MEDIATION ANALYSIS USING SATURATED AND RESTRICTED MODELS

Valerie Polad

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology and Neuroscience (Quantitative Psychology).

Chapel Hill
2025

Approved by:

Daniel J. Bauer

Kenneth A. Bollen

Patrick J. Curran

David M. Thissen

ABSTRACT

Valerie Polad: Mediation Analysis Using Saturated and Restricted Models
(Under the direction of Daniel Bauer)

Mediation analysis is widely used to evaluate causal mechanisms, yet the impact of saturated versus restricted model specifications remains understudied. Saturated models freely estimate all possible paths, while restricted models impose theoretically driven parameter constraints. Despite the flexibility of both regression and Structural Equation Modeling (SEM) to incorporate restrictions, applied researchers often default to saturated models—even when theoretical hypotheses do not justify the estimation of all paths. This approach may serve as a safeguard against bias due to model misspecification but also introduces potential drawbacks, including reduced statistical efficiency and lower statistical power for detecting specific indirect effects. This study examines the implications of model specification choices for a common serial mediation structure through a Monte Carlo simulation, assessing parameter estimation, standard errors, confidence intervals, and statistical power across varying sample sizes and effect size magnitudes. Findings suggest that model specification meaningfully impacts the power to detect indirect effects, with restricted models having greater power than overparameterized models. Improperly restricted models may, however, produce biased estimates of indirect effects. Theoretical and practical considerations for aligning model specification with hypothesized mechanisms are discussed.

To all my loved ones, here and gone. زندگی می‌گذرد

ACKNOWLEDGEMENTS

I have a village to thank here—bear with me. Patrick and Dan, thank you both so much for helping me stay afloat despite a series of challenging life events. Dan, thanks for reading a ridiculous number of proposal drafts, agreeing to chair my committee upon very short notice, and helping me get back on track. Ken and Dave, thank you both for your insightful feedback and guidance in this project. Amanda and Craig, thanks for dealing with undergrad me and continuing to be sources of support.

Jen, thank you for a million things, but—most importantly— for facilitating our attendance of the Eras Tour. Chris, thank you for all your wisdom, always. I'd also like to thank everyone in the Lab—your unwavering support has meant everything. Whether it's taking shifts being with me at the ER or having lunch conversations about the most random things, you're the highlight of my days in Davie.

Cecilia, Adilen, and Ariana, thank you for being my chosen family. Sam, thank you for being there for all of it, all the time; you mean the world to me. Eleanor and Justin, thanks for holding my PS5 hostage as motivation for me to finish this thesis. Jessica and Melody, thank you for being wonderful friends even from across the continent. To all my friends whose names wouldn't fit on this page, thank you for everything.

Finally, thank you to my family members for all of their sacrifices and being brave enough to leave everything behind to start over in a new country; I can't imagine how it must have felt. Malak, thanks for helping us both heal after my dad passed, and for being a great grand-meowther to Olive.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

ANOVA	Analysis of Variance
CFI	Comparative Fit Index
CI	Confidence Interval
DGM	Data Generating Model
ES	Effect Size
FM	Fitted Model
GEE	Generalized Estimating Equation
ML	Maximum Likelihood
OLS	Ordinary Least Squares
PME	Partially Mediated Effect
RMSEA	Root Mean Square Error of Approximation
SE	Standard Error
SEM	Structural Equation Modeling
B	Matrix of regression slopes (endogenous on endogenous)
i	Index for observation
I	Identity matrix
N	Sample size
R^2	Coefficient of determination (explained variance)
p	p-value (significance level)
t	T-statistic
Γ	Matrix of regression slopes (endogenous on exogenous)
Σ	Covariance matrix
α	Intercept/mean
θ	Vector of model parameters

β	Standardized regression coefficient
ζ	Vector of residuals
η_G^2	Generalized eta-squared
σ^2	Variance
χ^2	Chi-square
Ψ	Residual covariance matrix of endogenous variables

CHAPTER 1: INTRODUCTION

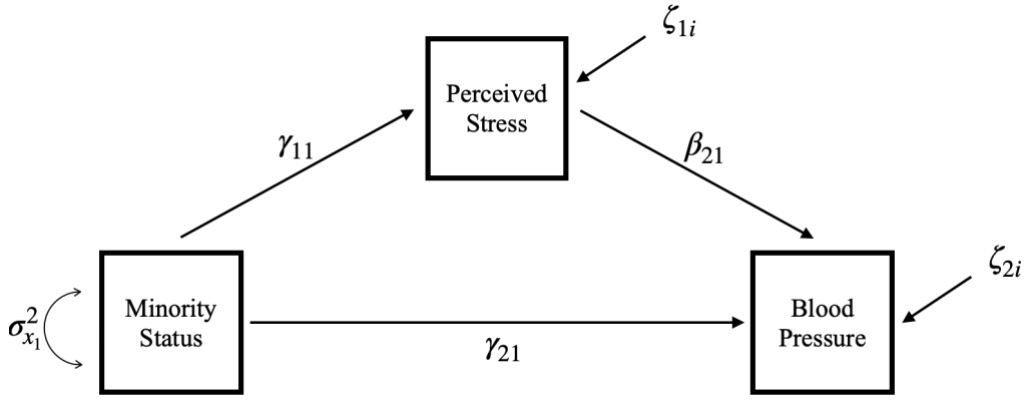
Mediation analysis is an important facet of psychological research (Imai et al., 2011; MacKinnon et al., 2007; VanderWeele, 2015); it allows researchers to identify, evaluate, and ultimately intervene upon causal mechanisms (Foster, 2010; Rucker et al., 2011; Sobel, 2000). This coincides with the four primary goals of psychological research: to observe, predict, explain, and manipulate (Yarkoni & Westfall, 2017). For example, health psychologists may be interested in *observing* blood pressure levels of individuals from minoritized populations, determining *predictors* of high blood pressure, and identifying underlying *explanatory* mechanisms involving those predictors. After evaluating such a mechanism using mediation analysis, researchers may then *intervene* at crucial points to lower blood pressure and reduce risk of cardiovascular diseases (Dolezsar et al., 2014; Larkin, 2008; Orom et al., 2017). To this end, mediation analysis is at the heart of preventative psychological research (Fairchild & MacKinnon, 2014; MacKinnon et al., 2002; O'Rourke & MacKinnon, 2019).

Mediation analysis models are inherently *multivariate* in nature (MacKinnon, 2007). For example, health psychologists may hypothesize that individuals from minoritized populations experience more perceived racial stress, and that this in turn leads to higher blood pressure levels (Dolezsar et al., 2014; Larkin, 2008; Orom et al., 2017). The corresponding mediation model equations and figure below align with this simple two-stage model (Edwards & Lambert, 2007), where perceived stress is both an outcome of minoritization and predictor of blood pressure. Notably, the equations together form a multivariate *system,* where blood pressure depends on perceived stress which depends on minority status.

$$Perceived\ Stress_i = \alpha_1 + \gamma_{11} Minority\ Status_i + \zeta_{1i} \quad (1)$$

$$Blood\ Pressure_i = \alpha_2 + \beta_{21} PerceivedStress_i + \gamma_{21} Minority\ Status_i + \zeta_{2i} \quad (2)$$

Figure 1 *Simple Mediation Model*



Although the concept of mediation can be traced back to Sewall Wright's work on path analysis (1934) and Otis Duncan's revival of Wright's work (1966), Kenny and Judd first provided a widely circulated definition of statistical mediation as a type of process analysis to "attempt to specify the causal chain responsible for the observed treatment effects" (1981a; 1981b). The increased prevalence and ubiquity of mediation analysis is, however, often attributed to Baron and Kenny's (1986) paper outlining the causal steps procedure, which requires all structural paths to be significant for mediation to be declared present. Certain requirements of that procedure, however, have since been deemed unnecessary for inferring the presence of mediation (Zhao et al., 2010). In the contemporary literature, mediation is often cast as a transmission of effects through intermediary variables.

According to MacKinnon and colleagues, statistical mediation refers to an intervening variable, a mediator, transmitting the effect of an independent variable to a dependent variable through an *indirect effect* (2002). The indirect effect of minority status on blood pressure is the product of the two constituent paths, $\gamma_{11} \times \beta_{21}$.[1] The *direct effect* is γ_{21}, and the *total effect* is the sum of the indirect and direct effects, equal to $\gamma_{11} \times \beta_{21} + \gamma_{21}$.[2] As described in more detail below, the notation used here (γ_{11}, β_{21} and γ_{21}) is consistent with how mediation models are expressed within a structural equation modeling (or path analysis) framework.

[1] In the mediation literature this is often referred to as *ab*, where *a* represents the first path and *b* represents the second (Hayes, 2022).
[2] In the regression-based literature, the direct effect is denoted as c' and the total effect as c.

Two conventional analytical approaches are frequently employed to fit mediation models. The first is to fit each equation separately using a linear regression model with the traditional ordinary least squares (OLS) estimator (i.e., Equations 1 and 2 are fitted in two separate regression models), from which estimates for the comprehensive mediation model are then assembled (Hayes, 2022). The regression method of obtaining estimates may thus be regarded as *fragmented* in nature. This fragmentation may be a vestige of the causal steps procedure originally proposed by Baron & Kenny (1986), where mediation was initially established by first verifying a significant relationship between the focal independent variable and outcome before moving on to independently test the other paths. In contrast, structural equation modeling (SEM) is a flexible procedure used to *simultaneously* fit the entire system of equations, thus aligning more closely with the fundamentally multivariate objective of mediation analysis and causal hypothesis testing (Edwards & Lambert, 2007; Imai et al., 2011). When all variables in the system are observed (not latent), the case considered in this thesis, the SEM is also known as manifest variable path analysis. Importantly, the SEM has the added benefit of allowing for the modeling of latent variables as well, but this thesis explores only manifest variables to maintain comparability to OLS regression modeling. Many potential estimators are available for the SEM, with *maximum likelihood (ML)* being the most commonly used in practice.

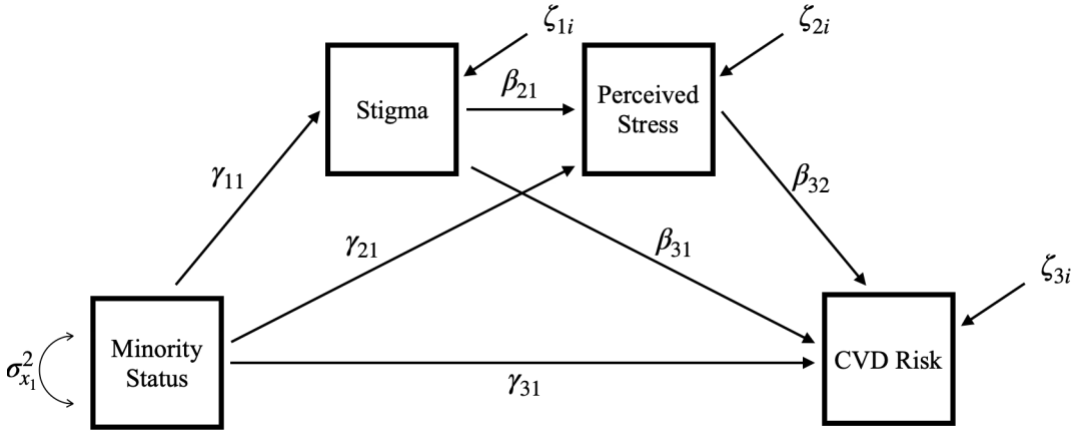
For a manifest-variables path analysis, OLS and ML will generate the same estimates for the raw coefficients (assuming equivalent model specification). In small samples, however, ML produces downwardly biased residual variance estimates. This in turn may lead to standard errors that are too small and inconsistent standardized coefficients (Firth, 1993). With sufficiently large samples, however, these differences become vanishingly small. Thus, my thesis focuses on a different aspect of mediation analysis that is often confounded with the chosen analytic approach, namely the use of saturated model specifications (virtually ubiquitous with the OLS regression approach) versus restricted model specifications (much more common with the SEM approach).

In a *saturated* model, all possible paths are estimated and all observed associations—including potentially spurious ones—are perfectly reproduced (Fossum, 2023; Hayes, 2022; Rohrer et al., 2022).

This is the most common form of model used by analysts implementing the OLS regression approach. In contrast, a *restricted* model implements constraints on the parameters of the model, for instance, constraining some paths to zero based on theory. Implementing such restrictions is more common within the SEM approach, within which their consistency with the data can be formally evaluated using model fit measures and modification indices (or, LaGrange multipliers) (Bollen, 1989b; Rockafellar, 1993; Sörbom, 1989). To be clear, this difference in the use of saturated versus restricted models is more a function of tradition and ease than a requirement of the modeling approach (e.g., one can fit a saturated model via SEM and one can implement some restrictions with the OLS regression approach, such as implicitly fixing a path to zero by omitting a predictor from a particular equation). Thus, although the use of saturated versus restricted models is highly associated with the chosen modeling approach, the distinction is equally relevant for users of either. My thesis will thus focus particularly on the SEM approach, for which restricted models are both more salient and easier to specify, though implications will pertain to users of the regression approach as well.

As an example of restricted versus saturated models, suppose a health psychologist hypothesizes that minority status results in higher rates of stigma, which increases levels of perceived stress, and ultimately leads to greater risk of cardiovascular disease (CVD) (Poteat et al., 2021; Figure 2). They also hypothesize that minority status directly impacts perceived stress, which also contributes to greater CVD.

Figure 2 *Saturated Theoretical Model*



In the conceptual diagram above, the model is *saturated* and the effect of minority status on CVD risk is decomposed into a direct effect and three indirect effects (now called *specific indirect effects*), where these continue to sum to the total effect (Alwin & Hauser, 1975; Bollen, 1987; Fox, 1980; Greene, 1977). The three modeled specific indirect effects are: (1) Minority Status → Stigma → Perceived Stress → CVD Risk, (2) Minority Status → Perceived Stress → CVD Risk, and (3) Minority Status → Stigma → CVD Risk. Particularly when using the regression modeling approach, it is common to evaluate the previously mentioned hypotheses with such a model. Importantly, though, the third specific indirect effect of Minority Status → Stigma → CVD Risk, which was *not* directly hypothesized, would still be evaluated. One could alternatively restrict the Stigma → CVD Risk path to zero, removing the third specific indirect effect, and the Minority Status → CVD Risk path to zero, removing the direct effect, to more directly map onto the hypothesized structure. Such restrictions are particularly straightforward to implement and test within the SEM approach to mediation analysis.

There are potential tradeoffs associated with fitting saturated versus restricted model specifications. For example, over-parameterized models are sometimes *intentionally* evaluated in an effort to mitigate model misspecification bias caused by omitting variables from an equation that may have non-zero effects in the population (Wilms et al., 2021). In other words, possible relations are evaluated in the event that they are true in the population—even if they are not directly hypothesized. On the other hand, incorporating additional parameters beyond those theoretically required may lead to a decrease in statistical power (Anderson et al., 2017) and increased potential for Type I errors (spurious findings). This study seeks to empirically evaluate these tradeoffs.

This thesis begins by first providing a more detailed overview of manifest-variable SEM, implementation of model restrictions, and testing mediational hypotheses. I then discuss common practices when evaluating mediation models in psychological research and potential consequences of including unnecessary paths versus excluding necessary ones. Finally, I set up initial hypotheses regarding the costs and benefits of fitting restricted or saturated mediation models, present a simulation

study to explore these hypotheses, provide and discuss results, and conclude with recommendations for applied researchers using mediation analysis.

SEM

The SEM approach to mediation analysis typically takes the form of *manifest* variable path analysis, in which all variables in the system are directly observed (e.g. not latent) and assumed to be measured without error. In an SEM framework, exogenous variables are those not predicted by any variables; endogenous variables are those that are predicted by any other variable in the system. As such, exogenous variables serve only as predictors whereas endogenous variables can serve as both outcomes *and* predictors. The equation for the path analysis model is then:

$$\mathbf{y}_i = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i, \qquad \boldsymbol{\zeta}_i \sim \text{MVN}(0, \boldsymbol{\Psi}) \tag{3}$$

In Equation 3, $\mathbf{y}_i$ is a $p \times 1$ vector of endogenous variables, $\mathbf{x}_i$ is a $q \times 1$ vector of exogenous variables, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of regression intercepts, $\mathbf{B}$ is a $p \times p$ matrix of regression slopes for endogenous variables regressed on other endogenous variables, $\boldsymbol{\Gamma}$ is a $p \times q$ matrix of regression slopes for endogenous variables regressed on exogenous variables, $\boldsymbol{\zeta}_i$ is a $p \times 1$ vector of residuals, and $\boldsymbol{\Psi}$ is a $p \times p$ covariance matrix of disturbances. SEM is most commonly fit using ML estimation (Wald, 1949), and, as such, is the approach considered in this thesis. ML operates by iteratively obtaining model-implied estimates for the *moment structures* that are most likely to give rise to the observed sample data.

Model-Implied Moment Structures

The hypothesized structure of a model implies specific patterns for the moments of the observed variables, in particular the means and covariances. The population model-implied mean vector and covariance matrix are denoted, respectively, as $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents a vector that contains the intercept, regression parameters, and residual variance of the model. The mean vector is partitioned into means for the endogenous (y) versus exogenous (x) variables; the covariance matrix is similarly partitioned into covariances among endogenous and exogenous variables:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{y}}(\boldsymbol{\theta}) \\ \boldsymbol{\mu}_{\mathbf{x}}(\boldsymbol{\theta}) \end{bmatrix} \qquad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{yy}}(\boldsymbol{\theta}) & \\ \boldsymbol{\Sigma}_{\mathbf{xy}}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{\mathbf{xx}}(\boldsymbol{\theta}) \end{bmatrix} \tag{4}$$

The full matrix expressions for the population model-implied mean and covariance structures are given by:

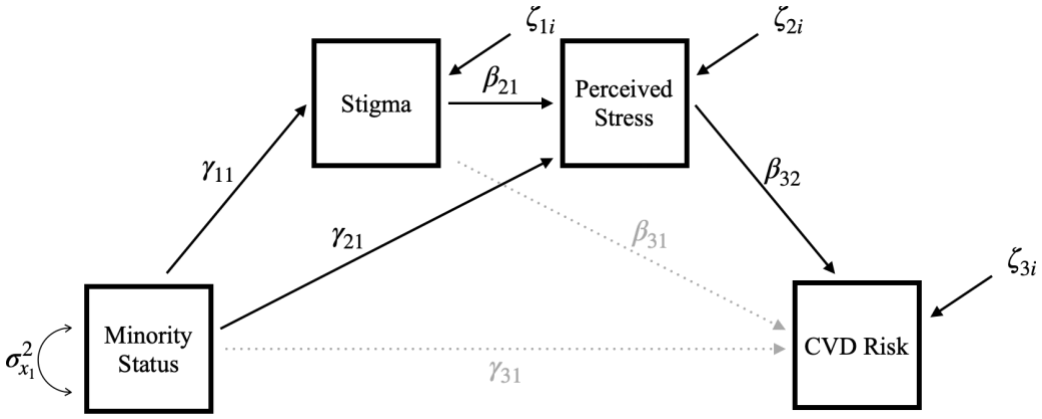
$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\alpha}+\boldsymbol{\Gamma}\boldsymbol{\mu}_{\mathbf{x}}) \\ \boldsymbol{\mu}_{\mathbf{x}} \end{bmatrix} \qquad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\Gamma}'+\boldsymbol{\Psi})(\mathbf{I}-\mathbf{B})^{-1'} & \\ \boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\Gamma}'(\mathbf{I}-\mathbf{B})^{-1'} & \boldsymbol{\Sigma}_{\mathbf{xx}} \end{bmatrix} \tag{5}$$

These can be derived using either covariance algebra or path tracing rules (Wright, 1934; Bollen, 1989). For the purposes of this thesis, $\boldsymbol{\mu}(\boldsymbol{\theta})$ is assumed to be saturated for all models; in other words, as many means and intercepts are included in the model as there are observed means, such that these will be perfectly reproduced. Instead, discussion of saturated and restricted models will focus on $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, which defines the *structural relations* among variables.

Model Restrictions

Restrictions on the covariance structure, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, often refer to constraining structural paths to zero. Such restrictions do not "allow" a variable to be regressed on another variable, or for them to covary directly. For instance, consider the modified model presented below that directly maps onto the original example's hypotheses.

Figure 3 *Restricted Theoretical Model*



This model—unlike the one in Figure 2— is *restricted*; some structural elements are set to zero and represented by omitted paths to denote that those relations cannot take on any other value; there is no direct effect of minority status on CVD risk, or direct relationship between stigma and CVD risk. Instead, the model explicitly evaluates only the two hypothesized specific indirect effects: (1) Minority Status →

Stigma → Perceived Stress → CVD Risk and (2) Minority Status → Perceived Stress → CVD Risk. The corresponding **Γ** and **B** matrices that define the structural relations are:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ 0 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ 0 & \beta_{32} & 0 \end{bmatrix} \tag{6}$$

In contrast, the *saturated* version (in Figure 2) does not have any 0s in **Γ** or the lower triangle of **B**, and all parameters are freely estimated.

Benefits of using a restricted model in an SEM framework include the ability to conduct tests of model fit in overidentified models, perform model comparisons among competing models, and evaluate the potential removal of restrictions using modification indices (Bollen, 1989b). Some similar restrictions can also be implemented through OLS regression, namely by omitting a predictor to remove its path, but these restrictions cannot be evaluated system wide (only equation-by-equation) with regression, becoming cumbersome for more complex models consistent with mediational hypotheses.

Effect Decomposition

The direct and indirect effects of interest in a mediational SEM may be easily obtained using matrix decompositions. In most cases, researchers are interested in the effects of a focal exogenous (*x*) variable on an endogenous (*y*) variable. The total effect for this relation is given as:

$$(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} \tag{7}$$

The direct effect is simply **Γ**, and the indirect effect is:

$$(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} - \boldsymbol{\Gamma} \tag{8}$$

The indirect effect in Equation 8, however, represents the *total* indirect effect. In other words, it is the sum of all indirect pathways from *x* to *y*. Effects for specific indirect pathways may also be decomposed using the procedure outlined below (Bollen, 1987).

Specific indirect effects are obtained by first deleting all rows and columns pertaining to paths *not* of interest in the structural coefficient matrix. Then, indirect effects are calculated using these modified matrices (Bollen, 1987). General matrix expressions for decomposing total, direct, and specific indirect

effects for path models are given in Table 1; see Bollen (1987) for a detailed discussion of their derivations.

Table 1 *Direct, Indirect, and Total Effects of* $\boldsymbol{x}_i$ *and* $\boldsymbol{y}_i$ *on* $\boldsymbol{y}_i$

	Effects of $\mathbf{x}_i$	Effects of $\mathbf{y}_i$
Direct	$\boldsymbol{\Gamma}$	$\mathbf{B}$
Indirect	$(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}-\boldsymbol{\Gamma}$	$(\mathbf{I}-\mathbf{B})^{-1}-\mathbf{I}-\mathbf{B}$
Total	$(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}$	$(\mathbf{I}-\mathbf{B})^{-1}-\mathbf{I}$

Note. $\mathbf{I}$ = identity matrix

While this method yields point estimates of the specific indirect effects, these purported effects must also undergo formal statistical significance testing.

Testing Mediational Hypotheses

Many procedures have been proposed for statistically evaluating mediational hypotheses that are applicable to both regression and SEM. Baron and Kenny's (1986) seminal paper outlining *the causal steps procedure* was one of the first. It assumed that three clauses must be true for mediation to be present:

1. The independent variable must be significantly related to the mediator.
2. The independent variable must be significantly related to the dependent variable.
3. The mediator must be significantly related to the dependent variable.

However, the second clause is unnecessary (Shrout & Bolger, 2002; Zhao et al., 2010). This is because the indirect and direct effects may have opposing signs, potentially canceling each other out and leading to insignificant total effects (failure of clause two). Thus, requiring the second clause reduces statistical power to detect mediation (MacKinnon et al., 2008). Instead, the indirect effect, or $\gamma_{11} \times \beta_{21}$ product, should be tested directly to determine the presence of mediation (Preacher & Hayes, 2004).

A traditional method for directly testing the indirect effect is the Sobel Test (also referred to as the delta method). This method assumes that the product term, $\hat{\gamma}_{11} \times \hat{\beta}_{21}$, is normally distributed (Sobel, 1982). However, the indirect effect is a product of coefficients with asymptotically normal sampling distributions, and the product of multiple normally distributed coefficients is *not itself* normally

distributed (Lomnicki, 1967). This violation of distributional assumptions leads to underpowered tests of the mediated effect when using the delta method (Shrout & Bolger, 2002). Another method, however, entirely sidesteps this limitation.

Percentile bootstrapped confidence intervals, a preferred method for inferring mediation, make no distributional assumptions (Bollen & Stine, 1990; Efron, 1987; Efron & Tibshirani, 1986; Preacher & Hayes, 2004, 2008; Shrout & Bolger, 2002). Instead, the observed sample data are randomly resampled with replacement thousands of times to form bootstrapped data sets. The hypothesized model is then fit to each bootstrapped data set and estimates of coefficients that comprise the product term are multiplied together and recorded, forming an *empirical* distribution of the indirect effect. To construct the confidence interval, the distribution of empirical indirect effect estimates is arranged in ascending order. Finally, the limits of the confidence interval are determined based on the desired significance level. For example, for a significance level of $p < .05$, the values below the 2.5^{th} percentile and 97.5^{th} percentile are excluded from the interval. If the resulting 95% confidence interval does not include 0, this indicates a statistically significant indirect effect.

There are also other methods for inferring the presence of a mediated effect, such as with Monte Carlo confidence intervals, Bayesian estimation, the distribution of the product, and other bootstrapping methods (MacKinnon et al., 2002; Preacher & Selig, 2012). Given its current regard as best practice and greater statistical power, my simulation study will focus on the percentile bootstrapped confidence interval (Hayes, 2022; Tibbe & Montoya, 2022). As I detail below, causal inference on the indirect effects requires additional conditions beyond statistical significance alone.

Causal Assumptions and Potential Outcomes

Certain assumptions are required to infer causality in mediation models. To begin, the cause must precede the effect (temporal precedence), the cause must be demonstrably related to the effect, and there must be no plausible alternatives (John Stuart Mill, 1856; Shadish et al., 2002). From its inception, Wright (1923) also cautioned that path analysis was not a method to *infer* causation, but rather to *quantify* the hypothesized causal relations.

These causal assumptions may be directly expressed in the *causal inference* or *potential outcomes* framework (Gonzalez et al., 2022; Imai et al., 2011; MacKinnon et al., 2020; Pearl, 2010a; VanderWeele, 2015). Here, we assume that there is a potential observed score (outcome) that would have resulted had the treatment condition been swapped with another. This serves as a counterfactual. For example, in a simple vignette-based experiment, there is a different potential outcome based on which experimental condition a participant is randomly assigned to. Crucially, each participant in a between-subjects design can only be in *one* treatment condition, and the other hypothetical outcomes must be mathematically *inferred.* There are six causal inference effects in total, defined as the difference between the counterfactuals: (1) total natural indirect effect, (2) pure natural indirect effect, (3) total indirect effect, (4) pure natural indirect effect, (5) controlled direct effect, and (6) total effect (Gonzalez et al., 2022). Additionally, the four most commonly cited assumptions in the causal mediation literature are:

1. No unmeasured confounders of the predictor → outcome relation.
2. No unmeasured confounders of the mediator → outcome relation.
3. No unmeasured confounders of the predictor → mediator relation.
4. No confounders of the mediator → outcome relation affected by predictor conditional on covariates.

There are certain scenarios where traditional mediation analysis is a special case of the potential outcomes framework (MacKinnon et al., 2020). Specifically, when there is not an interaction or other non-linear relation between x_{1i} and y_{1i}, the two are considered equivalent in the case of a binary x_{1i} and continuous y_{1i} (MacKinnon et al., 2020). In other words, when the relationship between y_{1i} and y_{2i} does not differ depending on the value of x_{1i} and the variables are linearly related, then using potential outcomes or traditional mediation analysis yields the same results. It is also more general than traditional mediation analysis such that it can account for differing functional forms and subsumes many commonly implemented mediation models (Holland, 1986). For a full treatment of the causal mediation framework, see Pearl (2010) and Vanderweele (2015).

Although the potential outcomes framework is a powerful tool for causal inference, my thesis focuses on traditional mediation methods due to current complexity in applying causal mediation models to more intricate mediation models (Gonzalez et al., 2022; Imai et al., 2011; MacKinnon et al., 2020; Pearl, 2010; VanderWeele, 2015). The models discussed in this thesis constitute a subset of causal mediation models where, to narrow scope to more traditional forms, no interactions are assumed.

Common Practice

Saturated mediation models are most common—perhaps because restricted specifications are less salient and more fragmented to implement when using the OLS approach, which remains dominant in the field (Fossum, 2023; Rohrer et al., 2022). The prevalence of such models may also be a precautionary measure against under-parameterization (Tomarken & Waller, 2005). Under-parameterization occurs when too few parameters are estimated (Wilms et al., 2021). Such a model is considered misspecified, resulting in biased estimates (Kolenikov, 2011; White, 1982). By incorporating additional parameters beyond those theoretically required, researchers reduce the likelihood of under-parameterization and consequent misspecification bias.

Over-parameterization, on the other hand, occurs when parameters are estimated that are in fact null in the population. The inclusion of extraneous parameters in a model diminishes the number of degrees of freedom available for evaluating model fit and requires a larger sample size to maintain the same level of statistical efficiency and power (Cohen, 1988, 1992). Therefore, adding parameters but not increasing sample size leads to notable reductions in power to detect small mediated effects that already had low power to begin with (Fritz & MacKinnon, 2007). Additionally, the estimation and testing of parameters that are actually null in the population presents an accumulating risk of committing Type I errors with subsequent interpretation of spurious results. The degree to which this occurs in commonly used mediation models remains unknown.

Finally, there is a possibility of mixing both over-specification and under-specification in the same model. In this scenario, the fitted model omits a necessary path but also includes an unnecessary one. The inclusion of this extra path may absorb the effect of the omitted path, depending on the

asymptotic covariance of the estimates and where exactly in the model the misspecification occurs (Curran et al., 1996; Kaplan & Wenger, 1993). However, this "effect absorption" may not hold if the two estimates are asymptotically uncorrelated (Kaplan & Wenger, 1993). For example, misspecification of paths leading from exogenous to endogenous variables may yield different effects on power and subsequent model respecification in relation to misspecification of paths among endogenous variables.

Aims of the Current Study

To recap, psychological mediational hypotheses are often evaluated using saturated models which may not always align with the hypothesized theoretical mechanism. Improperly saturated models reflect over-parameterization that may lead to increased Type I errors for null effects and decreased statistical power for non-null effects, especially when those effects are small. On the other hand, purposefully over-parameterized models may safeguard against under-parameterization, which is associated with model misspecification and biased results.[3] There is also a possibility of combining errors of over- and under-specification. Regardless of framework, the potential costs of estimating saturated models that do not directly align with theoretical hypotheses remains unclear. As such, this study aims to evaluate these potential costs and benefits through a Monte Carlo data simulation designed to test the following series of interrelated hypotheses:

- H1: Statistical efficiency for the specific indirect effect will be greater when the model is properly restricted as opposed to improperly saturated (over-parameterized). This difference will be more pronounced with small effect sizes and small *N*.
    a. Standard errors (SEs) for the pathways that comprise the specific indirect effect will be smaller when the model is properly restricted.
    b. 95% CI widths for the specific indirect effect will be narrower when the model is properly restricted.

[3] Restricted models in the SEM framework may be identified as improperly specified based on model fit measures and modification indices.

    c. Power to detect non-null specific indirect effects will be greater when the model is properly restricted relative to when it is improperly saturated.

- H2: Overparameterized models will result in increased family-wise Type I error rates (relative to .05) when multiple unnecessary paths are included in the fitted model.
- H3: Relative to properly saturated models, under-parameterized models will result in biased estimates and decreased confidence interval coverage of the specific indirect effect.
    a. Individual pathways that comprise the specific indirect effect will be biased, leading to greater bias in the indirect effect estimates relative to other conditions.
    b. Even when there are correct rejections of the null hypothesis, confidence interval coverage will be lower in the presence of biased estimates.
- H4: Misspecified models that combine errors of *over-* and *under-* parameterization (i.e., include some unnecessary effects while excluding other effects that should be included) will result in biased estimates of paths that comprise the specific indirect effect.

In the following section, I provide details on how I systematically tested the hypotheses above using a Monte Carlo data simulation study. The goal of evaluating these hypotheses was to provide researchers with a clear recommendation on whether there are any pitfalls regarding the use of saturated mediation models, and, if so, the extent of those potential deleterious effects.

CHAPTER 2: SIMULATION METHODS

In this section I describe my Monte Carlo simulation design, population data generating models, data generation process, fitted models, outcomes, and analysis plan. I varied several conditions such as sample size, population data generating and fitted models, and standardized effect size. Following discussion of data generation and models, I also provide operationalizations of my outcome measures. Finally, I present the simulation design matrix and the choice of meta-models and associated effect size measures to analyze results.

Simulation Design

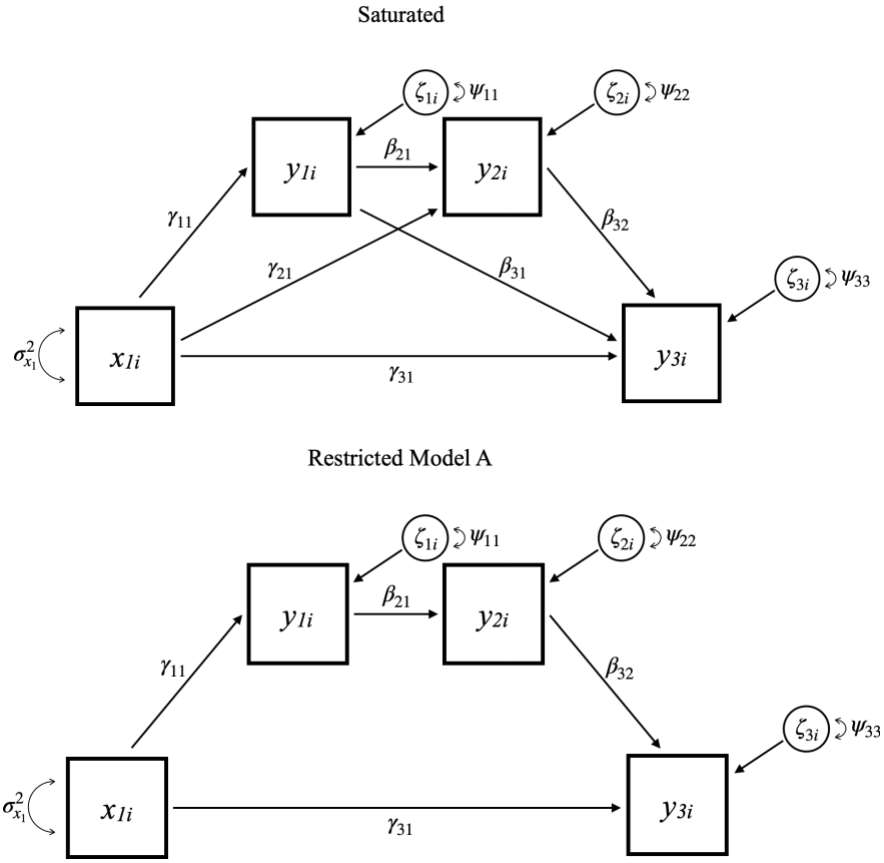
All data were generated and analyzed using R version 4.4.2 and the 'lavaan' (Rosseel et al., 2024), 'SimDesign' (Chalmers, 2015; Chalmers et al., 2024), 'mvtnorm' (Genz et al., 2024), 'MASS' (Ripley et al., 2025), 'matlib' (Friendly et al., 2024), 'doParallel' (Daniel et al., 2022), 'foreach' (Daniel et al., 2022), and 'rms' (Harrell, 2024) packages (R Core Team, 2025). Specific implementations of these packages are described below.

Population Data Generating Models

The two population models aligned with either a saturated or restricted structure (Figure 4). Both structures had one exogenous variable and three endogenous variables such that there was one focal predictor, two mediators, and one outcome. For the restricted structure (A), certain structural coefficients were set to 0 to represent omitted paths. All variables were assumed to be continuous, perfectly measured, and normally distributed with unit variance. Although many mediation simulation studies feature dichotomous exogenous predictors, previous research has shown that using a continuous focal predictor yields a slight power advantage for a mediation model of this general structure (Taylor et al., 2008). As such, continuously scaled predictors were used in the simulation as a "best case" scenario. Finally, this

serial mediation model was chosen due to its common usage based on a literature review of psychological studies with mediational hypotheses.

Figure 4 *Population Data Generating Structures*



Data Generation

Data were generated using the population model-implied mean vector and covariance matrix expressions in Equation 5. For both models, the mean vector was filled with 0s:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{9}$$

The covariance matrix was

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})(\mathbf{I}-\mathbf{B})^{-1'} & \\ \boldsymbol{\Sigma}_{xx}\boldsymbol{\Gamma}'(\mathbf{I}-\mathbf{B})^{-1'} & \boldsymbol{\Sigma}_{\mathbf{xx}} \end{bmatrix} \tag{10}$$

and was defined through specification of **Γ** and **B**, which varied in structure between models. For the saturated model, **Γ** and **B** were:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \tag{11}$$

For the restricted model (A), they were:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} \\ 0 \\ \gamma_{31} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ 0 & \beta_{32} & 0 \end{bmatrix} \tag{12}$$

For ease of interpretation, the marginal variance of each variable was set to 1 in the population, such that the elements of **Γ** and **B** represented standardized effects. To achieve this, $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ was set to 1, such that the single exogenous variable, *x*, had unit variance. Additionally, the elements of the **Ψ** matrix (the residual covariance matrix of the endogenous variables) were defined via covariance algebra to ensure that each endogenous variable had a total variance of one, as follows:

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & & \\ 0 & \psi_{22} & \\ 0 & 0 & \psi_{33} \end{bmatrix} \tag{13}$$

where

$$\psi_{11} = 1 - \gamma_{11}^2 \tag{14}$$

$$\psi_{22} = 1 - (\gamma_{21}^2 + \beta_{21}^2 + 2\gamma_{11}\beta_{21}\gamma_{21}) \tag{15}$$

$$\psi_{33} = 1 - (\beta_{32}^2 + \beta_{31}^2 + \gamma_{31}^2 + 2\beta_{32}\beta_{31}(\beta_{21} + \gamma_{21}\gamma_{11}) + 2\beta_{32}\gamma_{31}(\gamma_{21} + \beta_{21}\gamma_{11}) + 2\beta_{31}\gamma_{31}\gamma_{11}) \tag{16}$$

For Restricted Model A, γ_{21} and β_{31}were set to zero in Equations 15 and 16, leading several terms to drop out. Off-diagonals of **Ψ** were set to 0 to denote no correlated residuals among the variables. Numerical values for the elements within **Γ** and **B** were also varied between the conditions outlined below.

Once the population mean vector and covariance matrix were formed, they were provided as input values—alongside the desired sample size— to the "mvnorm()" function from the "mvtnorm" package (Genz et al., 2024). This function generates multivariate normal data with *N* observations

consistent with the supplied mean vector and covariance matrix. I generated test data using large sample sizes (N = 1,000,000) and various standardized effect combinations to confirm correct data generation and accurate recovery of the population mean vector and covariance matrix.

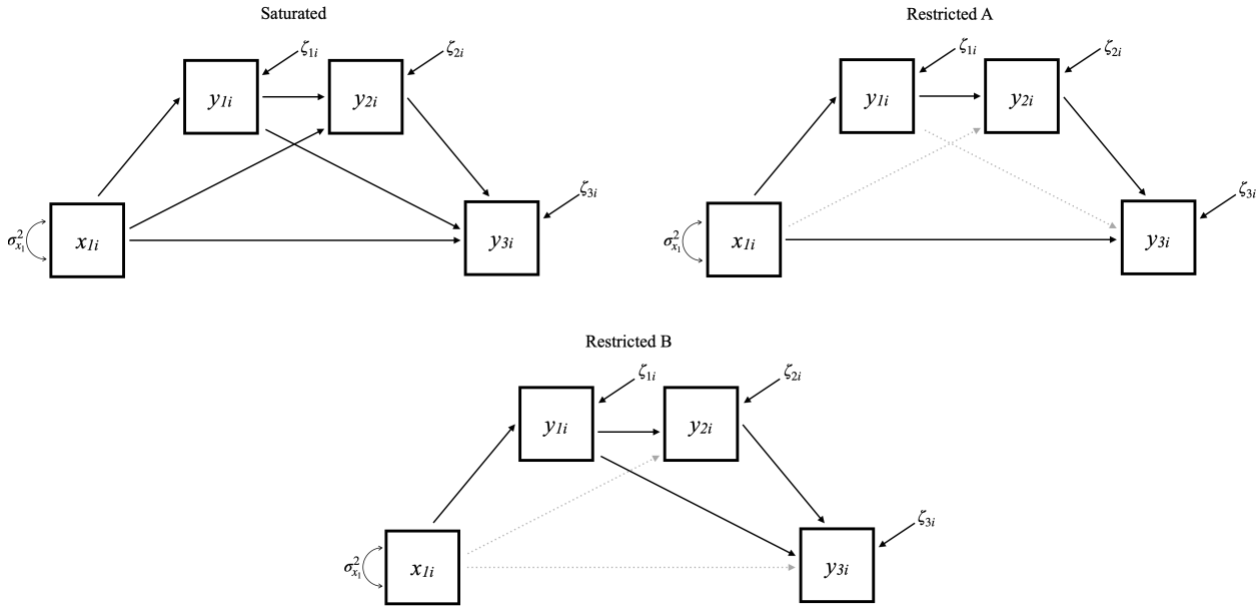
Fitted Models

Once data were generated, they were analyzed using three path analysis models: one saturated and two restricted (Figure 5). Of the two restricted models, only one (Restricted Model A) directly mapped onto the restricted data generating model; the other (Restricted Model B) represented a misspecified model regardless of whether the saturated or restricted model (A) were used to generate data. This model (Restricted B) was included to further explore combinations of *over-* and *under-* parameterization. For Restricted Model B, **Γ** and **B** were:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \tag{17}$$

All models converged and, as such, all replications were included in the simulation analyses.

Figure 5 *Fitted Models*



Conditions

I tested various sample sizes (*N*), standardized effect size values for the specific indirect effect of interest, and combinations of data generating and fitted models to evaluate my hypotheses. To maintain scope, the specific indirect effect of primary focus which was manipulated in magnitude was that of $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$. The two data generating models and three fitted models used are outlined above (Figures 4 and 5). Depending on the combination of population and fitted models, the conditions represented correctly specified models, *over-* or *under-* parameterization, or a combination of *over-* and *under-* parameterizations. Maximum likelihood was used for all model estimation.

I generated data of size 50, 100, and 200 as these sample sizes are representative of many psychological studies. Moreover, a sample size of 200 was specifically explored because standard errors (and efficiency) of the estimates are expected to reduce by a factor of ½ when a given sample size (here, 50) is quadrupled. An interim value, 100, was also examined to evaluate the rate of change in efficiency and power. I also varied the effect size of the specific indirect effect of interest such that γ_{21}, β_{21}, and β_{32} were all either .265 or .39. While .39 represents a "medium" effect size, .265 represents a value halfway between small (.14) and medium (Cohen, 1988, 1992; Fritz & MacKinnon, 2007). This value was chosen because using .14 for all links in the serial mediation model would have yielded a very small effect (.003), which would have only been detectable with very large sample sizes—something that is not the focus of this study (Fritz & MacKinnon, 2007). When multiplying the paths together, the standardized indirect effect was .019 for the small-medium effect and .059 for the medium effect. Although the partial mediated effect (PME) may also be used as a measure of effect size, simulation studies have demonstrated that the measure is unstable in several parameter combinations and has excess bias in small samples (Fairchild et al., 2009; MacKinnon et al., 2009a, 2009b).

Moreover, to ensure that the same R^2 value was maintained for y_3 in the restricted model (A), γ_{31}, the direct effect, was set to .36 and .44 for the small-medium (.265) and medium conditions (.39), respectively. These values represent larger direct effects that accounted for the lack of cross paths in Restricted Model A and ensured that the residual variance, an important facet of power, was held constant

and not confounded with the use of saturated versus restricted population data generating model. All other paths were held constant at .2 for the saturated model, and the four possible data generating models with specific coefficient values are given in Figure 6. In total, the study had a $2 \times 3 \times 2 \times 4$ factorial design, yielding 48 unique conditions (Table 2; see Appendix for full simulation design matrix).

Figure 6 *Population Data Generating Values*

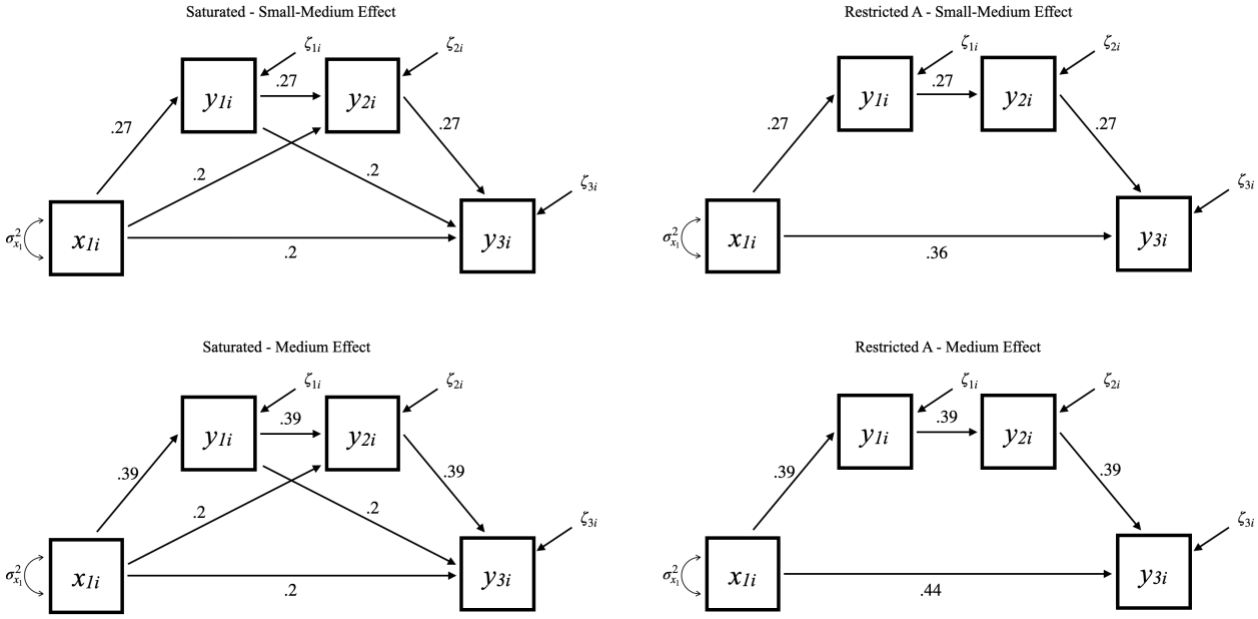


Table 2 *Simulation Conditions*

Data Generating Model	Fitted Model	Sample Size	Effect Size for $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$
Saturated	Saturated	50	Small-Medium (.265)
Restricted A	Restricted A	100	Medium (.39)
	Restricted B	200	

Note. Small-medium effect corresponds to R^2 of .234 for y_3; medium effect corresponds to R^2 of .391. Effect size reflects the same value for each individual pathway that comprises the specific indirect effect.

Procedure

Samples of size *N* were drawn from the population data generating model, with values of **B** and **Γ** dependent on the cell of the design matrix. Following this, all three sample path analysis models were fit using the 'sem()' function in 'lavaan' and estimated using maximum likelihood (Rosseel et al., 2024). Data were generated *once* per replication and *multiple* models were fit to the same data within a replication. The complete output was saved in a data frame and then extracted as follows.

Standardized point estimates and bootstrapped SEs were recorded for each individual pathway included in the fitted model. For the three possible specific indirect effects, the coefficients of the pathways that comprised each one were multiplied together and recorded. Their associated 95% percentile bootstrapped confidence intervals were obtained using 5000 bootstrapped samples and recorded as well.

Additionally, the model χ^2 test statistic and its associated *p*-value and degrees of freedom, the comparative fit index (CFI), and the root mean square error of approximation (RMSEA) were recorded for each of the three fitted models (Bentler, 1990; Bollen, 1989b; Steiger, 1990; West et al., 2012). These model fit statistics were extracted to provide insight into how often models would be considered acceptable based on conventional cutoff values (Hu & Bentler, 1999) and therefore used to make (potentially incorrect) substantive interpretations.

Data generation and analysis for 1000 replications per cell of the design were automated with the 'runSimulation()' function from the 'SimDesign' package using parallel processing across 10 CPU cores on a M1 Max Mac Studio (Chalmers et al., 2024). The 'RcppArmadillo', 'doParallel', 'foreach', 'tidyverse', 'tables', 'MASS', and 'matlib' packages served as dependencies for 'runSimulation' (Chalmers et al., 2024; Daniel, Corporation, et al., 2022; Daniel, Ooi, et al., 2022; Eddelbuettel et al., 2024; Friendly et al., 2024; Murdoch, 2024; Ripley et al., 2025; Wickham & RStudio, 2023).

Outcomes

I explored six primary outcomes: relative bias, standard errors, confidence interval width and coverage, and Type I error. Each of these outcomes were examined across all replications within a cell of the design matrix.

Relative Bias

Relative bias was calculated as follows for the specific indirect effect of interest and individual pathways that comprised that effect per replication.

$$Relative\ Bias\ =\ \frac{\text{Estimate}\ -\ \text{Population Value}}{\text{Population Value}}\ \times 100 \tag{18}$$

Smaller values are ideal, as they represent greater accuracy. Relative bias below 5% is often viewed as acceptable in small sample studies (Kline, 2023; Muthén & Muthén, 2002).

Standard Error

The standard error (SE) reflects the precision of an estimate; a lower value is more desirable and denotes greater efficiency. Bootstrapped SEs were extracted from the fitted 'lavaan' model object for each pathway that comprised the specific indirect effect of interest. Bootstrapped SEs were chosen given the relatively small sample sizes and to maintain methodological consistency.

Bootstrapped Confidence Interval Width

For the specific indirect effect of interest, confidence interval width was calculated as the difference between the upper and lower 95% percentile bootstrapped confidence limits within each replication. A narrower confidence interval suggests greater efficiency for the estimate of the indirect effect.

Bootstrapped Confidence Interval Coverage

Confidence interval coverage was determined as follows within each replication for the specific indirect effect of interest. A value of "1" was coded if the 95% confidence interval for the specific indirect effect contained the true population value, and a value of "0" if not. The overall confidence interval coverage for each cell of the design matrix was calculated as the proportion of times that the confidence interval contained the true population value across replications.

Statistical Power

I explored the power to detect the specific indirect effect of interest ($x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$). Significance was determined using 95% bootstrapped percentile confidence intervals due to non-normality of the product term. If the 95% confidence interval for the specific indirect effect contained 0, then the effect was deemed non-significant; if it did *not* contain 0, then the effect was coded as significant. Specifically, a value of "1" was assigned to a binary "significant" variable if the effect was significant and "0" if not for each replication. Power was then be calculated as the percentage of correct rejections of the null hypotheses across all replications per cell of the design matrix.

Type I Error

Type I error was calculated as the proportion of *incorrect* rejections of the null hypothesis for each pathway not included in the population data generating model but included in the fitted model across all replications of a given condition. The proportion was calculated using a similar binary "significant" variable as described above but significance was instead determined using a standard *z*-test of the estimate divided by the bootstrapped SE for the γ_{21} and β_{31} paths. Neither of these paths were included in Restricted Model A, and only β_{31} was included in Restricted Model B.

Moreover, the family-wise type I error rate, the focus of Hypothesis 2, was calculated as follows: if either γ_{21} or β_{31} were erroneously deemed significant, then a value of "1" was assigned to a family-wise "significant" variable, and a value of "0" was coded if not.

Model Fit

To assess model fit, researchers often attempt to gauge how well the model-implied moment structures reproduce the observed moments using a combination of absolute and relative fit measures. In the present research, the model χ^2 test statistic and its associated *p*-value and degrees of freedom, the CFI, and the RMSEA were recorded for each of the three fitted models.

The χ^2 test statistic evaluates how well the model-implied moment structure reproduces the population moment structure. The null hypothesis is given by:

$$H_o: \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}), \qquad \boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}) \tag{19}$$

The discrepancy between the model-implied estimated moments and observed moments is reflected in a *T*-statistic that asymptotically follows a chi-square distribution under the null hypothesis, from which one may obtain a *p*-value. This central χ^2 distribution has $k - t$ degrees of freedom, where k is the total number of observed (or known) moments, and t is the number of estimated parameters in the model. Contrary to traditional null hypothesis significance testing, the ideal outcome in many cases is *not* rejection—but rather retention—of the null hypothesis, enabling the researcher to conclude that the model fits the data well. Like any statistical test, the model χ^2 test is sensitive to sample size, leading to

potentially "excessive power" to detect trivial discrepancies with large *N*; it may also over-reject correct models (relative to the nominal rate) at small *N*, the *T* statistic not yet having obtained its asymptotic χ^2 distribution (Bentler & Bonett, 1980; Browne, 1984; McNeish, 2020; Yuan & Bentler, 2000). Given such concerns, the model χ^2 is often evaluated in conjunction with other measures of model fit, such as the CFI and RMSEA (Bentler, 1990; Bollen, 1989a; Steiger, 1990).

The *Comparative Fit Index (CFI)* compares the hypothesized model to a *highly restricted* baseline model in which no variables covary with any others. This index is a combination of four pieces of information: the *T*-statistic (T_h) and degrees of freedom (df_h) for the hypothesized model and the *T*-statistic (T_b) and degrees of freedom (df_b) for the baseline model. The equation is given by:

$$CFI = 1 - \frac{\max(T_h - df_h,\ 0)}{\max(T_b - df_b,\ 0)} \tag{20}$$

The root mean squared error of approximation (RMSEA) is given by (Steiger, 1990):

$$RMSEA = \sqrt{\frac{T_h - df_h}{df_h(N-1)}} \tag{21}$$

Lack of reliance on a baseline model and a known sampling distribution (for confidence interval estimation) are two strengths of the RMSEA.

When evaluating SEM models, it is regarded as good practice to present the *T*-statistic (with degrees of freedom and adjoining *p*-value), at least one of the relative fit indices, and the RMSEA (West et al., 2012).

Analysis Plan

To test my hypotheses, I compared the different combinations of population and fitted models, effect sizes, and sample sizes for the subset of the six possible outcomes measures described above. I used within-subjects ANOVAs for continuous outcomes such as relative bias and standard errors and generalized estimating equations (GEE) with logit link functions for binary outcomes such as confidence interval coverage and power. These modeling strategies account for inherently correlated data in within-subjects designs (Liang & Zeger, 1986). All independent variables were between-subject factors, except

for the within-subjects factor of Fitted Model, given all three were fit to each individual data set. Population model, fitted model, sample size, and effect size were dummy-coded, and all meta-models included an interaction between all factors to determine whether there were any higher-order effects.

In addition, I included measures of effect size to control for overpowered tests of effects in the meta-models due to the large number of replications. I computed generalized eta squared (η_G^2) for models with continuous outcomes such as relative bias. The use of η_G^2 is appropriate given the combination of between- and within-subjects factors (Bakeman, 2005; Olejnik & Algina, 2003). For binary outcomes such as power, type I error, and confidence interval coverage, I used odds-ratios as a measure of effect size. Measures of model fit were also investigated.

CHAPTER 3: SIMULATION RESULTS

The full simulation included all 36 cells of the design matrix and 1000 replications per cell, yielding 36,000 sets of results. Given very high power, effect size rather than significance was used to identify meaningful trends. For ANOVA meta-models, only effects with η_G^2 estimates of .01 or greater are discussed to account for overpowered tests. Based on Cohen (1988), η_G^2 values of .01, .06, and .14 correspond to small, medium, and large effects, respectively. Similarly, only effects with odds ratios notably different from 1 are discussed for GEE meta-models, as a value of 1 denotes no change in the odds. Specifically, odds ratios lower than .82 and greater than 1.22 were considered meaningful (Chen et al., 2010). All meta-models included all possible interactions between sample size (*N*), effect size (ES), data generating model (DGM), and fitted model (FM). Prior to evaluating the primary outcomes, model fit measures are examined to gauge what proportion of models would have reasonably been interpreted.

Preliminary Analyses: Model Fit

Model fit measures (Model χ^2, CFI, and RMSEA) were extracted for each combination of data generating and restricted fitted model across sample and effect sizes explored in the simulation. Saturated fitted models were not included because they do not have enough degrees of freedom to conduct traditional SEM tests of model fit.

Table 3 presents the proportions of models that would have been rejected based on conventional cutoff values (significant χ^2 test, CFI < .95, RMSEA > .05) (West et al., 2012). Figure 7 presents the proportion of *significant* model χ^2 tests, whereas Figures 8 and 9 present distributions of CFI and RMSEA values, respectively. The red vertical bar in Figure 8 is set at .90, and the blue vertical bar is set at .95. These values were chosen to visualize the proportion of values that would have been interpreted based off either liberal or conservative CFI conventional cutoff value. Similarly, the red vertical bar in Figure 9 is set at .08, and the blue vertical bar is set at .05 to represent common RMSEA cutoffs.

Especially in smaller sample sizes, a non-trivial proportion of misspecified models were not rejected by the χ^2 test (Figure 7), prompting potential acceptance and interpretation in practice. The CFI and RMSEA, however, were generally more conservative. Notably, the RMSEA even often rejected properly specified restricted models (Table 3, Figure 9). As sample size and effect size increased, the model fit measures correctly rejected misspecified models more often, particularly when considering the combination of Restricted Model A as the data generating process and Restricted Model B as the fitted model. Interestingly, effect size does not seem to exert a large effect on rejection rates when fitting an improperly restricted model to data from a saturated DGM. This may be due to the model χ^2 being much more sensitive to the sample sizes rather than effect sizes considered in this simulation. These results highlight the situations in which misspecified models may still be interpreted in these combinations of fitted and data generating models, even when considering measures of model fit.

Figure 7 *Proportion of Significant Model χ^2 Tests by Data Generating and Fitted Model*

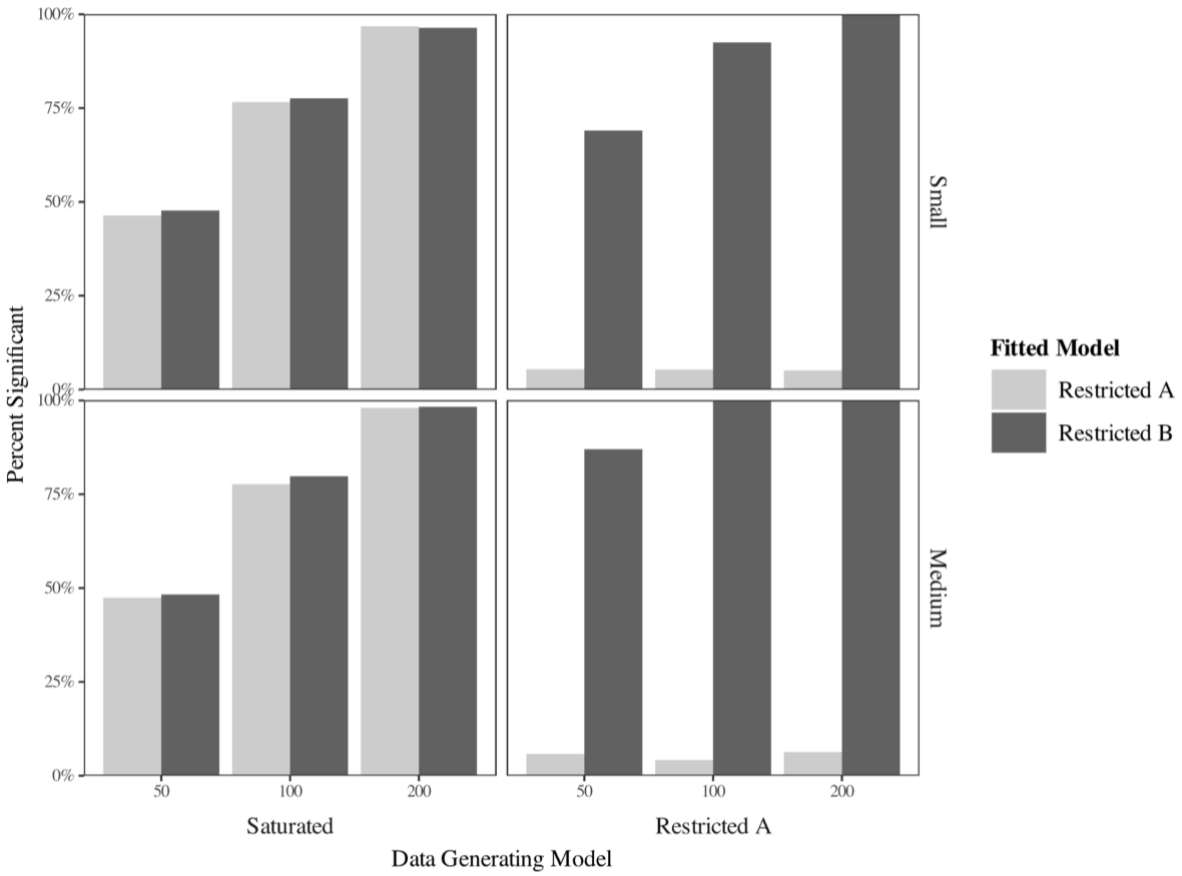


Table 3 *Percentages of Rejected Models Based on Conventional Model Fit Cutoff Values*

N	Effect Size	DGM	Fitted Model	χ^2 $p < .05$	CFI	RMSEA
50	.265	Saturated	Restricted A	46%	74%	82%
			Restricted B	48%	74%	83%
		Restricted A	Restricted A	5%	23%	33%
			Restricted B	69%	91%	93%
	.39	Saturated	Restricted A	47%	64%	82%
			Restricted B	48%	65%	84%
		Restricted A	Restricted A	6%	14%	32%
			Restricted B	87%	95%	98%
100	.265	Saturated	Restricted A	77%	86%	93%
			Restricted B	78%	88%	95%
		Restricted A	Restricted A	5%	15%	30%
			Restricted B	93%	98%	99%
	.39	Saturated	Restricted A	78%	74%	95%
			Restricted B	80%	76%	98%
		Restricted A	Restricted A	4%	4%	28%
			Restricted B	100%	100%	100%
200	.265	Saturated	Restricted A	97%	97%	100%
			Restricted B	96%	96%	99%
		Restricted A	Restricted A	5%	6%	23%
			Restricted B	100%	100%	100%
	.39	Saturated	Restricted A	98%	86%	100%
			Restricted B	98%	89%	100%
		Restricted A	Restricted A	6%	1%	23%
			Restricted B	100%	100%	100%

Note. The table contains proportions of models with Model χ^2 *p*-values less than .05, CFI values less than .95, and RMSEA estimates greater than .05.

Figure 8 *Distributions of CFI by Data Generating and Fitted Model*

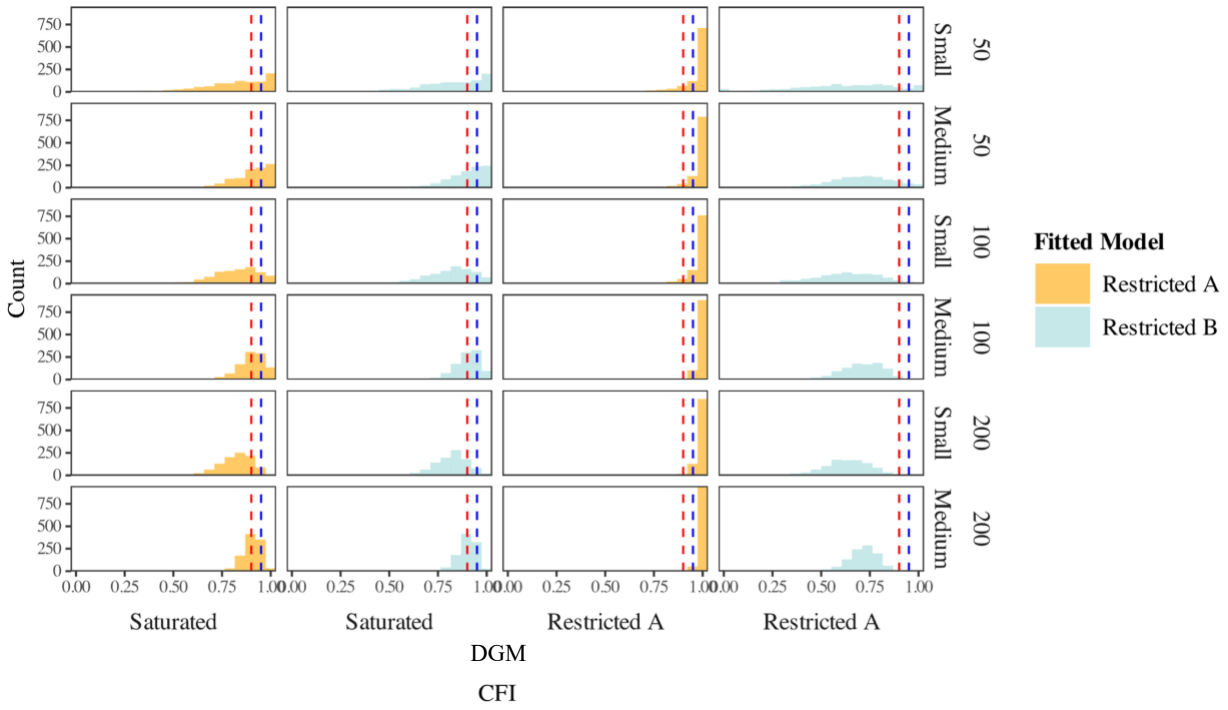
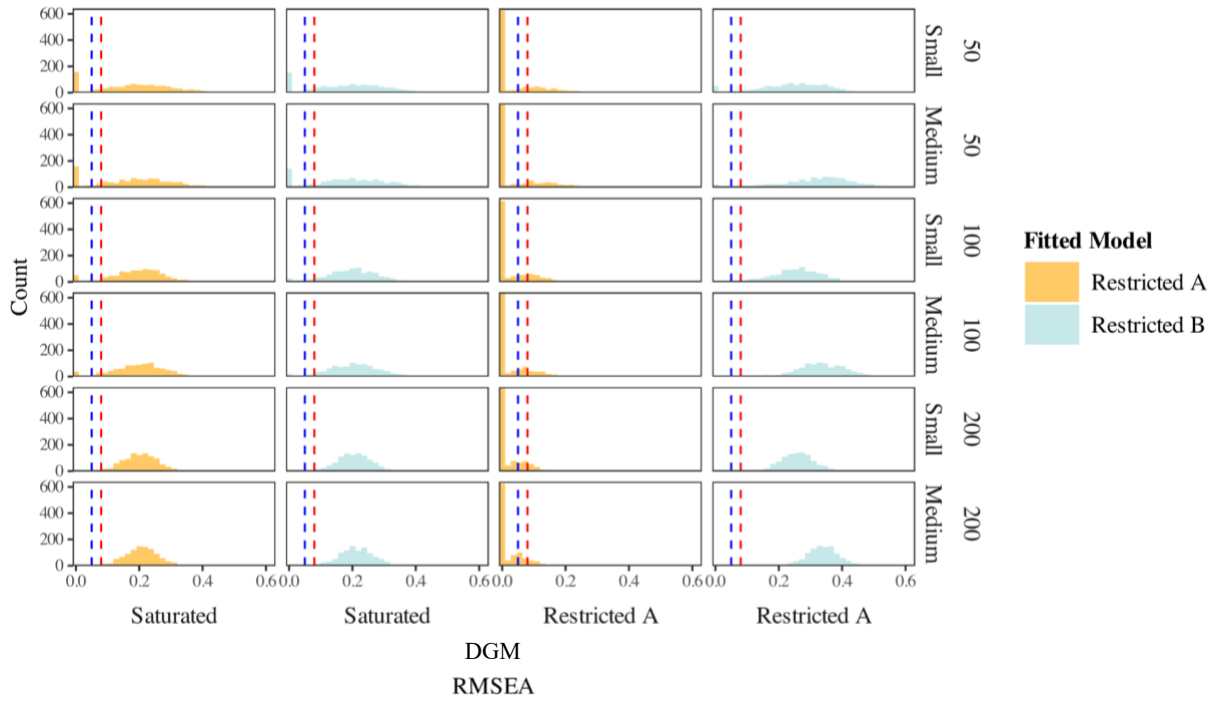


Figure 9 *Distributions of RMSEA by Data Generating and Fitted Model*



Hypothesis 1 (Efficiency Loss Due to Over-parameterization)

Hypothesis 1 explored statistical efficiency differences associated with fitting saturated versus restricted models. This was first evaluated by considering whether there were differences in SEs for the paths that comprised the specific indirect effect, whether confidence intervals were narrower, and whether power was greater when using a properly restricted model (Restricted Model A) relative to a saturated model. Table 4 presents η_G^2 effect size results for within-subjects ANOVAs for the SEs and CI widths. Cell values are contained in Tables 5 and 6.

Table 4 *Hypothesis 1 Efficiency Within-Subjects ANOVA η_G^2 Values by Outcome*

	$SE_{\gamma_{11}}$	$SE_{\beta_{21}}$	$SE_{\beta_{32}}$	CI Width
N	.73	.74	.74	.41
ES	.02	.01	.05	.36
DGM	-	.01	-	.04
FM	-	.03	.04	.01
N*ES	-	-	-	.03
N*DGM	-	-	-	-
N*FM	-	-	-	-
ES*DGM	-	-	-	-
ES*FM	-	-	-	-
DGM*FM	-	-	.02	.03
N*ES*DGM	-	-	-	-
N*ES*FM	-	-	-	-
N*DGM*FM	-	-	-	-
ES*DGM*FM	-	-	-	-
N*ES*DGM*FM	-	-	-	-

Note. All presented effects are significant at .05 level; omitted effects are $\eta_G^2 < .01$. N = sample size, ES = Effect size, DGM = data generating model, FM = Fitted model, CI = confidence interval.

Based on Table 4, data generating and fitted model resulted in meaningful differences in the standard errors for β_{21} and β_{32}, but not γ_{11}. In other words, the combinations of population and fitted models yielded differences in efficiency for the paths exclusively between endogenous variables involved in the indirect effect of interest. Specifically, the SE for β_{21} was .007 standard deviations lower when using a properly restricted as opposed to over-parameterized model, $t(11988) = 66.5$, $p < .001$. There was a similar trend for β_{32} ($b = -.005$, $t(11988) = 56.4$, $p < .001$) and the confidence interval for the indirect

effect (b = -.007, t(11988) = 31.8, p < .001). These effects, however, are generally considered very small. Unsurprisingly, Table 4 also shows that sample and effect size exerted large effects on the SEs and confidence interval width, with larger samples leading to smaller SEs and more narrow CIs, and larger effect sizes yielding the opposite effects.

Turning to power, for a medium effect size, GEE models showed the odds of detecting the specific indirect effect to be 31% lower when using an overparameterized saturated model relative to a properly restricted model, *OR* = .69, 95% CI = [.62, .76] (see Table 5, Figure 10). This discrepancy was slightly less when considering a small-medium effect (*OR* = .74, 95% CI = [.67, .82]) and, based on Figure 10, was most pronounced in medium power contexts (N = 50). Power was also generally higher for restricted models even when the DGM was saturated, but when the DGM was Restricted Model A, power was only greater when the fitted models' restrictions matched those in Restricted Model A. More of the total effect from x to y_3 was associated with the indirect effect upon the removal of the cross-paths (Restricted A) and both cross and direct paths (Restricted B), hence yielding greater power to detect the indirect effect.

Figure 10 Power of Indirect Effect by Combination of Data Generating and Fitted Model

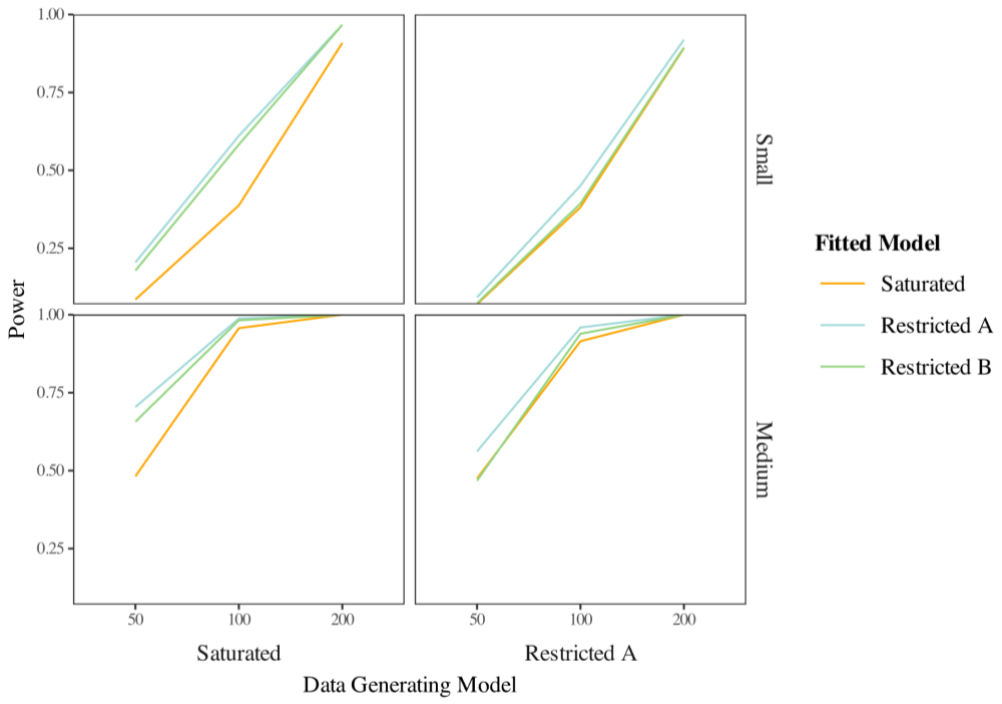


Table 5 *Outcomes of Interest for Specific Indirect Effect of* $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$

N	Effect Size	DGP	Fitted Model	Estimate	Power	CI Coverage	CI Width	Relative Bias
50	.019	Saturated	Saturated	.020	.09	.931	.089	**.06**
			Restricted A	.030	.21	.951	.108	**.61**
			Restricted B	.028	.18	.948	.104	**.52**
		Restricted A	Saturated	.019	.07	.931	.088	.02
			Restricted A	.019	.09	**.914**	.081	.04
			Restricted B	.019	.07	**.919**	.087	.03
	.059	Saturated	Saturated	.058	.48	.929	.152	-.02
			Restricted A	.085	.70	.936	.185	**.44**
			Restricted B	.078	.66	.946	.175	**.31**
		Restricted A	Saturated	.060	.48	.928	.155	.01
			Restricted A	.060	.56	**.912**	.138	.00
			Restricted B	.058	.47	**.920**	.148	-.01
100	.019	Saturated	Saturated	.019	.39	.930	.052	.02
			Restricted A	.028	.61	.932	.069	**.52**
			Restricted B	.027	.58	.946	.065	**.43**
		Restricted A	Saturated	.019	.38	.935	.052	.00
			Restricted A	.019	.45	.924	.050	.01
			Restricted B	.019	.39	**.915**	.052	.00
	.059	Saturated	Saturated	.059	.96	.944	.100	.00
			Restricted A	.086	.99	**.900**	.128	**.44**
			Restricted B	.079	.98	.924	.120	**.33**
		Restricted A	Saturated	.058	.92	.940	.100	-.02
			Restricted A	.059	.96	.938	.093	-.01
			Restricted B	.059	.94	.934	.100	.00
200	.019	Saturated	Saturated	.019	.91	.942	.034	.01
			Restricted A	.027	.97	**.902**	.047	**.48**
			Restricted B	.026	.97	.922	.044	**.41**
		Restricted A	Saturated	.019	.89	.946	.034	.00
			Restricted A	.018	.92	.950	.033	-.01
			Restricted B	.018	.89	.947	.034	-.01
	.059	Saturated	Saturated	.059	1	.941	.069	.00
			Restricted A	.085	1	**.793**	.090	**.44**
			Restricted B	.079	1	**.865**	.084	**.33**
		Restricted A	Saturated	.060	1	.942	.070	.02
			Restricted A	.060	1	.947	.065	.01
			Restricted B	.060	1	.946	.070	.02

Note. DGM = Data generating model, CI = confidence interval. Effect size values are true values for indirect effect. Estimates and confidence interval widths are across-replication means for each cell. Bolded relative bias values are those greater than .05.

Table 6 *Outcomes of Interest for Effects that Comprise* $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$

				γ_{11}				β_{21}				β_{32}			
N	Effect Size	DGM	Fitted Model	Estimate	SE	Relative Bias	Power	Estimate	SE	Relative Bias	Power	Estimate	SE	Relative Bias	Power
50	.265	Saturated	Saturated	.268	.137	.010	.51	.269	.140	.014	.49	.271	.139	.023	.51
			Restricted A	.268	.137	.010	.51	.324	.136	**.221**	.65	.325	.135	**.228**	.67
			Restricted B	.268	.137	.010	.51	.324	.136	**.221**	.65	.314	.137	**.183**	.63
		Restricted A	Saturated	.267	.138	.008	.51	.269	.144	.013	.48	.266	.137	.004	.51
			Restricted A	.267	.138	.008	.50	.270	.137	.019	.52	.266	.130	.005	.54
			Restricted B	.267	.138	.008	.51	.270	.137	.019	.52	.267	.145	.006	.48
	.39	Saturated	Saturated	.388	.133	-.004	.82	.387	.136	-.009	.80	.388	.133	-.006	.82
			Restricted A	.388	.133	-.004	.82	.462	.126	**.184**	.93	.465	.124	**.192**	.95
			Restricted B	.388	.133	-.004	.82	.462	.127	**.184**	.93	.428	.131	**.098**	.89
		Restricted A	Saturated	.391	.131	.002	.82	.394	.146	.009	.76	.388	.123	-.005	.87
			Restricted A	.391	.131	.002	.82	.392	.133	.004	.82	.390	.113	-.001	.91
			Restricted B	.391	.131	.002	.82	.392	.133	.004	.81	.384	.136	-.015	.80
100	.265	Saturated	Saturated	.266	.097	.003	.76	.273	.097	.029	.79	.260	.096	-.017	.77
			Restricted A	.266	.096	.003	.76	.326	.095	**.229**	.92	.315	.094	**.187**	.90
			Restricted B	.266	.096	.003	.75	.326	.095	**.229**	.92	.302	.096	**.140**	.87
		Restricted A	Saturated	.273	.096	.028	.80	.265	.101	.001	.74	.261	.094	-.015	.78
			Restricted A	.273	.096	.028	.80	.265	.096	.000	.78	.261	.090	-.015	.80
			Restricted B	.273	.096	.028	.80	.265	.096	.000	.78	.260	.100	-.019	.72
	.39	Saturated	Saturated	.388	.092	-.006	.99	.391	.095	.003	.98	.390	.091	.000	.99
			Restricted A	.388	.092	-.006	.99	.468	.089	**.200**	1	.467	.086	**.196**	1
			Restricted B	.388	.092	-.006	.99	.468	.089	**.200**	1	.433	.091	.109	1
		Restricted A	Saturated	.388	.092	-.006	.98	.385	.101	-.013	.96	.391	.086	.002	.99
			Restricted A	.388	.092	-.006	.98	.388	.092	-.005	.99	.391	.080	.003	1
			Restricted B	.388	.092	-.006	.98	.388	.092	-.005	.99	.394	.096	.011	.97
200	.265	Saturated	Saturated	.265	.068	.000	.98	.265	.069	.001	.97	.267	.067	.008	.97
			Restricted A	.265	.068	.000	.98	.318	.067	**.200**	1	.320	.066	**.209**	1
			Restricted B	.265	.068	.000	.98	.318	.067	**.200**	1	.309	.067	**.165**	1
		Restricted A	Saturated	.269	.068	.016	.97	.262	.071	-.010	.95	.263	.066	-.008	.97
			Restricted A	.269	.068	.016	.97	.261	.068	-.014	.97	.263	.063	-.007	.98
			Restricted B	.269	.068	.016	.98	.261	.068	-.014	.97	.261	.070	-.013	.95

Table 6 *Outcomes of Interest for Effects that Comprise* $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$

				γ_{11}				β_{21}				β_{32}			
N	Effect Size	DGM	Fitted Model	Estimate	SE	Relative Bias	Power	Estimate	SE	Relative Bias	Power	Estimate	SE	Relative Bias	Power
	.39	Saturated	Saturated	.389	.065	-.003	1	.391	.066	.003	1	.389	.064	-.003	1
			Restricted A	.389	.065	-.003	1	.469	.062	**.202**	1	.465	.060	**.192**	1
			Restricted B	.389	.065	-.003	1	.469	.062	**.202**	1	.432	.064	**.108**	1
		Restricted A	Saturated	.390	.065	.001	1	.392	.071	.005	1	.393	.060	.009	1
			Restricted A	.390	.065	.001	1	.392	.065	.006	1	.392	.056	.005	1
			Restricted B	.390	.065	.001	1	.392	.065	.006	1	.394	.067	.010	1

Note. DGM = Data generating model. Bolded relative bias values are those greater than .05.

Hypothesis 2 (Type I Error Rates)

Hypothesis 2 explored the family-wise Type I Error rate for paths not present in the population data generating model. This hypothesis considered settings where the true data generating model was Restricted Model A, but the fitted model was either over-parameterized (Saturated) or both *over*- and *under*-parameterized (Restricted Model B). Fitting the saturated model resulted in Type I Error rates close to the expected 5% level for both γ_{21} and β_{31}. As predicted, the family-wise error rates were also all above 10% for these models. For the model that was both over- and under-parameterized (Restricted B), the Type I Error rates for β_{31}were notably higher than 5% due to model misspecification (Table 7).

Table 7 *Outcomes of Interest for γ_{21} and β_{31}*

			γ_{21}			β_{31}			Family-wise
N	Effect Size	Fitted Model	Estimate	SE	Type I Error	Estimate	SE	Type I Error	Type I Error
50	.265	Saturated	.004	.145	.05	.000	.142	.07	.11
		Restricted B	-	-	-	.100	.145	.11	-
	.39	Saturated	-.007	.147	.06	.004	.134	.06	.12
		Restricted B	-	-	-	.175	.137	.28	-
100	.265	Saturated	-.003	.100	.07	-.001	.097	.06	.12
		Restricted B	-	-	-	.097	.100	.18	-
	.39	Saturated	.009	.101	.05	.001	.092	.05	.10
		Restricted B	-	-	-	.170	.096	.45	-
200	.265	Saturated	-.004	.071	.06	.001	.068	.06	.12
		Restricted B	-	-	-	.096	.070	.28	-
	.39	Saturated	.001	.071	.06	-.005	.065	.05	.10
		Restricted B	-	-	-	.165	.068	.67	-

Note. Restricted data generating model A for all presented results.

Hypothesis 3 (Consequences of Under-parameterization)

Hypothesis 3 investigated possible bias and differences in statistical efficiency for the specific indirect effect when using improperly restricted models relative to properly saturated models. Based on the ANOVA results in Table 8, the combination of data generating and fitted model yielded significant differences in the relative bias for β_{21} and β_{32}, but not γ_{11} (again isolated to pathways between endogenous variables). Specifically, relative to a properly saturated model, fitting Restricted Model A or B resulted in 21% greater relative bias (*SE* = .002, *t*(11988) = 122.9, *p* < .001) in estimates of β_{21}. The relative bias of β_{32} estimates was also 20% greater when fitting Restricted Model A (*SE* = .002, *t*(11988) = 131.4, *p* < .001) and 13% greater when fitting Restricted Model B (*SE* = .002, *t*(11988) = 82, *p* < .001). As hypothesized, there were also significant differences in the relative bias for the indirect effect dependent on data generating and fitted model combination due to downstream bias propagation (Table 8). Relative to a properly saturated model, fitting Restricted Model A resulted in 47% greater relative bias in the indirect effect (*SE* = .004, *t*(11988) = 126.2, *p* < .001), whereas fitting Restricted Modal B resulted in 38% greater relative bias (*SE* = .004, *t*(11988) = 92.7, *p* < .001).

Finally, assuming a medium effect size and a truly saturated DGM, confidence interval coverage for the indirect effect was significantly lower when using Restricted Model A (*OR* = 1.87, 95% CI = [1.30, 2.71]) relative to a properly saturated model at *N* = 100. Moreover, when considering a medium effect size at *N* = 200, confidence interval coverage was drastically lower when fitting either Restricted Model A (*OR* = 4.16, 95% CI = [2.91, 5.96]) or B (*OR* = 2.49, 95% CI = [1.73, 3.57]), Table 4, Figure 11.

Interestingly, coverage rates decreased as sample size increased when fitting under-parameterized models, even reaching as low as 79%. When *N* = 50, the two misspecified fitted models unexpectedly outperformed the correctly saturated one, but all coverage rates were still generally close to 95% (see Figure 11). This may be due to greater influence of misspecification bias in concentrations of confidence intervals as *N* increases. Confidence interval coverage is a product of both sampling error and bias; as sampling error and intervals widths decrease, bias in the estimates (however small) takes on a greater role in determining CI coverage of the true effect for misspecified models. In other words, CIs for

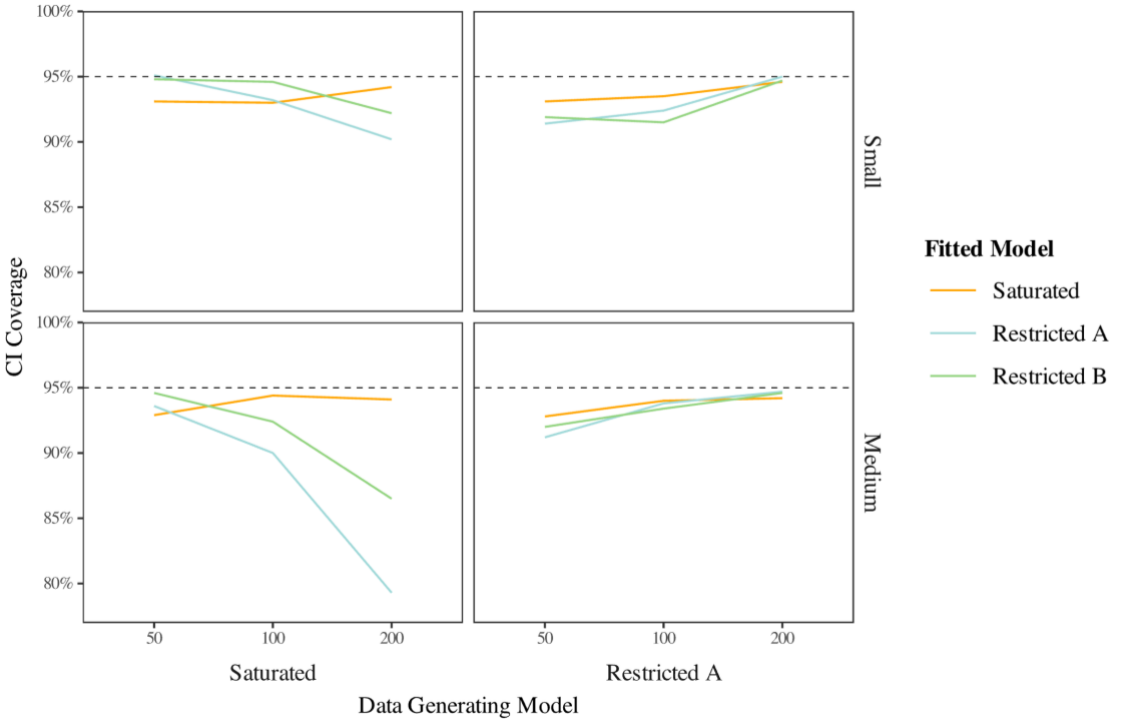
misspecified models initially cast a wide net due to high sampling variability, but then (with larger *N*) become more efficient and center on biased estimates, ultimately leading to markedly reduced confidence interval coverage of the true indirect effect. Confidence interval coverage values less than or equal to .92 were considered underestimated and are bolded in Table 6 (Bradley, 1978).

Table 8 *Hypothesis 3 and 4 Relative Bias Within-subjects ANOVA* η_G^2 *Values by Outcome*

	Indirect Effect	γ_{11}	β_{21}	β_{32}
N	-	-	-	-
ES	-	-	-	-
DGM	.04	-	.04	.03
FM	.02	-	.02	.02
N*ES	-	-	-	-
N*DGM	-	-	-	-
N*FM	-	-	-	-
ES*DGM	-	-	-	-
ES*FM	-	-	-	-
DGM*FM	.02	-	.02	.02
N*ES*DGM	-	-	-	-
N*ES*FM	-	-	-	-
N*DGM*FM	-	-	-	-
ES*DGM*FM	-	-	-	-
N*ES*DGM*FM	-	-	-	-

Note. All presented effects are significant at .05 level; omitted effects are $\eta_G^2 < .01$. N = sample size, ES = Effect size, DGM = Data generating model, FM = Fitted model, CI = Confidence interval.

Figure 11 Indirect Effect Coverage by Combination of Data Generating and Fitted Model



Hypothesis 4 (Over- and Under- Parameterization)

Hypothesis 4 considered the combination of Model 2 (Restricted A) as the DGM and Model 3 (Restricted B) as the fitted model, which represented both under-parameterization (missing the direct effect) and over-parameterization (extra cross path). Despite generally statistically significant meta-models, relative bias in the estimates of $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$, γ_{11}, β_{21}, or β_{32} did not exceed .05 when considering this model combination (Tables 5, 6, and 8).

CHAPTER 4: DISCUSSION

The goal of the present research was to evaluate the degree to which model specification choices in mediation analysis affect power, efficiency, type I error, and accuracy of estimates—particularly those of the indirect effect and the paths that comprise it. Specifically, the conditions explored model combinations that yielded correctly specified, over-parameterized, under-parameterized, and both over- and under-parameterized fitted models. Based on the simulation results, there are notable tradeoffs regarding power, bias, and confidence interval coverage for the indirect effect, arising from differences in SEs and relative bias for β_{21} and β_{32}, the two paths between exclusively endogenous variables included in the indirect effect. Summarizing across the results, I first discuss the costs and benefits of fitting an over-parameterized model, then those of an under-parameterized model, and finally the consequences of making both types of specification errors in the same model. I then discuss limitations, possible future directions, and general implications for applied researchers using mediation analysis to evaluate their hypotheses.

Fitting an over-parameterized saturated model resulted in inflated SEs for β_{21} and β_{32}, which led to wider confidence intervals and decreased power for detecting the indirect effect. In other words, there were notable decreases in statistical efficiency when estimating too many structural parameters. As expected, larger sample and effect sizes also led to increased power to detect the indirect effect regardless of model combination. The family-wise Type I error rate was also greater than 10% (relative to the oft-desired 5%) when estimating non-existent paths in over-parameterized saturated models. That is, there was an increased probability of incorrectly deeming either cross path as significant (when they were actually zero in the population) when fitting an over-parameterized saturated model. These costs to efficiency and power, however, come with a benefit with respect to bias. Specifically, estimates are

unbiased when fitting an over-parameterized saturated model given no non-zero paths are omitted (White, 1982).

Fitting a restricted model carries its own risks, however, due to the potential for omitting important structural paths, i.e., under-parameterization. In circumstances where the true population data generating model was saturated, estimating a model with improper restrictions led to greater relative bias in the estimates of β_{21}, β_{32}, and the indirect effect, but not γ_{11}. As expected, the bias in β_{21} and β_{32} due to misspecification accumulated across and chain and presented in the indirect effect. The estimates of γ_{11}, however, were not biased because it was exogenous and none of the omitted paths in either Restricted Model A or B affected both x and y_1 (Kaplan, 1990). This is also demonstrated by the covariance expression in Equation 14.

Lastly, fitting a model that is both over- and under-parameterized (Restricted Model B relative to Restricted Model A) did not lead to biased estimates of $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$, γ_{11}, β_{21}, or β_{32}. This is because bias caused by the omission of γ_{31}, the direct effect, was compensated for by the inclusion of β_{31}, a cross-path (Kaplan & Wenger, 1993). When referring to the covariance expressions in Equations 14, 15, 16, γ_{31} and β_{31} only affect the variance of y_3. When substituting values of 0 for the paths omitted in Restricted Model A, the variance of y_3 becomes:

$$\sigma^2_{y_3} = \beta^2_{32} + \gamma^2_{31} \tag{22}$$

For Restricted Model B, the variance of y_3 is modeled as:

$$\sigma^2_{y_3} = \beta^2_{32} + \beta^2_{31} \tag{23}$$

As shown in the variance expressions above, β_{31} directly compensates for γ_{31} when assuming constant values of $\sigma^2_{y_3}$ and β_{32}. The misspecification is restricted to this part of the model and does not affect estimates of the indirect effect. This follows previous work which demonstrated that the pattern of zero and non-zero elements in the estimated covariance matrix of the estimates (the inverse of the Fisher information matrix) determines whether misspecification errors will propagate to other areas of the structure (Kaplan, 1990; Kaplan & Wenger, 1993).

Interestingly, there also appears to be a paradoxical benefit of estimating a misspecified restricted model with small *N* when the true population model is saturated. In small *N* (e.g., 50), both misspecified restricted models outperformed the saturated model in terms of confidence interval coverage. That is, the 95% bootstrapped confidence intervals captured the true value of the indirect effect 95% of the time when the fitted models were improperly restricted but *not* when they were properly saturated. This relation held at $N = 100$ but reversed at $N = 200$ when considering the small-medium effect, with the saturated model approaching 95% coverage and the restricted models' coverage rates steadily declining. This decrease is much more salient when considering a medium effect size; while the restricted models still outperformed the saturated one at $N = 50$, the restricted models had drastically lower coverage rates at $N = 100$ and 200. In contrast, an over-parameterized saturated model had generally greater CI coverage than either restricted model when the true data generating model was restricted.

The general trend of the coverage rates for the properly saturated model converging at 95% as *N* increases is expected given the asymptotic properties of the percentile bootstrap CIs (Efron & Tibshirani, 1986). Interestingly, Restricted Model B, which aligned with neither data generating model, appeared to have greater coverage than Restricted Model A when the data generating model was saturated. Both models omitted two paths, but Restricted Model A restricted both cross-paths to 0 whereas Restricted Model B restricted 1 cross-path and the direct effect to 0. By restricting the direct effect to 0 in Restricted Model B, more of the total effect from $y_1 \rightarrow y_3$ was attributable to $y_1 \rightarrow y_2 \rightarrow y_3$, thus increasing both power and coverage for the indirect effect. A parameterization such as Restricted Model B may not be as common, however, as it is generally recommended to include the direct path for most theoretical mechanisms due to the possibility of unmodeled mediators and covariates (Hayes, 2022; MacKinnon et al., 2007).

Limitations

There are several limitations of this study. First, all variables were assumed to be continuous, linearly related, and perfectly measured. There were a limited number of sample and effect size combinations considered, but it appears that statistical power approached 1 at $N = 200$ in the conditions

with a medium effect. The general serial model structure considered in this thesis also does not include covariates; future work may explore how power and coverage fluctuate with the inclusion of additional covariates or mediators. Additionally, maximum likelihood, a large sample estimator, was used for all model estimation. While there may be differences in naïve standard errors relative to OLS due to the $N - p$ degrees of freedom correction for any given model equation (where p is the number of coefficients estimated, including intercepts), bootstrapped confidence intervals for the indirect effect do not rely on these SEs. Point estimates for the coefficients from ML and OLS are identical for the models considered here, and thus results pertaining to the indirect effect should generalize to mediation analyses conducted within an OLS regression modeling framework.

Finally, as demonstrated by the model fit measures, not all estimated models would have been reasonably interpreted based on conventional cutoffs. However, a non-negligible proportion of models would have been deemed "acceptable" even with stringent model χ^2, CFI, and RMSEA cutoffs and thus warrant discussion. As would be expected, the model χ^2 test correctly rejected more improper models as sample size (and degrees of freedom) increased, whereas the CFI and RMSEA were more conservative even when models were correctly specified. It is important to note that these model fit measures are only available for restricted ML SEM models and are not accessible when implementing an equation-by-equation OLS regression framework.

Recommendations for Applied Researchers

Researchers are encouraged to specify mediation models based on strong theoretical justification, rather than defaulting to saturated models that estimate all possible paths. When theoretically supported, a more parsimonious, restricted model can improve power to detect the indirect effect and reduce family-wise Type I error rates. While not directly examined here, modification indices may also be used to identify potential misspecifications and lead to theoretically justified model respecifications. That said, restricted models also carry the risk of imposing incorrect constraints, which can bias individual endogenous path estimates—particularly those contributing to the indirect effect. This, however, largely depends on the pattern of misspecification. All told, researchers should carefully consider which structure

is most appropriate for their hypothesized mechanism and then thoroughly evaluate measures of model fit instead of defaulting to saturated models.

Readily available macros for popular software packages increase accessibility for conducting mediation analyses using OLS by automating bootstrapping and providing saturated starting points for commonly encountered models. Importantly, PROCESS, a commonly used macro available for R, SAS, and SPSS, also allow users to impose specific constraints on structural paths by way of specifying custom matrices to facilitate estimation of restricted models in an OLS regression modelling framework (Hayes, 2022). It is unclear, however, how often researchers take advantage of this feature described in detail in the appendix of the accompanying book (Hayes, 2022).

Researchers are also strongly encouraged to estimate mediation models within a SEM framework, as ML provides measures of model fit and modification indices, which can help identify potential misspecifications when estimating restricted structures. While many models in this simulation may not have been interpretable with small sample sizes based on model fit measures, an OLS regression approach would not have provided such informative values in the first place. Consequently, it is possible that *all* misspecified models would have been interpreted in an OLS setting. Additionally, it is important to remember that naïve SEs—but not point estimates—are biased for ML in small samples. However, if the primary focus is on the indirect effect and the researcher is using percentile bootstrapped confidence intervals based on point estimates, then using path analysis with ML should present no significant drawbacks.

If considering open-source software, structural equation models may be estimated in R using 'lavaan', or also using Jamovi, a point-and-click graphical user interface for R (R Core Team, 2025; Rosseel et al., 2024; The jamovi project, 2025). There are also many other software packages, such as Mplus, SAS, JMP, and SPSS, that can be used to estimate SEM (IBM Corp., 2021; Muthén & Muthén, 2023; SAS Institute Inc., 2021b, 2021a).

Summary

In sum, model specification choices in mediation analysis represent nuanced tradeoffs between statistical efficiency, power, coverage rates, and bias. Over-parameterized saturated models offer potential protection against misspecification bias but at the cost of inflated standard errors, wider confidence intervals, and reduced power to detect indirect effects—especially in smaller samples. In contrast, restricted models have greater power and adequate Type I error rates when correctly specified, but carry the risk of bias in endogenous path estimates when incorrect restrictions are imposed. Notably, even in the presence of model misspecification, the indirect effect remained robust across several conditions, likely due to the structural isolation and separability of the misspecified paths. These findings align with prior theoretical work demonstrating that bias propagation depends on the pattern of covariances among parameters, and that not all misspecifications affect all estimates equally (Kaplan & Wenger, 1993). Ultimately, model specification should be guided by theoretical justification. As demonstrated in this study, thoughtful modeling choices can help balance the statistical tradeoffs of over- and under-specification, and even models with some degree of misspecification may still produce robust estimates of the indirect effect under certain conditions.

APPENDIX: SIMULATION DESIGN MATRIX

Table 1 *Simulation Design Matrix*

DGM	Fitted Model	Effect Size	Sample Size
Saturated	Saturated	.27	50
Saturated	Saturated	.39	50
Saturated	Saturated	.27	100
Saturated	Saturated	.39	100
Saturated	Saturated	.27	200
Saturated	Saturated	.39	200
Saturated	Restricted Model A	.27	50
Saturated	Restricted Model A	.39	50
Saturated	Restricted Model A	.27	100
Saturated	Restricted Model A	.39	100
Saturated	Restricted Model A	.27	200
Saturated	Restricted Model A	.39	200
Saturated	Restricted Model B	.27	50
Saturated	Restricted Model B	.39	50
Saturated	Restricted Model B	.27	100
Saturated	Restricted Model B	.39	100
Saturated	Restricted Model B	.27	200
Saturated	Restricted Model B	.39	200
Restricted Model A	Saturated	.27	50
Restricted Model A	Saturated	.39	50
Restricted Model A	Saturated	.27	100
Restricted Model A	Saturated	.39	100
Restricted Model A	Saturated	.27	200
Restricted Model A	Saturated	.39	200
Restricted Model A	Restricted Model A	.27	50
Restricted Model A	Restricted Model A	.39	50
Restricted Model A	Restricted Model A	.27	100
Restricted Model A	Restricted Model A	.39	100
Restricted Model A	Restricted Model A	.27	200
Restricted Model A	Restricted Model A	.39	200
Restricted Model A	Restricted Model B	.27	50
Restricted Model A	Restricted Model B	.39	50
Restricted Model A	Restricted Model B	.27	100
Restricted Model A	Restricted Model B	.39	100
Restricted Model A	Restricted Model B	.27	200
Restricted Model A	Restricted Model B	.39	200

Note. Effect size reflects the same value for each individual pathway that comprises the specific indirect effect of $x \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$.

REFERENCES

Alwin, D. F., & Hauser, R. M. (1975). The Decomposition of Effects in Path Analysis. *American Sociological Review*, *40*(1), 37–47. https://doi.org/10.2307/2094445

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-Size Planning for More Accurate Statistical Power: A Method Adjusting Sample Effect Sizes for Publication Bias and Uncertainty. *Psychological Science*, *28*(11), 1547–1562. https://doi.org/10.1177/0956797617723724

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384. https://doi.org/10.3758/BF03192707

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

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

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bollen, K. A. (1987). Total, Direct, and Indirect Effects in Structural Equation Models. *Sociological Methodology*, *17*, 37–69. https://doi.org/10.2307/271028

Bollen, K. A. (1989a). A New Incremental Fit Index for General Structural Equation Models. *Sociological Methods & Research*, *17*(3), 303–316. https://doi.org/10.1177/0049124189017003004

Bollen, K. A. (1989b). *Structural Equations with Latent Variables*. John Wiley & Sons.

Bollen, K. A., & Stine, R. (1990). Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability. *Sociological Methodology*, *20*, 115–140. https://doi.org/10.2307/271084

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Chalmers, P. (2015). *SimDesign: Structure for Organizing Monte Carlo Simulation Designs* (p. 2.18) [Dataset]. https://doi.org/10.32614/CRAN.package.SimDesign

Chalmers, P., Sigal, M., Oguzhan, O., & Ronkko, M. (2024). *SimDesign: Structure for Organizing Monte Carlo Simulation Designs* (Version 2.18) [Computer software]. https://cran.r-project.org/web/packages/SimDesign/index.html

Chen, H., Cohen ,Patricia, & and Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*, *39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cohen, J. (1992). Statistical Power Analysis. *Association for Psychological Science*, *1*, 98–101.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Daniel, F., Corporation, M., Weston, S., & Tenenbaum, D. (2022). *doParallel: Foreach Parallel Adaptor for the "parallel" Package* (Version 1.0.17) [Computer software]. https://cran.r-project.org/web/packages/doParallel/index.html

Daniel, F., Ooi, H., Calaway, R., Microsoft, & Weston, S. (2022). *foreach: Provides Foreach Looping Construct* (Version 1.5.2) [Computer software]. https://cran.r-project.org/web/packages/foreach/index.html

Dolezsar, C. M., McGrath, J. J., Herzig, A. J. M., & Miller, S. B. (2014). Perceived racial discrimination and hypertension: A comprehensive systematic review. *Health Psychology*, *33*(1), 20–34. https://doi.org/10.1037/a0033718

Eddelbuettel, D., Francois, R., Bates, D., Ni, B., & details, C. S. R. author. (2024). *RcppArmadillo: "Rcpp" Integration for the "Armadillo" Templated Linear Algebra Library* (Version 14.2.2-1) [Computer software]. https://cran.r-project.org/web/packages/RcppArmadillo/index.html

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*(1), 1–22. https://doi.org/10.1037/1082-989X.12.1.1

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*. https://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478410

Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, *1*(1), 54–75.

Fairchild, A. J., & MacKinnon, D. P. (2014). Using Mediation and Moderation Analyses to Enhance Prevention Research. In Z. Sloboda & H. Petras (Eds.), *Defining Prevention Science* (pp. 537–555). Springer US. https://doi.org/10.1007/978-1-4899-7424-2_23

Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R2 effect-size measures for mediation analysis. *Behavior Research Methods*, *41*(2), 486–498. https://doi.org/10.3758/BRM.41.2.486

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

Fossum, J. (2023). *Moderated Mediation Article Database*. https://www.jlfossum.com/moderated-mediation-article-database

Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, *46*(6), 1454–1480. https://doi.org/10.1037/a0020204

Fox, J. (1980). Effect Analysis in Structural Equation Models: Extensions and Simplified Methods of Computation. *Sociological Methods & Research*, *9*(1), 3–28. https://doi.org/10.1177/004912418000900101

Friendly, M., Fox, J., Chalmers, P., Monette, G., & Sanchez, G. (2024). *matlib: Matrix Functions for Teaching and Learning Linear Algebra and Multivariate Statistics* (Version 1.0.0) [Computer software]. https://cran.r-project.org/web/packages/matlib/index.html

Fritz, M. S., & MacKinnon, D. P. (2007). Required Sample Size to Detect the Mediated Effect. *Psychological Science*, *18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., Maechler, M., & Hothorn, T. (2024). *mvtnorm: Multivariate Normal and t Distributions* (Version 1.3-2) [Computer software]. https://cran.r-project.org/web/packages/mvtnorm/index.html

Gonzalez, O., Valente, M. J., Cheung, J., & MacKinnon, D. P. (2022). Mediation/Indirect Effects in Structural Equation Modeling. In *Handbook of Structural Equation Modeling* (pp. 409–426). Guilford Press.

Greene, V. L. (1977). An Algorithm For Total and Indirect Causal Effects. *Political Methodology*, *4*(4), 369–381.

Harrell, F. E. (2024). *rms: Regression Modeling Strategies* (Version 6.9-0) [Computer software]. https://cran.r-project.org/web/packages/rms/index.html

Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford Publications.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

IBM Corp. (2021). *IBM SPSS Statistics for Windows, Version 27.0*. IBM Corp. https://www.ibm.com/products/spss-statistics

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review*, *105*(4), 765–789. https://doi.org/10.1017/S0003055411000414

John Stuart Mill. (1856). *A System of Logic, Ratiocinative and Inductive: 1* (Vol. 1). Parker.

Kaplan, D. (1990). Evaluating and Modifying Covariance Structure Models: A Review and Recommendation. *Multivariate Behavioral Research*, *25*(2), 137–155. https://doi.org/10.1207/s15327906mbr2502_1

Kaplan, D., & Wenger, R. N. (1993). Asymptotic independence and separability in covariance structure models: Implications for specification error, power, and model modification. *Multivariate Behavioral Research*, *28*(4), 467–482. https://doi.org/10.1207/s15327906mbr2804_4

Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. Guilford Publications.

Kolenikov, S. (2011). Biases of Parameter Estimates in Misspecified Structural Equation Models. *Sociological Methodology*, *41*(1), 119–157. https://doi.org/10.1111/j.1467-9531.2011.01236.x

Larkin, K. T. (2008). *Stress and Hypertension: Examining the Relation between Psychological Stress and High Blood Pressure*. Yale University Press.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. https://doi.org/10.1093/biomet/73.1.13

Lomnicki, Z. A. (1967). On the Distribution of Products of Random Variables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *29*(3), 513–524. https://doi.org/10.1111/j.2517-6161.1967.tb00713.x

MacKinnon, D. (2007). *Introduction to Statistical Mediation Analysis*. Routledge. https://doi.org/10.4324/9780203809556

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology*, *58*(Volume 58, 2007), 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. *Psychological Methods*, *7*(1), 83.

MacKinnon, D. P., Taborga, M. P., & Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence*, *68*, 69–83. https://doi.org/10.1016/S0376-8716(02)00216-8

MacKinnon, D. P., Valente, M. J., & Gonzalez, O. (2020). The Correspondence between Causal and Traditional Mediation Analysis: The Link is the Mediator by Treatment Interaction. *Prevention Science : The Official Journal of the Society for Prevention Research*, *21*(2), 147–157. https://doi.org/10.1007/s11121-019-01076-4

MacKinnon, D., Yoon, M., Ryu, E., & Fairchild, A. (2009a). Evaluation of the proportion mediated measure of mediation. *Manuscript in Preparation*.

MacKinnon, D., Yoon, M., Ryu, E., & Fairchild, A. (2009b). Evaluation of the proportion mediated measure of mediation. *Manuscript in Preparation*.

McNeish, D. (2020). Should We Use F-Tests for Model Fit Instead of Chi-Square in Overidentified Structural Equation Models? *Organizational Research Methods*, *23*(3), 487–510. https://doi.org/10.1177/1094428118809495

Murdoch, D. (2024). *tables: Formula-Driven Table Generation* (Version 0.9.31) [Computer software]. https://cran.r-project.org/web/packages/tables/index.html

Muthén, L. K., & Muthén, B. O. (2002). How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Muthén, L. K., & Muthén, B. O. (2023). *Mplus User's Guide (Version 8.10)*. Muthén & Muthén. https://www.statmodel.com

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, *8*(4), 434–447. https://doi.org/10.1037/1082-989X.8.4.434

Orom, H., Sharma, C., Homish, G. G., Underwood, W., & Homish, D. L. (2017). Racial Discrimination and Stigma Consciousness Are Associated with Higher Blood Pressure and Hypertension in Minority Men. *Journal of Racial and Ethnic Health Disparities*, *4*(5), 819–826. https://doi.org/10.1007/s40615-016-0284-2

O'Rourke, H. P., & MacKinnon, D. P. (2019). The Importance of Mediation Analysis in Substance-Use Prevention. In Z. Sloboda, H. Petras, E. Robertson, & R. Hingson (Eds.), *Prevention of Substance Use* (pp. 233–246). Springer International Publishing. https://doi.org/10.1007/978-3-030-00627-3_15

Pearl, J. (2010a). An Introduction to Causal Inference. *The International Journal of Biostatistics*, *6*(2). https://doi.org/10.2202/1557-4679.1203

Pearl, J. (2010b). Causal Inference. *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, 39–58. https://proceedings.mlr.press/v6/pearl10a.html

Poteat, T., Gallo, L. C., Harkness, A., Isasi, C. R., Matthews, P., Schneiderman, N., Thyagarajan, B., Daviglus, M. L., Sotres-Alvarez, D., & Perreira, K. M. (2021). Influence of Stress, Gender, and Minority Status on Cardiovascular Disease Risk in the Hispanic/Latino Community: Protocol for a Longitudinal Observational Cohort Study. *JMIR Research Protocols*, *10*(5), e28997. https://doi.org/10.2196/28997

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717–731. https://doi.org/10.3758/BF03206553

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo Confidence Intervals for Indirect Effects. *Communication Methods and Measures*, *6*(2), 77–98. https://doi.org/10.1080/19312458.2012.679848

R Core Team. (2025). *R: A language and environment for statistical computing.* (Version 4.4.2) [Computer software]. R Foundation for Statistical Computing. www.R-project.org/

Ripley, B., Venables, B., Bates, D. M., ca 1998), K. H. (partial port, ca 1998), A. G. (partial port, & polr), D. F. (support functions for. (2025). *MASS: Support Functions and Datasets for Venables and Ripley's MASS* (Version 7.3-63) [Computer software]. https://cran.r-project.org/web/packages/MASS/index.html

Rockafellar, R. T. (1993). Lagrange Multipliers and Optimality. *SIAM Review*, *35*(2), 183–238. https://doi.org/10.1137/1035044

Rohrer, J. M., Hünermund, P., Arslan, R. C., & Elson, M. (2022). That's a Lot to Process! Pitfalls of Popular Path Models. *Advances in Methods and Practices in Psychological Science*, *5*(2), 25152459221095827. https://doi.org/10.1177/25152459221095827

Rosseel, Y., Jorgensen, T. D., Wilde, L. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Rockwood, N., Scharf, F., Du, H., Jamil, H., & Classe, F. (2024). *lavaan: Latent Variable Analysis* (Version 0.6-19) [Computer software]. https://cran.r-project.org/web/packages/lavaan/index.html

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation Analysis in Social Psychology: Current Practices and New Recommendations. *Social and Personality Psychology Compass*, *5*(6), 359–371. https://doi.org/10.1111/j.1751-9004.2011.00355.x

SAS Institute Inc. (2021a). *JMP® Pro 16*. SAS Institute Inc. https://www.jmp.com

SAS Institute Inc. (2021b). *SAS/STAT® 15.2 User's Guide*. SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/stat/index.html

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, *13*, 290–312. https://doi.org/10.2307/270723

Sobel, M. E. (2000). Causal Inference in the Social Sciences. *Journal of the American Statistical Association*. https://www.tandfonline.com/doi/abs/10.1080/01621459.2000.10474243

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*(3), 371–384. https://doi.org/10.1007/BF02294623

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr2502_4

Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2008). Tests of the Three-Path Mediated Effect. *Organizational Research Methods*, *11*(2), 241–269. https://doi.org/10.1177/1094428107300344

The jamovi project. (2025). *Jamovi* (Version 2.6) [Computer software]. https://www.jamovi.org

Tibbe, T. D., & Montoya, A. K. (2022). Correcting the Bias Correction for the Bootstrap Confidence Interval in Mediation Analysis. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.810258

Tomarken, A. J., & Waller, N. G. (2005). Structural Equation Modeling: Strengths, Limitations, and Misconceptions. *Annual Review of Clinical Psychology*, *1*(Volume 1, 2005), 31–65. https://doi.org/10.1146/annurev.clinpsy.1.102803.144239

VanderWeele, T. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

Wald, A. (1949). Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, *20*(4), 595–601.

West, S. G., Taylor, A. B., & wu, W. (2012). Model Fit and Model Selection in Structural Equation Modeling. *Handbook of Structural Equation Modeling*, 209–231.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, *50*(1), 1–25. https://doi.org/10.2307/1912526

Wickham, H., & RStudio. (2023). *tidyverse: Easily Install and Load the "Tidyverse"* (Version 2.0.0) [Computer software]. https://cran.r-project.org/web/packages/tidyverse/index.html

Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). Omitted variable bias: A threat to estimating causal relationships. *Methods in Psychology*, *5*, 100075. https://doi.org/10.1016/j.metip.2021.100075

Wright, S. (1923). Mendeleyan Analysis of the Pure Breeds of Livestock: I. The Measurement of Inbreeding and Relationship. *Journal of Heredity*, *14*(8). https://doi.org/10.1093/oxfordjournals.jhered.a102354

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yuan, K.-H., & Bentler, P. M. (2000). 5. Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data. *Sociological Methodology*, *30*(1), 165–200. https://doi.org/10.1111/0081-1750.00078

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257