0.5090	0.796	1	1	0.9924	0.5829	0.796	0.796	0.7899	0.4639
100	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.5179	**0.808**	1	1	0.9962	0.6014	**0.808**	**0.808**	**0.8049**	0.4859
100	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5861	0.74	1	1	1	0.8959	0.74	0.74	0.74	0.6629
100	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6549	0.682	1	1	1	0.9956	0.682	0.682	0.682	0.6789
500	0.14	0.14	0	0	0	0	0.019	0.1614	0.136	1	0.8455	0.0073	0	0.136	0.1149	0.0009	0
500	0.14	0.14	0	0.14	0.14	0.0196	0.019	0.1803	0.174	1	0.9022	0	0	0.174	0.1569	0	0
500	0.14	0.14	0	0.39	0.39	0.1521	0.019	0.2698	0.252	1	1	0.2857	0	0.252	0.252	0.0719	0
500	0.14	0.14	0	0.59	0.59	0.3481	0.019	0.3804	0.257	1	1	0.9377	0	0.257	0.257	0.2409	0
500	0.39	0.39	0	0	0	0	0.1521	0.3625	**0.991**	1	1	0.8466	0.0010	**0.991**	**0.991**	**0.8389**	0.0009
500	0.39	0.39	0	0.14	0.14	0.0196	0.1521	0.3772	**0.991**	1	1	0.9202	0.0020	**0.991**	**0.991**	**0.9119**	0.0019
500	0.39	0.39	0	0.39	0.39	0.1521	0.1521	0.4660	**0.984**	1	1	1	0.2113	**0.984**	**0.984**	**0.984**	0.2079
500	0.39	0.39	0	0.59	0.59	0.3481	0.1521	0.5550	**0.953**	1	1	1	0.9443	**0.953**	**0.953**	**0.953**	0.8999
500	0.59	0.59	0	0	0	0	0.3481	0.5067	**1**	1	1	1	0.6490	**1**	**1**	**1**	0.649
500	0.59	0.59	0	0.14	0.14	0.0196	0.3481	0.5192	**1**	1	1	1	0.7390	**1**	**1**	**1**	0.739
500	0.59	0.59	0	0.39	0.39	0.1521	0.3481	0.5900	**1**	1	1	1	0.9980	**1**	**1**	**1**	**0.998**
500	0.59	0.59	0	0.59	0.59	0.3481	0.3481	0.6599	**1**	1	1	1	1	**1**	**1**	**1**	**1**

Note. (-) Indicates cases when the estimate of the indirect effect was not statistically significant in any of the datasets, hence no data was considered. Joint test = The joint test was calculated by multiplying the probability of a statistically significant indirect effect (Empirical Power of the indirect effect) times the probability that confounding does not explain the results (Proportion of times the sensitivity values r_{e2e3} were higher than 0.05, 0.1, 0.3, 0.5).

```
                    ┌─────────────────────────────────┐
                    │      Summary of the project       │
                    └─────────────────────────────────┘


┌──────────────┐      ┌──────────────┐      ┌─────────────────────┐
│ 1. Simulation │──────│   X binary    │──────│ Total of 3 X 2 X 2 X 2 X 3 │
│    stage      │      │               │      │ X 3 (216) combinations │
└──────────────┘      └──────────────┘      │ of parameter values and │
                                             │   sample size.          │
                                             └─────────────────────┘
```

Divide the indirect effect estimate by its standard error and compare it with Z value 1.96.

Statistical significance of the indirect effect.

Comparing whether zero is in the confidence interval or not.

1. Normal theory
2. Percentile bootstrap method

Single mediator model

Empirical power
Aceptable values ≥ 0.80

Empirical Type I error rate
Acceptable values α (0.025 - 0.075)

2. Analysis stage

Divide the indirect effect estimate by its standard error and compare it with Z value 1.96.

Statistical significance of the indirect effect

Comparing whether zero is in the confidence interval or not.

1. Normal theory
2. Percentile bootstrap method

Single mediator model with a confounder

Empirical Power
Aceptable values ≥ 0.80

Empirical Type I error rate
Acceptable values α (0.025 - 0.075)

Significance test, Confidence intervals

L.O.V.E

Effect size (0.05, 0.1, 0.3, 0.5)

3. Sensitivity analyses

Significance test

Correlated residuals

Effect Size (0.05, 0.1, 0.3, 0.5)

Joint test

4. Applied data examples

Memory Study

ATLAS intervention

CHAPTER 6

APPLIED DATA EXAMPLES

Two datasets were used to evaluate confounding on the M to Y relation: 1) a data

set from a memory study, and 2) a data set from the Athletes Training and Learning to

Avoid Steroids (ATLAS) intervention.

Datasets

Memory study

In this study, 44 participants were randomly assigned to two conditions:

participants were asked to repeat a list of words (primary rehearsal) or make images of

the words (secondary rehearsal). In the end, participants were asked to report the number

of words they remembered. In the data set, X represents random assignment to

conditions, coded 0 (primary rehearsal) and 1 (secondary rehearsal). M is the response to

the question about how much each participant made images, scored from 1 (not at all) to

9 (definitely), and Y represents the total number of words that people remembered (out of

20). More detail about this study is described in MacKinnon, Valente, and Wurpts

(2018).

Athletes Training and Learning to Avoid Steroids (ATLAS) intervention

This intervention was designed to prevent steroid use among high school athletes.

The participants were 1,506 adolescent football players from 31 high schools in two

states: Oregon and Washington. The participants were randomly assigned to two

conditions: An intervention group (that received 14 sessions of the prevention program)

and a control group (that received a pamphlet with information about steroid use). More detail about this project is described in Goldberg et al. (1996).

Memory study

Left Out Variables Error (L.O.V.E) Method

Three correlations were estimated from the data: 1) the correlation between the outcome and the independent variable (r_{yx}), 2) the correlation between the mediator and the independent variable (r_{mx}), and 3) the correlation between the outcome and the mediator (r_{ym}). The estimates of those correlations were: $r_{yx} = 0.337$, $r_{mx} = 0.623$, $r_{ym} = 0.497$. The correlation between the independent variable and the omitted variable (r_{xu}) was fixed to zero. Figure 14 shows a plot for all the values of r_{um} and r_{uy} that make the standardized indirect effect zero.

Figure 14

Plot of the L.O.V.E Sensitivity to Confounding Method (Memory Study)
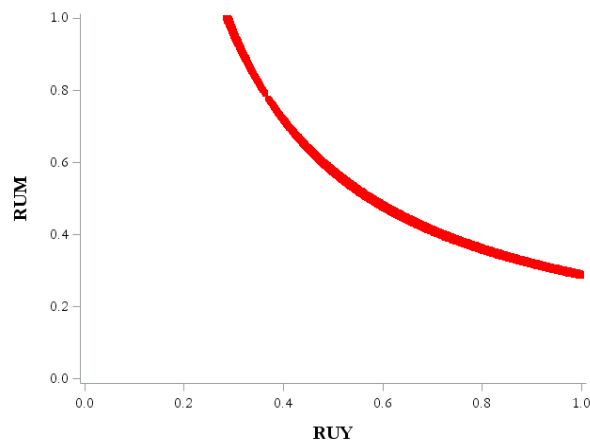
Significance test for the product of correlations ($r_{um} * r_{uy}$)

A significance test for $r_{um} * r_{uy}$ was performed. First, all values of r_{um} and r_{uy} that

make the standardized indirect effect zero were selected. Then, all cases where $r_{um} = r_{my}$

were chosen; in the memory data set there were five pairs of values that met the criterion

(See Table 10). The test was performed for the smallest value of $r_{um} * r_{uy} = 0.285$.

Table 10

Conditions of r_{mu} and r_{uy} That Make the Standardized Indirect Effect Zero, and Also Are Equal ($r_{um} = r_{uy}$)

r_{yx}	r_{mx}	r_{ym}	r_{um}	r_{uy}	$r_{um} * r_{uy}$
0.337	0.623	0.497	0.534	0.534	0.285
0.337	0.623	0.497	0.535	0.535	0.286
0.337	0.623	0.497	0.536	0.536	0.287
0.337	0.623	0.497	0.537	0.537	0.288
0.337	0.623	0.497	0.538	0.538	0.289

The variance and covariances, variances of the correlations, standard error and Z

value of $r_{um} * r_{uy} = 0.285$ were estimated with the following equations:

$$var(r_{um}) = \frac{(1 - r_{um}^2)^2}{N} = 0.011$$

$$var(r_{my}) = \frac{(1 - r_{my}^2)^2}{N} = 0.012$$

$$var(r_{uy}) = \frac{(1 - r_{uy}^2)^2}{N} = 0.011$$

$$cov(r_{um}, r_{my}) = \frac{\frac{1}{2}(2r_{my} - r_{um}r_{my})(1 - r_{uy}^2 - r_{um}^2 - r_{my}^2) + r_{uy}^3}{N} = .005$$

$$cov(r_{um}, r_{uy}) = \frac{\frac{1}{2}(2r_{my} - r_{um}r_{uy})(1 - r_{um}^2 - r_{my}^2 - r_{uy}^2) + r_{my}^3}{N} = .004$$

$$cov(r_{uy}, r_{my}) = \frac{\frac{1}{2}(2r_{um} - r_{my}r_{uy})(1 - r_{um}^2 - r_{my}^2 - r_{uy}^2) + r_{um}^3}{N} = .005$$

53

$$var_{r_{um}*r_{uy}} = r_{um}^2 * var(r_{uy}) + 2 * cov(r_{um}, r_{uy}) * r_{um} * r_{uy} + r_{uy}^2 * var(r_{um})$$
$$= .009$$
$$se = \sqrt{var_{r_{um}*r_{uy}}} = 0.095$$
$$Z = \frac{r_{um} * r_{uy}}{se} = 2.996$$

Effect size

The amount of confounding was assessed by the effect size of the product of the

correlations ($r_{um} * r_{uy} = 0.285$). The product of the correlations was greater than 0.05

(half small effect) and greater than 0.1 (small effect). But it was neither greater than 0.3

(medium effect) nor greater than 0.5 (large effect).

Correlated residuals method

Significance test for the correlation between residuals (r_{e2e3})

A significance test was performed for the value of r_{e2e3} that made the indirect effect

zero (0.389). The t-value was greater than the t$_{critical}$, at an alpha level of 0.05. Also, zero

was not included in the 95% confidence interval [0.104, 0.615].

Table 11

Significance Test (r_{e2e3}) (Memory Study)

r_{e2e3}	N	t	$t_{critical}$	z'	$SE_{z'}$	95% LCLz'	95% UCLz'	95% LCLr	95% UCLr
0.389	44	2.739	2.018	0.411	0.156	0.105	0.717	0.104	0.615

Effect size

The amount of confounding was assessed by the effect size of the correlation

between residuals ($r_{e2e3} = 0.389$), which was greater than 0.05 (half small effect),

greater than 0.1 (small effect), and greater than 0.3 (medium effect), but it was not greater

than 0.5 (large effect).

Figure 15

Plot of the Correlated Residuals Sensitivity to Confounding Method (Memory Study)
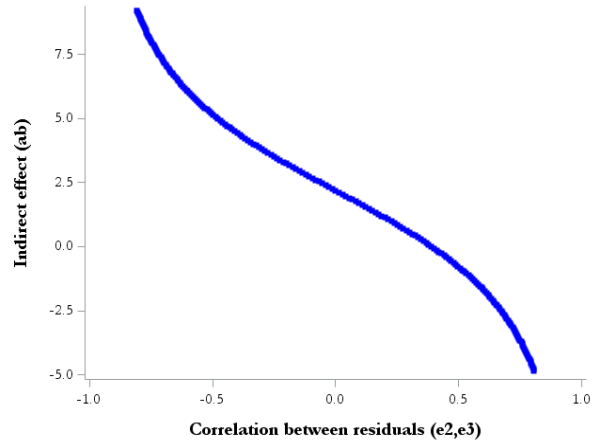
The correlated residuals method was also performed in R environment (R Core Team, 2020) using the mediation package (Tingley et al., 2014). Figure 16 shows how the Average Mediation Effect change as a function of the sensitivity parameter (ρ). The value of ρ that made the ACME zero was 0.39. This is the same value that was obtained with the SAS program.

Figure 16

Plot of the Correlated Residuals Sensitivity to Confounding Method
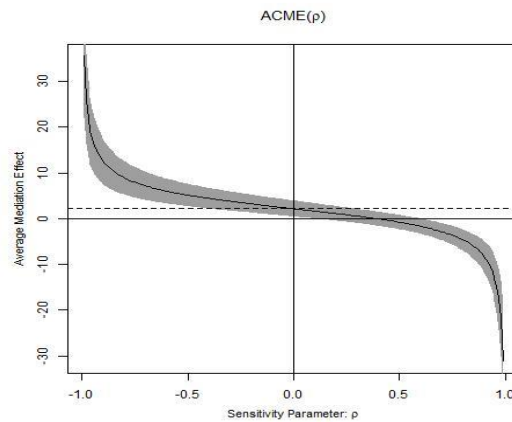
Athletes Training and Learning to Avoid Steroids (ATLAS) intervention

Left Out Variables Error (L.O.V.E) Method

Three correlations were estimated from the data: 1) the correlation between the outcome and the independent variable (r_{yx}), 2) the correlation between the mediator and the independent variable (r_{mx}), and 3) the correlation between the outcome and the mediator (r_{ym}). The estimates of those correlations were: $r_{yx} = 0.235$, $r_{mx} = 0.085$, $r_{ym} = 0.222$. The correlation between the independent variable and the omitted variable (r_{xu}) was fixed to zero. Figure 15 shows a plot for all the values of r_{um} and r_{uy} that make the standardized indirect effect zero.

56

Figure 17

Plot of the L.O.V.E Sensitivity to Confounding Method (ATLAS Intervention)
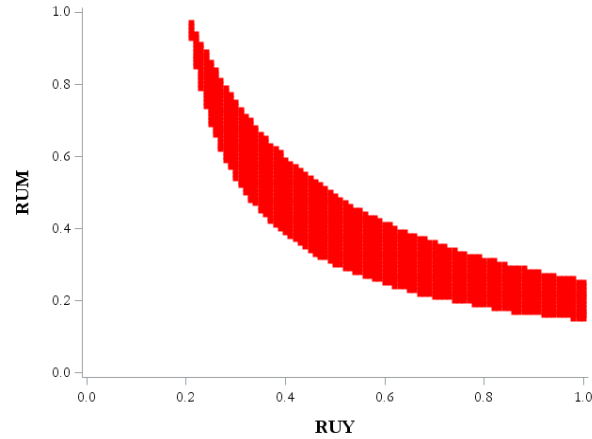
Significance test for the product of correlations ($r_{um} * r_{uy}$)

A statistical significance test for $r_{um} * r_{uy}$ was performed. First, all values of r_{um}

and r_{uy} that make the standardized indirect effect zero were selected. Then, all cases where

$r_{um} = r_{my}$ were chosen; in this dataset there were ten pairs of values that met the criterion

(See Table 11). The test was performed for the smallest value of $r_{um} * r_{uy} = 0.16$.

Table 12

Conditions of r_{um} and r_{uy} That Make the Standardized Indirect Effect Zero, and Also Are Equal ($r_{um} = r_{uy}$) (ATLAS Intervention)

r_{yx}	r_{mx}	r_{ym}	r_{um}	r_{uy}	$r_{um} * r_{uy}$
0.235	**0.085**	**0.222**	**0.40**	**0.40**	**0.160**
0.235	0.085	0.222	0.41	0.41	0.168
0.235	0.085	0.222	0.42	0.42	0.176
0.235	0.085	0.222	0.43	0.43	0.184
0.235	0.085	0.222	0.44	0.44	0.193
0.235	0.085	0.222	0.45	0.45	0.202
0.235	0.085	0.222	0.46	0.46	0.211

r_{yx}	r_{mx}	r_{ym}	r_{um}	r_{uy}	$r_{um} * r_{uy}$
0.235	0.085	0.222	0.47	0.47	0.220
0.235	0.085	0.222	0.48	0.48	0.230
0.235	0.085	0.222	0.49	0.49	0.240

The variance and covariances, variances of the correlations, standard error and Z value of $r_{um} * r_{uy} = 0.16$ were estimated with the following equations:

$$var(r_{um}) = \frac{(1 - r_{um}^2)^2}{N} = .0005$$

$$var(r_{my}) = \frac{(1 - r_{my}^2)^2}{N} = .0007$$

$$var(r_{uy}) = \frac{(1 - r_{uy}^2)^2}{N} = .0005$$

$$cov(r_{um}, r_{my}) = \frac{\frac{1}{2}(2r_{my} - r_{um}r_{my})(1 - r_{uy}^2 - r_{um}^2 - r_{my}^2) + r_{uy}^3}{N} = .0002$$

$$cov(r_{um}, r_{uy}) = \frac{\frac{1}{2}(2r_{my} - r_{um}r_{uy})(1 - r_{um}^2 - r_{my}^2 - r_{uy}^2) + r_{my}^3}{N} = .00008$$

$$cov(r_{uy}, r_{my}) = \frac{\frac{1}{2}(2r_{um} - r_{my}r_{uy})(1 - r_{um}^2 - r_{my}^2 - r_{uy}^2) + r_{um}^3}{N} = .0002$$

$$var_{r_{um}*r_{uy}} = r_{mu}^2 * var(r_{uy}) + 2 * cov(r_{um}, r_{uy}) * r_{um} * r_{uy} + r_{uy}^2 * var(r_{um})$$
$$= .000217226$$

$$se = \sqrt{var_{r_{um}*r_{uy}}} = 0.014$$

$$Z = \frac{r_{um} * r_{uy}}{se} = 10.855$$

Effect size

The amount of confounding was assessed by the effect size of the product of the correlations ($r_{um} * r_{uy} = 0.16$). The product of the correlations was greater than 0.05 (half small effect) and greater than 0.1 (small effect). However, it was neither greater than 0.3 (medium effect) nor greater than 0.5 (large effect).

Correlated residuals method

Significance test for the correlation between residuals (r_{e2e3})

A significance test was performed for the value of r_{e2e3} that made the indirect effect zero (0.229). The t-value was greater than the $t_{critical}$, at an alpha level of 0.05. Also, zero was not included in the 95% confidence interval [0.174, 0.282].

Table 13

Significance Test (r_{e2e3}) (ATLAS Intervention)

r_{e2e3}	N	t	$t_{critical}$	z'	$SE_{z'}$	95% LCLz'	95% UCLz'	95% LCLr	95% UCLr
0.229	1188	8.118	1.961	0.233	0.029	0.176	0.290	0.174	0.282

Effect size

The amount of confounding was assessed by the effect size of the correlation between residuals ($r_{e2e3} = 0.229$), which was greater than 0.05 (half small effect) and greater than 0.1 (small effect). However, it was neither greater than 0.3 (medium effect) nor greater than 0.5 (large effect).

Figure 18

Plot of the Correlated Residuals Sensitivity to Confounding Method (ATLAS Intervention)
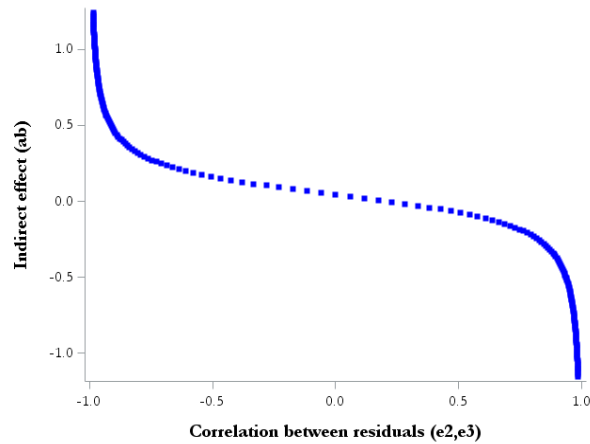
The correlated residuals method was also performed in R environment (R Core

Team, 2020) using the mediation package (Tingley et al., 2014). Figure 19 shows how

the Average Mediation Effect change as a function of the sensitivity parameter (ρ). The

value of ρ that made the ACME zero was 0.21. This value was close but not the same as

the one that was obtained with the SAS program (0.229).

Figure 19

Plot of the Correlated Residuals Sensitivity to Confounding Method
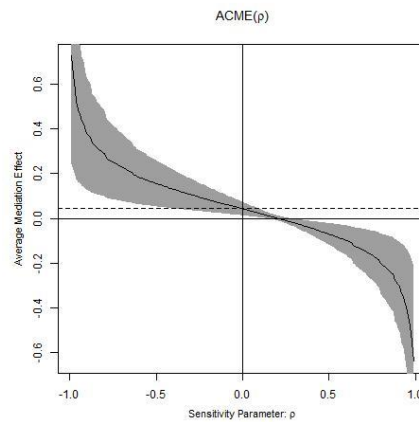
CHAPTER 7

DISCUSSION

There are four no-confounding assumptions to enhance a causal interpretation of

the effects in mediation analysis: 1) no unmeasured $X–Y$ confounding, (2) no unmeasured

$M–Y$ confounding, (3) no unmeasured $X–M$ confounding; and (4) no $M–Y$ confounder

affected by the exposure (Pearl, 2001; Valeri & Vanderweele, 2013). If X represents

random assignment to treatment and control conditions, assumptions 1 and 3 hold.

However, randomization of the treatment X does not guarantee assumptions 2 and 4. In

other words, if X is randomized, the effect of M to Y does not represent a causal effect

unless assumptions 2 and 4 are satisfied. Methods to assess confounding of M to Y were

the focus of this master's project.

Different techniques have been developed to enhance a causal interpretation of

the M to Y effect (see MacKinnon & Pirlott, 2015; Pirlott & MacKinnon, 2016; Valente

et al., 2017). For instance, sensitivity analyses are a group of methods that help deal with

confounding bias in mediation analysis when there is no measure of the potential

confounders. Researchers might omit relevant variables in their models. The goal of this

project was to understand how these methods work and to compare their accuracy. Lee et

al. (2021) proposed a list of guidelines while reporting mediation analyses of randomized

trials and observational studies. The authors suggested that researchers routinely report

any sensitivity analyses that were used to explore the causal assumptions, including

assumptions about confounding. Recommendations to investigate confounding in

mediation analysis and the probability that there are unmeasured confounders of the M to

Y relation in most studies motivated this work. It is important to evaluate how sensitivity to confounding methods work if researchers are encouraged to use and report them regularly.

In the current study, a test for significance and an effect size comparison were developed for two sensitivity to confounding analyses: the L.O.V.E. method and the correlated residuals method. The significance tests for the L.O.V.E. method and the correlated residuals method were not informative about whether confounding bias explains the results of a statistically significant indirect effect. The tests for significance of $r_{um} * r_{uy}$ and r_{e2e3}, suggested that in datasets where the indirect effect was statistically significant, the Null Hypothesis (H0: $r_{um} * r_{uy} = 0$ and H0: $r_{e2e3} = 0$) was always rejected. In other words, testing for confounding bias corresponded to the significance test of the indirect effect (ab). Suppose the significance test for confounding and the significance test for the indirect effect (ab) are the same. In that case a researcher could conclude that performing sensitivity to confounding would not be necessary, and the recommendation for researchers would be to look at the effect size of the indirect effect. The correspondence between sensitivity analysis results and tests of statistical significance has not been described in the previous research literature. That the effect size of the indirect effect is critical for sensitivity to confounding in mediation analysis is consistent with the first of Hill's considerations for a causal relation, which emphasizes the strength of the effect. Hill argued that "strong associations are more likely to be causal than weak associations because, if they could be explained by some other factor,

the effect of that factor would have to be even stronger than the observed association"

(Rothman & Greenland, 2005, p. 1)

In contrast, the effect size comparison was more informative about whether confounding bias explains the results of a statistically significant indirect effect. The values for $r_{um} * r_{uy}$ and r_{e2e3} were compared to whether they were greater than 0.05, 0.1, 0.3, and 0.5. Finally, a joint test was developed to assess the accuracy of the L.O.V.E method and the correlated residuals method; this was performed with the aim of making the results more comprehensible. The joint test was calculated by multiplying the probability of finding a statistically significant indirect effect times the probability that confounding does not explain the results.

Since the test for significance of the sensitivity parameters ($r_{um} * r_{uy}$ and r_{e2e3}) did not work as expected. We would not recommend its use. However, the general recommendation for researchers would be to focus on the effect size of the sensitivity parameters instead. Researchers are encouraged to use theory and prior knowledge to evaluate whether it is likely to find a confounder with such effect sizes. However, the effect size measures for a confounder to reduce a mediated effect to zero are also closely related to the effect size of the mediated effect.

CHAPTER 8

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

The current study has some limitations and suggestions for future research. First, only two sensitivity to confounding methods were selected for this project. A future study would include other sensitivity to confounding methods that have been developed for mediation analysis and would compare them in a simulation study. For example, VanderWeele (2010) developed a sensitivity to confounding analysis, which is an extension of the Cornfield et al. (2009) method for binary confounders. This method evaluates the effect of the binary confounder on Y, for individuals with the same value of the mediator, and the difference in the prevalence of the confounder variable U, for people with the same value of M in both the treatment and control groups. Also, recent advancements include the development of sensitivity analyses that allows the assessment of the no-omitted-confounders assumption and the no-measurement error assumption jointly (Fritz et al., 2016; Liu & Wang, 2020), a sensitivity analysis that allows assessing confounding bias for growth curve mediation models (Tofighi et al., 2019), and a sensitivity analysis that assesses confounding bias in a Bayesian mediation analysis (McCandless & Somers, 2019). However, these methods are extensions of the L.O.V.E. and correlated residuals methods so they would be closely related to the effect size evaluated for confounder bias.

Second, the simulation study though extensive, tested only 32 conditions, limiting the ability to generalize the effect size cutoffs found in this study (0.3 and 0.5) to other conditions. A future study would include more conditions with additional sample sizes

(e.g., N = 50, 250, and 1000), conditions when partial mediation is present ($c' \neq 0$), different signs of parameters, and conditions where the confounding paths have differential effects ($d \neq e$).

Third, the current project showed a correspondence between the test for significance of the indirect effect and the test for significance of confounding. A future direction involves assessing whether the test for significance of $r_{um} * r_{uy}$ and r_{e2e3} and the test for significance for the indirect effect (ab) give the same or different conclusions in terms of Empirical Type I Error Rates and Empirical Power. This could be accomplished by performing the significance test for $r_{um} * r_{uy}$ and r_{e2e3} in all the conditions, calculate the Empirical Type I Error Rates when the true indirect effect is zero ($ab = 0$), and calculate the Empirical Power when the true indirect effect is different from zero ($ab \neq 0$). In this way, the tests based on residuals can be compared to the Empirical Type I Error Rates and Empirical Power of the test for significance of ab. There is a context in which it would be important to suggest that researchers test the statistical significance of the mediated effect with residuals rather than the usual method. Testing for mediation based on the residuals or L.O.V.E. method would raise awareness of the possibility of confounding in any mediation analysis.

In addition, the cutoffs for the effect size comparison were selected arbitrarily. For instance, as mentioned in Cox et al. (2013), there is no scale for evaluating whether the confounding effect is small, medium, or large. Hence, the cutoffs were selected for the effect size of a correlation 0.05 (half small effect), 0.1 (small effect), 0.3 (medium effect), 0.5 (large effect), which might be problematic because the sensitivity parameter

value for the L.O.V.E. method is not a correlation, but it is a product of correlations

$(r_{um} * r_{uy})$. A future direction would be to develop more informative cutoffs to decide

whether confounding bias explains the indirect effect results. This would involve testing

many conditions and choosing the cutoffs that are required to make the Empirical Type I

Error Rates fall in the adequate range [0.025-0.075], and Empirical Power values fall

above 0.80.

Finally, these sensitivity to confounding analyses methods are focused on the

confounding effect that is needed to make the estimate of the indirect effect zero.

However, a future direction would include assessing confounding bias in terms of the

statistical significance of the indirect effect (the amount of confounding needed to make

the indirect effect non-significant). Another future direction is to focus on assessing

confounding bias in terms of b path (amount of confounding needed to make b

coefficient zero).

In conclusion, the inferences of the mediation analysis results rely on several

assumptions, including no-confounding assumptions. Even though the causal

interpretation of the M to Y relation is challenging, different techniques have been

developed to improve interpretation of causal mediation. In this study, a statistical

significance and an effect size comparison were developed to help understand how these

methods work. Even though the significance test did not provide additional information,

the effect size comparison may be a better approach to assess whether confounding

explains the results of a statistically significant indirect effect. Testing the significance

for unmeasured confounding and evaluating effect sizes of possible confounder effects

66

are two methods that increase awareness of the possible confounding of the M to Y

relation in mediation analysis.

REFERENCES

Albert, J. M., & Wang, W. (2014). Sensitivity analyses for parametric causal mediation. *Biostatistics*, *16*(2), 339–351. https://doi.org/10.1093/biostatistics/kxu048.

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173.

Bradley, J. V. (1978). Robustness. *British Journal of Mathematics and Statistical Psychology*, *31*, 144–151.

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

Cox, M. G., Kisbu-Sakarya, Y., Miocevic, M., & MacKinnon, D. P. (2013). *Sensitivity plots for confounder bias in the single mediator model*. *37*(5), 405–431. https://doi.org/10.1177/0193841X14524576.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53. https://doi.org/10.1146/annurev-soc-071913-043455.

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*(5), 681–697. https://doi.org/10.1080/00273171.2016.1224154.

Goldberg, L., Elliot, D., Clarke, G. N., MacKinnon, D. P., Moe, E., & Zoref, E. (1996). Effects of a multidimensional anabolic steroid prevention intervention: the Adolescents Training and Learning to Avoid Steroids (ATLAS) program. *JAMA - Journal of the American Medical Association*, *276*, 1555–15562.

Hafeman, D. M. (2011). *Practice of epidemiology confounding of indirect effects : A*

sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome. *174*(6), 710–717. https://doi.org/10.1093/aje/kwr173.

Harring, J. R., McNeish, D. M., & Hancock, G. R. (2017). Using phantom variables in structural equation modeling to assess model sensitivity to external misspecification. *Psychological Methods*, *22*(4), 616–631. https://doi.org/10.1037/met0000103.

Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71. https://doi.org/10.1214/10-STS321.

James, L. R. (1980). The unmeasured variables problem in path analysis. *Journal of Applied Psychology*, *65*(4), 415–421. https://doi.org/10.1037//0021-9010.65.4.415.

Lee, H., Cashin, A. G., Lamb, S. E., Hopewell, S., Vansteelandt, S., Vanderweele, T. J., … Henschke, N. (2021). A guideline for reporting mediation analyses of randomized trials and observational studies: the AGReMA statement. *JAMA - Journal of the American Medical Association*, *326*(11), 1045–1056. https://doi.org/10.1001/jama.2021.14075.

Liu, X., & Wang, L. (2020). The impact of measurement error and omitting confounders on statistical inference of mediation effects and tools for sensitivity analysis. *Psychological Methods*, *26*(3), 327–342. https://doi.org/10.1037/met0000345.

Lovis-McMahon, D., & MacKinnon, D. (2014). *Confound it: assessing the impact of confounders on mediation.*

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

MacKinnon, D. P. (2022). *Introduction to statistical mediation analysis.* (2nd ed.). New York, NY: Routledge.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181.

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*(1), 30–43. https://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*(4), 654–671. https://doi.org/10.1037/met0000174.

Mauro, R. (1990). Understanding L.O.V.E. (left out variables error): A method for estimating the effects of omitted variables. *Psychological Bulletin*, *108*(2), 314–329. https://doi.org/10.1037/0033-2909.108.2.314.

McCandless, L. C., & Somers, J. M. (2019). Bayesian sensitivity analysis for unmeasured confounding in causal mediation analysis. *Statistical Methods in Medical Research*, *28*(2), 515–531. https://doi.org/10.1177/0962280217729844.

Meinert, C. L., & Tonascia, S. (1986). *Clinical trials: design, conduct, and analysis*. New York: Oxford University Press.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus (Version 7.2) [Computer Software]*. Los Angeles, CA.: Muthén & Muthén.

Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin*, *118*(1), 155–164. https://doi.org/10.1037/0033-2909.118.1.155.

Pearl, J. (2001). Direct and indirect effects. In J. Breese & D. Koller (Eds.), *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence* (pp. 411–420). San Francisco, California: Morgan Kaufmann.

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Physiology & Behavior*, *66*, 29–38.

https://doi.org/doi:10.1016/j.jesp.2015.09.012

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49*(1), 37–47. https://doi.org/10.1007/BF02294204.

Rothman, K. J., & Greenland, S. (2005). Causation and causal inference in epidemiology. *American Journal of Public Health*, *95*(S1), S144–S151. https://doi.org/10.2105/AJPH.2004.059204.

Tofighi, D., Hsiao, Y. Y., Kruger, E. S., MacKinnon, D. P., Lee Van Horn, M., & Witkiewitz, K. (2019). Sensitivity analysis of the no-omitted confounder assumption in latent growth curve mediation models. *Structural Equation Modeling*, *26*(1), 94–109. https://doi.org/10.1080/10705511.2018.1506925.

Valente, M. J., Pelham, W. E., Smyth, H., & MacKinnon, D. P. (2017). Confounding in statistical mediation analysis: What it is and how to address it. *Journal of Counseling Psychology*, *64*(6), 659–671. https://doi.org/10.1037/cou0000242.

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*. *18*(2), 137–150. https://doi.org/10.1037/a0031034.

Vanderweele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, *21*(4), 540–551. https://doi.org/10.1097/EDE.0b013e3181df191c.

Vanderweele, T. J., & Chiba, Y. (2014). Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. *Epidemiology Biostatistics and Public Health*, *11*(2). https://doi.org/10.2427/9027.

APPENDIX A

DERIVED COVARIANCE MATRIX OF X, M, Y AND U

APPENDIX A

DERIVED COVARIANCE MATRIX OF X, M, Y AND U

$$
\begin{array}{c}
 & \quad X \qquad\qquad\qquad M \qquad\qquad\qquad\qquad Y \qquad\qquad\qquad\qquad U \\
\begin{array}{c} X \\ M \\ Y \\ \\ U \end{array}
\left[
\begin{array}{cccc}
\sigma_X^2 & & & \\
a\sigma_X^2 & a^2\sigma_X^2 + d^2\sigma_U^2 + \sigma_{\epsilon_M}^2 & & \\
c'\sigma_X^2 + ba\sigma_X^2 & ac'\sigma_X^2 + a^2 b\sigma_X^2 + d^2 b\sigma_U^2 + & c'^2\sigma_X^2 + 2c'ba\sigma_X^2 + b^2 a^2\sigma_X^2 + b^2 d^2\sigma_U^2 + & \\
& de\sigma_U^2 + b\sigma_{\epsilon_M}^2 & b^2\sigma_{\epsilon_M}^2 + 2ebd\sigma_U^2 + e^2\sigma_U^2 + \sigma_{\epsilon_Y}^2 & \\
0 & d\sigma_U^2 & bd\sigma_U^2 + e\sigma_U^2 & \sigma_U^2
\end{array}
\right]
\end{array}
$$

1. $Cov\,(X,X) = \sigma_X^2$

2. $Cov\,(X,M) = Cov(X,\; aX + dU + \epsilon_M)$

$= \;\; Cov(X, aX) + Cov(X, dU) + \cancel{Cov(X, \epsilon_M)}$

$= \;\; a\,Cov(X,X) + d Cov(X,U) + 0$

$= \;\; a\sigma_X^2 + d(0) + 0$

$= \;\; a\sigma_X^2$

3. $Cov\,(M,M) = Cov\,(aX + dU + \epsilon_M,\; aX + dU + \epsilon_M)$

$= \;\; Cov\,(aX, aX) + Cov\,(aX, dU) + \cancel{Cov\,(aX, \epsilon_M)} + Cov\,(dU, aX) +$

$Cov\,(dU, dU) + \cancel{Cov\,(dU, \epsilon_M)} + \cancel{Cov\,(\epsilon_M, aX)} + \cancel{Cov\,(\epsilon_M,\; dU)} + Cov\,(\epsilon_M, \epsilon_M)$

$= \;\; a^2 Cov\,(X,X) + ad Cov\,(X,U) + 0 + ad Cov\,(U,X) +$

$d^2\,Cov\,(U,U) + 0 + 0 + 0 + \sigma_{\epsilon_M}^2$

$= \;\; a^2\sigma_X^2 + 2ad\,Cov\,(X,U) + d^2 Cov\,(U,U) + \sigma_{\epsilon_M}^2$

$= \;\; a^2\sigma_X^2 + 2ad(0) + d^2\sigma_U^2 + \sigma_{\epsilon_M}^2$

$= \;\; a^2\sigma_X^2 + 0 + d^2\sigma_U^2 + \sigma_{\epsilon_M}^2$

$= \;\; a^2\sigma_X^2 + d^2\sigma_U^2 + \sigma_{\epsilon_M}^2$

4. $Cov\,(X,Y) = Cov\,(X, c'X + bM + eU + \epsilon_Y)$

$= \;\; Cov\,(X, c'X) + Cov\,(X, bM) + Cov\,(X, eU) + Cov\,(X, \epsilon_Y)$

$= \;\; c'\,Cov\,(X,X) + b\,Cov\,(X,M) + e\,Cov\,(X,U) + $ ~~$Cov\,(X,\epsilon_Y)$~~

$= \;\; c'\sigma_X^2 + b\,Cov\,(X,M) + e\,Cov\,(X,U) + 0$

$= \;\; c'\sigma_X^2 + ba\sigma_X^2 + e(0) + 0$

$= \;\; c'\sigma_X^2 + ba\sigma_X^2$

5. $Cov\,(M,Y) = Cov\,(aX + dU + \epsilon_M, c'X + bM + eU + \epsilon_Y)$

$= \;\; Cov\,(aX, c'X) + Cov\,(aX, bM) + Cov\,(aX, eU) + $ ~~$Cov(aX,\epsilon_Y)$~~ $+ Cov\,(dU, c'X) + Cov\,(dU, bM) + Cov\,(dU, eU) + $ ~~$Cov\,(dU,\epsilon_Y)$~~ $+$ ~~$Cov\,(\epsilon_M, c'X)$~~ $+ Cov\,(\epsilon_M, bM\;) ** +$ ~~$Cov\,(\epsilon_M, eU)$~~ $+$ ~~$Cov\,(\epsilon_M,\epsilon_Y)$~~

$= \;\; ac'\,Cov\,(X,X) + ab\,Cov\,(X,M) + ae\,Cov\,(X,U) + 0 + dc'\,Cov\,(U,X) + db\,Cov\,(U,M) + de\,Cov\,(U,U) + 0 + 0 + Cov\,(\epsilon_M, bM\;) ** + 0 + 0$

$= \;\; ac'\sigma_X^2 + ab(a\sigma_X^2) + ae(0) + 0 + dc'\,Cov(0) + db(d\sigma_U^2) + de\sigma_U^2 + 0 + 0 + b\sigma_{\epsilon_M}^2 + 0 + 0$

$= \;\; ac'\sigma_X^2 + a^2 b\sigma_X^2 + 0 + 0 + 0 + d^2 b\sigma_U^2 + de\sigma_U^2 + 0 + 0 + b\sigma_{\epsilon_M}^2 + 0 + 0$

$= \;\; ac'\sigma_X^2 + a^2 b\sigma_X^2 + d^2 b\sigma_U^2 + de\sigma_U^2 + b\sigma_{\epsilon_M}^2$

6. $Cov\,(Y,Y) = Cov\,(c'X + bM + eU + \epsilon_Y, c'X + bM + eU + \epsilon_Y)$

$= \;\; Cov\,(c'X, c'X) + Cov\,(c'X, bM) + Cov\,(c'X, eU) + $ ~~$Cov\,(c'X,\epsilon_Y)$~~ $+ Cov\,(bM, c'X) + Cov\,(bM, bM) + Cov\,(bM, eU) + $ ~~$Cov\,(bM,\epsilon_Y)$~~ $+ Cov\,(eU, c'X) + Cov\,(eU, bM) + Cov\,(eU, eU\;) + $ ~~$Cov\,(eU,\epsilon_Y)$~~ $+ $ ~~$Cov\,(\epsilon_Y, c'X)$~~ $+ $ ~~$Cov\,(\epsilon_Y, bM)$~~ $+ $ ~~$Cov\,(\epsilon_Y, eU)$~~ $+ Cov\,(\epsilon_Y, \epsilon_Y)$

$= \;\; c'^2 Cov\,(X,X) + c'b\,Cov\,(X,M) + c'e\,Cov\,(X,U) + 0 + bc'\,Cov\,(M,X) + b^2 Cov\,(M,M) + be\,Cov\,(M,U) + 0 + ec'\,Cov\,(U,X) + eb\,Cov\,(U,M) + e^2 Cov\,(U,U) + 0 + 0 + 0 + 0 + Cov\,(\epsilon_Y, \epsilon_Y)$

$= \;\; c'^2\sigma_X^2 + 2c'b\,Cov\,(X,M) + 2c'e\,Cov\,(X,U) + b^2 Cov\,(M,M) + 2eb\,Cov\,(U,M) + e^2\sigma_U^2 + \sigma_{\epsilon_Y}^2$

$= \;\; c'^2\sigma_X^2 + 2c'b(a\sigma_X^2) + 2c'e(0) + b^2(a^2\sigma_X^2 + d^2\sigma_U^2 + \sigma_{\epsilon_M}^2) + 2eb(d\sigma_U^2) + e^2\sigma_U^2 + \sigma_{\epsilon_Y}^2$

$= \;\; c'^2\sigma_X^2 + 2c'ba\sigma_X^2 + b^2 a^2\sigma_X^2 + b^2 d^2\sigma_U^2 + b^2\sigma_{\epsilon_M}^2 + 2ebd\sigma_U^2 + e^2\sigma_U^2 + \sigma_{\epsilon_Y}^2$

7. $Cov\,(X,U) = \;0$

8. $Cov\,(M,U) = Cov\,(aX + dU + \epsilon_M, U)$

$= \;\; Cov\,(aX, U) + Cov\,(dU, U) + Cov\,(\epsilon_M, U)$

$= \;\; a\,Cov\,(X,U) + d\,Cov\,(U,U) + $ ~~$Cov\,(\epsilon_M, U)$~~

$= \;\; a(0) + d\sigma_U^2 + 0$

$= \;\; d\sigma_U^2$

9. $Cov\,(Y, U) = Cov\,(c'X + bM + eU + \epsilon_Y, U)$

$= Cov\,(c'X, U) + Cov\,(bM, U) + Cov\,(eU, U) + \cancel{Cov\,(\epsilon_Y, U)}$

$= c'Cov\,(X, U) + bCov\,(M, U) + eCov\,(U, U) + \cancel{Cov\,(\epsilon_Y, U)}$

$= c'(0) + b(d\sigma_U^2) + e\sigma_U^2 + 0$

$= 0 + bd\sigma_U^2 + e\sigma_U^2 + 0$

$= bd\sigma_U^2 + e\sigma_U^2$

10. $Cov\,(U, U) = \sigma_U^2$

$** \; Cov\,(\epsilon_M, bM) = b[Cov\,(\epsilon_M, aX + dU + \epsilon_M)]$

$= b[\cancel{Cov\,(\epsilon_M, aX)} + \cancel{Cov\,(\epsilon_M + dU)} + Cov\,(\epsilon_M, \epsilon_M)]$
$= b[\cancel{(\epsilon_M, aX)} + \cancel{Cov\,(\epsilon_M + dU)} + Cov\,(\epsilon_M, \epsilon_M)]$
$= b\sigma_{\epsilon_M}^2$

APPENDIX B

ANALYTICAL PROGRAM

APPENDIX B

ANALYTICAL PROGRAM

```
*SECTION 1;
 *This section generates the true regression coefficients;
data a;
input a b cp d e N;
 do a=.1;
 do b=.5;
 do cp=0;
 do d=-1 to 1 by .1;
 do e=-1 to 1 by .1;
 do n=50;
 c=a*b+cp;
 TRUEAB=a*b;
*This section computes true variances and covariances based on residual
error variance equal to 1;
ERROR=1;ERRORM=1;ERRORY=1;ERRORU=1;
BMX=A;BYM=B;BYX=CP;BMU=D;BYU=E; NOBS=N;
EMOD1=(ERROR)**2;
EMOD2=(ERRORm)**2;
EMOD3=(ERRORy)**2;
EMOD4=(ERRORu)**2;
VXX=EMOD1;
CXM=BMX*VXX;
CXY=BYX*VXX+BYM*BMX*VXX;
CMM=BMX*BMX*VXX+BMU*BMU*EMOD4+EMOD2;
CMY=BMX*BYX*VXX+BMX*BMX*BYM*VXX+BMU*BMU*BYM*EMOD4+BMU*BYU*EMOD4+BYM*EMOD2
;
CYY=BYX*BYX*VXX+2*BYX*BYM*BMX*VXX+BYM*BYM*BMX*BMX*VXX+BYM*BYM*BMU*BMU*EMO
D4+BYM*BYM*EMOD2+2*BYU*BYM*BMU*EMOD4+BYU*BYU*EMOD4+EMOD3;
CXU=0;
CMU=BMU*EMOD4;
CYU=BYM*BMU*EMOD4+BYU*EMOD4;
CUU=EMOD4;
*This section computes the population regression coefficients based on
the true variances and covariances;
A = (CUU*CXM)- (CXU*CMU) / (VXX*CUU)- (CXU*CXU);
D = (VXX*CMU)- (CXU*CXM) / (VXX*CUU)- (CXU*CXU);
CP =(CMU*CMY*CXU + CXM*CMU*CYU - CMU*CMU*CXY - CXU*CYU*CMM - CXM*CMY*CUU
+  CXY*CMM*CUU)/
(CMU*CXU*CXM + CXM*CMU*CXU - CXU*CXU*CMM - CXM*CXM*CUU - CMU*CMU*VXX +
CMM*CUU*VXX);
B =(CXU*CYU*CXM - CMY*CXU*CXU + CMU*CXU*CXY - CXM*CXY*CUU - CMU*CYU*VXX +
CMY*CUU*VXX)/
(CMU*CXU*CXM + CXM*CMU*CXU - CXU*CXU*CMM - CXM*CXM*CUU - CMU*CMU*VXX +
CMM*CUU*VXX);
E = (CXM*CYU*CXM - CXM*CMY*CXU - CMU*CXM*CXY + CXU*CXY*CMM + CMU*CMY*VXX
- CYU*CMM*VXX)/
```

77

```
((-CMU)*CXU*CXM - CXM*CMU*CXU + CXU*CXU*CMM + CXM*CXM*CUU + CMU*CMU*VXX -
CMM*CUU*VXX);
*This section computes population correlations among X, M, U and Y;
RMX = CXM/SQRT(VXX*CMM);
RYM = CMY/SQRT(CMM*CYY);
RYX = CXY/SQRT(VXX*CYY);
RUM = CMU/SQRT(CMM*CUU);
RUY = CYU/SQRT(CYY*CUU);
RUX = CXU/SQRT(VXX*CUU);
*This section computes multiple correlations among the predictors;
R2XMU = (RMX*RMX + RUX*RUX - 2*RMX*RUX*RUM) / (1-RUM*RUM);
R2MXU = (RMX*RMX + RUM*RUM - 2*RMX*RUX*RUX) / (1-RUX*RUX);
R2UMX = (RUM*RUM + RUX*RUX - 2*RUM*RUX*RMX) / (1-RMX*RMX);
**********************************************************************;
*This section computes the population STANDARIZED regression coefficients
based on the true correlations;
CPRSTD =(RYX*(1-RUM**2)+RYM*(RUX*RUM-RMX)+RUY*(RMX*RUM-
RUX))/(1+2*(RMX*RUM*RUX)-RUX**2-RUM**2-RMX**2);
BSTD  =(RYM*(1-RUX**2)+RYX*(RUM*RUX-RMX)+RUY*(RMX*RUX-
RUM))/(1+2*(RMX*RUX*RUM)-RUM**2-RUX**2-RMX**2);
ASTD  =(RMX-RUM*RUX)/(1-RUX**2);
/*Observed values of a,b, cpr */
CPRBIASEDSTD=(RYX-RYM*RMX)/(1-RMX**2);
BBIASEDSTD=(RYM-RYX*RMX)/(1-RMX**2);
ABIASEDSTD=RMX;
/*Obtaining the bias of each coefficient, a is unbiased because RUX=0*/
BIASCPRSTD=CPRBIASEDSTD-CPRSTD;
BIASBSTD=BBIASEDSTD-BSTD;
BIASASTD=ABIASEDSTD-ASTD;
*The following code computes standardized true ab, standardized biased
ab, and the bias of the indirect effect;
TRUEABSTD=ASTD*BSTD;
BIASEDABSTD=ABIASEDSTD*BBIASEDSTD;
BIASABSTD=BIASEDABSTD-TRUEABSTD;
*This section computes the MSE;
MSE1 = CMM - d*d*CUU - a*a*VXX - a*d*CXU*CXU;
MSE2 = CYY - cp*cp*VXX - b*b*CMM - e*e*CUU - 2*cp*b*CXM - 2*cp*e*CXU -
2*b*e*CMU;
*This section computes the standard errors of the estimate;
*se1 = 1/(N-3);
*se2 = 1/(N-4);
*True standard errors;
sa = sqrt (((MSE1) / (N-1)) * (1/VXX/(1-RUX**2)));
sd = sqrt (((MSE1) / (N-1)) * (1/CUU/(1-RUX**2)));

scp = sqrt (((MSE2) / (N-1)) * (1/VXX /(1-R2XMU)));
sb = sqrt  (((MSE2) / (N-1)) * (1/CMM /(1-R2MXU)));
se = sqrt  (((MSE2) / (N-1)) * (1/CUU /(1-R2UMX)));

sab= sqrt(a*a*sb*sb+b*b*sa*sa);

*This section computes the regression coefficients' t-values (a, b,c', d,
e, ab);
ta = A/ sa;
```

```
td = D/ sd;
tcp = CP/scp;
tb = B/ sb;
te = E/ se;
tab= TRUEAB/sab;


*This section computes the regression coefficients' p-values (a, b,c', d,
e, ab);
pa = 1-probnorm(ta);
pd = 1-probnorm(td);
pcp = 1-probnorm(tcp);
pb = 1-probnorm(tb);
pe = 1-probnorm(te);
pab = 1-probnorm(tab);
*********************************************************************;
*SECTION 2;
*This section computes the regression coefficients based on the variances
and covariances of the single medition model;
ABIASED  = CXM/VXX;
CPBIASED = (CMM*CXY - CXM*CMY) / (VXX*CMM - CXM*CXM);
BBIASED  = (VXX*CMY - CXM*CXY) / (VXX*CMM - CXM*CXM);

BIASCP=CPBIASED-CP;
BIASB=BBIASED-B;
BIASA=ABIASED-A;

BIASEDAB= ABIASED * BBIASED;
BIAS = BIASEDAB-TRUEAB;
*This section computes the population variance of the errors in the
equation where X predicts M;
*se1p= 1/(N-2);
*This section computes the population variance of the errors in the
equation where X and M predict Y;
*se2p= 1/(N-3);
*This section computes the MSE;
MSE1p = CMM - ABIASED*ABIASED*VXX;
MSE2p = CYY - BBIASED*BBIASED*CMM - CPBIASED *VXX -
BBIASED*CPBIASED*CXM*CXM;

*True variances of the estimators (a', b' and c'') when residual error
variance equals 1;
sap  = sqrt ((MSE1p) / ((N-1)  * (VXX)));
scpp = sqrt (((MSE2p) / (N-1)) * (1/VXX/(1-RMX**2)));
sbp  = sqrt (((MSE2p) / (N-1)) * (1/CMM/(1-RMX**2)));
sabp = sqrt(ABIASED*ABIASED*sbp*sbp+BBIASED*BBIASED*sap*sap);


*This section computes the regression coefficients' t-values (a', b', c',
ab');
tap = ABIASED/ sap;
tbp = BBIASED/ sbp;
tcpp = CPBIASED/ scpp;
tabp = BIASEDAB/sabp;
```

```
*This section computes the regression coefficients' p-values (a', b',
ab');
pap = 1-probnorm(tap);
pbp = 1-probnorm(tbp);
pcpp = 1-probnorm(tcpp);
pabp = 1-probnorm(tabp);

output;
end;
end;
end;
end;
end;
end;
cards;
0 0 0 0 0 50
;

proc print data=a;
run;
```

APPENDIX C

L.O.V.E. METHOD PERFORMED IN SAS

APPENDIX C

L.O.V.E. METHOD PERFORMED IN SAS

```
TITLE 'SIMULATION OF CONFOUNDING with a LOVE plot of bias in ab when true
standardized ab is close to zero';
/*RYX is the observed correlation between X and Y;
RMX is the observed correlation between X and M;
RYM is the observed correlation between M and Y;
*/
DATA SIM;
INPUT RYX RMX RYM;
DO RUM=0 TO 1 BY .001;
DO RUY=0 TO 1 BY .0001;
DO RUX=0 ;
/* True values of a,b,cpr that account for the influence of a potential
confounder U*/
CPR=(RYX*(1-RUM**2)+RYM*(RUX*RUM-RMX)+RUY*(RUX*RUM-RUX))/(1+2*(RMX*RUM*RUX)-
RUX**2-RUM**2-RMX**2);
B=(RYM*(1-RUX**2)+RYX*(RUM*RUX-RMX)+RUY*(RMX*RUX-RUM))/(1+2*(RMX*RUX*RUM)-RUM**2-
RUX**2-RMX**2);
A=(RMX-RUM*RUX)/(1-RUX**2);
/*Observed values of a,b, cpr */
CPRBIASED=(RYX-RYM*RMX)/(1-RMX**2);
BBIASED=(RYM-RYX*RMX)/(1-RMX**2);
ABIASED=RMX;
/*Obtaining the bias of each coefficient, a is unbiased because RUX=0*/
BIASCPR=CPRBIASED-CPR;
BIASB=BBIASED-B;
BIASA=ABIASED-A;
*The following code computes standardized true ab, standardized biased ab, and
the bias of the indirect effect;
TRUEAB=A*B;
BIASEDAB=ABIASED*BBIASED;
BIASAB=BIASEDAB-TRUEAB;
RTRUEAB= round(TRUEAB,.01);
OUTPUT;
END;
END;
END;
/* the numerical values of correlations below are from the word spring 2012 data
set and should be changed by the user;
The order of correlations should be ryx, rxm, and rmy for the program to produce
the correct plot*/
CARDS;
0.30933 0.62300 0.49759
;
RUN;

DATA FORZERO;
SET SIM; IF RTRUEAB=0;

proc gplot data=FORZERO;
/*The title for the plot can be changed by the user*/
TITLE1 "Observed ab as a function of RUY and RUM when true ab is close to zero";
plot RUM*RUY=RTRUEAB/overlay;
symbol1 interpol=join r=1 c=red l=1 value=0;
run;
```

APPENDIX D
PHANTOM VARIABLE METHOD PERFORMED IN MPLUS

APPENDIX D
PHANTOM VARIABLE METHOD PERFORMED IN MPLUS

TITLE: Phantom variables;
DATA: FILE = "F:\memory data.csv";
VARIABLE: NAMES ARE X M Y;
ANALYSIS: ESTIMATOR =ML;

MODEL:
phantom BY;
phantom@1;
X ON phantom@0;
M ON X phantom@1.744;
Y ON M X phantom@1.744;

MODEL INDIRECT:
Y IND X;

APPENDIX E
PHANTOM VARIABLE METHOD PERFORMED IN SAS

PHANTOM VARIABLE METHOD PERFORMED IN SAS

```
title 'Phantom variable method to confounding';

data memory;
input  id x m y;
 cards;
 1      0      3      13
 2      0      7      11
 3      0      2      13
 4      0      1      15
 5      0      7      19
 6      0      4      13
 7      0      1      9
 8      0      1      9
 9      0      5      10
10      0      1      16
11      0      6      9
12      0      2      6
13      0      1      11
14      0      3      12
15      0      7      15
16      0      9      18
17      0      7      10
18      0      5      4
19      0      3      8
20      0      9      18
21      0      6      14
22      0      2      10
23      0      8      15
24      0      9      12
25      1      9      10
26      1      9      15
27      1      7      17
28      1      9      19
29      1      8      16
30      1      9      14
31      1      8      14
32      1      8      16
33      1      6      12
34      1      8      16
35      1      9      18
36      1      9      9
37      1      6      10
38      1      9      19
39      1      9      19
40      1      8      13
41      1      5      10
42      1      9      17
43      1      8      11
44      1      9      17;
```

```
/* Equation 3.1:
            Regress the outcome variable on the IV. That is, regress
            recall on memory technique. */
      proc reg;
      model y=x;

/* Equation 3.2:
            Regress the outcome variable on the IV and the
            mediator. That is, regress recall on memory technique and
imagery. */
      proc reg;
      model y=x m;

/* Equation 3.3:
            Regress the mediator on the IV. That is, regress imagery on
memory technique. */
      proc reg;
      model m=x;
run;

PROC CORR COV data = memory;
var x m y;
run;

PROC CALIS DATA=memory plots=pathdiagram METHOD=ML;
path
x      ===> m ,
x m   ===> y ,
phantom ===> m y = 0 0,
phantom ===> x = 0;

pvar
phantom = 1;
var x m y;
run;

PROC CALIS DATA=memory plots=pathdiagram METHOD=ML;
path
phantom x     ===> m = 1.744 a,
phantom x m   ===> y = 1.744 cp b ,
phantom ===> x = 0;

pvar
phantom = 1;
var x m y;
run;
```

APPENDIX F
CORRELATED RESIDUALS PERFORMED IN R

APPENDIX F
CORRELATED RESIDUALS PERFORMED IN R

```r
data1 <-read.table("memorydata.txt", header=TRUE, na.strings=".")

med.m <- lm(m ~ x, data = data1)
med.y <- lm(y ~ x + m, data = data1)

attach (data1)
x=as.character(x)
m=as.character(m)
library (mediation)

med.out <- mediate(med.m, med.y, sims = 2000, treat = "x", mediator = "m")
sens.out <- medsens(med.out, rho.by = 0.01, eps=.01)


jpeg ("imai plot.jpg")
plot (sens.out)
dev.off()

## png
##    2

detach(data1)
```

APPENDIX G
SIMULATION PROGRAM

APPENDIX G
SIMULATION PROGRAM

```sas
TITLE 'SIMULATION SINGLE MEDIATOR MODEL WITH A CONFOUNDER CAUSING M AND
                            Y. X-BINARY';
OPTIONS PS=59 LS=80 REPLACE NONOTES;
FILENAME NULLOG DUMMY 'C:\NULL';
PROC PRINTTO LOG=NULL;
RUN;

%MACRO SIMULATE(NREP,NOBS,B1,B2,B3,B4,B5,FILE,TYPE);
DATA SUMMARY;
SET _NULL_;
%DO I=1 %TO &NREP;
*GENERATE*;
DATA SIM;
DO I=1 TO &NOBS;
X = RAND('BERNOULLI', .5);
U = RAND('NORMAL', 0, 1);
M = &B1*X + &B2*U + RAND('NORMAL', 0, 1);
Y = &B3*X + &B4*M + &B5*U + RAND('NORMAL', 0, 1);
OUTPUT;
END;
*****************************************
****MODEL 0 (Y = X);
*****************************************
*Estimating the (Y = X) regression;
PROC REG DATA=SIM OUTEST=FILE TABLEOUT NOPRINT COVOUT;
Model0 : MODEL Y = X/ stb;
*Print model summary;
*PROC PRINT;
DATA A; SET FILE;
IF _TYPE_='PARMS'; C=X; MSE = _RMSE_*_RMSE_;
DROP MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X Y;
KEEP C MSE;
*Merging data Model 0;
DATA MODEL0; MERGE A;

*****************************************
****MODEL 1 (M = X U);
*****************************************
*Estimating the (M = X U) regression;
PROC REG DATA=SIM OUTEST=FILE TABLEOUT NOPRINT COVOUT;
Model1: MODEL M = X U/ stb;
*Print model summary;
*PROC PRINT;

*This code saves the value of a and u;
DATA B; SET FILE;
IF _TYPE_='PARMS';
A=X; D=U; MSE1 = _RMSE_*_RMSE_;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X U M;
KEEP A D MSE1;
```

```
*Saving the standard error of the coefficient a;
DATA B1; SET FILE;
IF _TYPE_='STDERR';
SEA=X;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X U M;
KEEP SEA;
*Merging data Model 1;
DATA MODEL1; MERGE B B1;
*PROC PRINT;
****************************************
****MODEL 2 (Y = X M U);
****************************************
*Estimating the (Y = X M U) regression;
PROC REG DATA=SIM OUTEST=FILE  TABLEOUT NOPRINT COVOUT;
Model2 : MODEL Y = X M U/ stb;
*Print model summary;
*PROC PRINT;
*This code saves the value of cp, b and e;
DATA C; SET FILE;
IF _TYPE_='PARMS';
CP=X; B=M; E=U; MSE2 = _RMSE_*_RMSE_;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ Intercept X M U Y;
KEEP CP B E MSE2;
*Saving the standard error of the coefficients biasedb and biasedc;
DATA C1; SET FILE;
IF _TYPE_='STDERR';
SECP=X;SEB=M;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ Intercept X M U Y;
KEEP SECP SEB;
*This code saves the variances and covariances of the sample;
PROC CORR COV DATA=SIM NOPRINT OUTP=COV;
VAR Y X M U;
*Print covariance matrix;
*PROC PRINT;
**This code saves the variances and covariances;
DATA F1; SET COV;
IF _NAME_ = 'Y' and _TYPE_ = 'COV'; VARY  = Y;
KEEP VARY;

DATA F2; SET COV;
IF _NAME_ = 'X' and _TYPE_ = 'COV'; VARX  = X;
KEEP VARX;

DATA F3; SET COV;
IF _NAME_ = 'M' and _TYPE_ = 'COV'; VARM  = M;
KEEP VARM;

DATA F4; SET COV;
IF _NAME_ = 'U' and _TYPE_ = 'COV'; VARU  = U;
KEEP VARU;

DATA F5; SET COV;
IF _NAME_ = 'X' and _TYPE_ = 'COV'; COVYX = Y;
KEEP COVYX;
```

```
DATA F6; SET COV;
IF _NAME_ = 'M' and _TYPE_ = 'COV'; COVYM = Y;
KEEP COVYM;

DATA F7; SET COV;
IF _NAME_ = 'U' and _TYPE_ = 'COV'; COVUY = Y;
KEEP COVUY;

DATA F8; SET COV;
IF _NAME_ = 'X' and _TYPE_ = 'COV'; COVMX = M;
KEEP COVMX;

DATA F9; SET COV;
IF _NAME_ = 'U' and _TYPE_ = 'COV'; COVUX = X;
KEEP COVUX;

DATA F10; SET COV;
IF _NAME_ = 'U' and _TYPE_ = 'COV'; COVUM = M;
KEEP COVUM;

DATA F11; SET COV;
IF _NAME_ = 'X' and _TYPE_ = 'CORR'; RYX  = Y;
KEEP RYX;

DATA F12; SET COV;
IF _NAME_ = 'M' and _TYPE_ = 'CORR'; RYM  = Y;
KEEP RYM;

DATA F13; SET COV;
IF _NAME_ = 'U' and _TYPE_ = 'CORR'; RUY  = Y;
KEEP RUY;

DATA F14; SET COV;
IF _NAME_ = 'X' and _TYPE_ = 'CORR'; RMX  = M;
KEEP RMX;

DATA F15; SET COV;
IF _NAME_ = 'U' and _TYPE_ = 'CORR'; RUX  = X;
KEEP RUX;

DATA F16; SET COV;
IF _NAME_ = 'U' and _TYPE_ = 'CORR'; RUM  = M;
KEEP RUM;

DATA MODEL2; MERGE C C1 F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13 F14
F15 F16;
*************************************
*************************************
*This section includes the misspecified single mediator model
which omits the U variable;

*************************************
****MODEL 3 (M = X) (Misspecified model)
```

```sas
*****************************************
*Estimating the (M = X) regression;
PROC REG DATA=SIM OUTEST=FILE TABLEOUT NOPRINT COVOUT;
Model3 : MODEL M = X/ stb;
*Saving the coefficient biaseda;
DATA D; SET FILE;
IF _TYPE_='PARMS';
bA=X; MSE3 = _RMSE_*_RMSE_;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M;
KEEP bA MSE3;
*Saving the standard error of the coefficient biaseda;
DATA D1; SET FILE;
IF _TYPE_='STDERR';
SEbA=X;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M;
KEEP SEbA;
*Merging data Model 3;
DATA MODEL3; MERGE D D1;
*****************************************
****MODEL 4 (Y = X M)
*Misspecified model
*****************************************
*Estimating the (Y = X M) regression;
PROC REG DATA=SIM OUTEST=FILE TABLEOUT NOPRINT COVOUT;
Model4 : MODEL Y = X M/ stb;
*Saving the coefficients biasedb and biasedc;
DATA E; SET FILE;
IF _TYPE_='PARMS';
bCP=X; bB=M; MSE4 = _RMSE_*_RMSE_;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M Y;
KEEP bCP bB MSE4;


*Saving the standard error of the coefficients biasedb and biasedc;
DATA E1; SET FILE;
IF _TYPE_='STDERR';
SEbCP=X;SEbB=M;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M Y;
KEEP SEbCP SEbB;
*Merging data Model 4;
DATA MODEL4; MERGE E E1;
*PROC PRINT;

*Starts Percentile Boostrap Single mediator model and a confounder;
%LET NBOOT=1000;

PROC SURVEYSELECT DATA=SIM NOPRINT OUT=OUT2 METHOD=URS SAMPSIZE=&NOBS
REP=&NBOOT OUTHITS;
RUN;

PROC REG DATA=OUT2 OUTEST=OUT3 NOPRINT;
BY REPLICATE;
MODEL Y = M X U;
MODEL M = X U;
```

94

```
DATA F; SET OUT3;
IF _MODEL_^='MODEL1' THEN DELETE;
B=M;
KEEP REPLICATE B;

DATA G; SET OUT3;
IF _MODEL_^='MODEL2' THEN DELETE;
A=X;
KEEP REPLICATE A;

DATA M; MERGE F G; BY REPLICATE;
AB=A*B;

PROC SORT DATA=M;
BY AB;
PROC UNIVARIATE DATA=M NOPRINT;
VAR AB;
*PERCENTILE BOOTSTRAP;
DATA I; SET M;
IF _N_=(CEIL(.025*&NBOOT)) THEN CALL SYMPUT("LCL95", AB);
IF _N_=(CEIL(.975*&NBOOT)) THEN CALL SYMPUT("UCL95", AB);
RUN;

DATA J;
PBLCL95=&LCL95;
PBUCL95=&UCL95;
KEEP PBLCL95 PBUCL95;
*Finishes Percentile Boostrap Single mediator model and a confounder;


*Starts Percentile Boostrap Single mediator model;
%LET NBOOT=1000;

PROC SURVEYSELECT DATA=SIM NOPRINT OUT=OUT21 METHOD=URS SAMPSIZE=&NOBS
REP=&NBOOT OUTHITS;
RUN;

PROC REG DATA=OUT21 OUTEST=OUT31 NOPRINT;
BY REPLICATE;
MODEL Y = M X;
MODEL M = X;

DATA F1; SET OUT31;
IF _MODEL_^='MODEL1' THEN DELETE;
biasedB=M;
KEEP REPLICATE biasedB;

DATA G1; SET OUT31;
IF _MODEL_^='MODEL2' THEN DELETE;
biasedA=X;
KEEP REPLICATE biasedA;

DATA M1; MERGE F1 G1; BY REPLICATE;
```

95

```
biasedAB=biasedA*biasedB;

PROC SORT DATA=M1;
BY biasedAB;
PROC UNIVARIATE DATA=M1 NOPRINT;
VAR biasedAB;
*PERCENTILE BOOTSTRAP;
DATA I1; SET M1;
IF _N_=(CEIL(.025*&NBOOT)) THEN CALL SYMPUT("LCL95", biasedAB);
IF _N_=(CEIL(.975*&NBOOT)) THEN CALL SYMPUT("UCL95", biasedAB);
RUN;

DATA J1;
bPBLCL95=&LCL95;
bPBUCL95=&UCL95;
KEEP bPBLCL95 bPBUCL95;
*Finishes Percentile Boostrap Single mediator model;

*This code merges the different datasets that contain estimates from the
simulation replication;
DATA ALL;
      MERGE MODEL0 MODEL1 MODEL2 MODEL3 MODEL4 J J1;
RUN;

DATA TEST; SET ALL;
B1=&B1 ; B2=&B2 ; B3=&B3 ; B4=&B4; B5=&B5;
NOBS= &NOBS;
TYPE = &TYPE;
DATA NEW; SET SUMMARY;
DATA SUMMARY; SET NEW TEST;
DATA SIMDATA.SIM; SET WORK.SIM;
RUN;
%END;




/*This code includes the test for significance of the indirect effect for
Model 1;
MODEL 1:
M = X+ U; Y = X  + M + U;   */
DATA SIG;
SET SUMMARY;
TRUEAB = B1 * B4;
AB = A * B;
SOBEL = SQRT(A*A*SEB*SEB+B*B*SEA*SEA);
ZSOBEL = AB/SOBEL;
PSOBEL = 1 - PROBNORM(ZSOBEL);
NTLCL = AB - 1.96 * SOBEL;
NTUCL = AB + 1.96 * SOBEL;
ABSZSOBEL = ABS(ZSOBEL);

IF ABSZSOBEL > 1.96 THEN DECZSOBEL = 1;
```

```
ELSE DECZSOBEL = 0;

IF PSOBEL < 0.05 THEN DECPSOBEL = 1;
ELSE DECPSOBEL = 0;

IF NTLCL <= 0 <= NTUCL THEN DECNTCI = 0;
ELSE DECNTCI = 1;

IF PBLCL95 <= 0 <= PBUCL95 THEN DECPBCI = 0;
ELSE DECPBCI = 1;

/*Coverage Boostrap*/
IF PBLCL95 <= TRUEAB <= PBUCL95 THEN COV1 = 1;
ELSE COV1  = 0;
/*Imbalance Boostrap*/
IF TRUEAB < PBLCL95 THEN LEFT1 = 1;
ELSE LEFT1 = 0;

IF TRUEAB > PBUCL95 THEN RIGHT1 = 1;
ELSE RIGHT1 = 0;

/*Bias*/
BIAS1 = AB-TRUEAB;
/*Relative Bias*/
RBIAS1 = (AB-TRUEAB)/TRUEAB;
/*Standardized Bias*/
SBIAS1 = (AB-TRUEAB)/SOBEL;


/*This code includes the test for significance of the indirect effect for
Model 2;
MODEL 2 (Misspecified model):
M = X ; Y = X  + M */
bAB = bA * bB;
SOBELP = SQRT(bA*bA*SEbB*SEbB+bB*bB*SEbA*SEbA);
ZSOBELP = bAB/SOBELP;
PSOBELP = 1 - PROBNORM(ZSOBELP);
NTLCLP = bAB - 1.96 * SOBELP;
NTUCLP = bAB + 1.96 * SOBELP;
ABSZSOBELP = ABS(ZSOBELP);

IF ABSZSOBELP > 1.96 THEN DECZSOBELP = 1;
ELSE DECZSOBELP = 0;

IF PSOBELP < 0.05 THEN DECPSOBELP = 1;
ELSE DECPSOBELP = 0;

IF NTLCLP <= 0 <= NTUCLP THEN DECNTCIP = 0;
ELSE DECNTCIP = 1;

IF bPBLCL95 <= 0 <= bPBUCL95 THEN DECbPBCI = 0;
ELSE DECbPBCI = 1;

/*Coverage Boostrap*/
```

```
IF bPBLCL95 <= TRUEAB <= bPBUCL95 THEN COV2 = 1;
ELSE COV2 = 0;
/*Imbalance Boostrap*/
IF TRUEAB < bPBLCL95 THEN LEFT2 = 1;
ELSE LEFT2 = 0;

IF TRUEAB > bPBUCL95 THEN RIGHT2 = 1;
ELSE RIGHT2 = 0;

/*Bias*/
BIAS2 = bAB-TRUEAB;
/*Relative Bias*/
RBIAS2 = (bAB-TRUEAB)/TRUEAB;
/*Standardized Bias*/
SBIAS2 = (bAB-TRUEAB)/SOBELP;
RUN;

**This code generates frequency tables;
PROC MEANS DATA=SIG N MEAN;
VAR DECZSOBEL DECPSOBEL DECNTCI DECPBCI COV1 RIGHT1 LEFT1 BIAS1 RBIAS1
SBIAS1 DECZSOBELP DECPSOBELP DECNTCIP DECbPBCI COV2 RIGHT2 LEFT2 BIAS2
RBIAS2 SBIAS2;
RUN;

*This code prints out the summary of the estimates in all the
replications;
PROC MEANS DATA=SUMMARY N MEAN STD MIN MAX SUM VAR STDERR;
RUN;

%MEND;

**Trial;
%SIMULATE(NREP=10,NOBS=50,B1=0.5,  B2=0.9, B3=0, B4=0.2,   B5= -0.9,
FILE = TEMP, TYPE = 'BCCC');
```

APPENDIX H
RESULTS STUDY 1 (TRUE MODEL)

Bias

Figure 20

Bias of the Indirect Effect (ab). The Data Were Analyzed with a Single Mediator Model with a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
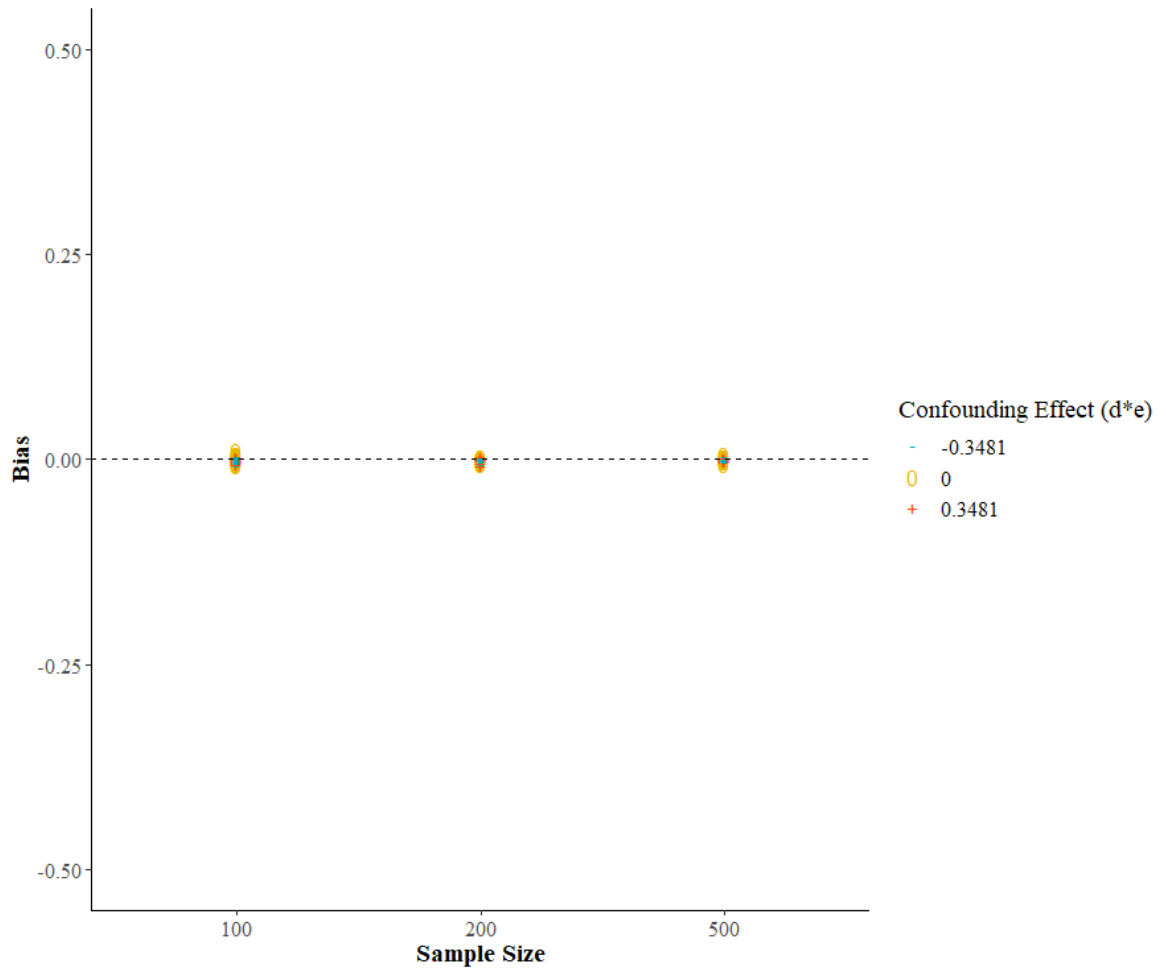
Figure 21

Relative Bias of the Indirect Effect (ab). The Data Were Analyzed With a Single Mediator Model with a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
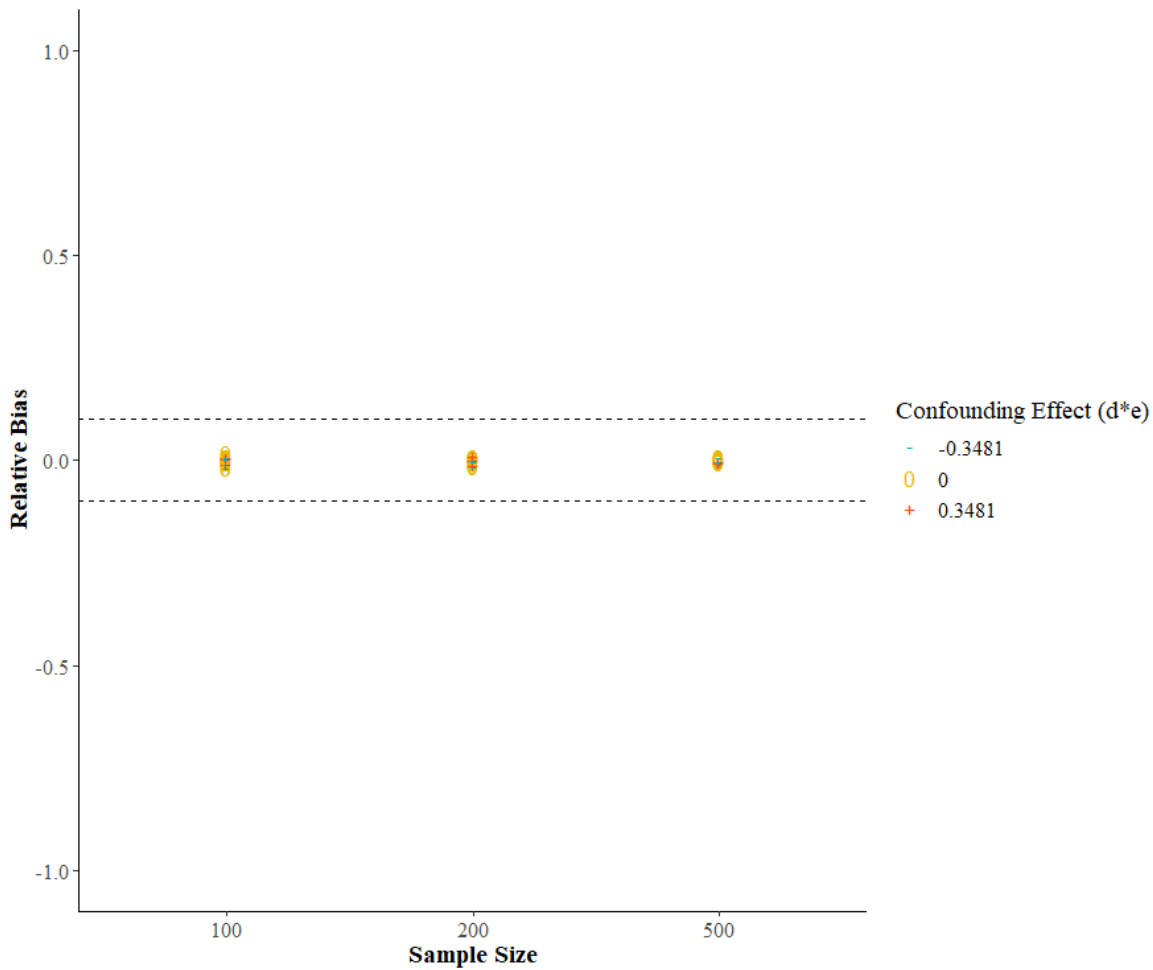
Note. Dashed lines indicate the adequate Relative bias range [-.10, .10].

Empirical Type I Error Rate

Figure 22

Empirical Type I Error Rate for the Indirect Effect (ab). Statistical Significance Was Assessed with a Z Test. The Data Were Analyzed With a Single Mediator Model With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
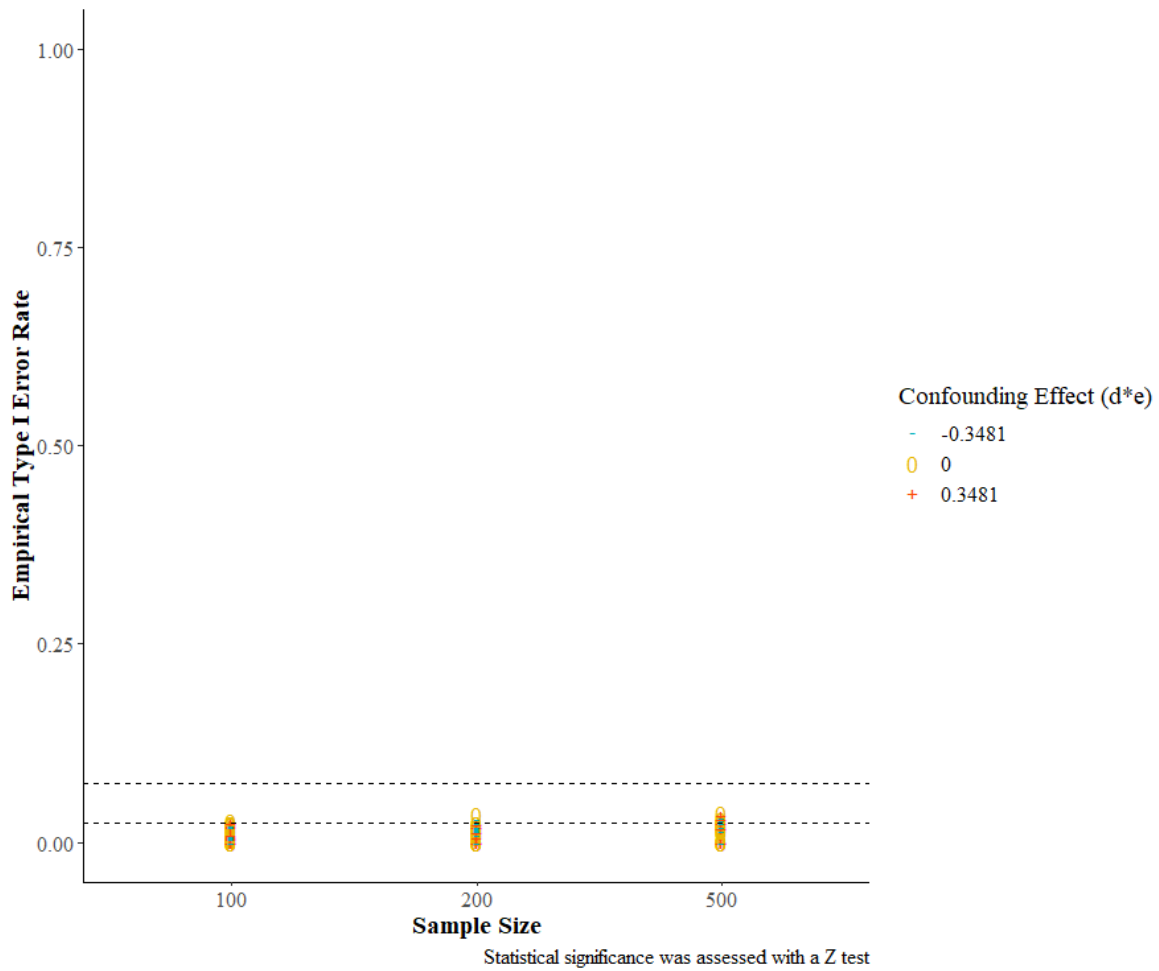
Note. Dashed lines indicate the adequate Empirical Type I Error Rate range [.025, .075].

Figure 23

Empirical Type I Error Rate for the Indirect Effect (ab). Statistical Significance Was Assessed Using Normal Theory Confidence Intervals. The Data Were Analyzed With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
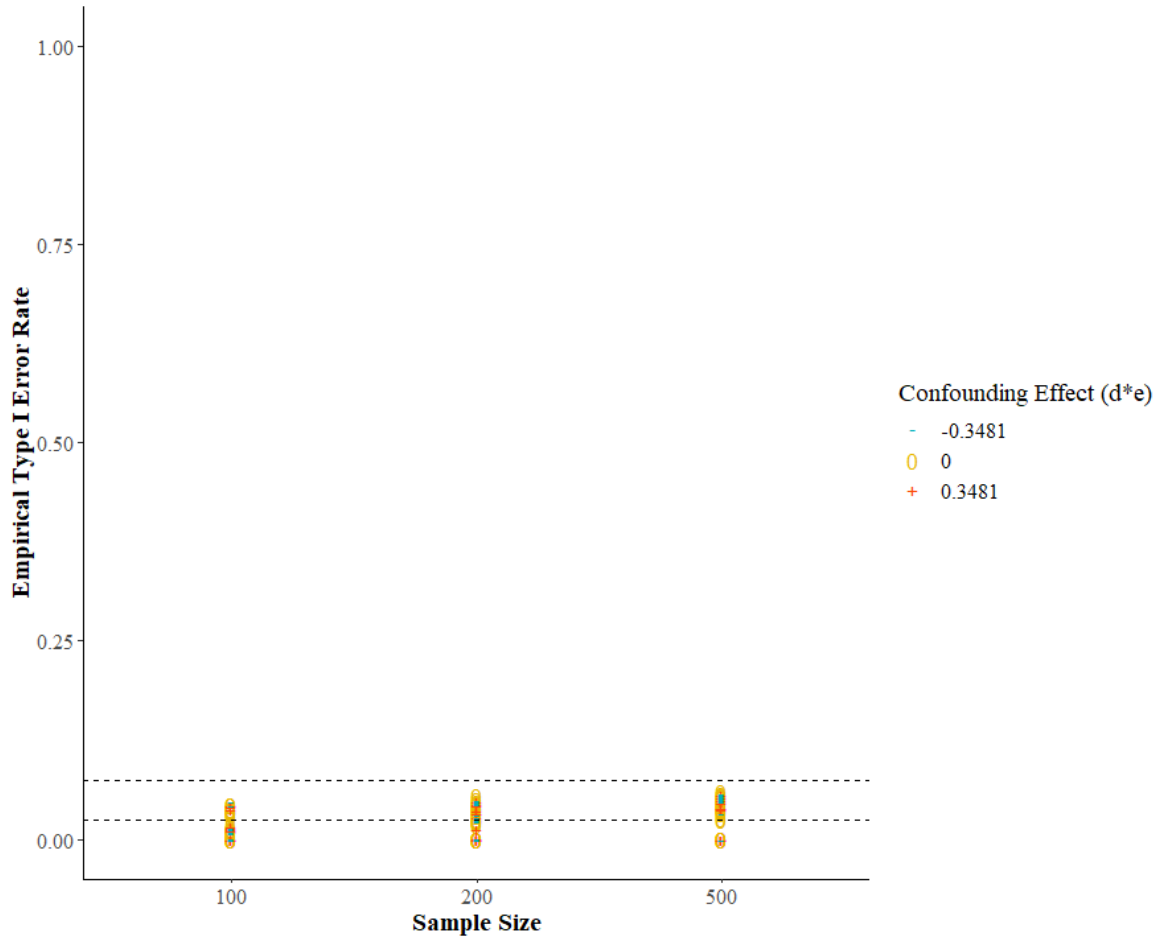
Statistical significance was assessed using Normal Theory Confidence Intervals

Note. Dashed lines indicate the adequate Empirical Type I Error Rate range [.025, .075].

Empirical Power

Figure 24

Empirical Power for the Indirect Effect (ab). Statistical Significance Was Assessed With a Z Test. The Data Were Analyzed With a Single Mediator Model With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
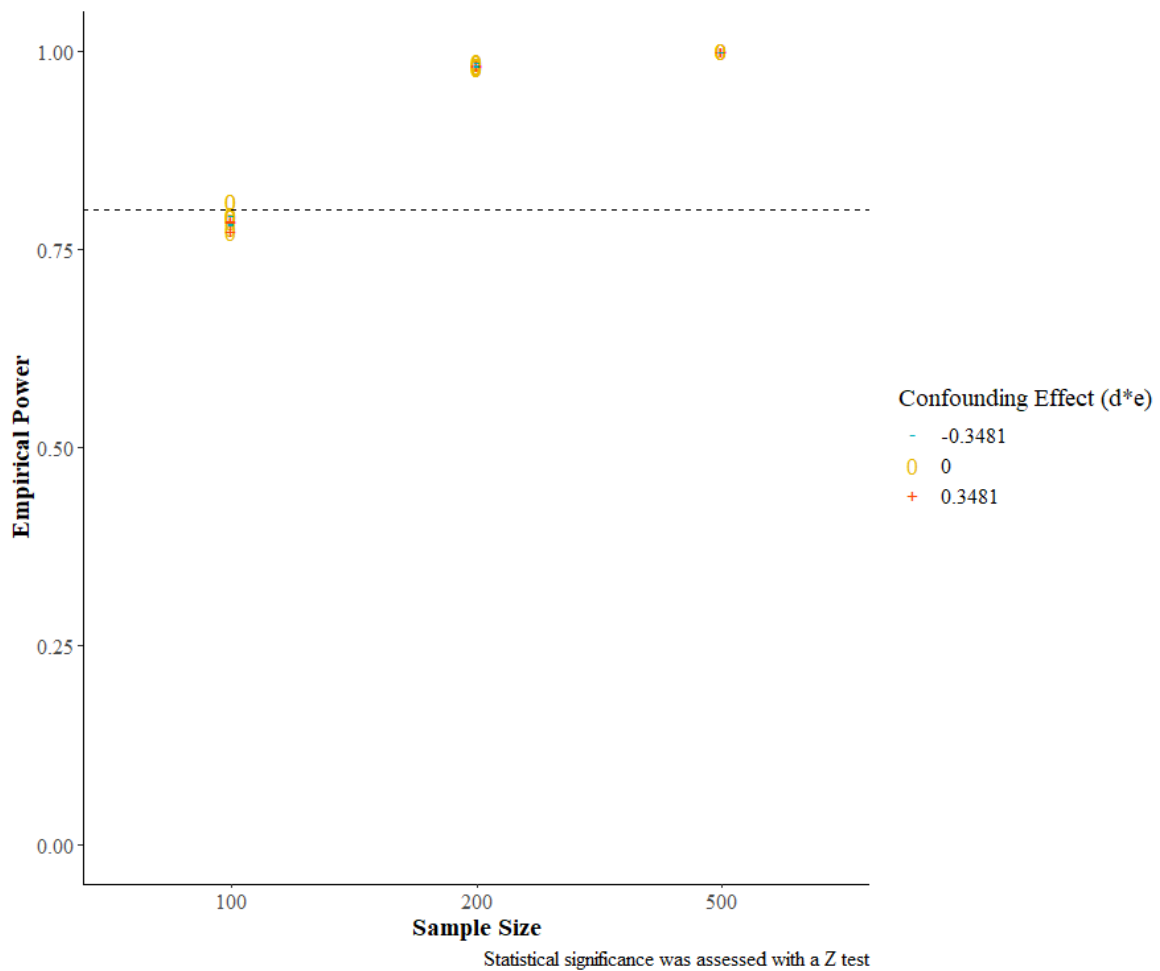
Note. The dashed line represents the conventional standard of 0.80 power.

Figure 25

Empirical Power for the Indirect Effect (ab). Statistical Significance Was Assessed Using Normal Theory Confidence Intervals. The Data Were Analyzed With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)
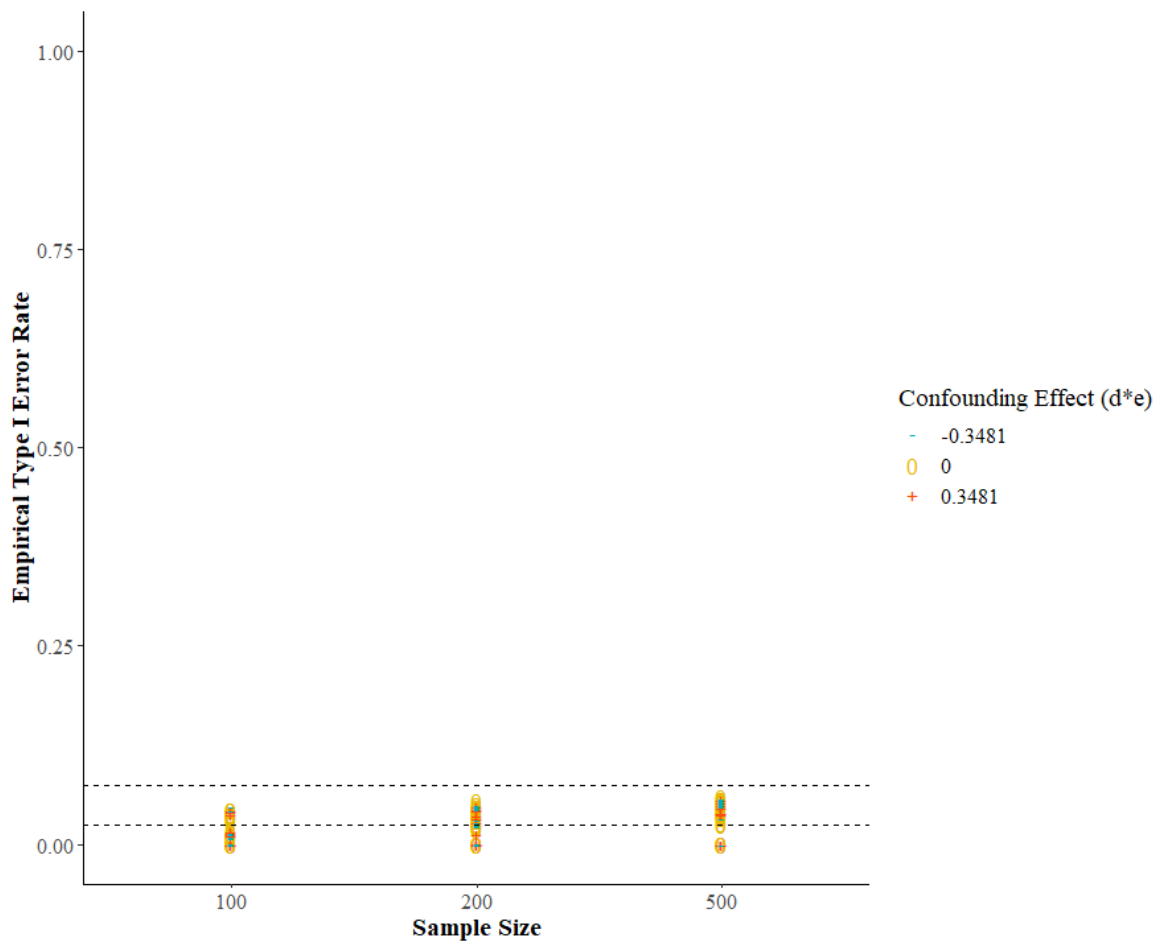
Statistical significance was assessed using Normal Theory Confidence Intervals

Note. The dashed line represents the conventional standard of 0.80 power.

Imbalance

Table 14

Imbalance Results. Data Analyzed with a Single Mediator Model With a Confounder (U) That Influences the Mediator (M) and the Outcome (Y)

							N= 100		N= 200		N= 500	
a	d	cp	b	e	ab	de	True to left	True to right	True to left	True to right	True to left	True to right
0	-0.59	0	0	0	0	0	**0.001***	**0.001***	**0.001***	**0.001***	**0***	**0.002***
0	0	0	0	-0.59	0	0	**0***	**0.002***	**0***	**0.002***	**0***	**0***
0	0	0	0	0	0	0	**0***	**0.001***	**0***	**0***	**0***	**0.001***
0	0	0	0	0.59	0	0	**0***	**0.001***	**0***	**0***	**0.001***	**0***
0	0.59	0	0	0	0	0	**0.002***	**0.001***	**0***	**0.001***	**0.002***	**0.002***
0	-0.59	0.59	0	0	0	0	**0***	**0.001***	**0.002***	**0***	**0***	**0***
0	0	0.59	0	-0.59	0	0	**0***	**0.003***	**0.002***	**0.001***	**0***	**0.001***
0	0	0.59	0	0	0	0	**0.001***	**0***	**0.001***	**0***	**0***	**0***
0	0	0.59	0	0.59	0	0	**0.001***	**0.003***	**0.001***	**0***	**0***	**0.001***
0	0.59	0.59	0	0	0	0	**0.003***	**0.001***	**0.001***	**0.001***	**0.001***	**0.001***
0	-0.59	0	0.59	0	0	0	0.023	0.027	0.026	0.025	0.023	0.029
0	0	0	0.59	-0.59	0	0	0.032	0.024	0.019	0.027	0.021	0.026
0	0	0	0.59	0	0	0	0.037	0.025	0.023	0.023	0.021	0.024
0	0	0	0.59	0.59	0	0	0.026	0.024	0.023	0.027	0.026	0.033
0	0.59	0	0.59	0	0	0	0.032	0.029	0.029	0.029	0.03	0.031
0	-0.59	0.59	0.59	0	0	0	0.019	0.032	0.025	0.031	0.036	0.019
0	0	0.59	0.59	-0.59	0	0	0.029	0.03	0.024	0.022	0.024	0.022
0	0	0.59	0.59	0	0	0	0.033	0.024	**0.038***	0.025	0.023	**0.042***
0	0	0.59	0.59	0.59	0	0	0.024	0.031	0.026	0.028	0.021	0.019
0	0.59	0.59	0.59	0	0	0	0.021	0.021	0.021	0.029	0.027	0.018
0.59	-0.59	0	0	0	0	0	0.019	0.021	0.035	0.019	0.015	0.023
0.59	0	0	0	-0.59	0	0	0.019	0.023	0.022	0.02	0.021	0.026
0.59	0	0	0	0	0	0	0.018	0.019	0.019	0.028	0.027	0.019
0.59	0	0	0	0.59	0	0	0.021	0.029	0.03	0.028	0.023	**0.042***
0.59	0.59	0	0	0	0	0	0.023	0.017	0.023	0.023	0.029	0.029

106

0.59	-0.59	0.59	0	0	0	0	0.023	0.015	0.024	0.019	0.033	0.02
0.59	0	0.59	0	-0.59	0	0	0.015	0.018	0.02	0.028	0.022	0.03
0.59	0	0.59	0	0	0	0	0.018	0.017	0.029	0.031	0.025	0.022
0.59	0	0.59	0	0.59	0	0	0.019	0.022	0.015	0.031	0.026	0.024
0.59	0.59	0.59	0	0	0	0	0.016	0.029	0.031	0.028	0.022	0.03
0.59	-0.59	0	0.59	0	0.3481	0	0.016	0.034	0.016	**0.039***	0.017	0.028
0.59	0	0	0.59	-0.59	0.3481	0	0.021	0.032	0.021	0.028	0.024	0.034
0.59	0	0	0.59	0	0.3481	0	0.018	0.033	0.02	0.032	0.018	0.03
0.59	0	0	0.59	0.59	0.3481	0	0.02	0.036	0.02	0.029	0.022	0.024
0.59	0.59	0	0.59	0	0.3481	0	0.019	0.036	0.014	**0.038***	0.025	0.031
0.59	-0.59	0.59	0.59	0	0.3481	0	0.018	0.031	0.027	0.026	0.02	0.027
0.59	0	0.59	0.59	-0.59	0.3481	0	0.02	0.036	0.016	0.029	0.017	0.02
0.59	0	0.59	0.59	0	0.3481	0	0.018	0.03	0.013	**0.039***	0.028	0.031
0.59	0	0.59	0.59	0.59	0.3481	0	**0.012***	0.031	0.018	0.031	0.017	0.028
0.59	0.59	0.59	0.59	0	0.3481	0	**0.008***	0.034	0.02	0.029	0.03	0.03
0	-0.59	0	0	-0.59	0	0.3481	**0.001***	0*	0.001*	0.003*	0*	0*
0	0.59	0	0	0.59	0	0.3481	**0.002***	0*	0*	0.001*	0*	0.001*
0	-0.59	0.59	0	-0.59	0	0.3481	**0***	0.002*	0.003*	0*	0*	0*
0	0.59	0.59	0	0.59	0	0.3481	**0.004***	0.001*	0*	0.002*	0*	0*
0	-0.59	0	0.59	-0.59	0	0.3481	**0.03**	0.025	0.026	0.022	0.031	0.027
0	0.59	0	0.59	0.59	0	0.3481	**0.024**	0.026	0.016	0.022	0.027	0.022
0	-0.59	0.59	0.59	-0.59	0	0.3481	**0.025**	0.04*	0.023	0.029	0.032	0.02
0	0.59	0.59	0.59	0.59	0	0.3481	**0.027**	0.027	0.027	0.027	0.037	0.022
0.59	-0.59	0	0	-0.59	0	0.3481	**0.015**	0.028	0.02	0.025	0.025	0.024
0.59	0.59	0	0	0.59	0	0.3481	**0.021**	0.022	0.031	0.023	0.021	0.026
0.59	-0.59	0.59	0	-0.59	0	0.3481	**0.021**	0.021	0.027	0.022	0.034	0.023
0.59	0.59	0.59	0	0.59	0	0.3481	**0.022**	0.02	0.021	0.026	0.028	0.029
0.59	-0.59	0	0.59	-0.59	0.3481	0.3481	**0.022**	0.029	0.021	0.032	0.019	0.035
0.59	0.59	0	0.59	0.59	0.3481	0.3481	**0.025**	0.046*	0.024	0.04*	0.023	0.036
0.59	-0.59	0.59	0.59	-0.59	0.3481	0.3481	**0.018**	0.038*	0.025	0.039*	0.02	0.037
0.59	0.59	0.59	0.59	0.59	0.3481	0.3481	**0.023**	0.03	0.028	0.038*	0.028	0.04*
0	-0.59	0	0	0.59	0	-0.3481	**0.002***	0.003*	0.001*	0.003*	0.003*	0.002*
0	0.59	0	0	-0.59	0	-0.3481	**0***	0.001*	0.001*	0*	0.001*	0.001*
0	-0.59	0.59	0	0.59	0	-0.3481	**0***	0.001*	0*	0*	0.001*	0*
0	0.59	0.59	0	-0.59	0	-0.3481	**0***	0*	0.001*	0*	0.002*	0.001*

0	-0.59	0	0.59	0.59	0	-0.3481	**0.026**	0.032	0.018	0.032	0.022	0.027
0	0.59	0	0.59	-0.59	0	-0.3481	**0.026**	0.03	0.031	0.026	0.035	0.029
0	-0.59	0.59	0.59	0.59	0	-0.3481	**0.03**	0.032	0.025	0.032	0.033	0.026
0	0.59	0.59	0.59	-0.59	0	-0.3481	**0.026**	0.036	0.027	0.024	0.025	0.029
0.59	-0.59	0	0	0.59	0	-0.3481	**0.021**	0.018	0.033	0.025	0.024	0.023
0.59	0.59	0	0	-0.59	0	-0.3481	**0.018**	0.023	0.025	0.023	0.032	0.039*
0.59	-0.59	0.59	0	0.59	0	-0.3481	**0.016**	0.022	0.03	0.031	0.028	0.033
0.59	0.59	0.59	0	-0.59	0	-0.3481	**0.026**	0.023	0.035	0.026	0.034	0.028
0.59	-0.59	0	0.59	0.59	0.3481	-0.3481	**0.014**	0.032	0.026	0.043*	0.015	0.038*
0.59	0.59	0	0.59	-0.59	0.3481	-0.3481	**0.017**	0.036	0.02	0.032	0.032	0.027
0.59	-0.59	0.59	0.59	0.59	0.3481	-0.3481	**0.014**	0.041	0.019	0.033	0.014	0.037
0.59	0.59	0.59	0.59	-0.59	0.3481	-0.3481	**0.023**	0.045	0.023	0.03	0.024	0.024

Note. Values marked with * are outside Bradley's (1978) robustness criterion (0.0125,0.0375)

APPENDIX I
RESULTS STUDY 1 (MISSPECIFIED MODEL)

Bias

Figure 26

Bias of the Indirect Effect (ab). The Data Were Analyzed With a Misspecified Model (Single Mediator Model)
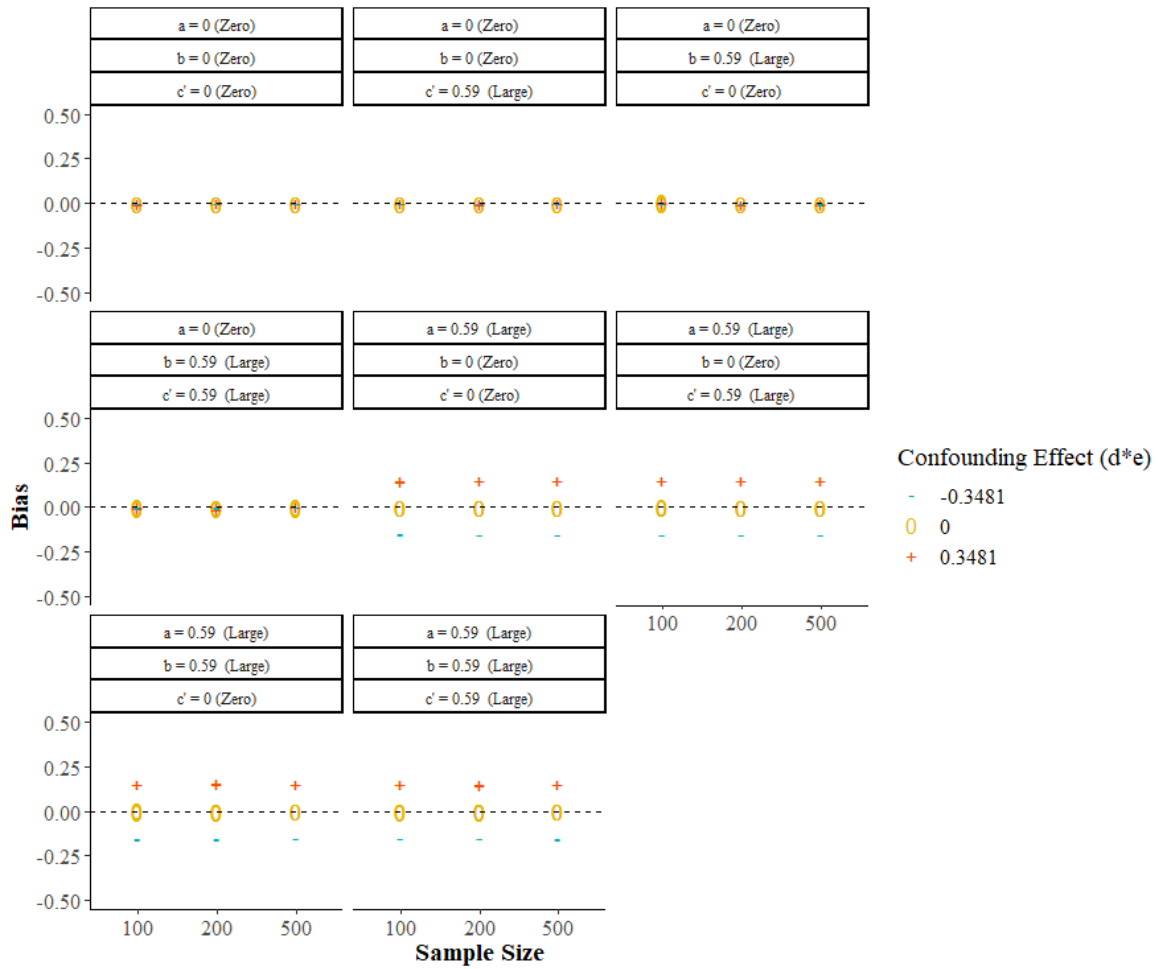
Figure 27

Relative Bias of the Indirect Effect (ab). The Data Were Analyzed with a Misspecified Model (Single Mediator Model)
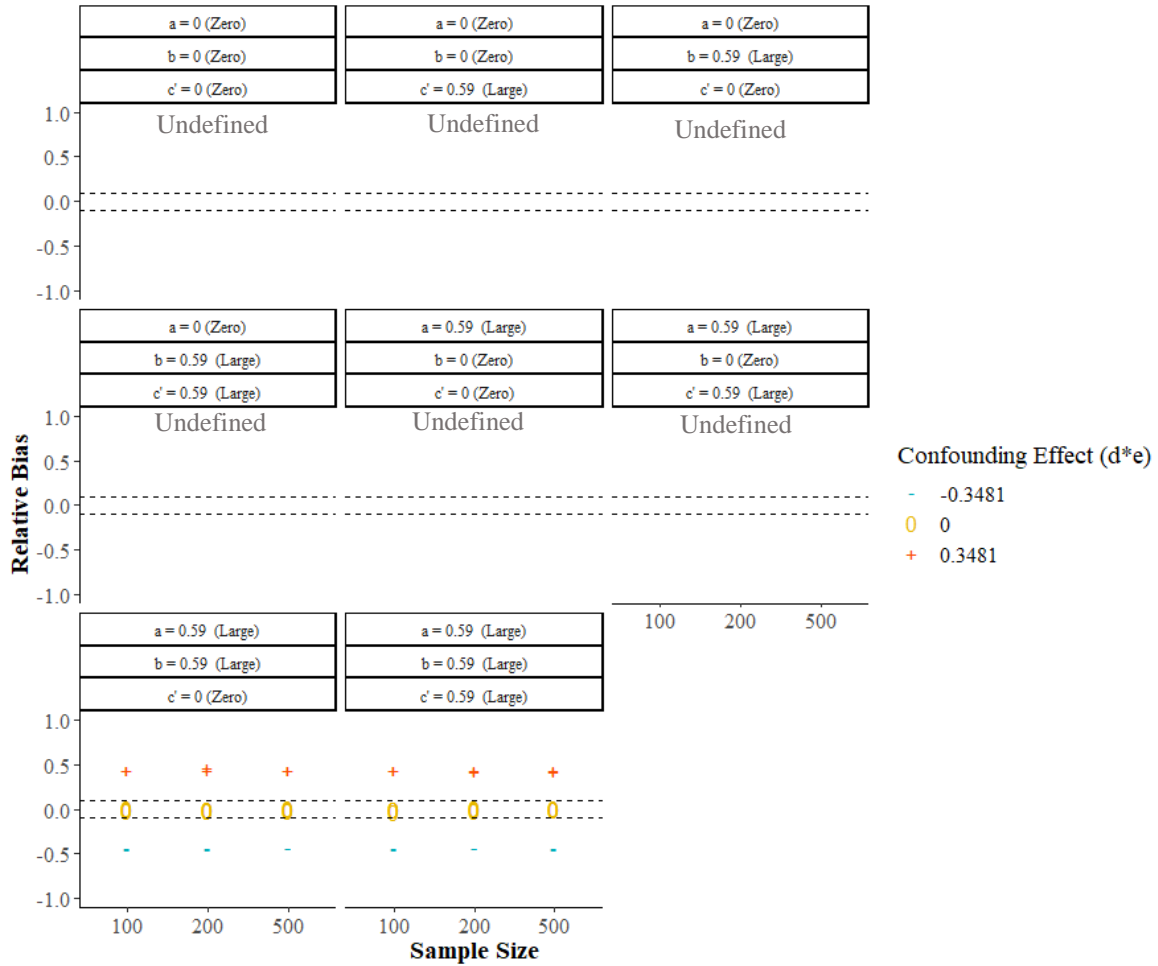
Note. Dashed lines indicate the adequate Relative bias range [-.10, .10]. The Relative bias formula requires to divide by the true indirect effect (ab). The undefined conditions resulted by diving by a true indirect effect equal to zero ($ab = 0$).

Type I Error Rates

Figure 28

Empirical Type I Error Rate for the Indirect Effect (ab). Statistical Significance Was Assessed with a Z Test. The Data Were Analyzed With a Misspecified Model (Single Mediator Model)
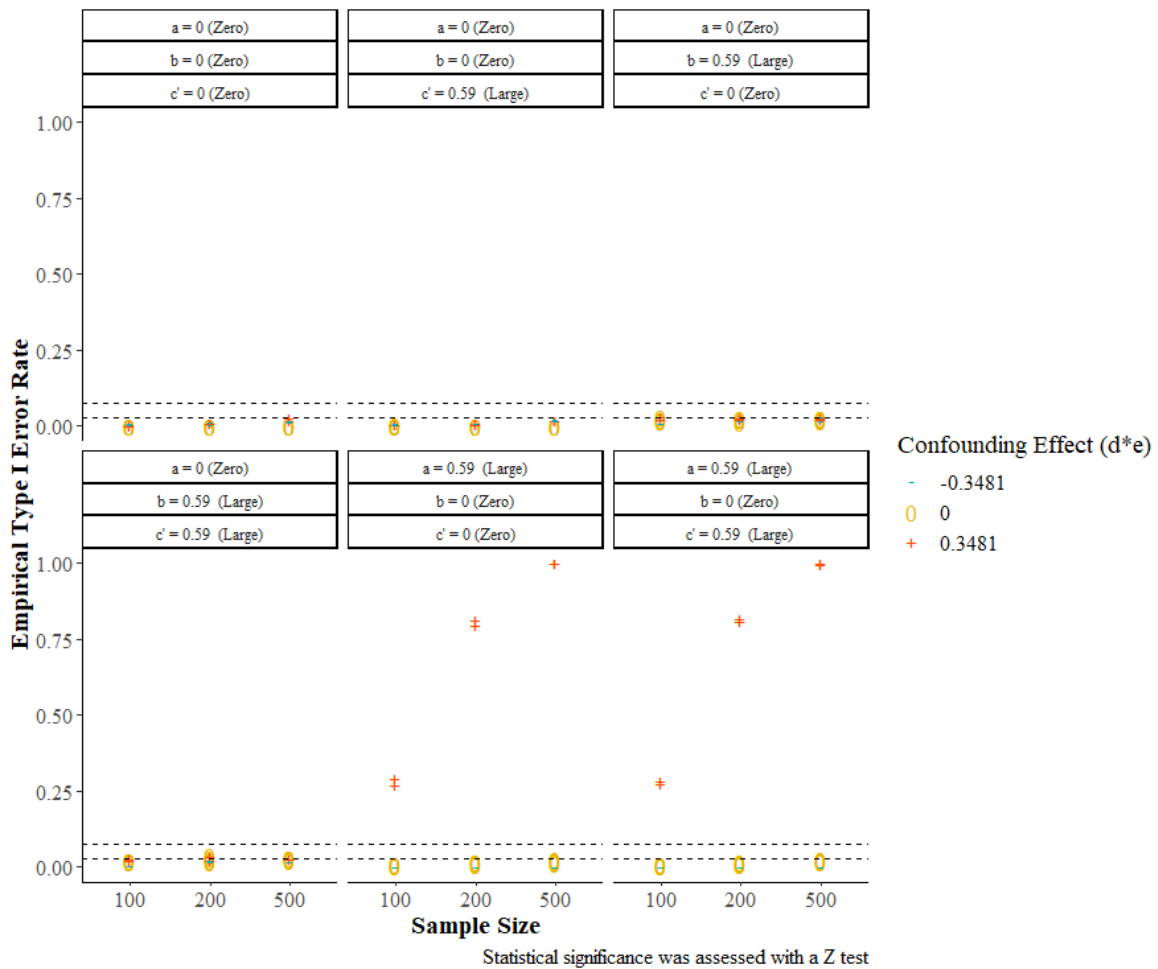
Note. Dashed lines indicate the adequate Empirical Type I Error Rate range [.025, .075].

Figure 29

Empirical Type I Error Rate for the Indirect Effect (ab). Statistical Significance Was Assessed Using Normal Theory Confidence Intervals. The Data Were Analyzed With a Misspecified Model (Single Mediator Model)

Statistical significance was assessed using Normal Theory Confidence Intervals
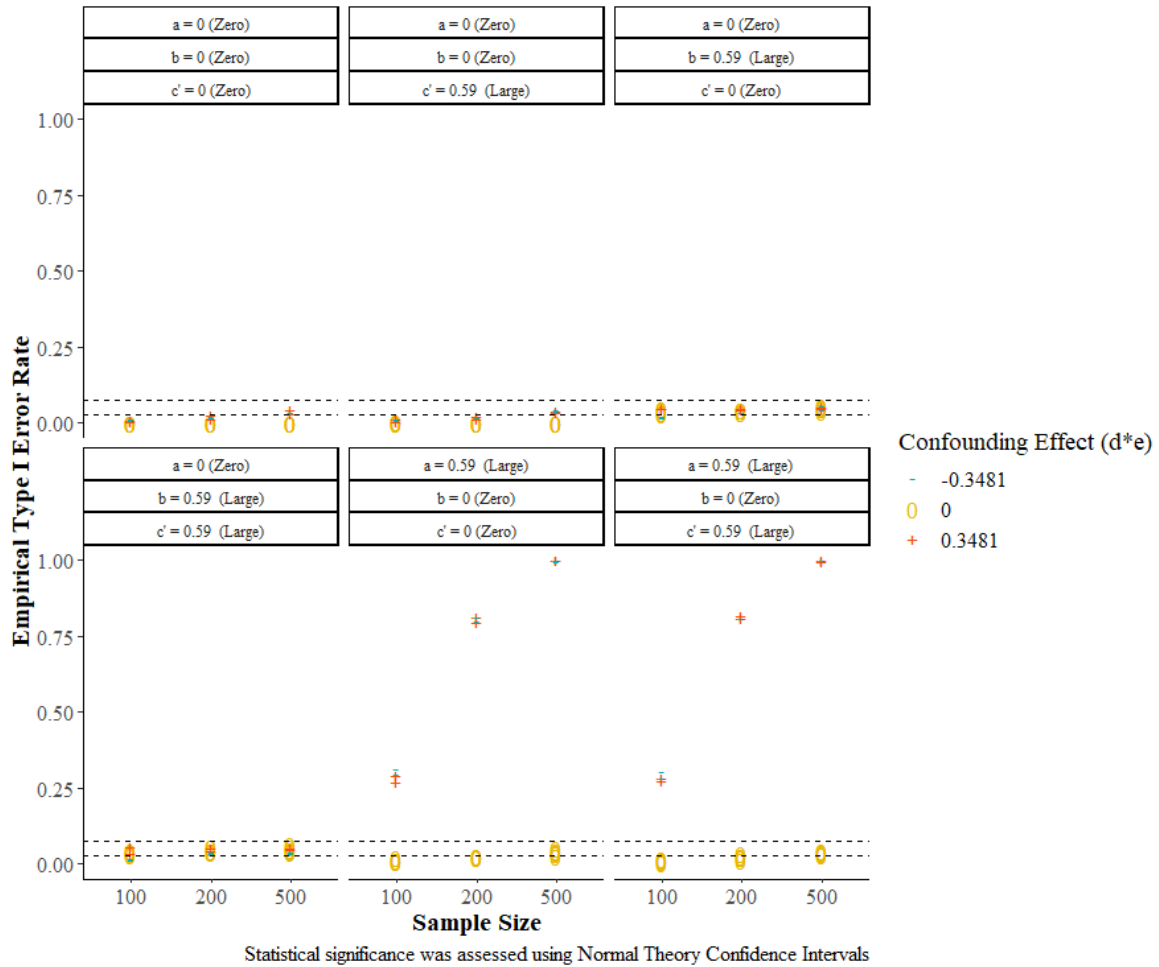
Note. Dashed lines indicate the adequate Empirical Type I Error Rate range [.025, .075].

Empirical Power

Figure 30

Empirical Power for the Indirect Effect (ab). Statistical Significance Was Assessed With a Z Test). The Data Were Analyzed with a Misspecified Model (Single Mediator Model)

Note. The dashed line represents the conventional standard of 0.80 power.

Figure 31

Empirical Power for the Indirect Effect (ab). Statistical Significance Was Assessed Using Normal Theory Confidence Intervals). The Data Were Analyzed With a Misspecified Model (Single Mediator Model)

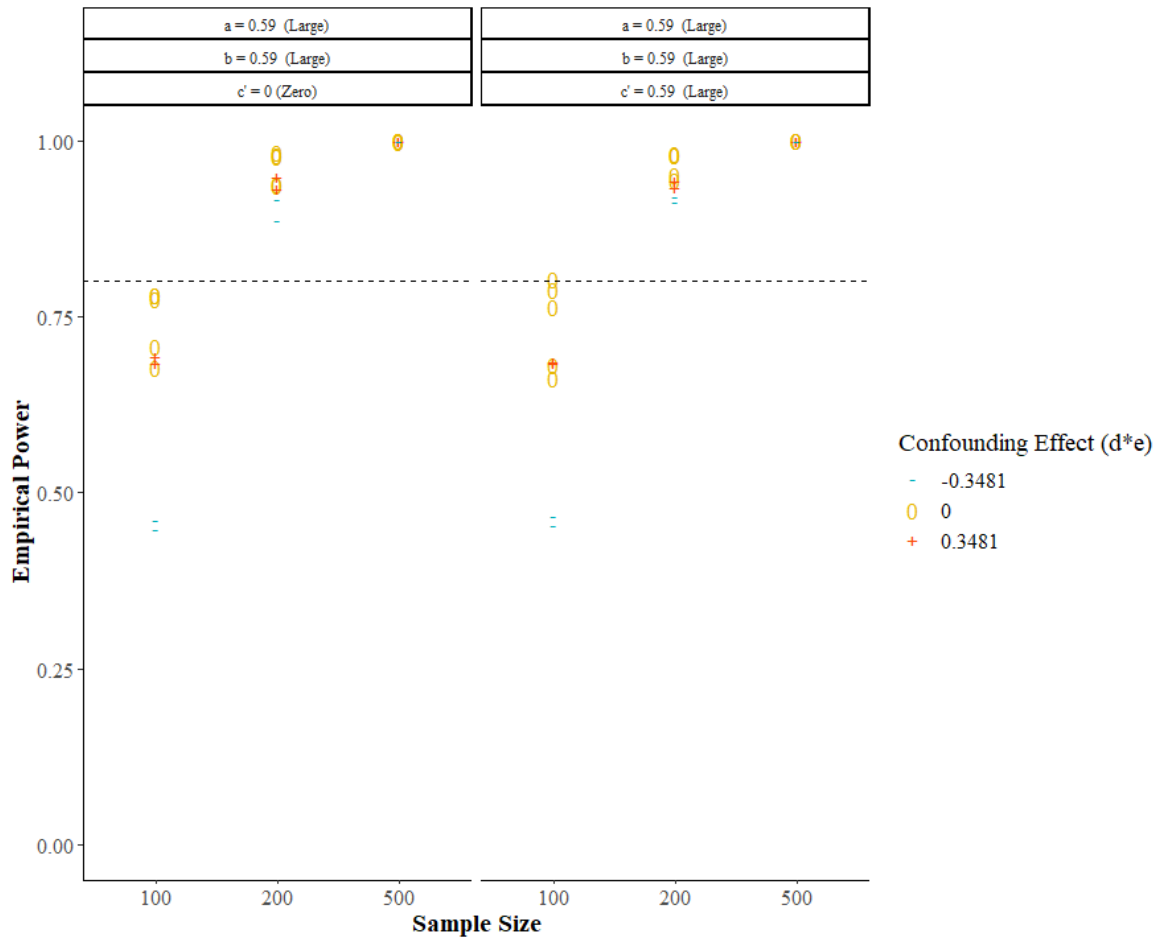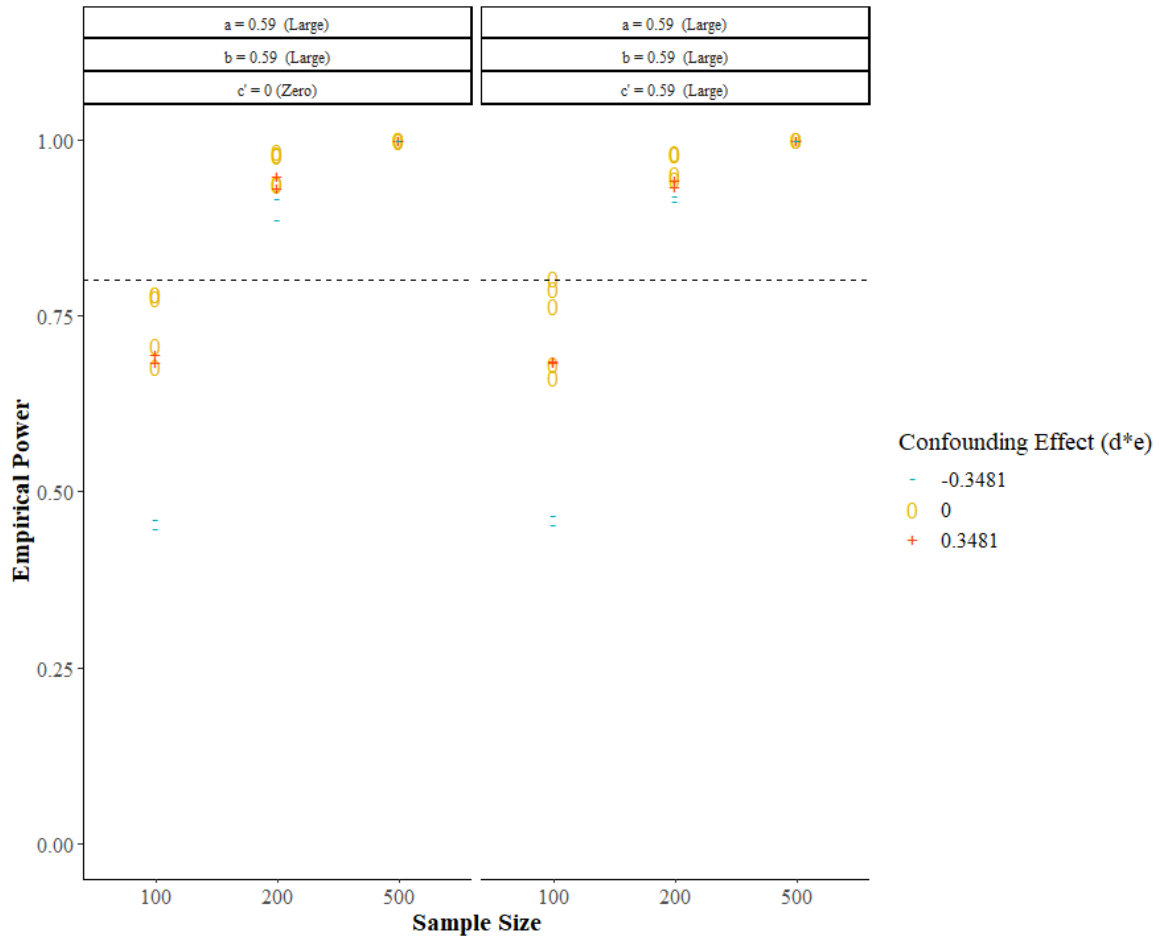*Note.* The dashed line represents the conventional standard of 0.80 power.

Table 15

Imbalance Results. Data Analyzed With a Misspecified Model (Single Mediator Model)

a	d	cp	b	e	ab	de	N= 100		N= 200		N= 500	
							True to left	True to right	True to left	True to right	True to left	True to right
0	-0.59	0	0	0	0	0	0*	0.002*	0*	0*	0*	0.002*
0	0	0	0	-0.59	0	0	0*	0.003*	0*	0.001*	0.001*	0*
0	0	0	0	0	0	0	0*	0.001*	0*	0*	0*	0.002*
0	0	0	0	0.59	0	0	0*	0*	0*	0*	0.001*	0.001*
0	0.59	0	0	0	0	0	0*	0.002*	0.001*	0.003*	0.001*	0.001*
0	-0.59	0.59	0	0	0	0	0.002*	0.002*	0*	0.001*	0*	0*
0	0	0.59	0	-0.59	0	0	0.001*	0.001*	0.001*	0.002*	0.001*	0.001*
0	0	0.59	0	0	0	0	0*	0*	0.001*	0*	0*	0*
0	0	0.59	0	0.59	0	0	0*	0.002*	0.001*	0.002*	0.001*	0.001*
0	0.59	0.59	0	0	0	0	0.001*	0*	0*	0*	0.001*	0.001*
0	-0.59	0	0.59	0	0	0	0.02	0.02	0.026	0.021	0.021	0.034
0	0	0	0.59	-0.59	0	0	0.032	0.022	0.017	0.028	0.022	0.024
0	0	0	0.59	0	0	0	0.034	0.021	0.028	0.024	0.023	0.019
0	0	0	0.59	0.59	0	0	0.026	0.024	0.022	0.024	0.025	0.032
0	0.59	0	0.59	0	0	0	0.023	0.028	0.024	0.026	0.032	0.024
0	-0.59	0.59	0.59	0	0	0	0.02	0.029	0.027	0.03	0.03	0.025
0	0	0.59	0.59	-0.59	0	0	0.026	0.036	0.023	0.022	0.027	0.028
0	0	0.59	0.59	0	0	0	0.035	0.022	0.04*	0.021	0.023	0.049*
0	0	0.59	0.59	0.59	0	0	0.023	0.027	0.027	0.029	0.022	0.017
0	0.59	0.59	0.59	0	0	0	0.025	0.025	0.015	0.028	0.032	0.018
0.59	-0.59	0	0	0	0	0	0.014	0.019	0.029	0.016	0.02	0.025
0.59	0	0	0	-0.59	0	0	0.017	0.017	0.02	0.022	0.017	0.036
0.59	0	0	0	0	0	0	0.019	0.024	0.017	0.026	0.023	0.017
0.59	0	0	0	0.59	0	0	0.024	0.033	0.029	0.029	0.031	0.033
0.59	0.59	0	0	0	0	0	0.02	0.006*	0.022	0.023	0.027	0.021
0.59	-0.59	0.59	0	0	0	0	0.024	0.014	0.021	0.023	0.033	0.018
0.59	0	0.59	0	-0.59	0	0	0.018	0.027	0.021	0.025	0.02	0.028
0.59	0	0.59	0	0	0	0	0.021	0.019	0.031	0.026	0.023	0.023
0.59	0	0.59	0	0.59	0	0	0.019	0.028	0.019	0.028	0.026	0.027
0.59	0.59	0.59	0	0	0	0	0.009*	0.02	0.023	0.029	0.036	0.023

0.59	-0.59	0	0.59	0	0.3481	0	0.015	0.037	0.018	0.032	0.023	0.028
0.59	0	0	0.59	-0.59	0.3481	0	0.018	0.034	0.024	0.033	0.025	**0.038***
0.59	0	0	0.59	0	0.3481	0	0.019	0.035	0.025	0.031	0.018	0.033
0.59	0	0	0.59	0.59	0.3481	0	0.016	0.028	0.02	0.034	0.022	0.021
0.59	0.59	0	0.59	0	0.3481	0	**0.011***	0.034	0.013	**0.045***	0.023	0.034
0.59	-0.59	0.59	0.59	0	0.3481	0	0.014	0.027	0.023	0.03	0.02	0.027
0.59	0	0.59	0.59	-0.59	0.3481	0	0.019	**0.048***	0.019	**0.039**	0.021	0.024
0.59	0	0.59	0.59	0	0.3481	0	0.015	0.031	0.017	0.037	0.032	0.032
0.59	0	0.59	0.59	0.59	0.3481	0	0.014	0.037	0.016	0.025	0.017	0.031
0.59	0.59	0.59	0.59	0	0.3481	0	**0.012***	0.031	0.019	0.022	0.024	0.032
0	-0.59	0	0	-0.59	0	0.3481	0.016	0.021	0.016	0.024	0.027	0.024
0	0.59	0	0	0.59	0	0.3481	0.013	**0.01***	0.023	0.029	0.036	0.023
0	-0.59	0.59	0	-0.59	0	0.3481	0.017	0.013	0.026	0.027	0.021	0.03
0	0.59	0.59	0	0.59	0	0.3481	**0.011***	0.02	0.02	0.029	0.025	0.025
0	-0.59	0	0.59	-0.59	0	0.3481	0.034	0.024	0.026	0.026	0.023	0.028
0	0.59	0	0.59	0.59	0	0.3481	0.023	0.032	0.029	0.025	0.02	0.034
0	-0.59	0.59	0.59	-0.59	0	0.3481	0.032	0.037	0.023	0.026	0.029	0.022
0	0.59	0.59	0.59	0.59	0	0.3481	0.027	0.017	0.035	0.02	0.028	0.028
0.59	-0.59	0	0	-0.59	0	0.3481	**0.483***	**0***	**0.888***	**0***	**1***	**0***
0.59	0.59	0	0	0.59	0	0.3481	**0.46***	**0***	**0.909***	**0***	**1***	**0***
0.59	-0.59	0.59	0	-0.59	0	0.3481	**0.499***	**0***	**0.9***	**0***	**1***	**0***
0.59	0.59	0.59	0	0.59	0	0.3481	**0.494***	**0***	**0.898***	**0***	**1***	**0***
0.59	-0.59	0	0.59	-0.59	0.3481	0.3481	**0.108***	**0.002***	**0.184***	**0.003***	**0.396***	**0***
0.59	0.59	0	0.59	0.59	0.3481	0.3481	**0.108***	**0.004***	**0.198***	**0.001***	**0.385***	**0***
0.59	-0.59	0.59	0.59	-0.59	0.3481	0.3481	**0.115***	**0.004***	**0.185***	**0.003***	**0.379***	**0***
0.59	0.59	0.59	0.59	0.59	0.3481	0.3481	**0.119***	**0.006***	**0.182***	**0.002***	**0.398***	**0***
0	-0.59	0	0	0.59	0	-0.3481	0.027	0.015	0.026	0.023	0.019	0.029
0	0.59	0	0	-0.59	0	-0.3481	0.021	0.02	0.028	0.017	0.028	0.024
0	-0.59	0.59	0	0.59	0	-0.3481	0.015	0.017	0.025	0.016	0.035	0.028
0	0.59	0.59	0	-0.59	0	-0.3481	0.017	0.016	0.025	0.019	0.02	0.024
0	-0.59	0	0.59	0.59	0	-0.3481	0.023	0.03	0.022	0.029	0.02	0.033
0	0.59	0	0.59	-0.59	0	-0.3481	0.021	0.028	0.023	0.028	0.036	0.029
0	-0.59	0.59	0.59	0.59	0	-0.3481	0.018	0.022	0.027	0.026	0.024	0.022
0	0.59	0.59	0.59	-0.59	0	-0.3481	0.017	0.024	**0.039***	0.029	0.029	0.028
0.59	-0.59	0	0	0.59	0	-0.3481	**0***	**0.49***	**0***	**0.9***	**0***	**1***

0.59	0.59	0	0	-0.59	0	-0.3481	**0***	**0.526***	**0***	**0.907***	**0***	**0.999***
0.59	-0.59	0.59	0	0.59	0	-0.3481	**0***	**0.499***	**0***	**0.903***	**0***	**1***
0.59	0.59	0.59	0	-0.59	0	-0.3481	**0***	**0.505***	**0***	**0.896***	**0***	**1***
0.59	-0.59	0	0.59	0.59	0.3481	-0.3481	**0***	**0.321***	**0***	**0.569***	**0***	**0.874***
0.59	0.59	0	0.59	-0.59	0.3481	-0.3481	**0***	**0.336***	**0***	**0.544***	**0***	**0.861***
0.59	-0.59	0.59	0.59	0.59	0.3481	-0.3481	**0.001***	**0.343***	**0***	**0.529***	**0***	**0.896***
0.59	0.59	0.59	0.59	-0.59	0.3481	-0.3481	**0***	**0.307***	**0***	**0.522***	**0***	**0.887***

Note. Values marked with * are outside Bradley's (1978) robustness criterion (0.0125,0.0375)

APPENDIX J
SAS PROGRAM TO ESTIMATE THE CORRELATION BETWEEN RESIDUALS
METHOD

SAS PROGRAM TO ESTIMATE THE CORRELATION BETWEEN RESIDUALS METHOD

```sas
LIBNAME SIMDATA "C:\Users\Documents\Study2_v2\N = 100, A = 0.59, B =
0.59, D = 0.59, E = 0.59_files_archivos\1.Simdata";
LIBNAME DATAOUT "C:\Users\Documents\Study2_v2\N = 100, A = 0.59, B =
0.59, D = 0.59, E = 0.59_files_archivos\2.Simdatacor";
LIBNAME DATAOUT1 "C:\Users\Documents\Study2_v2\N = 100, A = 0.59, B =
0.59, D = 0.59, E = 0.59_files_archivos\3.Corr";
TITLE 'SIMULATION SINGLE MEDIATOR MODEL WITH A CONFOUNDER CAUSING M AND
Y. X-BINARY';
OPTIONS PS=59 LS=80 REPLACE NONOTES;
FILENAME NULLOG DUMMY 'C:\NULL';
PROC PRINTTO LOG=NULL;
RUN;

%MACRO SIMULATE(NREP,NOBS,B1,B2,B3,B4,B5,FILE,TYPE,COND);

%DO I=1 %TO &NREP;
*GENERATE*;
DATA SIM&I&FILE;
DO I=1 TO &NOBS;
X = RAND('BERNOULLI', .5);
U = RAND('NORMAL', 0, 1);
M = &B1*X + &B2*U + RAND('NORMAL', 0, 1);
Y = &B3*X + &B4*M + &B5*U + RAND('NORMAL', 0, 1);
B1 = &B1;
B2 = &B2;
B3 = &B3;
B4 = &B4;
B5 = &B5;
REP = &I;
COND = &COND;
N = &NOBS;
OUTPUT;
END;
RUN;

DATA SIMDATA.&FILE.&I; SET SIM&I&FILE;
RUN;

%END;
%MEND;

/*POPULATION PARAMETERS*/
/*100*/
/*
B1 = a;
B2 = d;
B3 = cp;
B4 = b;
```

```
B5 = e;
*/

/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.14,  B2=0,      B3=0, B4=0.14,
B5= 0,   FILE = COND01, TYPE = 'BCCC', COND = '1');*/
/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.14,  B2=0.14,    B3=0,
B4=0.14,   B5= 0.14,  FILE = COND02, TYPE = 'BCCC', COND = '2');*/
/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.14,  B2=0.39,    B3=0,
B4=0.14,   B5= 0.39,  FILE = COND03, TYPE = 'BCCC', COND = '3');*/
/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.14,  B2=0.59,    B3=0,
B4=0.14,   B5= 0.59,  FILE = COND04, TYPE = 'BCCC', COND = '4');*/

/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.59,  B2=0,        B3=0,
B4=0.59,   B5= 0,      FILE = COND01, TYPE = 'BCCC', COND = '1');*/
/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.59,  B2=0.14,    B3=0,
B4=0.59,   B5= 0.14,  FILE = COND02, TYPE = 'BCCC', COND = '2');*/
/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.59,  B2=0.39,    B3=0,
B4=0.59,   B5= 0.39,  FILE = COND03, TYPE = 'BCCC', COND = '3');*/
/*%SIMULATE(NREP=1000, NOBS=100,    B1=0.59,  B2=0.59,    B3=0,
B4=0.59,   B5= 0.59,  FILE = COND04, TYPE = 'BCCC', COND = '4');*/

%MACRO CORRELATED(NREP, NOBS, FILE, COND);
%DO X=1 %TO &NREP;

DATA CORRES; SET _NULL_;

*This part saves the estimates a, b and cp for each replication. Also
it tests whether the indirect effect is statistically significant;

*************************************
****MODEL 1(X = M)
*************************************;
*Estimating the (M = X) regression;
PROC REG DATA = SIMDATA.&FILE.&X OUTEST=FILE TABLEOUT NOPRINT COVOUT;
Model1 : MODEL M = X/ stb;

*Saving the coefficient a;
DATA D; SET FILE;
IF _TYPE_='PARMS';
Asample=X; MSE1sample = _RMSE_*_RMSE_;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M;
KEEP Asample MSE1sample;


*Saving the standard error of the coefficients a;
DATA D1; SET FILE;
IF _TYPE_='STDERR';
SEAsample=X;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M;
KEEP SEAsample;

*************************************
****MODEL 2(Y = X M)
*************************************;
```

```
*Estimating the(Y  = X M) regression;

PROC REG DATA=SIMDATA.&FILE.&X OUTEST=FILE TABLEOUT NOPRINT COVOUT;
Model2 : MODEL Y = X M/ stb;

*Saving the coefficients b and c;
DATA E; SET FILE;
IF _TYPE_='PARMS';
CPsample=X; Bsample=M; MSE2sample = _RMSE_*_RMSE_;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M Y;
KEEP CPsample Bsample MSE2sample;


*Saving the standard error of the coefficients b and cp;
DATA E1; SET FILE;
IF _TYPE_='STDERR';
SECPsample=X; SEBsample=M;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M Y;
KEEP SECPsample SEBsample;

*Merging data;
DATA MODEL; MERGE D D1 E E1;

*This part performs the sesntivivity analysis varying the correlation
between residual variances;
%Do I = %SysEvalF(-3 * 10) %To %SysEvalF(3 * 10) ;
%Put I = %sysEvalF(&I / 10) ;
PROC CALIS DATA = SIMDATA.&FILE.&X METHOD=ML COV OUTEST=OUT
PLOTS=PATHDIAGRAM NOPRINT;

LINEQS
M = a X + D1,
Y = c X +  b M + D2;

STD
D1,
D2;

COV
D1 D2 = %sysEvalF(&I / 10);

DATA B; SET OUT;
IF _TYPE_='PARMS';
D1=_Parm1;
D2=_Parm2;
AB=A*B;
KEEP A B AB D1 D2;

DATA SE; SET OUT;
IF _TYPE_='STDERR';
SEA=A; SEB=B;
KEEP SEA SEB;

DATA COV;
```

```
COVE1E2= %sysEvalF(&I / 10);
KEEP COVE1E2;

DATA ALL; MERGE MODEL B SE COV;
CORRE1E2 = COVE1E2/ (SQRT(D1)*SQRT(D2));

*Test for significance a*b;
ABsample = Asample * Bsample;
SOBEL =
SQRT(Asample*Asample*SEBsample*SEBsample+Bsample*Bsample*SEAsample*SEAs
ample);
ZSOBEL = ABsample/SOBEL;
ABSZSOBEL = ABS(ZSOBEL);

IF ABSZSOBEL > 1.96 THEN DECZSOBEL = 1;
ELSE DECZSOBEL = 0;

RAB = abs (AB);
COND = &COND;
N = &NOBS;

DATA NEW;SET CORRES;
DATA CORRES; SET NEW ALL;

%END;
DATA DATAOUT.&FILE.&X; SET CORRES;
RUN;
%END;
%MEND;
/*NEP = 1000, NOBS AND CONDITION*/
/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND01, COND = '1');*/
/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND02, COND = '2');*/
/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND03, COND = '3');*/
/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND04, COND = '4');*/

/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND01, COND = '1');*/
/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND02, COND = '2');*/
/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND03, COND = '3');*/
/*%CORRELATED (NREP = 1000, NOBS = 100, FILE = COND04, COND = '4');*/


%MACRO CORR(NREP, FILE, COND);
%DO X=1 %TO &NREP;

DATA SUMMARY2; SET _NULL_;

PROC SORT DATA=DATAOUT.&FILE.&X;
   BY RAB;
RUN;

DATA CORRES1;
SET DATAOUT.&FILE.&X;
IF _N_ = 1;
COND = &COND;
```

```
REP = &X;
RUN;

DATA NEW; SET CORRES1;
DATA  SUMMARY2; SET SUMMARY2 NEW;

DATA DATAOUT1.&FILE.&X; SET SUMMARY2;
RUN;
%END;

%MEND;
/*NEP = 1000, CONDITION*/
/*100*/
/*%CORR(NREP=1000, FILE = COND01, COND = '1');*/
/*%CORR(NREP=1000, FILE = COND02, COND = '2');*/
/*%CORR(NREP=1000, FILE = COND03, COND = '3');*/
/*%CORR(NREP=1000, FILE = COND04, COND = '4');*/


/*%CORR(NREP=1000, FILE = COND01, COND = '1');*/
/*%CORR(NREP=1000, FILE = COND02, COND = '2');*/
/*%CORR(NREP=1000, FILE = COND03, COND = '3');*/
/*%CORR(NREP=1000, FILE = COND04, COND = '4');*/

PROC SQL;
SELECT catt('DATAOUT1.',MEMNAME) INTO : MEMNAMES separated by '  '
FROM DICTIONARY.TABLES
WHERE LIBNAME='DATAOUT1';
QUIT;

DATA WANT;
SET &MEMNAMES.;
/*t*/
ALPHA = 0.05;
P = 1 - ALPHA/2;
T = CORRE1E2*SQRT(N-2)/SQRT(1-CORRE1E2**2);
TABS = ABS (T);
TCRITICAL = TINV (P, N-2);
IF TABS > TCRITICAL THEN DECISIONT = 1;
ELSE DECISIONT = 0;
/*Condifence  intervals*/
ZP = ARTANH(CORRE1E2);
/*Zp1 = .5 *(log(1+CORRE1E2) - log(1-CORRE1E2));*/
SEp = 1 / SQRT (N-3);
LCLz = zp - 1.96 *  SEp;
UCLz = zp + 1.96 *  SEp;

LCLr = TANH (LCLz);
UCLr = TANH (UCLz);

IF LCLr < 0 < UCLr THEN DECISIONCI = 0;
ELSE DECISIONCI = 1;

ABSCORR = ABS(CORRE1E2);
```

124

```
IF ABSCORR  > 0.05 THEN EFFR05 = 1;
ELSE EFFR05 = 0;
IF ABSCORR  > 0.1 THEN EFFR1 = 1;
ELSE EFFR1 = 0;
IF ABSCORR  > 0.3 THEN EFFR3 = 1;
ELSE EFFR3 = 0;
IF ABSCORR  > 0.5 THEN EFFR5 = 1;
ELSE EFFR5 = 0;

RUN;

PROC MEANS DATA = WANT;
RUN;

PROC MEANS DATA = WANT;
CLASS DECZSOBEL;
RUN;

PROC SGPLOT DATA = WANT;
TITLE "Distribution of the correlation between residuals that make the
indirect effect 0";
HISTOGRAM CORRE1E2 / GROUP = DECZSOBEL;
RUN;
TITLE;
```

APPENDIX K
SAS PROGRAM TO ESTIMATE THE L.O.V.E METHOD

```
LIBNAME SIMDATA "C:\Users\Diana Alvarez Bartolo\Documents\Study2_v2\N =
100, A = 0.59, B = 0.59, D = 0.59, E = 0.59_files_archivos\1.Simdata";
LIBNAME DATAOUT "C:\Users\Diana Alvarez Bartolo\Documents\Study2_v2\N =
100, A = 0.59, B = 0.59, D = 0.59, E = 0.59_files_archivos\4.love";
TITLE 'SIMULATION L.O.V.E';
OPTIONS PS=59 LS=80 REPLACE NONOTES;
FILENAME NULLOG DUMMY 'C:\NULL';
PROC PRINTTO LOG=NULL;
RUN;

%MACRO CORR(NREP, NOBS, FILE, COND);
%DO X=1 %TO &NREP;
DATA LOVE; SET _NULL_;

*This part saves the estimates a, b and cp for each replication. Also
it tests whether the indirect effect is statistically significant;
*************************************
****MODEL 1(X = M)
*************************************;
*Estimating the (M = X) regression;
PROC REG DATA = SIMDATA.&FILE.&X OUTTEST=FILE TABLEOUT NOPRINT COVOUT;
Model1 : MODEL M = X/ stb;

*Saving the coefficient a;
DATA D; SET FILE;
IF _TYPE_='PARMS';
Asample=X; MSE1sample = _RMSE_*_RMSE_;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M;
KEEP Asample MSE1sample;

*Saving the standard error of the coefficients a;
DATA D1; SET FILE;
IF _TYPE_='STDERR';
SEAsample=X;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M;
KEEP SEAsample;

*************************************
****MODEL 2(Y = X M)
*************************************;
*Estimating the(Y  = X M) regression;

PROC REG DATA=SIMDATA.&FILE.&X OUTTEST=FILE TABLEOUT NOPRINT COVOUT;
Model2 : MODEL Y = X M/ stb;

*Saving the coefficients b and c;
DATA E; SET FILE;
IF _TYPE_='PARMS';
CPsample=X; Bsample=M; MSE2sample = _RMSE_*_RMSE_;
```

127

```
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M Y;
KEEP CPsample Bsample MSE2sample;

*Saving the standard error of the coefficients b and cp;
DATA E1; SET FILE;
IF _TYPE_='STDERR';
SECPsample=X; SEBsample=M;
DROP _MODEL_ _TYPE_ _NAME_ _DEPVAR_ _RMSE_ Intercept X M Y;
KEEP SECPsample SEBsample;

*Merging data;
DATA MODEL; MERGE D D1 E E1;

PROC CORR COV DATA=SIMDATA.&FILE.&X NOPRINT OUTP=COV;
VAR Y X M U;
RUN;

DATA CORR1; SET COV;
IF _NAME_ = 'X' and _TYPE_ = 'CORR'; RYX  = Y;
KEEP RYX;

DATA CORR2; SET COV;
IF _NAME_ = 'M' and _TYPE_ = 'CORR'; RYM  = Y;
KEEP RYM;

DATA CORR3; SET COV;
IF _NAME_ = 'X' and _TYPE_ = 'CORR'; RMX  = M;
KEEP RMX;

DATA R;
MERGE CORR1 CORR2 CORR3;

DATA COMPLETE;
MERGE MODEL R;
RUN;

DATA LOVE; SET COMPLETE;
DO CORUM=0 TO 1 BY .01;
DO CORUY=0 TO 1 BY .01;
DO CORUX=0;
/* True values of a,b,cpr that account for the influence of a potential
confounder U*/
CPR =(RYX*(1-CORUM**2)+ RYM*(CORUX*CORUM-RMX)+ CORUY*(CORUX*CORUM-
CORUX))/(1+2*(RMX*CORUM*CORUX)- CORUX**2-CORUM**2-RMX**2);
B =(RYM*(1-CORUX**2)+ RYX*(CORUM*CORUX-RMX)+ CORUY*(RMX*CORUX-
CORUM))/(1+2*(RMX*CORUX*CORUM)- CORUM**2-CORUX**2-RMX**2);
A =(RMX-CORUM*CORUX)/(1-CORUX**2);

/*Observed values of a,b, cpr*/
CPRBIASED=(RYX-RYM*RMX)/(1-RMX**2);
BBIASED=(RYM-RYX*RMX)/(1-RMX**2);
ABIASED=RMX;

/*Obtaining the bias of each coefficient, a is unbiased because RUX=0*/
```

```
BIASCPR=CPRBIASED-CPR;
BIASB=BBIASED-B;
BIASA=ABIASED-A;

*The following code computes standardized true ab, standardized biased
ab, and the bias of the indirect effect;
TRUEAB=A*B;
BIASEDAB=ABIASED*BBIASED;
BIASAB=BIASEDAB-TRUEAB;
RTRUEAB = round(TRUEAB,.01);

ABsample = Asample * Bsample;
SOBEL =
SQRT(Asample*Asample*SEBsample*SEBsample+Bsample*Bsample*SEAsample*SEAs
ample);
ZSOBEL = ABsample/SOBEL;
ABSZSOBEL = ABS(ZSOBEL);

IF ABSZSOBEL > 1.96 THEN DECZSOBEL = 1;
ELSE DECZSOBEL = 0;

OUTPUT;
END;
END;
END;
RUN;

DATA FORZERO; SET LOVE;
IF RTRUEAB=0;

COND = &COND;
N = &NOBS;
REP = &NREP;

*VARIANCE OF THE CORRELATIONS;
VRMU = ((1-CORUM*CORUM)*(1-CORUM*CORUM))/ N;
VRMY = ((1-RYM*RYM)*(1-RYM*RYM))/N;
VRUY = ((1-CORUY*CORUY)*(1-CORUY*CORUY))/ N;

*COVARIANCES AMONG THE CORRELATIONS;
CRMYRMU =(.5*(2*CORUY-CORUM*RYM)*(1-CORUY*CORUY-CORUM*CORUM-RYM*RYM) +
CORUY*CORUY*CORUY)/N;
CRMYRUY =(.5*(2*CORUM-RYM*CORUY)*(1-CORUM*CORUM-RYM*RYM-CORUY*CORUY) +
CORUM*CORUM*CORUM)/ N;

CRUYRMU = (.5*(2*RYM - CORUM*CORUY)*(1-CORUM*CORUM-RYM*RYM-CORUY*CORUY)
+ RYM*RYM*RYM)/N;
CORR = CORUM * CORUY;
CORVAR = ((CORUM**2)*VRUY)+(2*CRUYRMU*CORUM*CORUY)+ ((CORUY**2)*VRMU);
CORSE = SQRT(CORVAR);
CORZ = CORR/CORSE;
CORZABS = ABS (CORR/CORSE);

IF CORZABS > 1.96 THEN DECISIONZ = 1;
```

```
ELSE DECISIONZ = 0 ;

IF CORR > 0.05 THEN EFFR05 = 1;
ELSE EFFR05 = 0;

IF CORR > 0.1 THEN EFFR1 = 1;
ELSE EFFR1 = 0;

IF CORR > 0.3 THEN EFFR3 = 1;
ELSE EFFR3 = 0;

IF CORR > 0.5 THEN EFFR5 = 1;
ELSE EFFR5 = 0;

DATA ROW; SET FORZERO;
WHERE CORUM = CORUY;
RUN;

DATA ROW1; SET ROW;
IF _N_ = 1;
RUN;

DATA NEW; SET ROW1;
DATA ROW1; SET ROW1;

DATA DATAOUT.&FILE.&X; SET ROW1;
RUN;
%END;

%MEND;
%CORR(NREP = 1000, NOBS = 100, FILE = COND04, COND = '4');

PROC SQL;
SELECT catt('DATAOUT.',MEMNAME) INTO : MEMNAMES separated by ' '
FROM DICTIONARY.TABLES
WHERE LIBNAME='DATAOUT';
QUIT;

DATA WANT;
SET &MEMNAMES.;
RUN;

PROC MEANS DATA = WANT;
RUN;

PROC MEANS DATA = WANT;
CLASS DECZSOBEL;
RUN;

PROC SGPLOT DATA = WANT;
TITLE "Distribution of the product of correlations (rum*ruy)";
HISTOGRAM CORR / GROUP = DECZSOBEL;
RUN;
TITLE;
```