Planned Missing Data in Mediation Analysis

by

Amanda N. Baraldi

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2015 by the
Graduate Supervisory Committee:

Craig Enders, Co-Chair
David Mackinnon, Co-Chair
Leona Aiken
Jenn-Yun Tein

ARIZONA STATE UNIVERSITY

August 2015

ABSTRACT

This dissertation examines a planned missing data design in the context of mediational analysis. The study considered a scenario in which the high cost of an expensive mediator limited sample size, but in which less expensive mediators could be gathered on a larger sample size. Simulated multivariate normal data were generated from a latent variable mediation model with three observed indicator variables, M_1, M_2, and M_3. Planned missingness was implemented on M_1 under the missing completely at random mechanism. Five analysis methods were employed: latent variable mediation model with all three mediators as indicators of a latent construct (Method 1), auxiliary variable model with M_1 as the mediator and M_2 and M_3 as auxiliary variables (Method 2), auxiliary variable model with M_1 as the mediator and M_2 as a single auxiliary variable (Method 3), maximum likelihood estimation including all available data but incorporating only mediator M_1 (Method 4), and listwise deletion (Method 5).

The main outcome of interest was empirical power to detect the mediated effect. The main effects of mediation effect size, sample size, and missing data rate performed as expected with power increasing for increasing mediation effect sizes, increasing sample sizes, and decreasing missing data rates. Consistent with expectations, power was the greatest for analysis methods that included all three mediators, and power decreased with analysis methods that included less information. Across all design cells relative to the complete data condition, Method 1 with 20% missingness on M_1 produced only 2.06% loss in power for the mediated effect; with 50% missingness, 6.02% loss; and 80% missingess, only 11.86% loss. Method 2 exhibited 20.72% power loss at 80% missingness, even though the total amount of data utilized was the same as Method 1.

Methods 3 – 5 exhibited greater power loss. Compared to an average power loss of 11.55% across all levels of missingness for Method 1, average power losses for Methods 3, 4, and 5 were 23.87%, 29.35%, and 32.40%, respectively. In conclusion, planned missingness in a multiple mediator design may permit higher quality characterization of the mediator construct at feasible cost.

ACKNOWLEDGMENTS

I thank my committee members for their commitment to my academic progress and their invaluable input into this research. I especially thank my mentor, Dr. Craig Enders, for his patience, guidance, and training throughout my time at Arizona State University. I thank Dr. David MacKinnon for acting as a secondary mentor and for all the opportunities he has provided me through my work in the Research in Prevention Lab. I thank Dr. Leona Aiken for her unwavering support and mother-like care during my graduate studies. I thank Dr. Jenn-Yun Tien for her constant enthusiasm and encouragement. Finally, I wish to thank my family and friends for their steadfast support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Research methodology is a continually growing and expanding field with new analytical methods paving the way for variations from traditional research design. Two areas of particular importance to research pertaining to human participants are mediation analyses and statistical methods for accommodating missing data with unbiased results. Many areas of psychology and prevention research study mediated effects (also called indirect effects) to understand the causal chain of relations between three or more variables (e.g., $X \rightarrow M \rightarrow Y$). With mediation analysis, researchers can identify the causal mechanism by which one variable transmits an effect to another. Another important area of research, missing data methodology (e.g., MAR-based methods such as maximum likelihood and multiple imputation procedures), provides researchers with tools to provide estimates of the population parameters being studied when there are missing values.

The advent of modern missing data analyses such as maximum likelihood estimation and multiple imputation provides the opportunity to leverage purposeful missing data. Graham, Taylor and Cumsille (2001) state that researchers often ask, “Why would anyone ever want to plan to have missing data? Is it not always better to limit the amount of missing data one has?” Enders (2010) expresses a similar sentiment by stating that “researchers tend to view the idea of planned missing data with some skepticism and are often reluctant to implement this strategy.” Although the idea of planned missingness may feel counterintuitive, there are many advantages to carefully planned designs incorporating intentional missing data. These advantages include reduction in resources expended (e.g., time and money) and decreased respondent burden, while simultaneously maintaining the desired full scope of research. These designs also directly address practical limitations. For instance, researchers may be interested in a large scope of questionnaire items that is greater than the number of items that might be expected for participants to answer.

One assumption of mediation analysis is that the variables are reliable and valid (MacKinnon, 2008). Furthermore, the literature suggests that reliable and valid measures are important factors in having sufficient power to detect the mediated effect. Unfortunately, the necessity of valid and reliable measures can pose problems in the planning of research designs. In many cases, there are multiple measures of the same construct that researchers might incorporate into their studies. Because resources typically limit the choice of measures and number of participants in a given study, researchers often choose between collecting a large sample of data with an inexpensive measure or collecting a smaller sample of data with a more costly measure. For example, researchers interested in smoking behavior as a mediator may choose between using a less expensive self-report of smoking behavior (which may be underreported due to the undesirability of the behavior) or a more expensive measure of cotinine in saliva. In another example, body composition may be measured inexpensively as Body Mass Index (BMI) computed from self-reported height and weight versus a more accurate and expensive measure of body fat percentage using hydrostatic weighing.

To date, planned missing designs have not been applied to mediation analyses, but they are potentially useful in this context. To address issues of limited resources in a study that hypothesizes mediation when multiple measures of the same mediation construct are available, a researcher may utilize a two-method measurement design as proposed by Graham and colleagues (2006). The two-method measurement design is an innovative way to leverage the statistical benefits of collecting relatively inexpensive and less valid measures on a complete sample of participants and a more expensive, but more accurate, measures on a subsample of participants. In essence, this use of purposeful missing data may maximize both power and accuracy by "borrowing" information from a large sample of the inexpensive mediator(s) and a small subsample of the more expensive measure. Modern missing data analysis methods can be used to analyze this design. This research study evaluates a simplified variation of the Graham two-method measurement design extended to mediation analysis and evaluates other viable methods for incorporating planned missingness in mediation. Specifically, I evaluate the use of intentional missing data in a mediation design incorporating multiple measures of a mediation construct and compare five potential ways to analyze such data.

To delve into the potential for planned missing data in a mediation analysis, there are two distinct pieces of methodology that must be understood: mediation analysis and modern missing data analysis. The first chapter provides the foundational information required to understand the literature in the fields of mediation and missing data analysis. First, I provide a brief summary of mediation analysis. Next, I introduce modern missing data methods. To discuss these methods, I introduce the three missing data mechanisms – these mechanisms can be thought of as underlying assumptions that dictate the performance of a given missing data analysis method. Understanding of these mechanisms is crucial to a planned missing data design. With intentional missing data, the researcher has full control of the missing data mechanism. Next, I provide an overview of both maximum likelihood estimation and multiple imputation for use in missing data analyses. Finally, I provide a short introduction to planned missing data designs.

After laying the foundation of missing data and mediation designs in Chapter 1, Chapter 2 reviews the relevant literature on mediation analysis and planned missing data designs that inform my research. Finally, after reviewing the literature that informs my research, Chapter 3 describes a research study that evaluates the potential for the use of a planned missing data design with mediation analysis, Chapter 4 describes the results of the study, and Chapter 5 discusses the implications of the results.

Mediation Analysis

A mediator is an example of a so-called "third variable" that attempts to clarify or elaborate the relation between an independent variable (IV) and a dependent variable (DV). A mediating variable represents the intermediate member of a causal chain of relations, such that some independent variable, X, causes the mediator variable, M, and M causes the dependent variable, Y (MacKinnon, 2008). In prevention or treatment research, M may be used to understand the mechanism by which an intervention or treatment, X, produces the outcome, Y. Consider a variable, X, representing a dietary intervention, and the outcome variable, Y, measuring body mass index (BMI). A third mediating variable, M, representing dietary habits, may explain how the intervention alters BMI. The mediating variable is said to be the mechanism by which X causes Y (in this example dietary habits is the mechanism by which the intervention changes BMI). As another example, a smoking intervention program (X) may be theorized to increase knowledge of the health consequences of smoking cigarettes (M) which, in turn, may decrease cigarette consumption. The scope of mediation analysis goes well beyond prevention or treatment research and mediation has application in many fields such as psychology, business, and education (see Mackinnon, 2008 for examples).

In the OLS regression framework, the equation that quantifies the direct relationship between *X* and *Y* is below and is represented by path diagram on the left-side of Figure 1.

$$Y = \tau X + i_1 + e_1 \tag{1.1}$$

Here, τ quantifies the relationship between *X* and *Y* without accounting for the mediator. The coefficient, τ, is also referred to as the "total effect." The term i_1 represents the intercept and the term e_1 is the residual error.

The basic mediation model (single intermediate variable) consists of two causal paths (the path diagram on the right side of Figure 1). These causal paths are represented by two equations in the OLS regression framework. The first is the path from the manipulation, *X*, to the theoretical mediator, *M* (α in Figure 1). This path, α, quantifies how the independent variable, *X*, changes the mediator. This relationship can be expressed as follows below.

$$M = \alpha X + i_2 + e_2 \tag{1.2}$$

In this equation, α quantifies the relationship of *M* on *X*. The term i_2 represents the intercept and the term e_2 is the residual error. In the dietary intervention example, α represents the effect of the dietary intervention on dietary habits.

The second path of interest is the path from the mediator, *M*, to the outcome, *Y* (β in Figure 1). This path quantifies and represents the theory of how the mediator, *M*, is related to the outcome variable, *Y*, as represented by the following equation.

$$Y = \tau' X + \beta M + i_3 + e_3 \quad (1.3)$$

In this equation, β represent the relationship of Y on M partialling for the variable, X. In the dietary habit example, β would represent the effect of dietary changes on BMI, accounting for the intervention. The other two terms, i_3 and e_3, represent the intercept and residual terms, respectively. This set of relationships enables researchers to test if the relationship between X and Y is transmitted via the mediator, M. In a study where X is an experimental manipulation (such as a treatment or intervention versus a control), the mediation model measures whether the experimental manipulation changes the outcome through the mediating variable. The other regression coefficient in Equation 1.3, τ', represents the remaining relationship between X and Y after controlling for M. The coefficient, τ', is also known as the "direct effect" and, in the case of complete mediation, $\tau' = 0$. The current study focuses only on complete mediation.

Testing the Mediated Effect

The equations above provide the estimates necessary for determining the presence of a mediated effect. There exist a variety of methods for quantifying and testing the mediating effect. Historically, the Baron and Kenny (1986) causal steps test of mediation has been the most widely used approach to test for a mediated effect (Fritz & MacKinnon, 2007). The methodological literature suggests that this approach is underpowered relative to other tests of mediation (Fritz & MacKinnon, 2007; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). The low power in the causal steps approach is generally attributed to the requirement of a significant X to Y relation, especially in the case of complete mediation (MacKinnon, Fairchild, & Fritz, 2007). Many methodologists have suggested that the requirement that X and Y be related is not necessary; it is possible for mediation to exist even when the relationship between X and Y is not statistically significant (Hayes, 2009; MacKinnon, Krull, & Lockwood, 2000; Rucker, Preacher, Tormala, & Petty, 2011; Shrout & Bolger, 2002; Zhao, Lynch Jr., & Chen, 2010). Besides the causal steps test, other mediation procedures have been proposed including the MacArthur model (Kraemer, Kiernan, Essex, & Kupfer, 2008) and various modifications of the Baron and Kenny causal steps test of mediation (Mackinnon, 2008). One variation, the joint significance test, only requires that the α and β paths are statistically significant to provide support for mediation (MacKinnon, 2008).

The most recent methodological literature generally supports the use of the "product of coefficients" to estimate and assess the mediated effect. In the product of coefficients approach, the mediated effect is based on the product of coefficients α and β, from Equations 1.2 and 1.3 above. The statistical significance of the mediated effect, $\alpha\beta$, is based on this product and corresponding standard error/confidence intervals. Some researchers estimate mediation based on the total effect, τ, minus the direct effect, τ' (from Equations 1.1 and 1.3 above). For linear models without missing data, these two estimates of mediation, $\alpha\beta$ and $\tau - \tau'$, are numerically equivalent (Mackinnon & Dwyer, 1993). However, these two estimates of the mediated effect are not necessarily equivalent for non-linear models or models with missing data. Because the product of coefficients method is the most general method to estimate mediation applicable to a wide range of models, the methodological research focuses on this method. Accordingly, the current study relies on the distribution of the product. Furthermore, there is literature suggesting that in certain situations (including when not all variables are reliable), relying on the relationship between X and Y (either before controlling for the mediator as in the coefficient τ, or after controlling for the mediator as in the coefficient, τ') can lead to misleading or false conclusions in theory testing (Rucker et al., 2011). This is particularly true if there is a highly reliable *M*, but moderately reliable *X* and *Y*. This limitation of the $\tau - \tau'$ estimate of mediation provides further support for my focus on the product of coefficients approach.

With the product of coefficients approach, the mediated effect for a single intermediate mediating variable is estimated based on the product of the coefficients, $\alpha\beta$. By virtue of the fact that two coefficients go into creating the estimate of the mediated effect, additional complexities are added to the computation of the standard error as the standard error must incorporate the coefficients α and β, their corresponding standard errors, and the correlation between α and β (if non-zero). There are a variety of formulas available to esttimate the standard error of the mediated effect. The multivariate delta standard error (Sobel, 1982) is one of the most commonly used standard errors (MacKinnon, 2008; MacKinnon, Fairchild, et al., 2007). However, when α and β are correlated (as typically occurs with missing data), a more accurate standard error is based on the variance of the product of the α and β coefficients derived using a second-order Taylor series (Baron & Kenny, 1986; Mackinnon & Dwyer, 1993; MacKinnon, Fritz, Williams, & Lockwood, 2007). The resulting formula for the standard error of the mediated effect is below.

$$s_{\alpha\beta} = \alpha^2 s_\beta^2 + \beta^2 s_\alpha^2 + s_\alpha^2 s_\beta^2 \tag{1.4}$$

In this formula, s_α^2 and s_β^2 are the squared standard error of α and β, respectively.

One method to test the statistical significance of a mediated effect is to evaluate the ratio of the mediated effect to its standard error as compared to a standard normal distribution as is used for many test statistics that assume a normal distribution. Similarly, a method to create confidence limits for the mediated effect is to calculate them off of the following equations for the Lower Confidence Limit (LCL) and the Upper Confidence Limit (UCL).

$$LCL = \alpha\beta - (z_{Type1\ Error} s_{\alpha\beta}) \tag{1.5}$$

$$UCL = \alpha\beta + (z_{Type1\ Error} s_{\alpha\beta}) \tag{1.6}$$

The term, $z_{Type1\ Error}$, is the value of the z-statistic required for the confidence limit calculations. Under the assumption of a normal distribution, for a 95% confidence limit, $z_{Type1\ Error}$ = ±1.96.

Distribution of the Product. A major issue with the test of significance and confidence interval computation described in the preceding paragraph is that these tests assume the product of α and β is normally distributed. The distribution of the product of two random variables is not typically normal distributed and is most often asymmetric and highly kurtotic (Aroian, 1947; Craig, 1936). Consequently, power based on an assumption of normality (and corresponding symmetric confidence limits) typically results in underpowered tests of mediation (Fritz & MacKinnon, 2007; MacKinnon et al., 2002) . Methods that accommodate the distribution of the product provide more accurate significance tests and greater statistical power (MacKinnon et al., 2002; MacKinnon, Lockwood, & Williams, 2004). Broadly, there are two general ways to accommodate the distribution of the product: (1) using resampling methods such as bootstrapping or (2) relying on the critical values of the distribution of the product (which is typically non-normal). This research project focuses on significance testing relying on the distribution of the product. For more information on resampling methods see Mackinnon (2008), Preacher & Hayes (2008), Shrout and Bolger (2002).

A researcher can test the mediated effect and accommodate the potential non-normal distribution of the distribution of the product with the computation of asymmetric confidence limits. Calculation of confidence limits under normal distribution assumptions, as in Equations 1.6 and 1.7, uses the same *z*-value for both the upper and lower confidence limits. Because the mediated effect typically has a non-normal distribution, more accurate confidence limits for the mediated effect would require different critical values for the upper and lower confidence limits (i.e., different values of $\boldsymbol{z_{Type1\ Error}}$) to appropriately adjust for non-normality. Thus, Equations 1.6 and 1.7 are restated below to account for varying critical values for the upper and lower confidence limits.

$$LCL = \alpha\beta - (ProdLower_{Type1\ Error} s_{\alpha\beta}) \tag{1.7}$$

$$UCL = \alpha\beta + (ProdUpper_{Type1\ Error} s_{\alpha\beta}) \tag{1.8}$$

Here, $z_{Type1\ Error}$ in Equations 1.5 and 1.6 is replaced with $ProdLower_{Type1\ Error}$ and $ProdUpper_{Type1\ Error}$ to explicitly denote different critical values for the lower and upper confidence limits. One way to obtain the critical values based on the distribution of the product is through the use of a program called PRODCLIN (MacKinnon, Fritz, et al., 2007); this program was updated more recently (Tofighi & MacKinnon, 2011) to resolve some of the initial limitations of computing confidence intervals for the case when the coefficients are correlated (as would be expected in missing data analyses). These updates were implemented in both a new R Package and in the original PRODCLIN program; this study uses the updated PRODCLIN program. In PRODCLIN, the user inputs the estimates of α, s_α, β, s_β, the correlation between α and β (if non-zero as is typical with missing data), and the Type 1 error rate for the desired confidence interval. Based on this user input, the PRODCLIN program provide the critical values that rely on the non-normal distribution of the product and calculates the resulting asymmetric confidence interval. In sum, the current literature generally suggests estimating the mediated effect using product of coefficients approach and assessing the statistical significance with a method that relies on the (potentially) asymmetric and kurtotic distribution of the product.

Mediation in the Structural Equation Modeling Framework

Mediation analysis may also be evaluated in the structural equation modeling (SEM) framework. The SEM framework is more flexible than the regression framework and can accommodate more complex models including models that have more than one independent, mediating, and/or dependent variables. The SEM framework can also incorporate latent variables for X, M and/or Y; latent variables are unobserved but inferred from observed variables. In a latent variable model, these observed variables have a variety of names and are commonly referred to as indicator or manifest variables. Importantly, latent variables are free of random or systematic measurement error (Bollen, 1989, p. 11), although this rule is not without exception (e.g., single indicator latent variables). The current study considers a scenario where multiple measures of a mediation construct are available. These mediation variables can be modeled as indicators of a latent mediation variable. Figure 2 illustrates a mediation model with a latent mediator informed by three measures of the mediation construct, M_1, M_2, and M_3. In this model, X and Y are manifest variables. The model in Figure 2 corresponds with one of the analysis models in the current research study. I also use a variation of this model to generate simulated data.

Using this SEM approach, the model is defined by both a measurement model and a structural model. The measurement model defines the relationship of the observed variables to their latent variable constructs and the structural model summarizes the relationships between the latent variables. In Figure 2, instead of an observed mediation variable, a measurement model quantifies a latent mediation variable. Notice that the structural portion of the model is virtually the same as the manifest variable path model on the right side of Figure 1. Consistent with Figure 1 and Equations 1.2 and 1.3, the path quantifying the relationship between X and the mediator is α and the path quantifying the relationship between the mediator and Y, controlling for X is β. In SEM, the structural portion is often expressed using matrix equations, and the matrix equation that represents Figure 2 is:

$$\begin{bmatrix} \mathrm{LM} \\ \mathrm{Y} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta & 0 \end{bmatrix} \begin{bmatrix} \mathrm{LM} \\ \mathrm{Y} \end{bmatrix} + \begin{bmatrix} \alpha \\ \tau' \end{bmatrix} [X] + \begin{bmatrix} \zeta_M \\ \zeta_Y \end{bmatrix}. \tag{1.9}$$

In non-matrix form, this equation can be written as two equations that are virtually identical to Equations 1.2 and 1.3 with the exception that the latent variable, *LM*, replaces the manifest variable M. Replacing the manifest variable, *M*, with a measurement model with multiple indicators of a latent construct, *LM*, addresses issues of measurement error in the mediator.

Missing Data

Missing Data Mechanisms

Before delving into the idea of planned missing data in mediation analysis, it is necessarily to explore general analysis methods used when data are missing. A discussion on missing data methods must first begin with terminology that describes the so-called missing data mechanisms. I use a classification system of missing data mechanisms to describe the relationship between observed variables and the propensity for missingness (Enders, 2010; Rubin, 1976; Rubin & Little, 2002). Missing data analysis techniques are typically dictated by an assumption of a particular mechanism. The most stringent missing data mechanism is missing completely at random (MCAR). Data are MCAR if the propensity for missing data on some variable, *Y*, is completely unrelated to other variables in the data set and analysis and to the unobserved values of *Y* itself (i.e., every participant has the same probability of missing data). As an example, consider a scenario where a random set of questionnaires are lost due to an administrative mistake or a scenario where scheduling difficulties unrelated to a study (or any of the variables in the study) preclude follow-up. MCAR missingness is completely unsystematic and the resulting data represent a random subsample of the hypothetical complete-data. This is the missing data mechanism that is assumed for classic missing data techniques such as pairwise and listwise deletion; if the MCAR assumption is not met, deletion methods provide biased parameters estimate. Many methodologists argue that the MCAR mechanism is rarely met in practice (B. Muthén, Kaplan, & Hollis, 1987; Raghunathan, 2004). However, it is important to note that the planned missingness design used in this study produces an MCAR mechanism.

The second missing data mechanism, missing at random (MAR), describes data where the propensity for missing data on some variable, *Y*, is related to other observed variables but not the hypothetical value of *Y* if it had been observed. This potentially confusing name includes the word “random,” but, in fact, MAR describes a type of systematic missingness. In the MAR-mechanism, missingness is contingent on another observed variable or variables in the analysis, but is unrelated to the values of the incomplete variable(s). For example, consider a study interested in the relationship between stress and consumption of unhealthy snack food. Participants with high Body Mass Index (BMI) scores might feel embarrassed about their snack consumption and participants with higher BMI may be more likely to skip items about snack food consumption. Importantly, after controlling for BMI, no other variables predict missingness (i.e., two people with the same BMI score have the same probability of missing data on snack food question). The MAR mechanism is the assumption underlying the modern missing data techniques that will be focused on in this document.

Lastly, data are considered Missing Not at Random (NMAR), if missing values on *Y* are related to the value of *Y* that would have been obtained had *Y* been observed if they were not missing. In other words, missingness is systematically related to the hypothetical underlying values of the missing data. Consider a study evaluating sexual behavior in high school students. Students with high levels of sexual activity may be more likely to skip questions regarding sexual activity for fear of the repercussions that may occur if parents or teachers saw their answers. Although the NMAR mechanism is easily satisfied in terms of assumptions of missingness, it tends to be the most challenging mechanism from an analytical standpoint.

Formal definitions of the missing data mechanisms. Now that we have a broad conceptual understanding of the missing data mechanisms, I will go into more detail about the precise mathematical underpinnings of these mechanisms. The missing data mechanisms technically describe probability distributions for a missing data indicator, *R*.

This missing data indicator is a binary variable that denotes whether a score is observed (R = 1) or missing (R = 0) for some variable, Y. This variable, Y, is contained in a data set where some values are observed, Y_{obs}, and some values are missing, Y_{mis}. The general distribution to define a missing data mechanism is:

$$p(R|Y_{obs}, Y_{mis}, \phi). \tag{1.10}$$

Here, ϕ is a parameter or set of parameters that describes the relationship between the missing data indicator, R, and the data (both observed and unobserved). The symbol "|" can be interpreted as "conditional on". Thus this expression demonstrates that the probability of missing data indicator R is conditional on the observed data, the missing unobserved data, and some set of parameters. Rubin's definitions of missing data mechanisms can be differentiated based on the quantities to the right of the "conditional on" symbol.

Let's now consider the MCAR mechanism. Data are MCAR if missingness is unrelated to any of the observed or missing values in the data. Thus, when the MCAR mechanism is satisfied, the probability distribution in Equation 1.10 can be reduced as follows.

$$p(R|Y_{obs}, Y_{mis}, \phi) = p(R|\, \phi) \tag{1.11}$$

Here, because missingness is not related to any of the observed or missing data, Y_{obs} and Y_{mis} have no bearing on the probability of R and can be removed from the conditional statement. Said differently, each case has the same probability of missing data on Y. MCAR is the mechanism that has historically been assumed for many of the traditional missing data approaches (Enders, 2010). As a result, researchers often find the MCAR assumption the most convenient of the missing data mechanisms because traditional approaches tend to be easier to implement and many are unbiased under MCAR (e.g. deletion methods).

Next, consider the conceptual definition of the MAR mechanism. Data are considered MAR if the probability of missingness on some variable is related to other variables in the analysis but not related to the would-be values of the missing data had the variable been observed. Said differently, after partialling out all other variables, there is no relationship between R and Y_{mis}. For the MAR mechanism, our general probability distribution reduces to:

$$p(R|Y_{obs}, Y_{mis}, \phi) = p(R|\, Y_{obs}, \phi). \tag{1.12}$$

Equation 1.12 demonstrates that, given the observed values in the data set, the probability of R does not depend on the missing values. As such, we remove Y_{mis} from the conditional portion of the expression.

Finally, there is the NMAR mechanism; this mechanism is the most difficult to model. Recall that data are NMAR when missingness on some variable is dependent on the hypothetically complete variable. In other words, NMAR assumes all possible associations between the missing data indicator and the observed and missing data values. Consequently, the expression in Equation 1.10 does not reduce any further when representing the NMAR mechanism. From this probability distribution, we can see that the probability of missing data on some variable can depend on both the observed and missing values.

As noted, when discussing these mechanisms I have referred to both data and analysis. Technically missing data mechanisms are not characteristics of an entire data set (although you will often mistakenly see the mechanisms described in this fashion).

Instead, the mechanisms are assumptions that are specific to the variables included in a particular analysis or imputation model. In fact, depending on which variables are included in the analysis, the same data set may produce analyses that satisfy all three of the mechanisms.

Generally, the missing data mechanisms are untestable assumptions. Fortunately, in a planned missing data design, missingness is intentionally planned and the researcher has full control over the missingness mechanism (typically MCAR). Modern missing data methods, maximum likelihood estimation and multiple imputation, typically rely on the assumption of the MAR mechanism; I focus on maximum likelihood in this document. Note that because the MAR mechanism is a less stringent assumption than the MCAR mechanism (see Equations 1.11 and 1.12), MAR-based methods also provide unbiased estimates under the MCAR mechanism. Although traditional deletion methods (i.e., listwise and pairwise deletion) are unbiased when the MCAR mechanism is satisfied, the use of these MAR-based methods for when the MCAR assumption is satisfied is generally recommended because eliminating data is wasteful and can significantly reduce power (Enders, 2010). As a note, for planned missing data scenarios that will be discussed later, the researcher knows the exact mechanism underlying the planned missingess. Although the planned missing scenario I introduce later will satisfy the MCAR assumption, MAR-based analyses are still superior due to their potential for improving power.

Maximum Likelihood Estimation

In the 1970's, two "state of the art" missing data methods for MAR-based analyses (Schafer & Graham, 2002) were introduced in the methodological literature:

maximum likelihood estimation and multiple imputation (Beale & Little, 1975; Dempster & Laird, 1977; Rubin, 1978b). These missing data methods provide researchers with tools to estimate unbiased population parameters (under certain assumptions) when there are missing values. Furthermore, these methods enable researchers to include all available data in the analysis (unlike the traditional deletion methods of listwise and pairwise deletion) resulting in greater power than the traditional methods. As an important note, maximum likelihood estimation and multiple imputation are asymptotically equivalent and there is no reason to expect differences among the two approaches. This document and the research study presented focus only on maximum likelihood estimation in planned missing data designs, because maximum likelihood estimation requires reduced computing resources compared to multiple imputation.

At the broad level, maximum likelihood estimation identifies the population parameter values that most likely produced the sample data. Importantly, maximum likelihood estimation is a common complete-data analysis approach that readily extends to missing data. Before introducing the method in the context of missing data, I start by providing an overview of the basic estimation procedures in the context of a study with complete-data. I later expand the use of maximum likelihood estimation to missing data. In maximum likelihood analysis, the researcher is required to identify the type of distribution that the population data came from. Although maximum likelihood estimation is flexible enough to handle a wide variety of distributions (e.g. binomial distribution for binary outcomes), I focus on the most commonly assumed distribution in the social sciences: multivariate normal.

To begin, let's consider the probability density function for a multivariate normal distribution.

$$L_i = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}}\ e^{-.5(\mathbf{Y}_i-\ \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i-\boldsymbol{\mu})} \tag{1.13}$$

In Equation 1.13, L_i is the likelihood value; this value quantifies the likelihood (similar to a probability) that a set of scores from a particular study participant come from a normally distributed population with mean vector, $\boldsymbol{\mu}$, and covariance matrix, $\boldsymbol{\Sigma}$. The term, $\mathbf{Y}_i$, denotes a set of *k* observed scores. Euler's number is denoted as *e*; this value is a mathematical constant approximately equal to 2.718. The set of terms in fractional form to the left of *e* constitutes a scaling factor that makes the integral of the multivariate normal distribution equal to one. The critical portion of Equation 1.13 is in the exponent of *e*. The expression,

$$(\mathbf{Y}_i-\ \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i-\boldsymbol{\mu}) \tag{1.14}$$

is the matrix form of Mahalanobis distance, which can be interpreted as a squared *z*-score. To ease interpretation, let's consider this value in terms of its univariate form. A univariate expression for Mahalabonis distance is a squared *z*-score,

$$\left(\frac{Y_i-\mu}{\sigma}\right)^2 \tag{1.15}$$

where Y_i is an observed observation, μ is the arithmetic mean of all observations, and σ is the standard deviation of the observations. Equation 1.15 demonstrates that, in the univariate form, Mahalanaobis distance is a value that quantifies the standardized distance between an observed score and population mean. The interpretation from the multivariate standpoint remains largely the same. We interpret the multivariate Mahalanaobis distance (as in Equation 1.15) as a value that quantifies the distance of a given vector of observations, $\mathbf{Y}_i$, from the center (mean vector, $\boldsymbol{\mu}$) of the multivariate normal distribution. In practical terms, this implies that small deviations from a multivariate normal distribution with a particular mean vector and covariance matrix will minimize Mahalonobis distance. Because this distance is multiplied by –0.5 and is in exponential form, minimizing distances results in larger likelihood values.

As previously stated, the goal of maximum likelihood estimation is to determine the population parameter values (i.e., the mean vector and covariance matrix) that have the highest probability of producing the observed sample of data. In other words, we are looking for parameter values that maximize the likelihood values across the sample of participants. To determine the parameters that best maximize likelihoods in the entire sample (i.e., minimizes the sum of the squared *z*-scores), we need a value that summarizes the fit for all participants. To compute the sample likelihood, we would take the product of all the participants' likelihood values. It is the product of each individual likelihood value, rather than the sum of likelihood values, because likelihood values share the mathematical properties of probabilities. In probability theory, the joint probability for a set of independent events is the product of all the individual probabilities. Thus, a measure that summarizes fit for all participants would be the product of all the individual likelihood values. Taking a product of so many small likelihood values is not practical; the resulting total likelihood value tends to be an *extremely* small value. In fact, the value is often so small that it is difficult to work with and can be problematic due to rounding error inherent in computational devices.

To facilitate computation, it is conventional to convert the likelihood values to the log-likelihoods by taking the natural logarithm (i.e., a logarithm base *e*) of the likelihoods. The use of logarithms simplifies the computation of the sample likelihood because logarithms have the distinctive property that the log of a product is equal to the sum of the log of the individual terms. Consequently, the sample log-likelihood can be computed using summation rather than multiplication. The sample log-likelihood is expressed below.

$$logL = \sum_{i=1}^{N} log\left[\frac{1}{(2\pi)^{k/2}|\mathbf{\Sigma}|^{1/2}}\ e^{-.5(\mathbf{Y}_i - \boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu})}\right] \tag{1.16}$$

In this notation, the large sigma is an operator that denotes summation of each of the log-likelihood values from participants in the sample (for participants $1 \leq i \leq N$ where N is sample size and i denotes individual participants). Notice the expression in brackets is the likelihood expression from Equation 1.13.

The sample log-likelihood from Equation 1.16 can be further simplified by distributing the natural logarithm throughout the expression.

$$LogL = \sum_{i=1}^{N} -\frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}(\mathbf{Y}_i - \boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}) \tag{1.17}$$

Here, *LogL* quantifies the likelihood that the sample data came from a multivariate normal distribution with particular mean and variance/covariance values. Relying on this relationship, the maximum likelihood estimation routine determines which parameters of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ produce the highest log-likelihood values. The parameters of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ that yield the highest sample log-likelihood, are the estimated parameters that are most likely to have produced the sample data. For complete-data, software packages typically rely on properties of calculus to determine the parameters that maximize the sample log-likelihood. Specifically, the process determines the values of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ that yield a partial derivative (slope of the function) equal to zero. As we will see, with complicated models and missing data models, we are not able to analytically determine the maximum log-likelihood value and instead must rely on iterative algorithms with different parameters of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to determine the values that maximize the function.

Until now, I have discussed maximum likelihood in terms of a simple mean vector and covariance matrix. In actuality, maximum likelihood is far more flexible. Consider a typical regression equation regression *Y* on *X* and *Z*. In the estimated model, the predicted value of *Y* is the conditional mean given particular values of *X* and *Z*. Now, we can express the mean vector and the covariance matrix as a function of regression coefficients of *X* and *Z* as well as the residual covariance matrix. As the maximum likelihood model gets more complicated, the calculus required to determine an analytical solution that maximizes the log-likelihood expression is not as straightforward, and in most cases, unfeasible to compute analytically. As mentioned, software packages use iterative algorithms to determine a solution. Conceptually, the software is just auditioning different parameter values until it finds the parameters that are most likely to have produced that particular data set.

Maximum likelihood with missing data. Now that we have established the fundamentals of maximum likelihood analyses with no missing data, let's turn to using maximum likelihood estimation when there are missing data. Importantly, not much changes and the estimation procedure remains largely the same. Data do not need to be complete in order to use maximum likelihood estimation; instead all the available information can be used. In Equation 1.17, I demonstrated the sample log-likelihood for a complete-data analysis. The sample log-likelihood for incomplete multivariate normal data is largely the same.

$$LogL = \sum_{i=1}^{N} -\frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}_i| - \frac{1}{2}(\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) \qquad (1.18)$$

The distinctive difference between Equations 1.17 and 1.18 is the subscript *i* attributed to the number of scores, k_i, mean vector, $\boldsymbol{\mu}_i$, and covariance matrix, $\mathbf{\Sigma}_i$. This subscript *i* denotes a particular individual (or case) in the sample. The addition of this subscript adds additional flexibility so that the size and contents of the matrices for a person *i* depend on the variables with observed data for person *i*. Each participant may have a different number of *k* complete observations, and the number of items in the mean vector and covariance matrix varies depending on the available data. With missing data, the log-likelihood functions differ for individuals depending on the observed data, effectively using all available information to estimate the parameters.

To illustrate how the sample log-likelihood is computed using maximum likelihood estimation with missing data, let's consider a data set measuring variables *A*, *B* and *C*. In this data set, some participants are missing information on *B*, some on *C*, and some on both *B* and *C*. There are four missing data patterns: (1) complete *A*, *B* and *C*, (2) complete *A* and *B*, missing *C*, (3) complete *A* and *C*, missing *B*, and (4) complete *A*, missing *B* and *C*. For each of the missing data patterns, there is a log-likelihood function; here we have four log-likelihood functions. The individual log-likelihood values for participants with complete-data (Pattern 1) would be:

$$logL_{Pattern1_i} = -\frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log\begin{bmatrix}\sigma_A^2 & \sigma_{A,B} & \sigma_{A,C}\\ \sigma_{A,B} & \sigma_B^2 & \sigma_{B,C}\\ \sigma_{A,C} & \sigma_{B,C} & \sigma_C^2\end{bmatrix}$$
$$-\frac{1}{2}\left(\begin{bmatrix}A_i\\ B_i\\ C_i\end{bmatrix} - \begin{bmatrix}\mu_A\\ \mu_B\\ \mu_C\end{bmatrix}\right)^T \begin{bmatrix}\sigma_A^2 & \sigma_{A,B} & \sigma_{A,C}\\ \sigma_{A,B} & \sigma_B^2 & \sigma_{B,C}\\ \sigma_{A,C} & \sigma_{B,C} & \sigma_C^2\end{bmatrix}^{-1} \left(\begin{bmatrix}A_i\\ B_i\\ \mathrm{C}_i\end{bmatrix} - \begin{bmatrix}\mu_A\\ \mu_B\\ \mu_C\end{bmatrix}\right). \tag{1.19}$$

The individual log-likelihood values for the subset of participants with complete-data for variables *A* and *C*, but missing data for variable *B* (Pattern 2) would be:

$$logL_{MissingPattern2_i}$$
$$= -\frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log\begin{bmatrix}\sigma_A^2 & \sigma_{A,C}\\ \sigma_{A,C} & \sigma_C^2\end{bmatrix}$$
$$-\frac{1}{2}\left(\begin{bmatrix}A_i\\ C_i\end{bmatrix} - \begin{bmatrix}\mu_A\\ \mu_C\end{bmatrix}\right)^T \begin{bmatrix}\sigma_A^2 & \sigma_{A,C}\\ \sigma_{A,C} & \sigma_C^2\end{bmatrix}^{-1} \left(\begin{bmatrix}\mathrm{A}_i\\ C_i\end{bmatrix} - \begin{bmatrix}\mu_A\\ \mu_C\end{bmatrix}\right) \tag{1.20}$$

The individual log-likelihood values for the subset of participants with complete-data for variables *A* and *B*, but with missing data on variable *C* (Pattern 3), would largely be the same as Equation 1.20 above, but would replace the *C*'s with *B*'s. Finally, the individual log-likelihood values for the participants with complete-data for variable *A* and missing data for variables *B* and *C* (Pattern 4), is expressed below.

$$logL_{MissingPattern3_i}$$
$$= -\frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log[\sigma_A^2]$$
$$-\frac{1}{2}([A_i] - [\mu_A])^T[\sigma_A^2]^{-1}([A_i] - [\mu_A]) \tag{1.21}$$

Combining the log-likelihoods for all the participants gives us the sample log-likelihood. This sample log-likelihood can be thought of as the sample sum from each of the patterns and can be partitioned in the log-likelihood for complete cases and the log-likelihood for missing cases. For this example, the sample log-likelihood is represented below. Recall that Pattern 1 is the complete-data pattern and Patterns 2 – 4 represent three missing data patterns.

$$logL = \sum_{i=\mathrm{i}_1}^{n_1} logL_{Pattern1_i} + \left[\sum_{i=\mathrm{i}_1}^{n_2} logL_{Pattern2_i} + \sum_{i=\mathrm{i}_1}^{n_3} logL_{Pattern3_i} + \sum_{i=\mathrm{i}_1}^{n_4} logL_{Pattern4_i}\right] \tag{1.22}$$

Here, n_j represents the number of participants making up a particular data pattern subgroup and $i_1$represents the first participant belonging to that subgroup. We can also think of Equation 1.22 in terms of the sample log-likelihood being comprised of two components: a contribution from complete cases and a contribution from incomplete cases. The first term in Equation 1.22 is the portion of the sample log-likelihood that comes from complete-data; the trailing three terms in brackets are the portion of the sample log-likelihood that comes from incomplete-data. This partitioning illustrates an important aspect of maximum likelihood analysis; namely, that maximum likelihood uses all available information to estimate parameters. Because listwise deletion would only include complete cases, the last 3 terms in Equation 1.22 would be eliminated. The inclusion of these additional terms in the sample log-likelihood is what allows maximum likelihood estimation to provide unbiased estimates when the MAR mechanism is satisfied.

With missing data models, maximum likelihood requires the use of iterative data algorithms. There are a wide variety of potential algorithms, but the general procedures of the algorithms are similar (Enders, 2010). First the program takes a set of starting values that provide an initial guess of the parameter estimates. Then, the program iteratively auditions sets of parameters until it finds the set of values that maximizes the sample log-likelihood. Importantly, although the parameters in the description above are an unstructured mean vector and covariance matrix, we can substitute in the expectations based on a model with additional predictors or parameters.

Planned Missing Data Designs

An important extension of missing data methodological research is the use of planned missing data designs that rely on these modern missing data methods. Planned missing designs seek to minimize some of the prohibitive demands of large-scale research (e.g., respondent burden, time, money) while maintaining the desired scope of the research. Because researchers are often wary of including purposefully including missing data, I start by assuaging any discomfort researchers might have with the idea of intentional missing data by demonstrating that the classic randomized experiment can be viewed from a planned missing data perspective as explicated by Rubin's causal model (R. J. Little & Rubin, 2002; Rubin, 1974, 1978a, 2005; West & Thoemmes, 2010).

Randomized Experiment as a Planned Missing Data Design

In the Rubin causal model for a basic randomized controlled study with one treatment and one control condition, some treatment, T, is given to a participant, p_i, resulting in the observation of outcome $T(p_i)$. If the ability existed to simultaneously give the control condition, C, to "the same participant at the same time and in the same context" (West & Thoemmes, 2010), we would also observe the outcome $C(p_i)$. If we are able to compare these two outcomes, we could compute the causal or treatment effect for a given participant as the difference between the two outcomes, and the total treatment effect as the differences between the group means across sample N for the two outcomes:

$$\sum_{i=1}^{N} \frac{T(p_i) - C(p_i)}{N} \tag{1.23}$$

Although giving the same participant a different intervention under the exact same conditions provides for an elegant and straightforward estimation of the causal effect, in reality, research scenarios with humans preclude the ability to replicate the exact same conditions on the same participant. Once the first condition is administered, the participant is changed by virtue of having received the first treatment and by the passage of time. As a result, each participant in a classic randomized study is only assigned to one condition and researchers only observe the outcome(s) for one condition for each participant. The other half of the data are missing and these values are the unobserved responses we would have observed had the participant been able to simultaneously be in both conditions. With randomized designs, the unobserved missing responses for each participant's unassigned condition satisfy the MCAR mechanism (i.e., the assignment to condition, and the resulting probability of missing value on an unassigned condition, is purely random and unrelated to other variables, on average). The true experiment provides an unbiased estimate of the average causal effect in the population (at least asymptotically) assuming certain assumptions are met (e.g., proper randomization was obtained ensuring independence of covariates at baseline, full treatment adherence, no attrition, and stable-unit treatment value assumption). Thus the classic randomized experimental design, arguably a gold-standard of research design, is actually a planned missing data design.

Historical Background of Planned Missing Data Designs

The ability to analyze data using modern missing data methods affords new opportunities to implement research designs with intentional missing data to address issues of both researcher and participant burden. As an example, a researcher may be interested in a large scope of questionnaire items that are greater than the number of items that can be reasonably expected for participants to answer. Rather than limiting the number of questionnaire items, a planned missing data design that relies on the concept of "matrix sampling" could be used whereby each participant only answers a subset of the complete set of research questions. Thus the research collects data on the full set of questionnaire items while simultaneously reducing respondent burden. To lay the foundation of planned missingness, I provide some brief historical background. Planned missing data designs fit into a general class of *efficiency-of-measurement* designs (or simply "efficiency designs"; Graham et al., 1996) and borrow heavily from the efficiency design literature. Generally, the efficiency of a design is the extent to which a design can achieve the objective of a study while minimizing expenditure of time and/or money (Dodge, 2006, p. 127; Graham, 2012). One design is considered more efficient than another if the design uses fewer resources but can answer the same research question with the same precision. Typically, designs are made to be more efficient by both careful selection of the independent variables and determination of the best ways to allocate observations/participants. Broadly, many efficiency designs fall into one of three classes: (1) factorial designs, (2) optimal designs and (3) matrix sampling approaches. It should be noted that these classes are not mutually exclusive and many efficiency designs can be categorized into multiple classes.

Much of the current planned missing data work has evolved from the class of efficiency designs known as matrix sampling. In a type of matrix sampling known as "multiple matrix sampling," items are divided into subsets and different subsets of items are administered to different subgroups of participants. The research on multiple matrix sampling mostly originated from educational testing research (Gonzalez & Eltinge, 2007). According to Shoemaker (1973), as early as the 1950's researchers at Educational Testing Services (ETS) were considering this type of testing design. One of the classic designs that came out of educational testing research is the Balanced Incomplete block (BIB; Beaton et al., 1987; Johnson, 1992; Kaplan, 1995). In the BIB spiral design, like other multiple matrix sampling techniques, the item pool is split amongst participants so that each participant only answers a subset of all possible questions. In fact, Beaton and colleagues (1987) note that "it is not necessary or even desirable that each individual student take the entire battery of exercises" but can still gain information about all pairs of association. Specifically, the BIB spiral design uses seven forms (although other educational testing designs have been proposed with differing number of forms). The design is "balanced" in three ways: (1) the same number of participants responds to each item, (2) every pair of items appears together in at least one testing booklet, and (3) fatigue issues are mitigated because items appear in different orders on different forms. In educational testing scenarios, these large designs can be quite useful and have a variety of benefits both methodologically (e.g., ability to estimate higher order effects) and administratively (e.g., decrease in concerns of cheating because students will not have the same test). Although the BIB spiral design has many benefits in educational testing, this design has practical limitations in terms of application to social sciences. Designs stemming from the educational testing literature generally require complicated administration and very large sample sizes. In the social sciences, as compared to large scale educational testing, sample sizes are considerably smaller and seven or more questionnaire forms may pose a logistical challenge.

Modern Planned Missing Data Designs

John Graham and colleagues propose a variation on multiple matrix sample designs known as the 3-form design (Graham et al., 1996; Graham, Hofer, & Piccinin, 1994; Graham et al., 2001). The 3-form design is the close cousin of the BIB spiral design, but has features making the 3-form design more useful in social science research. Namely, the design is less cumbersome administratively because it only requires three forms. In this design, the entire set of research items (typically questionnaire items, but any type of item is appropriate) are divided into four blocks. The researcher then creates three item forms with each of these forms only including three out of the four blocks of items. The three forms are randomly assigned to participants. As a result, any given participant does not provide data on one of the blocks, but the research collects information from at least some of the participants on all of the items in the four blocks. Table 1 shows the most commonly implemented version of the 3-form design (Graham, 2012). Notice that there are four "Blocks" named *X*, *A*, *B*, and *C*. These blocks denote a set of variables (e.g., items/questions). The rows in Table 1 correspond to the three forms in the design and each of these three forms is distributed to one of three subgroups of participants. As demonstrated in the table, all three forms contain the variables in Block *X* plus the variables in two of the other Blocks. Variables in Blocks *A*, *B*, and *C* are each missing one time from one of the three forms. To clarify exactly how this design is implemented, let's consider the scenario where students in a classroom respond to a research survey. Because of school scheduling time constraints, participants can only be reasonably expected to respond to 75 items. However, the researcher is interested in collecting a total of 100 items. Using the 3-form design, the researcher can collect data on these 100 items even though any given participant will only need to respond to 75 items. To implement the design, the researcher would decide which items should belong to Block *X*, *A*, *B*, and *C* and then she would create three forms as in Table 1. Although I am talking specifically of items, blocks may also be composed of entire questionnaires such as a commonly used scale for a construct of interest. The blocks are combined (as in Table 1) to create three forms and participants are randomly assigned to each form. Typically an equal (or close to equal) number of participants responds to each form, but this need not be the case. To illustrate a very general 3-form design, Table 2 shows an example of the questions that might comprise a given form if the researcher evenly distributed the items across blocks. Each of the four blocks contain 25 items and each of the forms contain three of the 25 blocks resulting in only 75 items on any given form. For instance, Form 2 would include items 1 – 25 (Block *X*), 26-50 (Block *A*), items 76-100 (Block *C*) totaling 75 items. Thus although the total item pool is 100 items, each participant is only expected to answer only the 75 that are contained in his/her version of the form. One aspect of particular importance in the 3-form design is that the distribution of versions of the form is *randomized* across the sample of participants. Thus, the resulting pattern of missingness satisfies the MCAR mechanism. Importantly, analysis requires "modern" MAR-based missing data methods to use all available information to estimate parameters of interest (Baraldi & Enders, 2010; Graham, 2012).

In another design proposed by Graham and colleagues (1996), referred to as the "two-method measurement" design, researchers use multiple measures of a construct. A variation of this two-method measurement design informs my research study and I describe this design in more detail in Chapter 2. As a brief introduction, the two-method measurement design involves collecting data on two types of measures – generally inexpensive measures and more expensive measures. Researchers collect data on inexpensive measures that may have less construct validity and reliability for the entire sample. Data is collected on more expensive, but more valid, measure of the same construct for a *subset* of the sample – this is the planned missingness of the two-method measurement. For example, researchers interested in smoking behavior may collect self-report data on smoking behavior from a large sample (which may be underreported due to the undesirability of the behavior), in addition to the more expensive and more reliable measure of cotinine in saliva in a small subsample. The two-method measurement design provides an innovative way to maximize sample size and construct validity while simultaneously controlling for cost by using multiple measures of the same construct and only providing the more expensive measures to a subsample of the participants. Furthermore, because the expensive measures provide greater construct validity, these expensive measures can be used to model bias (i.e., lack of construct validity) in the inexpensive measure resulting in more valid statistical conclusions (Graham, Taylor, Olchowski, & Cumsille, 2006). The issue of having reliable and valid measures is particularly salient in mediation analysis, because the literature elucidates that reliable and valid measures are important factors in having sufficient power to detect the mediated effect. Thus, coupling the two-method measurement design with mediation analyses may be an advantageous strategy in certain research scenarios.

Study Goals

My research study investigates the use of intentional missing data for a mediation analysis that incorporates multiple measures of the same mediation construct. For purposes of description, I consider two types mediating variables: expensive and inexpensive mediators. The basic premise is that expensive measures often incorporate high construct validity, high reliability, and high assessment cost and can be thought of as the measure that would be used in a scenario with unlimited resources. The inexpensive mediating variables are less expensive and often less reliable measures of the same construct. My specific design will consider a scenario where *X* and *Y* are manifest variables and three measures of the mediator are available: one expensive and two inexpensive. With complete data, the most straightforward way to analyze the data would be with a latent variable mediation model as depicted in Figure 2. In a planned missing scenario, the researcher could collect complete data on two inexpensive measures of *M*, but would collect the expensive measure for only a subset of the participants. This missing data design can be analyzed using maximum estimation to incorporate all available data. As with complete data, the analysis model may be a latent variable mediation model. This study evaluates the power to detect the mediated effect with a latent variable mediation model that accounts for missing data using maximum likelihood estimation. In addition, the missing data literature describes another approach that might be appropriate for such a data structure. Specifically, if the expensive measure is a variable that might have been used in a scenario with unlimited resources, it makes sense to evaluate a model with this expensive variable as the mediator and the inexpensive mediation variables serving as "auxiliary" variables. I describe auxiliary variables in more detail in the next chapter, but broadly, auxiliary variables are variables that are not of direct research interest, but are included in a missing data model to reduce the uncertainty caused by missing data and increase the precision of model estimates. For completeness, the study also compares three other models that do not include all the measures of the mediator. In Chapter 2, I will review the pertinent literature that informs and supports my research study.

CHAPTER 2

LITERATURE REVIEW

The importance of statistical power is well known. In the social and behavioral sciences, statistical power plays a central role in the planning and design of research studies and the interpretation of results (Davey & Savla, 2009). A priori power analyses are often done prior to a study to determine the sample sizes needed to detect an effect. Not surprisingly, because missing data typically result in estimates with decreased power, concerns about adequate power are prevalent in the methodological literature on planned missing data (e.g., Graham et al., 2001, 2006; Jia et al., 2014; Mistler & Enders, 2012). In terms of mediation analysis, Fritz and MacKinnon (2007) note that a common question researchers ask is “How many subjects do I need to achieve adequate power when testing for mediation?” Mediation can provide crucial information about the mechanism by which one variable transmits effects to another. If mediation studies are underpowered, researchers may miss out on detecting important mechanisms relating two variables. In fact, power is so important to a study that hypothesizes mediation, numerous methodological studies have addressed this issue (e.g., Fritz & MacKinnon, 2007; Hoyle & Kenny, 1999; MacKinnon et al., 2002; Mallinckrodt, Todd, Wei, & Russell, 2006; Preacher & Hayes, 2008; Taylor, MacKinnon, & Tein, 2007). Further complicating concerns about power in mediation, there is evidence to suggest that measurement error may add bias to the mediated effect and/or reduce the power to detect a true mediated effect. To mitigate the impact of measurement error, researchers may choose to use highly reliable measures. Unfortunately, the best measures are typically associated with higher costs. To accommodate the higher cost, it is possible that a researcher might need to compromise sample size. Because power to detect the mediated effect is influenced by both the quality of the measured variables and sample size, cost limitations may place constraints on designing a research study that maximizes power. My research attempts to address this issue by exploring the possibility of utilizing planned missing data design whereby a large sample of participants provides information on less expensive, but possibly less reliable, mediating variables and a smaller subsample of participants provide information on a more expensive, but often more reliable measure of the mediator. By collecting data in this way, a researcher might be able to control costs (i.e., not every participant is measured on the expensive variable), increase accuracy (i.e., highly reliable expensive measure is incorporated to provide less biased results), and increase power (i.e., larger sample size measured on inexpensive measure). It is hoped that my research will help elucidate the conditions in which a planned missing data design for mediation analysis with multiple measurement might be useful.

This chapter reviews the relevant literature that informs the current research study. First, I briefly review statistical power of the mediated effect. I then expand that discussion to consider the implication of power when the mediator is measured with error. Next, I discuss the literature on planned missing data designs relevant to the study. This includes a section on auxiliary variables and reviews the planned missing data literature that has properties similar to my research design. Following the discussion of planned missingness, I provide a brief overview of the missing data literature that reviews missingness specifically in the context of mediation. Although this literature does not explicitly refer to planned missing data designs, the results of these missing data studies under the assumption of a MCAR mechanism are directly applicable to a planned missing data design which assumes a missingness that satisfies the MCAR mechanism. To conclude the chapter, I provide a summary of the justification for the current research study.

Statistical Power of the Mediated Effect

Varying methods for assessing mediation have been proposed (e.g., Baron and Kenny causal steps test, MacArthur model of mediation, and the joint significance test; MacKinnon, 2008). Consequently, much of the literature concerned with statistical power of the mediated effect focuses on comparing the power of the different methods to testing mediation (e.g., Fritz & MacKinnon, 2007; MacKinnon et al., 2002; Mallinckrodt et al., 2006). One simulation study evaluated fourteen methods for assessing the mediated effect found that the commonly used causal steps approach to mediation (Baron & Kenny, 1986) and tests of mediation based on normal distribution theory are severely underpowered relative to tests of mediation that rely on the distribution of the product (MacKinnon et al., 2002). Confirming this finding, a study by Fritz and Mackinnon (2007) determined that to achieve 0.8 power using the causal steps approach when there is complete mediation ($\tau' = 0$) and small effect sizes for the α and β paths, a research study would require 21,000 participants. The low power in the causal steps approach is generally attributed to the requirement of a significant X to Y relation, especially in the case of complete mediation (MacKinnon, Fairchild, et al., 2007). For this reason, the joint significance test, which only requires the α and β paths to be significant, has higher power than the causal steps test (MacKinnon et al., 2002). Power based on an assumption of normality (and corresponding symmetric confidence limits) will also typically result in underpowered tests of mediation (Fritz & MacKinnon, 2007; MacKinnon et al., 2002).

Because the distribution of the product is not normally distributed, detection of the mediated effect based on normal distribution theory is underpowered; the lack of power is due to the distribution of the product being asymmetric and highly kurtotic. Accordingly, power to detect the mediated effect based on the distribution of the product (e.g., bootstrap tests and distribution of the product asymmetric confidence intervals) has been shown to have more accurate Type 1 error rates and better statistical power as compared to other tests of mediation. Because the literature clearly supports the use product of coefficients approach using significance testing that relies of the distribution of the product (Fritz & MacKinnon, 2007; MacKinnon, Fritz, et al., 2007; MacKinnon et al., 2004; Mallinckrodt et al., 2006; Preacher & Hayes, 2008; Tofighi & MacKinnon, 2011), this research study employs the asymmetric confidence interval test based on PRODCLIN (MacKinnon, Fritz, et al., 2007).

Although power is enhanced by using the appropriate method for testing the mediated effect (and most tests of power in mediation are most concerned with comparing these methods), tests of the mediated effect still tend to require large sample sizes to have sufficient power. Fritz and MacKinnon (2007) determined that for various combinations of small, medium and large effect sizes of the α or β paths corresponding to Cohen's (1988) criteria of percent variation accounted for by the predictors, the required sample sizes to achieve power equal to 0.8 range from $N = 35$ to $N = 549$. The lowest end of the range, $N = 35$, occurs only when both α and β correspond to a large effect size which is rare in many areas of research. For mediated effects where either the α or β paths correspond to a small effect size, sample size requirements range from $N = 401$ to $N = 539$; these are much larger sample sizes than we often see in psychology. Not surprisingly, when Fritz and MacKinnon (2007) surveyed two psychology journals (*Journal of Consulting and Clinical Psychology* and *the Journal of Applied Psychology* from 2000 to 2003), they found that 75% of the studies that included a small α or β effect size had less than .8 power to find a mediated effect. As an aside, I would be remiss not to mention that power issues are not just limited to mediation analyses; many studies in psychology are underpowered and it would be unfair to expect mediation to be any different (reviewer comment referenced in a footnote in Fritz & MacKinnon, 2007). In fact, the power to detect a mediated effect between two variables is greater than the power to text the simple bivariate association in many conditions (Shrout & Bolger, 2002). Even though mediation has this interesting property of increased power in some scenarios, mediation analyses still tend to be underpowered. Furthermore, if the mediator contains measurement error, as expected in most realistic research scenarios, power to detect the mediated effect may be further decreased.

The Effect of Measurement Error on Power of the Mediated Effect

It is clear from the literature that mediation often requires large sample sizes. However, the research on power in mediation has largely been focused on scenarios that assume an ideal situation where X, M, Y, and any additional covariates are measured without error. In fact, one assumption of mediation analysis in the OLS regression framework is that predictor variables are measured without error (Cohen, Cohen, West, & Aiken, 2002; MacKinnon, 2008). This begs the question, what happens to the power to detect the mediated effect when variables are measured with error? This is an important question because, in real research applications, measurement error is virtually unavoidable. Measurement error refers to any immaterial factors, both systematic and unsystematic, that contribute to the measured scores on a variable that are not related to the theoretical construct of interest. Much has been written on the topic of measurement error (for example see the classic texts Crocker & Algina, 2006 or Lord & Novick, 1968), and broadly the two major aspects of measurement error are validity and reliability (MacKinnon, 2008). The concept of validity has evolved overtime in the psychometric literature (Algina & Penfield, 2009), and validity generally refers to the extent that a measure actually measures what it purports to measures. Reliability refers to the consistency a measure. In classical test theory, reliability is quantified as the proportion of variability in a measure directly related to the true score measured without error (Cohen et al., 2002). Although both validity and reliability are important components of measurement error, I will rely on reliability as representative of measurement error for purposes of this discussion and resulting research project for two reasons: (1) reliability is more straightforward to quantify and model than validity, and (2) it can be reasonably assumed that a program of research has generated valid measures of the theoretical constructs of interest (MacKinnon, 2008). Semantically, I will use reliability and measurement error interchangeably.

To begin the discussion of unreliability in mediation, let's first consider what we know about the impact of unreliable predictor variables in OLS regression. As stated, one assumption when estimating regression coefficients in OLS regression is that the predictor variables are measured without error (i.e., they are perfectly reliable). When one or more independent variables have any degree of unreliability, the estimates of the regression coefficients and their standard errors will be biased. In a regression equation with only one IV, bias in the independent variable will result in an attenuated relation between the independent and dependent variables. That is, when assessing the magnitude of regression coefficient relating a single predictor with measurement error to a single outcome (as in Equations 1.1 and 1.2), on average, the regression coefficient will be smaller in magnitude in the sample than it is in the population if the independent variable is measured with some error. In terms of the mediation equations introduced in Chapter 1, this implies that if there is error in X in Equation 1.2, the coefficient, α, will be attenuated (making the rather unrealistic assumption that no other covariates are included). When there are multiple IVs in a regression equation, as in Equation 1.3 or in any equations that include additional covariates, the effects of measurement error on the estimation of the regression coefficients is less obvious; the absolute value of the predictors may decrease or increase (Cohen, et al., 2003). This means that the coefficient, β, may be biased, but it is not immediately obvious in which direction the bias will occur. Note that the bias related to measurement error only occurs when the measurement error is in one of the IV's. When measurement error is in the DV, the error does not affect the value of the unstandardized regression coefficients (measurement error in the DV will affect the value of the standardized coefficients). From this, we can conclude that measurement error in the mediator would not affect the value of α, because the meditator is the outcome variable in Equation 1.2. Although measurement error in the dependent variable will not bias estimates, this error typically results in increased residuals and standard errors resulting in a decrease in the power to detect an effect (Cohen et al., 2002). Consequently, measurement error in M will increase the standard error of α, making it hard to find a statistically significant effect even if one exists. Given the research that demonstrates that you are most likely to find significant mediated effects when both α

and β are significant (MacKinnon, 2008), this is another way in which measurement error in the mediator could decrease power.

The potential impact of measurement error in the mediator as described above is substantiated by the literature. Hoyle and Kenny (1999) demonstrated analytically that as the reliability of the mediator departs from 1.0, the observed effect of β is underestimated and, in some cases, the observed effect of τ' is overestimated. Baraldi and MacKinnon (2011) demonstrated comparable results in a small simulation study. As a reminder, these effects hold only when there are no covariates with error in the model. Darlington (1990) notes that covariates with considerable measurement error can also affect regression coefficients in an unpredictable ways. Perhaps somewhat surprisingly, recent research suggests that simultaneously omitting confounders and using unreliable variables may offset each other, resulting in a relatively unbiased estimate of the population mediated effect (Fritz, Kenny, & MacKinnon, 2015).

Other research on measurement error in mediation comes from the field of epidemiology. One study investigated measurement error in the mediator for logistic regression models assuming *M* is a normally distributed variable and *Y* is binary variable (le Cessie, Debeij, Rosendaal, Cannegieter, & Vandenbroucke, 2012). This particular study was most interested in the influence of measurement error on detecting partial mediation via the effect size of τ' to emulate a scenario where researchers need to know if there are further mediating mechanisms that need be identified. For example, if a mediator only partially mediates the relationship between a DNA characteristic and a disease, another potential pathway relating the DNA characteristic and a disease must exist and further research on mechanisms of action is needed. Illustrating the importance of this question, an epidemiological study on lung cancer was interested in to what extent certain genetic variants and lung cancer were mediated by smoking (VanderWeele, Asomaning, et al., 2012). The mediator, smoking, was measured by self-report which is known to have high measurement error. Thus, understanding how a mediator with measurement error would affect conclusions about partial versus complete mediation is important. A major contribution of the study by le Cessie and colleagues (2012) is that their consideration of measurement error was not just limited to classical test theory and instead considered a variety of different forms of measurement error. For example, researchers included a condition of measurement error from intraindiviual variation over time. The error from intraindividual variation assumes that the value of a measure is comprised of both permanent factors (noninvariant components) and temporary factors (variant components). Another interesting consideration was the effect of measurement error that varies in M depending on the value of X. This is a type of error that might occur when different measurement instruments are used to measure M depending on the value of X. For example, if X is a measure of blood type and M is a measure of a clotting factor, and the instrument to measure clotting factor is different depending on blood type, measurement error in M will vary depending on the value of X. Thus, rather than limiting evaluation of the impact of measurement error to measurement error as defined by classical test theory, the authors evaluated the effects of a variety of types of measurement error using both analytic work and a simulation study. Not surprisingly, the results suggest that measurement error (in all forms described by the paper) may bias the estimate of τ'. Specifically, the researchers found that the partialled direct effect, τ', is biased positively if the direction of the relationship between X and M is the same as the direction of the relationship between *M* and *Y*. For effects in opposite directions, τ' is negatively biased as a result of measurement error.

VanderWeele, Valeri and Ogburn (2012) expanded the research by le Cessie and colleagues (2012) to consider the influence of measurement error on the mediator to detect the mediated effect (as opposed to the previous focus on τ'). Under the assumption of measurement error in the classical test theory framework, the researchers found that in a logistic regression model measurement error in the mediator with a normal distribution results in bias of the mediated effect towards a null odds ratio of one. Like the previous work, they found that if the direct and indirect effects are in the same direction, the bias of the direct effect is away from the null odds ratio of one. Interestingly, these results do not necessarily hold for all types of mediators; specifically a multinomial mediator with measurement error will not follow the patterns found by this research.

Potential Corrections for Measurement Error. There are a variety of potential remedies to address issues of measurement error. For example, Cohen, Cohen, West and Aiken (2002) note that a correlation matrix corrected for unreliability can be used to estimate the regression coefficients instead of the raw data. The corrected correlation matrix relies on the known attenuation of the correlation between variables with measurement error. If the reliability of two variables, *J* and *K*, is denoted as r_{JJ} and r_{KK}, respectively, then the observed correlation, r_{JK}, is a function of the true correlation, r_{JK_T}, and the measures of reliability, $r_{JK} = r_{JK_T}\sqrt{r_{JJ}r_{KK}}$. Using this relationship, observed correlations can be corrected for unreliability and the updated correlation matrix can be used to estimate the regression model (Cohen et al., 2002). Another solution for a single mediator model with only one measure of *M* is to replicate the manifest variable model in the latent variable framework and incorporate a correction in the residual error variance by constraining this variance to one minus the reliability of the measure times the variance of the measure (MacKinnon, 2008; Stephenson & Holbert, 2003). Simulations suggest that this method is better than not providing any correction for measurement error, but that latent variable models outperform this method (Stephenson & Holbert, 2003).

Le Cessie and colleagues (2012) also provide correction formulas to directly apply to the observed τ' coefficient. Further research expanded these correction formulas for generalized linear models that can accommodate an *XM* interaction (VanderWeele, Valeri, et al., 2012). Three correction approaches from the measurement error literature were evaluated: measure of moments, regression calibration estimators, and a simulation-based approach. Simulation studies suggest that regression calibration worked best across all scenarios considered (VanderWeele, Valeri, et al., 2012). At a basic level, the method of regression calibration uses a validation sample to model a correction for the biased estimates. Conceptually, this is not unlike what happens in a planned missing data model with two types of measures of the mediator.

All of these methods provide an adjustment for the mediated effect. However, successful use of the correction methods that utilize correction formulas requires having a very good estimate of the reliability. For most measures, precise estimates of reliability are not available (Cohen et al., 2002). To the extent that estimates of reliability are incorrect, these corrections will not eliminate all bias and could even introduce new bias. Consequently, it has been suggested that sensitivity analyses be used when there are no available gold standard or validation samples available (Valeri, Lin, & VanderWeele, 2014). The notion of using a validation sample with a gold standard measure to correct estimates using variables measured with error is not that unlike a planned missing data design.

Although many researchers prefer to use regression based tests (Fritz & MacKinnon, 2007), structural equation modeling can also be used to assess mediation and, if latent variables are incorporated, these models can potentially mitigate some of the measurement error (Cole & Maxwell, 2003; Holmbeck, 1997; Kenny, Kashy, & Bolger, 1998). Structural equation modeling that only includes manifest variables with measurement error will still be biased. As stated by MacCallum (1995, p. 21), "The presence of such error in the measurements will contaminate estimates of model parameters." Latent variables can be incorporated into structural equation models to attempt to disentangle measurement error from the variables. For example, consider the mediation model in Figure 2. In this model, three observed measures of a mediator, M_1, M_2, and M_3, inform the latent mediator variable, *LM*. Structural equation models with latent variables partition unique variance and random error from variance shared by a set of measures of the same construct. The resulting associations between latent variables are free of measurement error (Hoyle, 1991). To accomplish this task, latent variable models require multiple measures of the mediating construct. In some cases, multiple measures are not available or logistic constraints might prohibit collection of multiple measures. Another potential drawback of using SEM with latent variables is that these models tend to require larger sample sizes. Hoyle and Kenny (1999) evaluated the use of latent variable models in a scenario with three measures of a mediation construct for three levels of reliability: .60, .75, and .90. The researchers found that for the lowest reliabilities investigated (.60 and .75), a sample size of at least N = 200 is needed to have sufficient power to detect the mediated effect.

As the literature demonstrates, to maximize power to detect the mediated effect, researchers should maximize sample size and reliability of the measures and, ideally, collect multiple measures in order to use a latent variable model. Thus, researchers are left in a situation where for a single mediator model they either need a perfectly reliable measure of the mediator or multiple measures of the mediator so that they can model error in the structural equation modeling framework. Because reliable measures are often the most expensive, and collecting multiple measures, even if less reliable, also increases expense, this requirement may come at the expense of a reduced number of participants. My research seeks to address some of these issues using a planned missing data design that evaluates a latent variable model as one solution for unreliable measures.

Planned Missing Data

Modern missing data analyses, such as maximum likelihood estimation, provide researchers the opportunity to leverage purposeful missing data to their advantage. Used correctly, planned missing data designs can address many practical problems. These designs can mitigate both respondent burden and researcher expense; researcher expense includes any researcher resources used for the study such as time and money. My research study investigates the use of mediation analysis that incorporates two types of measures in a planned missing design. Although such a design has not yet been evaluated in the literature, there are aspects of the general literature on missing data and planned missing designs that are applicable to the multiple measures mediation design I am studying. This section will review the pertinent literature. First, I introduce auxiliary variables; these are variables that can be used to increase power in a planned missing data design and are incorporated in two of my analysis models. Next, I review the pertinent literature on planned missing data. Specifically, I narrow the focus to planned missing data designs that include latent variables as these scenarios most closely mimic my research design. I pay particular attention to the two-method measurement design as this design most closely represents one of the designs in the current research study.

Auxiliary Variables

A discussion on missing data is not complete without mentioning auxiliary variables. In modern missing data analysis, auxiliary variables are variables that are not part of the research question, but can be incorporated to predict the propensity for missing data and/or predict the incomplete analysis variables. Said differently, auxiliary variables are variables that would not be included in a complete data analysis, but may be useful in a missing data analysis and are included solely for the purpose of improving the missing data procedures. Including these auxiliary variables can reduce nonresponse bias and improve power. Consequently, researchers generally recommend an “inclusive” strategy of including as many auxiliary variables as possible. In a planned missing data design, where the underlying mechanism is known to be MCAR, non-response bias is a mitigated concern. Instead, auxiliary variables may be useful in increasing power in planned missing designs. Furthermore, if there is additional unplanned missing data, these auxiliary variables can help meet the requisite MAR assumption. Methodologists note that improvements in efficiency also may occur with the use of auxiliary variables (Rubin, 1996; Collins, et al., 2001). When auxiliary variables are highly correlated with incomplete analysis variables in the model, auxiliary variables restore some of the power lost to missing data. Even when the auxiliary variables are not highly correlated, the worst case scenario for the inclusion of auxiliary variables is “neutral, and at best extremely beneficial” (Collins, et al., 2001, p. 348). Thus, in a situation where correlated measures of a mediation construct are available, these variables could restore some of the power lost in a planned missing design with missingness on the mediator. In fact, in terms of detecting mediation, research has explicitly demonstrated that that auxiliary variables can increase the power to detect the mediated effect (Zhang & Wang, 2013).

The literature describes three ways to include auxiliary variables when using maximum likelihood estimation in missing data. In the saturated correlates approach, auxiliary variables are incorporated via a series of correlations and the analysis model variables and/or residual terms (Enders, 2010; Graham, 2003). Generally, the auxiliary variables must correlate with the manifest explanatory variables, other auxiliary variables, and the residual terms of the indicators of the latent outcome variables. The saturated correlates approach is the method I use in my research study. Besides the saturated correlates approach, there are two other strategies to incorporate auxiliary variables using maximum likelihood estimation: extra dependent variable model and the two stage approach. The extra dependent variable model incorporates an auxiliary variable as an additional dependent variable and correlates the residuals of the dependent variables (Graham, 2003; Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997). Simulations suggest that this model performs similarly to the saturated correlates model in terms of parameter bias, but the saturated correlates model performs better in terms of model fit (Graham, 2003). Savalei and Bentler (2009) proposed a two-stage approach to application of auxiliary variables whereby a population covariance matrix is estimated from a model and this covariance matrix is used in a complete data maximum likelihood function to find parameter estimates. This is specifically the method that Zhang and Wang (2013) used to demonstrate that auxiliary variables can increase power to detect the mediated effect. The two-stage approach has drawbacks compared to the saturated correlates approach. Namely, it adds complexity in that it requires a single value of sample size of *N* and this may bias standard errors (Enders, 2010; Enders & Peugh, 2004). There are correction formulas to address these issues, but these corrections are not automated by all software packages (EQS is an exception). Because the saturated correlated approach is automated by the software package M*plus*, this is the method incorporated in the study

Planned Missing Data Designs that Inform the Research Study

Next, I will review the planned missing designs that are related to my research by virtue of the inclusion of a latent variable. Recall from Chapter 1 that many planned missing data designs rely heavily on the notion of matrix sampling. Matrix sampling is often in the form of random assignment of items (or waves) to participants. For example, the 3-form design discussed in Chapter 1 utilizes the idea of matrix sampling. Although this design is extremely flexible and customizable, Table 1 shows the most commonly implemented version of the 3-form design (Graham, 2012). In this design, the entire set of research items are divided into four blocks and then three item forms are created each including three of the four blocks of items (Block *X* plus the variables in two of the other blocks). Variables in Blocks *A*, *B* and *C* are each missing on time from one of the three forms. A recent methodological study examined the use of the 3-form design for three latent variable models (Jia et al., 2014). Specifically, the researchers evaluated a cross-sectional confirmatory factor analysis model (see Figure 3), a two-time point confirmatory factor analysis model, and a three time point mediation model. Note the similarities in Figure 3 with a cross-sectional latent variable mediation model. The main difference is that a confirmatory factor analysis model lacks the directional paths linking the variables. In this study, the researchers evaluated a planned missing data design using a 3-form design whereby all participants provided data at all measurement occasions (when applicable), but do not provide information on all measures. In the cross-sectional confirmatory factor analysis model, each participant provides data for only two of the three indicators for all three latent variables. Even though the conception of the design differs from mine in that it utilizes a 3-form design, aspects of this design parallel my design. My design has three indicators of a latent mediation variable and some participants do not provide information on one of the indicators. Similarly, this research also has three indicators, but each participant only provides information for two of the three indicators. Similar parallels also exist between my research design and the two-time point confirmatory factor analysis model and the three time point mediation model. Simulation studies evaluating these models found a few important results. For smaller sample sizes, maximum likelihood approaches perform better in terms of required sample sizes, parameter bias, and convergence rates than multiple imputation providing further support for my selection of maximum likelihood estimation in my research. This study also demonstrates the importance of a measure called fraction of missing information (FMI). This is a measure of the proportion of information lost due to missing data. When using a 3-form design with a latent variable model, it is important to conduct a power analysis to learn the FMI on the model design to determine the sample size needed for the model. This is particularly important because unlike conventional wisdom, a larger sample size is not always uniformly better; the sample size is contingent on the complexity of the analytic model and the FMI (Jia et al., 2014).

Matrix sampling techniques have also been applied to longitudinal studies to create so-called "wave missing designs" (Graham et al., 2001; Mistler & Enders, 2012; Rhemtulla, Jia, Wu, & Little, 2014). In these designs, instead of participants being assigned to miss sets of items, participants are assigned to miss one or more measurement occasions. A longitudinal planned missing design is relative to planned missing data in mediation analysis in two distinct ways. First, by definition, ideal mediation research would incorporate a longitudinal design. Thus, the fact that planned missingness may be used in longitudinal research is relevant to a discussion of mediation analysis. Second, and most importantly, aspects of planned missing data design with longitudinal data are similar to one of the designs in my research study. For example, consider the latent variable mediation model in Figure 2; this model has three indicators loading onto a latent construct. Likewise, a linear growth model in the linear framework is also a latent variable model. For example, the model in Figure 4 is an SEM measurement model with three waves. Notice the nuts and bolts are virtually the same as in Figure 2. In Figure 2, there are three measures of a mediating construct loading onto a latent variable and in Figure 4 there are three measurement waves loading onto a latent variable. There are, of course, distinct differences between the models. For example, rather than freely estimating some of the loadings relating the indicators (observed measures) to the latent construct, the loadings in a growth model are typically fixed to demonstrate growth. However, the general pieces are still the same and, because of the similarities between this model and one of the models in my research design, the performance of wave missing design warrants some discussion.

One of the earliest variations of wave missing designs is the accelerated time-lag design. In an accelerated time-lag design, each participant is measured at two waves, but the amount of time between these waves varies across participants (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; McArdle & Woodcock, 1997). Conceptually, the resulting data matrix includes complete-data at the first wave for all participants. Each participant has one additional wave (depending on the given time lapse), but is missing data on every other potential time lapsed wave collection. In this example, there are extremely high rates of missingness, yet this method can be used successfully to demonstrate growth over time using latent growth curve techniques (McArdle et al., 2002; McArdle & Hamagami, 1992; McArdle & Woodcock, 1997). The success of this method provides further support to the utility of planned missing data designs.

Graham and colleagues (Graham et al., 2001) simulated a hypothetical growth model with five waves and empirically tested and illustrated the use of seven different wave missing patterns . As would be expected, as the percentage of missing data increased and the number of observed data points decreased, the standard errors for the coefficients of interest increased. However, it is of particular interest that for the particular set of conditions simulated, standard errors increased at a faster rate for complete case reduction than they did in a planned missing data design. In other words, as *N* decreased, complete case analyses decreased in power more rapidly than planned missing data designs. From this, it is reasonable to conclude that planned missing data designs were somewhat more robust against decreasing sample sizes. This finding is further strengthened by analyses that put a cost value on each data point. Given the selected conditions, the missing data designs had more statistical power than the complete case designs that costs the same (Graham et al., 2001).

The major lesson from wave missing data designs is that these designs have the capability of being advantageous to the researcher. As a note, the literature also shows that planned missing data designs are not uniformly advantageous and performance of these designs is very design specific. Not surprisingly, the advice about planned missing designs is not always consistent when it comes from studies on wave missing designs interested in different components of the model, using different population parameters, and with missing data rates. For example, Rhemtulla et al. (2014) and Graham et al. (2001) found somewhat conflicting results. In Graham and colleagues' simulation studies (2001), they found that wave missing designs resulted in higher power per observation than complete-data designs. Rhemtulla et al. (2014) found this not to be the case for variances and covariances among the slopes in a LGM model. This conflicting advice is more evidence that the optimal design is dependent on the exact model specification, relation among variables, and parameters of interest, and corresponding effect sizes. It is clear that there is no "one size fits all" approach to wave missing data designs. This is true in general of planned missing data designs. I also expect the same to be true in a mediation design.

Another type of efficiency design that incorporates planned missing data proposed by Graham and colleagues is the two-method measurement design (Graham & Shevock, 2012; Graham et al., 2006; Rhemtulla & Little, 2012). This is the design that is most aligned with my research study and provided inspiration for my model. As a note, the two-method measurement design is a bias correction design. In my line of study, I evaluate only a basic design that does not include a bias factor. Still, given the similarities between this design and one of the designs evaluated in the current research, I provide details on the two method measurement design.

The two-method measurement design involves collecting data on two types of measures – generally inexpensive measures and more expensive measures. Researchers collect data on inexpensive measures that may have less construct validity for the entire sample. Data is collected on more expensive, but more valid, measure of the same construct for a *subset* of the sample – this is the planned missingness of the two-method measurement. To understand the importance of this design, consider that in most research scenarios there are a variety of measures that may be used to test a construct of interest. In health and social sciences, these measures range from a simple self-report to more comprehensive assessments such as clinician observations or biological sampling. In many research situations, there often exist better measures of the construct with improved reliability and/or construct validity. However, these "better" measures may be more costly in terms of time and other resources (e.g., money, materials, equipment). For example, Graham and Shevock (2013) note that in smoking research, although self-reports of smoking continue to be common, researchers also use methods such as measures of cotinine (a metabolite of nicotine) in saliva. Here, the less valid measure is the virtually costless self-report of smoking and the more valid measure is the more expensive biological measure of smoking. VanderWeele and colleagues (VanderWeele, Asomaning, et al., 2012) provide an example of smoking as a mediator measured with error. In health studies, body composition relative to healthy norms is often of interest.

An inexpensive, but less valid measure relies on self-report of height and weight to compute a standardized numeric measure of a person's body composition, the Body Mass Index (BMI). A better measure of body composition is determining percentage of body fat using hydrostatic weighing but this technique is prohibitively expensive. In sum, although many measures exist for any one particular construct, often the more expensive measures are most valid. Ideally, researchers would routinely use the better measures, but cost constraints are often a limiting factor in research designs. For example, assuming a fixed budget, choosing expensive measures may significantly reduce the number of participants in the sample. On the other hand, using a less expensive measure typically allows for increased sample size, but the low construct validity may pose a problem in both results and interpretation. Weighing and balancing these issues is something that must be taken into account in designing a research study.

The two-method measurement design provides an innovative way to maximize sample size and construct validity while simultaneously controlling for cost by using multiple measures of the same construct and only providing the more expensive measures to a subsample of the participants. With modern missing data analysis methods, these designs leverage the statistical benefits of collecting inexpensive measures on a larger sample of participants and a more expensive, but more accurate, measures on a smaller subset. Furthermore, because the expensive measures provide greater construct validity, these expensive measures can be used to model bias (i.e., lack of construct validity) in the inexpensive measure resulting in more valid statistical conclusions (Graham & Shevock, 2012; Graham et al., 2006).

The two-method measurement model is a two factor response bias-correction model' this model borrows heavily from the multitrait-multimethod tradition. Response bias is bias as relates to construct validity. The degree that a measure is “valid” is based on the degree to which it measures the construct purports to measure. On the other hand, the extent that a measure measures something different than intended reflects the extent that the measure is not valid (i.e., biased). For example, in a measure that assesses dietary behaviors, the valid part of the measure is the extent that the measure captures actual dietary behaviors. The biased part of the measure may be the systematic underreporting of unhealthy dietary behaviors that is often observed (e.g., Braam, Ocké, Bueno-de-Mesquita, & Seidell, 1998; Lafay et al., 1997). The two-method measurement design allows researchers to model this type of systematic bias in the analysis by including both a factor getting at construct and a factor getting at bias in the measurement model. Readers may recognize familiarity of the notion of response bias with other terms the literature such as halo errors (Hoyt, 2000) and correlational errors (Berman & Kenny, 1976; Graham & Collins, 1991).

The general bias correction model must have at least two measures of the inexpensive but potentially flawed measure of the construct of interest and one (or more) expensive measure. This expensive measure should be a highly valid measure, but for actual application, Graham and Shevock (2012) assert that the expensive measure simply needs to be the preferred measure and not necessarily a perfectly valid measure. In terms of a missing data design, this measure can be thought of as the measure that would have been collected if expense (e.g., time, cost, etc.) was of no consideration.

To illustrate the bias correction model, Figure 5 shows the Bias Correction model presented by Graham et al. (2006) and also discussed elsewhere (e.g., Garnier-Villarreal, Rhemtulla, Mijke, & Little, Todd D., 2014; Graham & Shevock, 2012; Rhemtulla & Little, 2012). The structural equation model in Figure 5 shows a latent construct with three indicators. Returning to the smoking example, the two biased indicators might be two smoking self-reports. The unbiased measure is a saliva measure of cotinine. In this model, there are two factors allowing for correlation between the biased measures. One factor in the IV (“bias” in the figure) represents response bias and the other factor (“construct” in the figure) is the common constructs. For purposes of the smoking example, bias represents systematic self-report bias. Importantly, to model bias, there must be at least two measures that are thought to have a common source of bias. Although Figure 5 shows only one unbiased measure, more than one may be used. The dependent variable can take on any form (of course, within the boundaries of the rules of structural equation modeling). As another option, rather than estimating a separate bias factor, an alternative model estimates the residual covariance between the two self-report items (Graham et al., 2006; Kenny, 1976; Marsh, 1989). My research is concerned only with additional error rather than bias, thus, evaluated model does not include the bias factor.

Important to this model if it includes a bias factor, there must be at least one (presumably) unbiased measure in order to estimate a separate factor for bias. That means that the unbiased/expensive measure is crucial to disentangling the biased parts of the biased measures from the unbiased parts. Notice that the model I just described doesn’t mention anything about missing data. In reality, this is a model that can be used in a complete-data scenario. However, in the two-method measurement design, only a random subsample of participants are given the more expensive/unbiased measure to create a planned missing data design that satisfies the MCAR mechanism. By using modern missing data techniques, the data are analyzed using all available information.

Graham and colleagues (Graham et al., 2006) explored the benefit of using two-method measurement designs and described this research again more recently (Graham & Shevock, 2012). The benefit of this design stems largely from leveraging the power obtained by a larger sample size with the inexpensive measures and the information gained from the sub sample of participants with data on both the expensive and inexpensive measures. Simulations placing a dollar amount per participant have shown this method's utility. In a scenario with a 5:1 cost ratio where for every one less participant presented with the expensive measure subsample, five more participants can be added to the inexpensive measure total sample. In a scenario where all IV indicators have loadings of .5, the DV indicators have loadings of .7, the response bias factor accounts for 25% of the variation in self-reports and the common factor accounts for 49% of the variation, it is shown that this method generally produces much higher "effective *N*" than it would have had only the expensive measure for a smaller sample size been used (Graham et al., 2006). "Effective *N*" is a tool used to compare to complete-case designs costing the same as the optimal two-method measurement design configuration. Effective *N* is the number of participants needed in a complete case scenario to achieve the same power as obtained in the two-method measurement design. As an example, a simulation demonstrated that a two-method measurement design that collected inexpensive measures for 1200 participants and expensive measures for 120 participants provided statistical tests with power equivalent to a sample size 1.81 times larger than the comparable complete-case design. Effective *N* is most influenced by the cost ratio between the inexpensive and expensive measures and the size of the relationship between the IV and the DV. For smaller effect sizes, the ratio of Effective *N* to complete case *N* for the same dollar amount, demonstrates this methods potential utility in scenarios with small effect sizes (Graham & Shevock, 2012). In general, like all planned missing designs, the exact benefits of such a design are conditional on the anticipated parameters and researchers are encouraged to simulate conditions in order to create the optimal planned missing design. Even so, it is clear from the body of literature that in many research scenarios, planned missing data designs have the potential to enhance research findings.

Missing Data and Mediation

In this section, I review missing data literature that particularly evaluates missingness in the context of mediation. Recall that researchers may perform mediation analyses in the SEM framework or the OLS regression framework (which is really just a specialized case of the SEM framework). Although there has been a significant amount of literature on modern missing data analyses in both the SEM and regression frameworks (e.g., Baraldi & Enders, 2010; Enders, 2010; Rubin & Little, 2002; Schafer & Graham, 2002), only a handful of research papers have focused on mediation specifically (e.g., .Enders, Fairchild, & MacKinnon, 2013; Wang, Zhang, & Tong, 2014; Wu & Jia, 2013; Zhang & Wang, 2013). Perhaps not surprisingly, a common practice in the substantive literature is to use deletion methods to analyze mediation effects using complete data methods (Zhang & Wang, 2013). Although the papers that have been published addressing mediation analyses using modern missing data methods do not assess performance under the paradigm of planned missingness, research on missing data that assumes the MCAR mechanism is relevant to planned missing data designs as these designs also typically assume the MCAR mechanism. One recent example of missing data analysis in mediation extended work on Bayesian mediation methods (Yuan & MacKinnon, 2009) to a general Bayesian missing data handling approach for mediation manifest variable models (Enders et al., 2013). There are two major benefits of accommodating missing data (and complete data) in the Bayesian framework in terms of power. First, the Bayesian approach does not require distributional assumptions so the fact that the mediated effect has a non-normal distribution is not a concern. Second, prior information from earlier research can be incorporated. Simulations suggest that, from a frequentist standpoint, incorporating prior information improves power to detect mediation effects with missing data (Enders et al., 2013). The results of the Bayesian missing data analyses suggest that even with a non-informative prior, coverage and power estimates are comparable to the use of maximum likelihood estimation with bias corrected bootstrap (Enders et al., 2013). A review of literature looking at PRODCLIN suggests that the distribution of product method performs similarly to the bootstrap (Fritz & MacKinnon, 2007; Tofighi & MacKinnon, 2011) suggesting the appropriateness of using PRODCLIN in my simulation study. Another study compared maximum likelihood approaches to multiple imputation approaches to mediation analyses with missing data (Zhang & Wang, 2013). This research found that for the studied conditions, under MCAR mechanism, both the deletion methods (listwise and pairwise deletion) and modern missing data methods (multiple imputation and maximum likelihood estimation) could capture the mediated effect with little bias even with a small sample size and high missing data rates. However, only the modern methods of maximum likelihood and multiple imputation obtained good coverage values, and these methods had much increased power to detect the mediated effect. None of these findings are particularly surprising given the current state of the literature on missing data (e.g., Enders, 2010; Schafer & Graham, 2002).

Justification for Research

The discussion in this chapter illuminates a few key points that justify the current research study. Generally, we know that unreliable measures can have treacherous results on parameter estimates and tests of significance. Because the mediated effect relies on the distribution of the product of two coefficients and corresponding standard error, estimates of the mediated effect may be biased and/or underpowered if α and/or β are biased and/or if the corresponding standard errors are too large. We also know that even with perfect reliability, tests of the mediation effect are often underpowered and often require sample sizes larger than those generally seen in psychology and related fields. Moreover, there is a great deal of literature suggesting that planned missing data designs analyzed using modern missing data methods has the potential to create an efficient design that maximizes resources. Specifically, it has been shown that a two-method measurement design utilizing multiple measures of the same construct can have greater accuracy power than a complete data design in some scenarios. Consequently, a logical next step is to investigate the utility of a planned missing data design to address logistical constraints of mediation analysis (i.e., limited resources), while maximizing power to detect the mediated effect. In a traditional mediation research design, an ideal scenario would include data collection of multiple measures of the mediator or a single mediator with near perfect reliability all while simultaneously using a very large sample size. None of these components come for free – there is typically a cost associated with each of these items. Measures with high reliability tend to have high cost, and come at the expense of number of participants. Likewise, collecting multiple mediators may also be more expensive. The researcher could consider including more participants in the study, but that might mean sacrificing the quality of a measure or reducing the number of measures of the mediation construct. Thus my research evaluates the utility of using purposeful missing data to leverage resources in mediation analyses.

CHAPTER 3

METHODS

This dissertation evaluates the utility of a planned missing data design in a mediation analysis that incorporates multiple mediators of the same construct. Specifically, I empirically evaluate research designs that address limited resources with purposeful missing data on one of the mediation variables. This study uses Monte Carlo simulations to determine the empirical power to detect the mediated effect under a variety of simulation conditions and analysis methods. I address the question of how to maximize the power to detect mediation with multiple mediators including an expensive measure that incorporates planned missingness. The associations among the mediators are defined based on correlation matrix of the mediating variables and corresponding measurement model.

The study uses simulations with data generated from a variety of population parameters for manifest X and Y variables and a latent mediator variable, LM, with three observed indicator variables, M_1, M_2, and M_3. Planned missingness is implemented on variable M_1. When using multiple mediators with missing data, there are a variety of ways to analyze the data, and I evaluate five approaches. First, data are analyzed using a latent variable mediation model where all three of the mediators are indicators of a latent mediation construct (Method 1). Next, I consider auxiliary variable models where one of the mediation variables, M_1, is the mediator of interest, and one or both of the other mediating variables, M_2 and M_3, serve as auxiliary variables. As described in Chapter 2, auxiliary variables are variables that are not part of the research question, but can be incorporated to predict the propensity for missing data and/or predict the incomplete analysis variables. In the auxiliary variable models evaluated in the study, the less reliable but inexpensive measures of the mediator serve as auxiliary variables. In a planned missing data design such as in this study, where the underlying mechanism is known to be MCAR, highly correlated auxiliary variables may be useful in increasing power (Collins et al., 2001). In this particular simulation study where M_1 is missing, the highly correlated variables M_2 and M_3 would restore some of the power lost in a planned missing data design. The ability for auxiliary variables to increase the power to detect the mediated effect was also demonstrated empirically (Zhang & Wang, 2013). The second and third methods considered in the study are auxiliary variable models with one model utilizing two auxiliary variables (Method 2 incorporates M_1 as mediator and M_2 and M_3 as auxiliary variables) and one model utilizing only one auxiliary variable (Method 3 incorporates M_1 as mediator and M_2 as an auxiliary variable). The fourth and fifth methods of interest consider a situation where only the expensive mediation measure, M_1, is included; M_1 is the variable that includes planned missingness when applicable. A model including only M_1 may be analyzed using MAR-based maximum likelihood estimation missing data approaches (Method 4) or using a deletion method such as listwise deletion (Method 5). The main outcome of interest is the comparison of empirical power across the models. Type 1 errors, bias, and confidence interval coverage are also considered.

A Monte Carlo simulation study is used to study the performance of the varying analyses. Factors that were hypothesized to affect the performance of the methods include: (1) sample size, (2) planned missing data rate of the expensive measure, M_1, (3) the correlation among M_1, $M_{2,}$ and $M_{3,}$ (4) magnitude of α path, (5) magnitude of the β

path, and (6) analysis method. Each of these factors were varied in a simulation study and crossed in full factorial design.

It was expected that the missing data analysis models that incorporate all three mediation measures (latent variable model with three indicators and single mediator model with two auxiliary variables) would perform better than the other models by virtue of the availability of additional data. Like most statistical methods, I also expected that all evaluated methods would perform better as sample sizes increase and missing data rates decrease. Because of the literature suggesting that very large sample sizes are needed to estimate power for a small mediated effect size (e.g., Fritz & MacKinnon, 2007), I also anticipated that power would increase rather dramatically as a function of effect size. Of particular interest was whether or not a model incorporating all three measured mediators performs better as a latent variable or an auxiliary variable missing data model. To date, no study makes such a comparison. For methods that incorporate all three mediation measures, the auxiliary variable model (Method 2) estimates more parameters than the latent variable mediation model (Method 1); this difference might explain any observed differences in the performance of these methods.

Population Generation Model

Data were generated based on the data generation model in Figure 6. In this model, observed variables are X, M_1, M_2, M_3 and Y. Variable M_1 is the "expensive" mediation variable and variables M_2 and M_3 are the "inexpensive" mediation variables. All variables are assumed to be multivariate normal with a mean of zero (as if they had been "centered" or deviated around the mean). Although many mediation examples in the prevention and treatment literature often use a dichotomous X variable, I chose to use multivariate normal independent variables for greater generalizability. Supporting this decision, a study comparing methods to test mediation effects found that power for models in which X was binary was not noticeably different than power with a continuous X variable when effect sizes are equal (MacKinnon et al., 2002).

Although the research question assumes variables X and Y are manifest variables, they are generated as single manifest indicators latent variables with loadings and residual variances of one and zero, respectively. Using the standard SEM approach, the data generation model is defined by both a measurement model and a structural model. The measurement model defines the relationship of the observed variables to their latent variable constructs and the structural model summarizes the relationships between the latent variables. Equations 3.1 – 3.2 below provide a measurement model in matrix form for the observed variables in Figure 6. For clarity, I use notation consistent with Bollen (1989).

$$\mathbf{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \tag{3.1}$$

$$\mathbf{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{3.2}$$

The matrix, **x**, in Equation 3.1 is a matrix of the observed indicators of the latent exogenous variables. In the model depicted in Figure 6, there is only one indicators of the exogenous latent variable, *LX*. Thus the **x** matrix is simply the observed variable, X.

$$\mathbf{x} = [x_1] = [X] \tag{3.3}$$

The matrix, **y,** in Equation 3.2 is a matrix of the observed indicators of the latent endogenous variables. In this model, the **y** matrix is comprised of the observed mediation and outcome variables, M_1, M_2, M_3 and Y.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ Y \end{bmatrix} \tag{3.4}$$

The Lambda Matrices, $\mathbf{\Lambda}_X$ and $\mathbf{\Lambda}_y$, are matrices of coefficients relating the manifest variables to the latent variables.

$$\mathbf{\Lambda}_X = [1] \qquad \mathbf{\Lambda}_y = \begin{bmatrix} \lambda_1 = 1 & 0 \\ \lambda_2 = \lambda & 0 \\ \lambda_3 = \lambda & 0 \\ 0 & 1 \end{bmatrix} \tag{3.5}$$

Notice that $\mathbf{\Lambda}_X$ has a single value of one, corresponding with the creation of the *X* latent variable with a single manifest indicator with a loading of one. This loading (coupled with a residual variance of zero) creates a latent variable that is identical to a normally distributed manifest variable. The matrix, $\mathbf{\Lambda}_y$, includes the path coefficients for the mediation latent variable indicators (λ_1, λ_2, and λ_3 in column 1, rows 1-3). The $\mathbf{\Lambda}_y$ matrix also includes a path coefficient of one (in column 2, row 4) to denote the one-to-one relation of the *Y* indicator with the latent variable of *Y* creating a latent variable identical to a normally distributed manifest variable. In order to facilitate comparison across the different analysis methods, the factor loading relating M_1 to LM, λ_1, is constrained to one ($\lambda_1 = 1$). This ensures that the structural paths of all analysis models are on the same metric for comparison. Additionally, because M_1's loading equals one, it sets up a scenario where M_1 is perfectly reliable; the ideal scenario. As will be described in the study design, the values of λ_2 and λ_3 in the $\mathbf{\Lambda}$ matrices are constrained to be equal such that $\lambda_2 = \lambda_3$. Thus, only one value of lambda is included in the population generation model and subscripts are no longer needed ($\lambda_2 = \lambda_3 = \lambda$).

The covariance matrices are denoted by the $\boldsymbol{\theta}_\delta$ matrix for the covariance matrix of $\boldsymbol{\delta}$ in Equation 3.1, and the $\boldsymbol{\theta}_\varepsilon$ matrix for the covariance of $\boldsymbol{\varepsilon}$ in Equation 3.2.

$$\boldsymbol{\theta}_\delta = [0] \qquad \boldsymbol{\theta}_\varepsilon = \begin{bmatrix} \sigma^2_{\varepsilon_1} = 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\varepsilon_2} = 1 - \lambda^2 & 0 & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_3} = 1 - \lambda^2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.6}$$

Here, $\boldsymbol{\theta}_\delta$ only has one term, zero, corresponding to the residual variance of zero for the *X* variable. Again, this is the value that creates a latent variable, *LX* that is identical to the observed *X* variable. The matrix, $\boldsymbol{\theta}_\varepsilon$, has three potentially non-zero values: the residual variances of the three mediation indicators. For reasons that will be described in the next section, $\sigma^2_{\varepsilon_1}$ is constrained to zero and $\sigma^2_{\varepsilon_2}$ and $\sigma^2_{\varepsilon_3}$ is constrained to $1 - \lambda^2$. The value in the last row and column is zero to denote the residual variance of zero for the observed *Y* variable. The first portion of Appendix B provides the expanded equations (in non-matrix form) for the measurement model of population generation model.

Now that I have defined the measurement portion of the data generation model, I will explicate the structural portion of the population generation model. The structural portion is expressed by the following matrix equation:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}. \tag{3.7}$$

The matrix equation representing the structural design can also be written out in full matrix form as in terms of Bollen (1989) notation:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{22} \end{bmatrix} [\xi_1] + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}, \tag{3.8}$$

and can be re-expressed in notation consistent with Figure 6 as:

$$\begin{bmatrix} \mathrm{LM} \\ \mathrm{LY} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta & 0 \end{bmatrix} \begin{bmatrix} \mathrm{LM} \\ \mathrm{LY} \end{bmatrix} + \begin{bmatrix} \alpha \\ \tau' \end{bmatrix} [X] + \begin{bmatrix} \zeta_M \\ \zeta_Y \end{bmatrix}. \tag{3.9}$$

The last component of Equation 3.7 is the Zeta Matrix, $\boldsymbol{\zeta}$. This matrix denotes the latent errors in the equations. The Psi Matrix, $\boldsymbol{\Psi}$, is the covariance matrix of $\boldsymbol{\zeta}$ and includes the residual variances of LM and LY.

$$\boldsymbol{\Psi} = \begin{bmatrix} \sigma^2_{\zeta_M} & 0 \\ 0 & \sigma^2_{\zeta_Y} \end{bmatrix} \tag{3.10}$$

As will be described in the next section, the values in the Psi Matrix, $\boldsymbol{\Psi}$, are constrained based on a function of the model coefficients, α and β, resulting in the variances of LM and LY being equal to one. The second portion of Appendix B provides the expanded equations (in non-matrix form) for the structural portion of population generation model.

Manipulated Factors

Population Parameters

The population parameters were manipulated to vary both the correlation matrix of the M_i's (and corresponding loadings in the measurement model) and the magnitude of the mediated effect size via the α and β paths. For the measurement portion of the model, as stated previously, the X and Y variables are single indicators of latent variables with loadings of one and residual variances of zero. In order to facilitate comparison across the different analysis methods and ensure all analysis methods are on the same metric for comparison, the factor loading relating M_1 to LM, λ_1, is constrained to one. This constraint mimics the ideal scenario of an expensive measure where M_1 is perfectly reliable. The factor loadings for M_2 and M_3, are constrained to be equal and represented as λ. The value of λ was varied to reflect pairs of low, medium and high loadings for the two inexpensive measures, M_2 and M_3. Factor loadings corresponding with low, medium, and high loadings were chosen to be 0.4, 0.6, and 0.8, respectively. These three factor loading conditions can also be expressed as three corresponding correlation matrices. In general, the correlation between any of the two mediating indicator variables, M_j and M_k, in the measurement model can be expressed as:

$$r_{jk} = \frac{\lambda_j \lambda_k \sigma_{LM}^2}{\sigma_{M_j} \sigma_{M_k}}. \tag{3.11}$$

In the population generation model, the standard deviations of the mediation variables, σ_{M_j} and σ_{M_k}, are equal to one. As mentioned, and to be described in more detail later, the variance of the latent mediation variable, *LM*, was also constrained to one. Thus Equation 3.11 reduces to $r_{jk} = \lambda_j \lambda_k$. In the low loading scenario, the correlation between M_1 and the two other mediation variables is $r_{M_1 M_2} = r_{M_1 M_3} = 1 * \lambda = .4$, reflecting that the loading of M_1 is constrained to one and the loadings of M_2 and M_3 are equal to λ. Based on the constraints of the data generation model, the loadings of M_2 and M_3 are equivalent to the value of the correlation of M_1 with M_2 and M_3. In the low loading scenario, the correlation between M_2 and M_3 is $r_{M_2 M_3} = \lambda^2 = .4^2 = .16$. The residual variance for M_2 and M_3 is $1 - \lambda^2$. Similarly we can calculate the correlations for the medium and high factor loading conditions. Table 3 provides a summary of the standardized factor loadings, residual variances and resulting correlations among the three mediation variables for each of the three measurement model conditions.

Next, I manipulated the structural portion of the data generation model to represent varying combinations of the α and β path coefficients to represent different mediation effect sizes. I did not manipulate the direct effect, τ', and constrained this direct effect to zero to represent complete mediation. Although it is often unrealistic to expect complete mediation with a single mediation construct (Baron & Kenny, 1986), I chose to only evaluate conditions with complete mediation for two reasons. First, the addition of partial mediation adds complexities to creating comparable effect sizes across conditions. Second, and importantly, previous research suggests that partial mediation does not significantly affect power of the mediated effect when using tests of significance that rely on the product of the distribution (Fritz & MacKinnon, 2007; MacKinnon et al., 2004). Because the main outcome of interest in this study is power, it is not crucial that the simulation incorporates partial mediation; avoiding the assessment of partial mediation eliminates complications in comparing the performance of the methods across different effect sizes.

In all conditions, $\phi_X = \sigma_X^2 = 1$. In order to retain the correlation structure of M_1, M_2, and M_3 in the population generation covariance matrix, $\sigma_{\zeta_M}^2$ was constrained to $1 - \alpha^2$ so that the variance of the mediation latent variable, *LM*, is equal to one[1]. For simplicity, I likewise constrained $\sigma_{\zeta_Y}^2 = 1 - \beta^2$ so that the variance of latent *Y*, *LY*, is also equal to one.[2]

By constraining the variances of the latent variables to one, parameters α and β are equivalent to correlations. Values for parameters α and β are based on Cohen's definition of zero, small, medium, and large effect sizes for the Pearson product-moment correlation coefficients: r = 0.0, 0.1, 0.3, and 0.5, respectively (Cohen, 1988). Because the latent variables *LM* and *LY* have variances of one (created as described in the

---

[1]Based on the derivation of the covariance, $\mathrm{Var}(LM) = \mathrm{a}^2\sigma_X^2 + \sigma_{\zeta_M}^2$. Since $\sigma_X^2 = 1$ in all conditions, $\mathrm{Var}(LM) = \mathrm{a}^2 + \sigma_{\zeta_M}^2$. Thus, to constrain the variance of the latent mediator, *LM*, to one, $\sigma_{\zeta_M}^2 = 1 - \alpha^2$.

[2] Based on the derivation of the covariance, $\mathrm{Var}(LY) = \beta^2(a^2\sigma_X^2 + \sigma_{\zeta_M}^2) + 2\beta c' a\sigma_X^2 + c'^2\sigma_X^2 + \sigma_{\zeta_Y}^2$. Substituting $\sigma_X^2 = 1$ and $\sigma_{\zeta_M}^2 = 1 - a^2$, $\mathrm{Var}(LY) = \beta^2 + 2\beta\tau' a + \tau'^2 + \sigma_{\zeta_Y}^2$. Thus, to constrain the variance of the latent *Y* variable, *LY*, to one, $\sigma_{\zeta_Y}^2 = 1 - \beta^2 - 2\beta\tau' a - \tau'^2$. In complete mediation, as in this study, the last two terms of the $\sigma_{\zeta_Y}^2$ expression are zero and $\sigma_{\zeta_Y}^2$ can be reduced to $\sigma_{\zeta_Y}^2 = 1 - \beta^2$.

previous paragraph), in a complete mediation condition with τ' = 0, parameters α and β correspond directly to Cohen's benchmarks of zero, small, medium and large correlations. Thus, parameters α and β are fully crossed for the values α = 0.0, 0.1, 0.3, 0.5 and β = 0.0, 0.1, 0.3, 0.5.

Other Manipulated Factors

Sample Size. Four sample sizes were evaluated: 100, 200, 500, and 1000. These were chosen to represent a wide range of sample sizes and correspond with sample sizes often seen in psychology and prevention literature. In particular, the chosen sample sizes correspond with sample sizes often evaluated in the mediation literature (e.g., Iacobucci, Saldanha, & Deng, 2007; MacKinnon et al., 2002). I included the N = 1000 condition in order to examine the asymptotic (i.e. large sample) properties of the analysis methods.

Missing Data Rates. Four missing data rates on M_1 were evaluated: 0%, 20%, 50% and 80%. These correspond to 100%, 80%, 50% and 20% of participants providing data on the expensive measure. The 0% missing data scenario (i.e., complete data) serves as a benchmark to evaluate power loss due to planned missingness. To impose the missing data rates, I used the complete data sets generated with all combinations of the parameters and sample sizes described as above and imposed a missing completely at random (MCAR) mechanism by randomly deleting scores on M_1 (the expensive variable). Specifically, a random number from a uniform distribution ranging from zero to one was generated for each observation in a data set. Cases were sorted based on this random number, and the bottom 20%, 50% and 80% of the cases had the score for M_1 deleted corresponding with the manipulated missing data rates.

Analysis Methods. The analysis method to assess the mediated effect was also a manipulated factor. When using multiple mediators in a planned missing data design, there are a variety of approaches available to analyze the data including a latent variable mediation models, a mediation model among manifest variables with the expensive measure as the mediator and either one or both of the inexpensive measures as auxiliary variables, and a manifest variable model using only the expensive mediator analyzed using maximum likelihood estimation that accounts for missing data or with listwise deletion. For full consideration of possible methodological scenarios, I included these five analysis methods in the study. For all methods, I used M*plus* with maximum likelihood estimation. Recall from Chapter 1 that there are multiple methods to determine the standard error and the resulting significance of the mediated effect. For purposes of this simulation study, I utilized the distribution of the product approach using asymmetrical confidence intervals as implemented with PRODCLIN (MacKinnon, et al., 2007). I chose to use the distribution of the product approach because it has been empirically shown to be more accurate than many of the other commonly used approaches such as those that rely on a normal distribution assumption (MacKinnon et al., 2004), but the distribution of the product approach is not as computer intensive as boot-strap strategies (e.g., bias-corrected bootstrap, percentile bootstrap). Furthermore, the bias-corrected bootstrap method may have inflated Type 1 error rates, and, particularly for smaller sample sizes, may result in coverage less than the desired 95% (Tofighi & MacKinnon, 2011).

The first analysis method (Figure 7, Method 1) was a latent mediation model analysis approach. Method 1 considers the scenario where all three of the mediating variables are available (expensive variable for the subset of the data and two inexpensive measures for the entire sample). I also analyzed the data in M*plus* using a MAR-based maximum likelihood estimation approach with the three observed mediator variables serving as indicators of a latent mediation construct. The complete data version of this model (condition where missing data rate = 0%) serves as a benchmark for comparison of the various analysis approaches.

The second analysis method (Figure 7, Method 2) also incorporates all three mediating variables. This approach utilized a MAR-based maximum likelihood estimation approach with an expensive mediator, M_1, as the sole mediator in the model, and M_2 and M_3 serving as auxiliary models using the saturated correlates approach as described by Graham (2003). In this approach, the auxiliary variables are incorporated with a series of correlations that do not alter the substantive interpretation of the parameter estimates. For this mediation model, the auxiliary variables must correlate with the manifest explanatory variables, other auxiliary variables, and the residual terms of the indicators of the latent outcome variables. Figure 7, Method 2, depicts this analysis model.

Next, I evaluated another auxiliary variable approach (Figure 7, Method 3); this method incorporated only two mediation variables and applies an auxiliary variable approach to a scenario where only one inexpensive mediator is available. Like Method 2, Method 3 uses a saturated correlates approach for inclusion of auxiliary variables. However, unlike Method 2, Method 3 only includes one auxiliary variable (i.e., the model only incorporates mediating variables (M_1 and M_2). As a note, a latent variable model similar to Method 1 would also potentially work in a scenario that only incorporates variables M_1 and M_2. However, such a model requires additional constraints. In a model where the mediating latent variable has only two indicators, both M_1 and M_2 would need to correlate with another variable in the model, have both loadings constrained to be equal, or have a fixed error variance. I chose not to incorporate this model into my research because it would not generalize well to real data where the expensive measure does not have perfect reliability and the choice of initial constraints might be somewhat arbitrary.

Next I applied a manifest variable model with M_1 as the sole mediator and no auxiliary variables using maximum likelihood estimation (Figure 7, Method 4). This mimics a scenario where X and Y are collected on a large sample and an expensive mediator is collected on a subsample. Finally, I applied a manifest variable single mediator analysis to subsample of expensive variables only (M_1) using M*plus* with listwise deletion (Figure 7, Method 5). As a note, because the population generation model was estimated with a loading of one and a residual variance of zero, there was a possibility of there being issues estimating some of the models. To address this potential issue, I constrained the residual variance of M_1 to a small number close to zero (i.e., $\sigma^2_{\varepsilon_1} = .001$).

Methods 1 and 2 serve to demonstrate power that would be achieved when all three of the mediating variables are available. Because there were two viable methods for incorporating these variables, the comparison of these methods will help guide analytical decisions in these scenarios. Method 3 incorporated only two mediating variables and provides information on the viability of only collecting two mediating measures. Method 4 only included one mediator, M_1, but was analyzed with the MAR-based maximum likelihood estimation. Method 5 also included only one mediator, but used listwise deletion. This model can also be thought of as the analysis and reduced *N* that might have been used had the researcher only collected data on the expensive measure.

Data Generation

Data were generated in SAS 9.3 software based on the data generation model in Figure 6. Each design cell has 2000 replications. Data were generated by algebraically computing the model-implied population covariance matrix from the defined manipulated population parameters. For an algebraic expression of the population covariance matrix, see Appendix C. The SAS program in Appendix D automated the process of computing the population covariance matrix based on the chosen population parameters and requires choosing values for twelve parameters: $\lambda_{1,}, \lambda_2, \lambda_3, \sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_2}, \sigma^2_{\varepsilon_3}, \alpha, \beta, \tau', \phi_X, \sigma^2_{\zeta_M}, \sigma^2_{\zeta_Y}$. In this simulation study, λ_1, $\sigma^2_{\varepsilon_1}$, τ', and ϕ_X were fixed values, $\sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_2}\sigma^2_{\varepsilon_3}$, $\sigma^2_{\zeta_M}$ and $\sigma^2_{\zeta_Y}$ were values solely dependent on the other population parameters, and all other parameters were varied as described in the manipulated factors section. Data were generated using the SAS IML procedure by generating five standard normal variables and then using the Cholesky decomposition technique to impose the desired correlation structure based on the population covariance matrix.

Outcomes

The 1,536,000 artificial data sets (2000 replications for each cell of a 4 sample sizes $\times$4 missing data rates$\times$4 α coefficients$\times$4 β coefficients $\times$3 λ's denoting mediation variable correlation matrices) were analyzed using all five of the analysis methods described in the previous section and illustrated in Figure 7. The first step in analysis of outcomes was to analyze the rates of non-convergence and improper solutions from when the maximum likelihood algorithm fails to converge or results in a non-positive definite covariance matrix. High rates of non-convergence were not expected for these simulation conditions, particularly because of the constraint on the residual error, but it was expected that high non-convergence would be more likely to occur for high rates of missing data. Any cases with improper solutions would be identified and removed before computing the other outcome variables. Finally, I evaluated the outcomes of interest: bias, empirical power, and confidence interval coverage. This study is particularly interested in the empirical power to detect the mediated effect.

Bias

Bias is a measure of systematic deviation from the true population parameter. Because the missing data mechanism is MCAR and all analysis methods used have been empirically demonstrated to be unbiased for data that satisfy the MCAR mechanism, it was expected that the procedures would produce negligible levels of bias. Nevertheless, it is prudent to evaluate a measure of bias. The raw bias of a parameter is the difference between the average estimates of the parameter of interest across the 2000 simulation replications and the true population parameter estimate. Note that because the factor loading relating M_1 to LM was constrained to one, all models have the same population α paths, β paths, and resulting mediated effects ($\alpha\beta$). This ensures that all analysis models were on the same metric for comparison. As pointed out by Collins, Schafer & Kam (2001), with a large number of replications it is easy for raw bias to have statistical significance even though it may not be practically significant. To address this, I used a measure suggested by Collins et al. (2001) of *standardized bias*. Standardized bias is a measure of bias that expresses bias in standard error units. Standardized bias was calculated for $\alpha\beta$ effects as follows:

$$SB_{\alpha\beta} = \frac{\overline{\alpha\beta} - \alpha\beta}{ESD}. \tag{3.12}$$

Here, $\overline{\alpha\beta}$ is the average estimate of the mediated effect across all converged replications and $\alpha\beta$ is the true population mediated effect. ESD is an empirical standard deviation of the estimates of $\alpha\beta$ from the complete data conditions. Because the empirical SE from complete data may vary slightly across simulation conditions, I use the within-cell ESD from the complete data (M_1 missing data rate = 0%) maximum likelihood estimation approach as in Method 1, Figure 7 across all conditions.

Empirical Power

The empirical power to detect the α path, β path, and mediated effect ($\alpha\beta$) for each condition is calculated by determining the proportion of non-zero effect size replications in which a significant result was obtained using the distribution of the product approach implemented using PRODCLIN as previously described (MacKinnon, Fritz, et al., 2007; Tofighi & MacKinnon, 2011). Because power is only relevant for non-zero effects, power was not evaluated for situations where the population mediated effect, $\alpha\beta$, is zero; this includes any instance of α and/or β equal to zero. In conditions where the population value of $\alpha\beta$ equals zero, empirical power was replaced with Type 1 error rates. Type 1 error was calculated by determining the proportion of replications where the asymmetric CI does not include zero.

Confidence Interval Coverage

I also evaluated the 95% confidence interval coverage for the mediated effect. Confidence interval coverage was determined by computing the proportion of replications in which the asymmetric confidence intervals contain the true population parameter. Ideally, coverage would be close to 95% for a .05 alpha level. The extent to which coverage deviates from 95% reflects inaccuracies in the standard errors. If the standard errors are too large, confidence intervals capture the population parameters too frequently and coverage rates are larger than 95%. Conversely, if standard errors are too small, confidence intervals do not capture the population parameter as frequently as they should and coverage rates drop below 95%. Because of these properties of coverage, coverage provides a benchmark for assessing the accuracy of the standard error estimates because it is directly related to Type 1 error inflation (e.g., 90% coverage implies a two-fold increase in Type I errors).

Recap of Expectations

As stated in the introduction of this chapter, I expected that the missing data analysis model that incorporate all three mediation measures (Figure 7, Methods 1 and 2) would perform better than the other models by virtue of the availability of additional data. As with most methods, I also expected that all methods would perform better as sample sizes increase and missing data rates decrease. Because there are two viable missing data methods for incorporating multiple measures of the same mediation construct, one goal of this study is to compare latent variable mediation models with auxiliary variable models to help guide analytical decisions in the future. This research seeks to provide guidance to researchers interested in utilizing planned missing data designs in mediation analyses.

CHAPTER 4

RESULTS

To begin, I checked the accuracy of the data generation by confirming accurate missing data rates, checking for non-convergence and improper solutions, and computing the average complete-data parameter estimate within each design cell. To check for non-convergence, I evaluated the minimum and maximum parameter estimates in each design cell to ensure all values were reasonable estimates of the population parameters and also confirmed that M*plus* did not provide any negative variances. There was no evidence of either non-convergence or improper solutions. Next, I turned to the outcomes of interest as described in the previous chapter: bias, empirical power, Type 1 error, and confidence interval coverage.

Bias

Bias is a measure of systematic deviation from the true population parameter. In this study, the missing data mechanism is MCAR. Previous literature clearly indicates that all of the analysis procedures evaluated produce unbiased results when the MCAR mechanism is satisfied (e.g., as described in Enders, 2010). As expected, the analysis procedures produced negligible levels of bias. Table 4 shows the descriptive statistics of raw bias for each missing data rate collapsed across all other factors (i.e., α, β, λ and N). Table 4 demonstrates that the values of raw bias range from -0.0001 to 0.0009. Raw bias across all design cells (last row in Table 4) had a trivial mean, median and standard deviation ($Mean_{\overline{\alpha\beta}-\alpha\beta}$ = 0.0003, $Median_{\overline{\alpha\beta}-\alpha\beta}$ = -0.0003, $SD_{\overline{\alpha\beta}-\alpha\beta}$ = 0.0442). Correspondingly, standardized bias (calculated using Equation 3.12 from the previous chapter) also indicated negligible levels of bias. As described previously, I used the within-cell Empirical Standard Error (ESD) from the complete data condition (M_1 missing rate = 0% which provides equivalent results across all methods) to standardize raw bias; the resulting value provides the proportion of a standard error the estimate falls above or below the true parameter value. Table 5 shows descriptive statistics for standardized bias by method and missing rate collapsed across all other factors (i.e., α, β, λ and N). Previous literature suggests that a design cell with bias above an absolute value of 0.40 is of practical significance (Collins et al., 2001). As Table 5 demonstrates, standardized bias values across all conditions were well under the threshold that might be considered problematic; even the cells with the minimum and maximum bias were well below an absolute value of 0.40. The design cell with the largest bias (-0.337) occurred in the 80% missing data rate condition with the maximum likelihood analysis method that incorporated no addition mediators (Method 4 in Figure 7) where $\alpha = 0$, $\beta = 0$, $\lambda = .8$, and $N = 100$. A standardized bias value of -0.337 suggests that, on average, the estimate falls about one third of a standard error below the true parameter value. In general, the 80% missing data rate condition had the greatest bias. Appendix E provides the standardized bias with an 80% missing rate for all sample sizes, non-zero effect sizes, and analysis conditions. Because the bias outcome performed as expected and demonstrated minimal bias, the function of estimating bias served mainly as a data generation check and I did not explore further trends.

Empirical Power

The empirical power to detect the mediated effect ($\alpha\beta$) for each condition was calculated by determining the proportion of non-zero effect size replications in which a significant result was obtained using the distribution of the product approach; the distribution of product approach was described in the previous chapter. In other words, power was determined as the percentage of replications that did not include zero in the asymmetric confidence interval. For conditions where the population value of $\alpha\beta$ was equivalent to zero, Type 1 error rates were evaluated in lieu of power and discussed later.

Power varied widely depending on design cell with power estimates ranging from nearly zero (0.60%) to 100.00%. As would be expected, power tended to increase with increasing mediation effect size ($\alpha\beta$), increasing sample size (N) and decreasing missing data rate. Power also varied as a function of analysis method within a given design cell. Consistent with expectations, power is the greatest for analysis methods that include all three mediators. Specifically the latent variable model (Method 1, Figure 7) demonstrates the most empirical power followed by the model with two auxiliary variables (Method 2, Figure 7). Power decreases with analysis methods that include less information with the other methods showing decreasing power in the following order: one auxiliary variable model (Method 3, Figure 7), maximum likelihood estimation model with no additional mediators (Method 4, Figure 7), and listwise deletion (Method 5, Figure 7). This pattern of decreasing power among the five methods was consistent in every design cell with the exception of cells that had floor or ceiling effects (i.e., power at or near 0 or 100% across all methods). Table 6 shows the power for each method collapsed across all other factors. The average power across all design cells was 57.14% (last row in Table 6). For reference, Appendix F provides the power results for all design cells.

To probe the effects of the manipulated factors on power, I ran a variety of logistic regression analyses on the replications where the population mediated effect, $\alpha\beta$, was not equal to zero. In total, there were 4,320,000 replications where population $\alpha\beta$

was non-zero. However, because all five analysis methods produced identical results with no variability for design cells in the complete data condition, I limited the logistic regression analyses to the 3,240,000 replications with a non-zero missing data rate (i.e., 20%, 50% or 80% missing on variable M_1). The outcome variable in the logistic regression analysis was coded zero when the asymmetric confidence interval of $\alpha\beta$ contained zero (i.e., non-significant results) and coded one when the asymmetric confidence interval did not contain zero (i.e., significant results at the .05 significance level). In order to simplify analyses, all factors were treated as between-subjects factors, including the analysis method factor which is technically a within-subjects factor. Additionally, for ease of analysis, the manipulated factors were effect coded as categorical variables. I started by running a logistic regression analysis using the full factorial up to a six-way interaction ($\alpha \times \beta \times \lambda \times N \times$ Missing Rate $\times$ Method; $\chi^2(192) = 148$, $p = 0.99$). In order to better understand the interactions among the simulated factors, I ran a model using effect coding that included all lower order terms up to and including all three-way interactions. I also chose to combine factors α and β into one mediated effect size factor, $\alpha\beta$. I chose this combination because a preliminary assessment of the data suggested that the mediated effect size, $\alpha\beta$, provided similar results regardless of the pattern of α and β comprising the mediated effect. In other words, power results were similar small α and medium β versus medium α and small β and so forth. Table 7 provides the results of this logistic regression analysis treating the mediated effect size as one factor for an analysis with all lower order terms up to and including all three-way interactions. The Table is grouped by order of terms and ranked by the size of the Wald χ^2 joint test. Because the factors were effect coded, the lower-order terms represent the conditional effects at the mean. All three-way and two-way interactions from this analysis were significant. The only significant conditional main effect was the mediation effect, $\alpha\beta$.

Due to the number of significant effects in the logistic analysis, and because so many cells in the design were constrained by floor and ceiling effects (i.e., very low power or power reaching 100%), it was difficult to identify trends. Only a limited range of cells provided sufficient variation to evaluate power trends. Consequently, it became fruitful to look at effects graphically. Because this research is concerned with method, I focused on effects that included method as a factor and evaluated all two and three-way interactions. In a graphical investigation, I identified two prominent two-way interactions. Specifically, there appeared to be a method by λ interaction and a method by missing rate interaction. Evaluation of three-way interactions suggested that these two-way interactions were modified by a third dimension. The notable three-way interactions that demonstrated a noticeable change in power by method, and supported by the size of the Wald joint test and non-centrality parameters (shown in Table 7), were Method$\times\lambda\times$Missing Rate, Method$\times\alpha\beta\times$Missing Rate, and Method$\times\alpha\beta\times\lambda$. The remainder of this section will focus on these three three-way interactions.

Method×λ×Missing Rate

Table 8 provides percent power to detect the mediated effect for the Method $\times\lambda\times$Missing Rate interaction. This table shows the average percent power value for each combination of λ, missing rate, and method averaged across N, α, and β. For reference purposes, power for the complete data condition is shown in the first row of data.[3] Recall that λ is the loading used in the data generation model for indicators M_2 and M_3 and values of λ directly correspond with the correlations among the mediators. As shown in Table 3, values of λ = 0.4, 0.6 and 0.8 for M_2 and M_3 are equivalent to the correlations of M_2 and M_3 with M_1 ($r_{M_1M_2} = r_{M_1M_3}$ = 0.4, 0.6 and 0.8, respectively). As demonstrated by Table 8, there is a Method by λ interaction effect whereby increasing λ results in increasing power for the missing data methods that incorporate M_2 and M_3 (Methods 1 – 3, the latent variable model and the auxiliary variable models). Not surprisingly, power for methods that don't include M_2 and M_3 (Methods 4 and 5, ML with no additional mediators and listwise deletion, respectively) was not functionally improved by increased λ. The missing data rate further modifies the method by λ interaction; the impact of λ on power is amplified for higher missing data rates. This three-way interaction is graphically presented in Figure 8. This figure depicts the interaction between method and λ paneled by missing data rate for M_1. This figure demonstrates that for 80% missing (bottom panel of Figure 8), power is extremely variable for the five methods when λ is large (λ = 0.8), but this variability is reduced considerably when λ is small (λ = 0.4). When the missing rate is 20% (top panel of Figure 8) there is minimal variability of power as a function of λ. Figure 9 is an alternative visual presentation of these results and shows a bar graph of the percent power for each analysis method for the different values of λ paneled by the missing data rate. The solid bars denote power for λ = 0.4, the striped bars denote λ = 0.6,

[3] Note that in complete data conditions, all methods provide nearly identical estimates for α, β, and the corresponding standard errors. Consequently, percent power values for the five methods are identical through the first decimal place for complete data, and it is only necessary to provide power estimates for one of the analysis methods. In all tables that include complete data results, I present results for Method 1, the latent variable model.

and the grey bars denote $\lambda = 0.8$. Note that the replications with 20% missing data (the left panel) show limited variability across missing data methods and ranges of λ whereas 80% missing data (the right panel) demonstrate significant changes in power conditioned on the analysis method and value of λ.

Method×αβ×Missing Rate

Table 9 presents power for the Method×$\alpha\beta$×Missing Rate interaction. Complete data results are provided in the first row for comparison. Recall that I collapsed the α and β factors into a single factor (e.g., $\alpha = 0.1$ and $\beta = 0.3$ were treated the same as $\alpha = 0.3$ and $\beta = 0.1$), resulting in a single effect size factor with six levels: 0.01, 0.03, 0.05, 0.08, 0.15, and 0.25. This table shows the average percent power of each combination of mediated effect size ($\alpha\beta$), missing data rate, and method averaged across sample size and λ. For reference, complete data results are included in the first row of data. There is a method by $\alpha\beta$ interaction where the differences in the performances of the five methods change depending on effect size. This two-way interaction is further moderated by the missing data rate of M_1; the interaction between method and effect size on power is contingent on the missing data rate. This three-way interaction is clearer with graphical inspection. Figure 10 provides graphs showing the interaction between $\alpha\beta$ and method on power paneled by missing data rate. Notice that for the 20% missing data rate (the top panel) there is no two-way interaction: all methods effectively produce the same level of power for the 20% missing data condition. For higher rates of missing data, there is an interaction between method and mediated effect size. Differences among the performance of the methods for a given effect size are particularly salient when the missing data rate is 80% (bottom panel of Figure 10). Some of the observed interaction may be driven by floor and ceiling effects. For example, notice that for $\alpha\beta$ = .25, missing data rates 20% and 50% demonstrate nearly 100% power and no differentiation among the methods; this may be due to a ceiling effect. Differential performance for the effect size when both α and β are large, $\alpha\beta$ = .25, only occurs at a missing data rate of 80%. Figure 11 is an alternative visual presentation of these results and shows a bar graph of the percent power for each analysis method for three select mediated effect sizes ($\alpha\beta$ = 0.01, 0.09 and 0.25, corresponding with $\alpha = \beta$ for small, medium and large effect sizes, respectively) paneled by missing data rate on M_1. The black bars denote power for $\alpha\beta$ = 0.01, the striped bars denote power for $\alpha\beta$ = 0.09, and the grey bars denote power for $\alpha\beta$ = 0.25. Notice that the relation between method and power for a given mediated effect size is relatively consistent across the methods for a missing data rate of 20% (left panel). For a missing data rate of 80%, it is clear that there is an interaction between method and effect size because the power for a particular effect size varies depending on method. To illustrate these differences, consider that at a 20% missing data rate, power only ranges from 22.93 – 26.71% (range of 3.78) for $\alpha\beta$ = 0.01 across the analysis methods, from 87.80 – 90.42% (range of 2.62) across the methods for $\alpha\beta$ = 0.09, and from 99.69 – 99.84% for $\alpha\beta$ = 0.25 (range of 0.15). These ranges become more extreme at 80% missing data with the medium sized mediated effect ($\alpha\beta$ = 0.09) showing the most differences between methods. For an 80% missing data rate, power ranges from 2.71 – 16.71% (range of 13.56) across the methods for $\alpha\beta$ = 0.01, from 49.97 – 78.70% across the methods for $\alpha\beta$ = 0.09 (range of 28.73), and from 80.03 – 96.93% (range of 16.90) across the methods for $\alpha\beta$ = 0.25. This variation in ranges supports a method by $\alpha\beta$ by missing rate interaction.

Method$\times\alpha\beta\times\lambda$

Table 10 provides power for the Method$\times\alpha\beta\times\lambda$ interaction. This table shows the average percent power value for each combination of mediated effect size $\alpha\beta$, λ, and method averaged across all non-zero missing data rates and sample size. Recall that λ directly corresponds to the correlations among the mediators. This three-way interaction suggests that the interaction between method and mediated effect size (as previously described) is moderated by λ. Figure 12 depicts the interaction of method and mediated effect size on power paneled by λ. This particular three-way interaction doesn't appear to be as strong as the other interactions discussed. As shown in the top panel of Figure 12, for small λ (λ = 0.4 corresponding with small correlations among the mediators), differences among the methods are not very pronounced across all effect sizes. For larger λ (λ = 0.6 and 0.8), differences among the methods become more pronounced. Not surprisingly, methods that do not incorporate the additional mediators (making the size of λ irrelevant) do not perform as well. There also appears to be a ceiling effect for λ = 0.8 as shown in the last panel of Figure 12. For most mediated effect sizes, there is differentiation among the methods when λ = 0.8. However, the influence of λ by method on power is negligible for a large mediated effect size ($\alpha\beta$ = 0.25) and power is greater than 92% across all methods. Figure 13 presents an alternative graphical representation of this three way interaction. This figure depicts a bar graph of percent power for each analysis method for three values of $\alpha\beta$ paneled by the values of λ. The solid bars denote power for $\alpha\beta$ = 0.01 (α = β = 0.1), the striped bars denote $\alpha\beta$ = 0.09 (α = β = 0.3), and the grey bars denote $\alpha\beta$ = 0.25 (α = β = 0.5). The interaction is not particularly clear from this graph, but the potential ceiling effect can be seen by the grey bars indicating the power for $\alpha\beta$ = 0.25. These bars have less variation relative to the black bars representing $\alpha\beta$ = 0.01.

Type 1 Error Rates

For conditions where the population values of $\alpha\beta$ equal zero (i.e., α and/or β equal zero), Type 1 error rates were evaluated in lieu of power. Because I used the 5% significance level in estimating the confidence intervals of the mediated effect, it is expected that 5% of the samples will yield intervals that do not contain zero when the population mediated effect is zero (i.e., $\alpha\beta = 0$). Asymmetric confidence interval estimates calculated using the product of coefficients approach described previously when the population mediated effect is zero yielded an empirical Type I error rate averaged across all cells of 3.40%; Type 1 error rates across all cells ranged from 0.00 – 9.90%. Based on a sample size of 2000 (2000 replications per design cell), I calculated the standard error expected for the proportion 0.05. Then I formed a $\pm 1.96 \times SE$ interval around the nominal rate of 0.05. This resulted in a range of 0.0404 to0 .0596. From this calculation, I conclude that values with honest error rates would expect to yield Type 1 error rates ranging from 4.04% to 5.96%; this is the range we might expect to see if the Type 1 error rate was actually 5%. For replications with complete data, 54.76% of the Type 1 error rates were within this range and complete data had a mean Type 1 error rate of 3.51% (M = 3.51, SD = 2.11). For 20% missingness, 51.43% of the Type 1 error rates were within the honest error rate range (M = 3.42, SD = 2.11). For 50% missingness, 45.24% of the Type 1 error rates were within the range (3.34, SD =2.21), and for 80% missingness, 35.00% were within the range (M = 3.35, SD = 2.41). For conditions where population parameters $\alpha = \beta = 0$ (as opposed to conditions where only α ***or*** β equivalent to zero), there were no instances of Type 1 error rates within the honest error rate range of from 4.04% to 5.96% across all design cells; these conditions where both population values of α and β equivalent to zero had a mean Type 1 error rate extremely close to zero (M = 0.23, SD = 0.25).

To better understand the effect of the manipulated factors on Type 1 error rates, I further probed using logistic regression. I conducted a logistic regression analysis with the manipulated factors as predictors including all three-way interactions and corresponding lower order terms on the 2,520,000 analysis results for where population $\alpha\beta = 0$. In order to simplify analysis and to be consistent with the logistic regression used for power analysis, all factors were treated as between-subjects factors, including analysis method which is technically a within-subjects factor. For ease of analysis, the manipulated factors were effect coded as categorical variables. The treatment of the mediated effect size in this analysis warrants a comment. Assessment of Type 1 errors is only pertinent to data with α or β equivalent to zero. When α and β are included as separate factors, the resulting logistic model is a fractional factorial requiring more complex estimation; the resulting design is not fully crossed because Type 1 errors are only pertinent when at least α or β is equivalent to zero, all levels of α (0.0, 0.1, 0.3 and 0.5) are not fully crossed with all levels of β (0.0, 0.1, 0.3 and 0.5). To remedy this, I combined factors α and β into a single factor. Combined, and including only conditions where at least α or β is equivalent to zero, the effect size factor has 7 potential levels: (1) $\alpha = \beta = 0.0$, (2) $\alpha = 0.0$ and $\beta = 0.1$, (3) $\alpha = 0.1$ and $\beta = 0.0$, (4) $\alpha = 0.0$ and $\beta = 0.3$, (5) $\alpha = 0.3$ and $\beta = 0.0$, (6) $\alpha = 0.0$ and $\beta = 0.5$, (7) $\alpha = 0.5$ and $\beta = 0.0$. A review of the Type 1 error rates suggested that Type 1 error rates for α ***or*** β at a particular effect size (small =

0.1, medium = 0.3, and large = 0.5) were virtually equivalent; in other words, α = 0.0 and β = 0.1 provides similar Type 1 error rates as α = 0.1 and β = 0.0 and likewise for medium and large effect sizes of α and β. Thus, conditions with a non-zero path of α or β = 0.1 (Levels 2 and 3 above) produced similar Type 1 error rates across all other factors. Similarly, conditions with a non-zero path of α or β = 0.3 (Levels 4 and 5) produced similar Type 1 error rates and α or β = 0.5 (Levels 6 and 7) produced similar Type 1 error rates. To simplify analyses, the effect size factor was collapsed to four levels: (1) $\alpha = \beta$ = 0.0, (2) α ***or*** β = 0.1, (3) α ***or*** β = 0.3, (4) α ***or*** β = 0.5.

The results of the logistic regression are in Table 11. I focus mainly on the three-way interactions as the important main effects are subsumed in the presentation of the three-way interactions. Interpretation of the lower order terms is based on effect coding, thus lower order terms are evaluated at the mean. Three of the three-way interactions had extremely large Wald χ^2 values (Wald χ^2 > 99) and large non-centrality parameters (λ > 75). Of these three interactions, two interactions included the method effect, the effect of most interest to this research study. I used these three-way interactions to parse out interactions that were meaningful for presentation purposes. I present results in terms of a Method $\times$ $\alpha\beta$ $\times$ Missing Rate interaction and a Method $\times$ N $\times$ Missing Rate interaction.

Method $\times$ $\alpha\beta$ $\times$ Missing Rate

Table 12 presents the Type 1 error rates of Effect Size $\times$ Method $\times$ Missing Data Rate. Bold values in Table 12 represent values within the honest range of 4.04% to 5.96% as previously described. To ease presentation of results, Figure 14 displays the results graphically showing the Type 1 error rates of method crossed with effect size (size of α or β) paneled by missing rate on M_1. Effect size indicates the size of the non-zero α or β effect (effect size equal to zero for $\alpha = \beta = 0.0$). The grey brands represent the range 4.04% - 5.96%, the values we might expect to see if the Type 1 error rate was actually 5%. A three-way interaction means that the two-way interaction of method by effect size differs by missing data rate. For a 20% missing data rate, all methods performed comparably for a given effect size. With increasing missing data rates, there was a greater distinction in Type 1 error rates across methods and this distinction is magnified as the α or β paths increase. The differential performance of the different methods was particularly acute for a high missing data rate (missingness = 80%) and a small sample size (N = 100). Also notice that the Type 1 error rates within or very close to the expected range (represented by the grey band) for a 5% Type 1 error rate only when α or β = 0.3 or 0.5 (corresponding with an α or β path that is medium or large).

Method × *N* × Missing Rate

Table 13 presents Type 1 error rates based on Method × *N* × Missing Rate interaction. Bold values in Table 13 represent values with honest error rates ranging from 4.04% to 5.96%. Notice only two cells contain values within this range (complete data for N = 1000 and the model with two auxiliary variables with an 80% missing data rate for N = 100), demonstrating that these Type 1 error rates are somewhat lower than the expected value of 5%. To ease presentation of results, I provide Figure 15 showing the Type 1 error rates of method crossed with sample size (N) paneled by the missing data rate. Because of the low rate of values within the honest error range, I do not include grey bands to depict this range as the grey band would be largely outside the scope of the graph. The three-way interaction means that the method by sample size (N) interaction varies across missing data rate. As demonstrated by Figure 15, the relationship between method and sample size on Type 1 error rates differs depending on the missing rate. At a 20% missing data rate, there is a minimal *N* by method interaction; methods perform nearly identically for a particular sample size, with minor differentiation for small sample sizes. For larger missing data rates (50% and 80%), there is a clear method by sample size interaction. For example, the Method 4 (ML with no additional mediators) trajectory is much less steep than the Method 1 (latent variable model) trajectory. The interaction becomes more acute for a missingness rate of 80%. For small samples (N = 100) with a high missing data rate (80% missing), the various methods produced large ranges of Type 1 error rates (0.00% - 9.90%). For large samples (N = 1000), the range of Type 1 error rates across methods for 80% missingness was reduced considerably (.05% -6.75%).

Confidence Interval Coverage

I assessed confidence interval coverage by computing the proportion of replications where the 95% asymmetric confidence interval contained the true mediation population parameter, $\alpha\beta$. If the estimated asymmetric confidence intervals are accurate, confidence interval coverage for a .05 alpha level should equal 95%. In contrast, if the asymmetric confidence intervals are too narrow, confidence intervals will not capture the population parameter as frequently as they should, and coverage rates will drop below 95%. From a practical standpoint, coverage provides a benchmark for assessing the accuracy of the asymmetric confidence intervals because it directly related to Type I error inflation (e.g., a 90% coverage value suggests a twofold increase in Type I errors, an 85% coverage value reflects a threefold increase, and so on). Based on a sample size of 2000 (2000 replications per design cell), I calculated the standard error expected for the proportion 0.95. I then formed a ±1.96×*SE* interval around the nominal rate of 0.95. This resulted in a range of 0.9405 to 0.9596. From this computation, I conclude that values with honest confidence intervals would expect to yield confidence interval percentages ranging from 94.05 to 95.96%; this is the range we might expect to see if the confidence interval was actually 95%.

Coverage was evaluated only for replications where the mediated effect was non-zero; when the mediated effect is zero, coverage is directly related to Type 1 error (i.e., Coverage = 100% – Type 1 error %). Across all design cells where both $\alpha\beta$ and the missing rate are non-zero, confidence interval coverage had a mean of 94.90% and ranged from 89.75 to 99.80%. For complete data, coverage performed exactly as expected with a mean of 95.00% and a range of 93.00 to 99.25%. For replications with missing data (i.e., missing rates 20%, 50% and 80%), coverage had a minimum value of 89.75% and a maximum value of 99.80%. Across the five analysis methods, there was not a lot of variation with coverage. Table 14 shows the average percent coverage for each analysis method collapsed across all other factors (only for $\alpha\beta > 0$ and missing rate > 0). Although some cells had coverage values outside of the honest range for 95% coverage values (i.e., 94.05 to 95.96% as described), the aggregate values for coverage by method as presented in Table 14 all fall within the expected range for 95% confidence interval coverage.

Even though coverage generally performed as expected and desired, I further probed the effects of the manipulated factors on coverage. Similarly to the previous logistic regression to understand the effects of the manipulated factors on power to detect the mediated effect, I limited the logistic regression analyses for coverage to the

3,240,000 replications with both a non-zero mediated effect and a non-zero missing data rate (i.e., 20%, 50% or 80% missing). The outcome variable in the logistic regression analysis was coded zero when the asymmetric confidence interval for the mediated effect did not contain the population value of $\alpha\beta$ and coded one when the asymmetric confidence interval for the mediated effect did contain the true population value of $\alpha\beta$. As in previous analyses, all factors were treated as between-subjects factors, including the analysis method factor which is technically a within-subjects factor. Additionally, for ease of analysis, the manipulated factors were effect coded as categorical variables. I started by running a logistic regression analysis using the full factorial up to a six-way interaction ($\alpha \times \beta \times \lambda \times N \times$Missing Data Rate $\times$Method; $\chi^2(192) = 239.63$, $p = 0.01$). Consistent with the power section, I also ran a model that only included all lower order terms up to and including all three-way interactions. As previously, I ran the logistic regression on the 3,240,000 replications in two ways: (1) treating α and β as separate factors and (2) combining α and β into one mediated effect factor, $\alpha\beta$. As previously, I provide the results for this second analysis treating the mediated effect as a single factor. Table 15 provides the results of this logistic regression analysis grouped by order of terms and ranked by the size of the Wald χ^2 joint test. Because the factors were effect coded, the lower-order terms represent the conditional effect at the mean. Although overpowered, not all effects were significant. Based on the results, I further investigated three three-way interactions that include the method factor: Method$\times\alpha\beta\times$Missing Rate, Method$\times\alpha\beta\times N$, and Method$\times N \times$Missing Rate. However, because there was limited variation in the design cells and a majority of the coverage values were in the range expected for a true 95% coverage, I only briefly focus on the confidence interval coverage interactions.

Method×$\alpha\beta$×Missing Rate

Table 16 shows the 95% confidence interval coverage values for the method by $\alpha\beta$ by missing data rate interaction; each cell is averaged across all values of sample size (N) and λ. All values were close to the desired 95% value. Bold values in Table 16 represent values within the previously described honest range of 94.05 to 95.96%. Figure 16 shows the method by missing rate interaction paneled by mediation effect size, $\alpha\beta$. This Figure makes the interaction more clear. The graph shows that the interaction between method and mediation effect size is moderated by missing data rate. The interaction between method and mediated effect is nearly non-existent for a missing data rate of 20%, but quite apparent for a missing data rate of 80%. From this graph, it is also clear that for lower missing data rates (missing rates = 20 and 50%), coverage values fall in the honest range for all mediated effect values where $\alpha\beta > 0.01$. However, for the high missing data rate (missing rate = 80%), very few coverage values fall in this range.

Method×$\alpha\beta$×N

Table 17 shows the 95% confidence interval coverage values for the method by $\alpha\beta$ by N interaction; each cell is averaged across all other factors (i.e., missing data rate and λ). Consistent with the logistic regression analysis, complete data conditions were not included when aggregating values across missing data rates. Bold values in Table 17 represent values within the previously described honest range of 94.05 to 95.96%. Figure 17 provides a graphical depiction of the method by sample size (N) interaction paneled by mediated effect size ($\alpha\beta$). Here, it is clear that as N increases, all design cells converge towards 95%. However, the effect of method on coverage varies differentially depending on sample size. At smaller sample sizes, coverage is more likely to be out of the honest range. Furthermore, this two-way interaction is moderated by mediated effect size. Depending on the mediated effect size, coverage values may be above or below the desired range for small sample sizes.

Method×*N*×Missing Rate

The last three-way interaction considered in terms of coverage is the method by N by missing rate interaction. Table 18 provides the percent coverage rates for method by N by missing rate averaged across all other factors (i.e., $\alpha\beta$ and λ). Bold values in Table 18 represent values within the honest range of 94.05 to 95.96%. Note that only two conditions fall outside of this range: missing rate of 80% for N = 100 for both auxiliary variable models. Figure 18 provides a graphical depiction of the method by sample size (N) interaction paneled by missing data rate. The top two panels (missingness = 20% and 50%) show a slight interaction between method and sample size, N. However, when missingness is high (missingness = 80%, bottom panel), the interaction is particularly salient. Not surprisingly, small sample sizes coupled with large missing data rates magnify the differential performance of the analysis methods. For missingness = 80% and a small sample size (N = 100), there is considerable differentiation among the methods in terms of coverage values.

CHAPTER 5

DISCUSSION

Psychologists and other social scientists often use mediation analyses to investigate mechanisms of change. However, the sample sizes required to achieve sufficient power to detect a mediated effect are frequently quite high. Modern research in psychology and social sciences also entails increasingly sophisticated measurement of mediating mechanisms, but these measurements are often expensive. Researchers with limited budgets may be forced to choose between using expensive measures on a small sample of participants or less expensive measures on a larger sample of participants. One area of methodological research, modern missing data analysis, has demonstrated that, in some situations, carefully planned missing data designs that employ modern analysis techniques may optimize utilization of resources. The current study investigated the use of intentional missing data designs for a mediation analysis that incorporates multiple measures of the same mediation construct. My research considered two classes of measures of a mediating variable: expensive and inexpensive. Specifically, the current study considered a scenario where researchers are most interested in an expensive measure of the mediating variable, but they are unable to afford the required sample size to ensure adequate statistical power to detect the mediated effect. However, the researchers can potentially collect additional, less expensive, mediators on a larger sample of participants.

As described in detail in Chapter 3, the study used simulations with data generated from a variety of population parameters for manifest X and Y variables and a latent mediator variable, LM, with three observed indicator variables, M_1, M_2, and M_3.

Planned missingness was implemented under the missing completely at random (MCAR) mechanism on variable M_1 to mimic a scenario where M_1 is the expensive mediator that cannot be collected for the full sample of participants. I then evaluated five approaches of incorporating the available measures into a mediation analysis: (Method 1) a latent variable mediation model where all three mediators are indicators of a latent mediation construct, (Method 2) an auxiliary variable model where one of the mediation variables, M_1, is the mediator, and the other mediation variables, M_2 and M_3, are auxiliary variables, (Method 3) an auxiliary variable model where M_1 is the mediator, and M_2 serves as a single auxiliary variable, (Method 4) an analysis using maximum likelihood estimation, including all available data but incorporating only one mediator, M_1, and (Method 5) an analysis using listwise deletion and incorporating only one mediator, M_1. The goal of this study was to evaluate the potential use of planned missing data designs with mediation analysis by evaluating the empirical power to detect the mediated effect using five different analysis methods under a variety of simulation conditions. Specifically, the research aimed to address the question of what the best method is for incorporating additional inexpensive measures into a cross-sectional mediation analysis with planned missing data.

Summary and Discussion of the Results

Chapter 4 describes the results in detail. As expected given that all analyses satisfied the MCAR mechanism, there was minimal bias in the results. Design cells with 80% missing data on M_1 demonstrated the most bias. I also evaluated confidence interval coverage for conditions where the population value of the mediated effect, $\alpha\beta$, was non-zero. In most situations, coverage values were within the range of values expected if the true coverage rate in the population was 95%. Coverage most often deviated from the desired 95% value for conditions with high missing data. I evaluated Type 1 error rates for conditions where the population mediated effect, $\alpha\beta$, was zero. When both mediated effect population coefficients, α and β, were zero, Type 1 error rates were extremely close to zero for all missing data rates. For conditions where the population mediated effect was zero, but either α or β were non-zero, Type 1 error rates were closer to the expected value of 5%. As the missing data rate increased, Type 1 error rates tended to decrease.

The main outcome of interest was empirical power to detect the mediated effect; empirical power was evaluated for conditions where the population mediated effect, $\alpha\beta$, was non-zero. The average empirical power across all design cells for the complete data condition was 66.46%. The empirical power values observed for complete data conditions were somewhat lower than the values found in previous literature (MacKinnon et al., 2002; Fritz & MacKinnon, 2007), but these differences can be ascribed to the population generation model and the specific definition of small, medium, and large effect sizes between the current and the previous studies. Unlike the current study which used standardized regression coefficients (equivalent to correlations) to define small, medium, and large effects for α and β (i.e., 0.1, 0.3, and 0.5, respectively), MacKinnon and colleagues (2002) defined small, medium, and large effect sizes based on proportion of variation accounted for in the dependent variable. The specific data generation procedure of the previous literature resulted in larger standardized α and β coefficients for small, medium, and large effect sizes (i.e., 0.139, 0.363, 0.508 for conditions with complete mediation). Consequently, empirical power values from these previous simulation studies were higher. Table 19 compares the complete data power estimates from this study to the results from MacKinnon and colleagues (2002) for mediated effect sizes where $\alpha = \beta$ for small, medium, and large α and β coefficients.

As expected based on previous literature (e.g., Graham, et al., 2001), empirical power to detect the mediated effect decreased when missing data was introduced; as the percent of missing data in the mediator increased, empirical power to detect the mediated effect decreased. However, as compared to the reduction in the amount of observed data on M_1, the decrease in power was less than might be expected. This is consistent with the literature that suggests that the proportion of information lost due to planned missing data (i.e., the proportion of missing data observations) is often larger than the amount of power lost due to planned missing data (Graham et al., 2001; Graham et al., 2006; Jia et al., 2014). Although there was 20% missingness on M_1, the latent variable mediation model (Method 1) only demonstrated a reduction in power of 2.06% across all design cells relative to the complete data condition. The condition with 50% missingness using the latent variable mediation model (Method 1) resulted in a 6.02% reduction in power across all design cells. Similarly, for 80% missingness, there was only an 11.86% reduction in power across all design cells using the latent variable mediation model (Method 1). The other analysis methods did not perform as well as the latent variable mediation model. For the 80% missing data condition, the model with two auxiliary variables (Method 2) demonstrated an average reduction of power of 20.72% relative to the power for complete data. Method 2 had lower empirical power than Method 1 even though the total amount of data utilized in the two methods was the same. Not surprisingly, Methods 3 through 5 exhibited larger reductions in power for the mediated effect, given that they were, in fact, based on fewer variables. The model with one auxiliary variable (Method 3) demonstrated an average reduction of power of 23.87%; the model with no additional mediators using maximum likelihood estimation (Method 4) demonstrated an average reduction of power of 29.35%; and listwise deletion (Method 5) demonstrated an average reduction of power of 32.40%. Consistent with expectations, power was the greatest for analysis methods that included all three mediators and power decreased with analysis methods that included less information (i.e., fewer mediators).

The fact that maximum likelihood estimation only performed marginally better than listwise deletion (3.03% more power averaged across all design cells) was somewhat surprising given previous missing data literature that suggests that using maximum likelihood estimation results in greater statistical power and efficiency than listwise deletion (e.g., Enders & Bandalos, 2001; Little & Rubin, 2002; Schafer, 1997; Schafer & Graham, 2002). As described in Chapter 2, Zhang and Wang (2013) specifically evaluated modern missing data methods in the context of mediation and found that maximum likelihood estimation had increased power to detect the mediated effect compared to listwise deletion. The similarities between the aggregated empirical power values of maximum likelihood estimation (Method 4) and listwise deletion (Method 5) may be somewhat driven by the many design cells that had floor and ceiling effects in the results (i.e., power approaching zero or 100%). For conditions that showed the greatest variability among the methods, the differences between the performance of maximum likelihood estimation and listwise deletion in terms of empirical power were more notable. The largest difference between the two methods was found in the design cells where $\alpha = \beta = 0.5$ (large effect size) and $N = 100$ (using results from $\lambda = 0.8$). For this condition, maximum likelihood estimation achieved power that was 12.60% higher than power attained with listwise deletion. Although there was only a small power advantage with maximum likelihood over listwise deletion, when aggregated across all design cells, in some conditions, maximum likelihood estimation may have an appreciable increase in power over listwise deletion. The increase in power of maximum likelihood estimation over listwise deletion is consistent with the missing data literature.

The results of the simulation study also demonstrated that, in terms of accounting for empirical power, the analysis method (i.e., Methods 1 – 5) interacted with missing data rate, mediated effect size, and the correlation among the mediators. The specific performance of each analysis method in terms of power depended on the interactions among these factors. Specifically, the missing data rate modified both the interaction between method and the correlation among the mediators (λ) and the interaction between method and mediated effect size ($\alpha\beta$). The correlation among the mediators (λ) modifies the interaction between method and mediated effect size ($\alpha\beta$).

Latent Variable Mediation versus Auxiliary Variable Mediation Models

One issue this research aimed to address was whether or not a model incorporating all three measured mediators would perform better as a latent variable model including all three measures of the mediator in a single latent variable specification or an auxiliary variable missing data model with missingness on the variable in the mediated path and with two complete auxiliary variables to support estimation of the mediated effect. The results demonstrated that the latent variable mediation model (Method 1) outperformed the auxiliary variable model with two auxiliary variables (Method 2) in terms of power. Coverage and Type 1 error rates were comparable across the two methods.

Although the latent variable mediation model performed better in terms of power to detect the mediated effect, aspects of the simulation design may explain the results. First, the data generation model was a latent variable mediation model. Consequently, it is not surprising that the analysis model that most closely matches the data generation model (Method 1) performed better. Furthermore, the latent variable mediation model (Method 1) had fewer parameters than the model with two auxiliary variables (Method 2) and was more parsimonious. In other words, the latent variable mediation model accurately represents the relations in the data using fewer degrees of freedom.

To further understand the difference in performance between Methods 1 and 2, it is useful to investigate how the population generation model used in the current study may have affected these results. Specifically, the model-implied covariance matrix of the data generation model always resulted in the partial correlation of Y and M_2 controlling for M_1 equal to zero ($r_{YM_2.M_1} = 0$). Likewise, the partial correlation of Y and M_3 controlling for M_1 is also equal to zero ($r_{YM_3.M_1} = 0$). Given M_1 in the model prediction Y, no further increment in in prediction from M_2 or M_3 was possible.

To demonstrate this, consider the population model implied covariance matrix of the manifest variables based on the population generation model as shown in Appendix C. The constraints chosen for this simulation study simplify the model implied covariance matrix for in Appendix C to:

$$\begin{array}{cc|ccccc} & & \underline{X} & \underline{M_1} & \underline{M_2} & \underline{M_3} & \underline{Y} \\ X & & 1 & & & & \\ M_1 & & \alpha & 1 & & & \\ M_2 & & \alpha\lambda & \lambda & \lambda^2+\sigma_\varepsilon^2 & & \\ M_3 & & \alpha\lambda & \lambda & \lambda^2 & \lambda^2+\sigma_\varepsilon^2 & \\ Y & & \beta\alpha+\tau' & \tau'\alpha+\beta & \lambda(\tau'\alpha+\beta) & \lambda(\tau'\alpha+\beta) & 1 \end{array} \quad (5.1)$$

The partial correlation of Y with M_2 controlling for M_1 can be expressed as follows.

$$r_{YM_2.M_1} = \frac{r_{YM_2} - r_{YM_1} r_{M_2M_1}}{\sqrt{(1 - r_{YM_1}^2)(1 - r_{M_2M_1}^2)}} \tag{5.2}$$

Because the variances of the latent variables *LM* and *LY* are both equal to one resulting in the model-implied variances of M_1 and Y being equivalent to one, the three correlations required for the formula in Equation 5.2 are as follows.

$$r_{YM_2} = \frac{Cov(Y, M_2)}{\sigma_Y \sigma_{M_2}} = \frac{\boldsymbol{\lambda}(\tau' \alpha + \beta)}{\sqrt{\lambda^2 + \sigma_\varepsilon^2}} \tag{5.3}$$

$$r_{YM_1} = \frac{Cov(Y, M_1)}{\sigma_Y \sigma_{M_1}} = \tau' \alpha + \beta \tag{5.4}$$

$$r_{M_2M_1} = \frac{Cov(M_2, M_1)}{\sigma_{M_2} \sigma_{M_1}} = \frac{\lambda}{\sqrt{\lambda^2 + \sigma_\varepsilon^2}} \tag{5.5}$$

Substituting the expressions from Equations 5.3 – 5.5 into the numerator of Equation 5.2 demonstrates that, given the population generation model, the partial correlation between Y and M_2 controlling for M_1 will always be zero (regardless of the values of α, β, or τ').

$$r_{YM_2.M_1} = \frac{\lambda(\tau' \alpha + \beta)}{\sqrt{\lambda^2 + \sigma_\varepsilon^2}} - \frac{(\tau' \alpha + \beta)\lambda}{\sqrt{\lambda^2 + \sigma_\varepsilon^2}} \tag{5.6}$$

Similarly, the same algebra demonstrates that the partial correlation between Y and M_3 controlling for M_1 will also always be zero given the population generation model. As a result of the data generation model, after controlling for M_1, both M_2 and M_3 have no correlation with Y. As a note, the above algebraic manipulations took into account the constraints placed on the variances of *LM* and *LY* (and correspondingly the variances of M_1 and Y) so that the variances are equivalent to unity. These constraints simplified the algebraic proof, but it can also be demonstrated that the partial correlations of the two inexpensive mediators with Y controlling for M_1 are zero regardless of whether the variance constraints on LM and LY are included.

Careful examination of the covariance matrix and corresponding partial correlation calculations suggest that the assumption of M_1 being perfectly reliable with a loading of one and a residual variance of zero was the critical component driving the population partial correlations of the Y with the inexpensive mediators to be zero. In other words, generating data assuming perfect reliability of M_1 resulted in the following expression holding true at the population level: $r_{YM_2.M_1} = r_{YM_3.M_1} = 0$. However, if M_1 had not been generated as perfectly reliable and, consequently, the residual variance of M_1 was non-zero (i.e., $\sigma^2_{\varepsilon_{M_1}} > 0$) then the population partial correlations of the two inexpensive mediators with Y controlling for M_1 would not necessarily be zero.

The fact that the data were generated with partial correlations equal to zero ($r_{YM_2.M_1} = r_{YM_3.M_1} = 0$) has important ramifications on interpreting the results. The model that incorporates two auxiliary variables (Method 2) includes the partial correlations of Y with both M_2 and M_3 in the model specification. Thus, given the data generation, the auxiliary variable model (Method 2) is at a power disadvantage relative to the latent variable model (Method 1) because the auxiliary variable model requires these two additional parameters in the model (i.e., $r_{YM_2.M_1}$ and $r_{YM_3.M_1}$); both of these additional parameters have true population values equivalent to zero. This issue is specific to the current population data generation model and the somewhat tenuous assumption that M_1 has perfect reliability. Had I not constrained $\sigma^2_{\varepsilon_{M_1}}$ to zero, the population partial correlation of Y with the inexpensive mediators (M_2 and M_3) after controlling for the expensive mediator (M_1) would not necessarily be zero, and these non-zero partial correlations may improve power to detect the mediated effect.

The difference in the number of parameters between the latent variable mediation model (Method 1) and the auxiliary variable mediation model (Method 2) raises an important point. Even when the partial correlations between Y and the inexpensive mediators are non-zero, the additional number of parameters in the auxiliary variable mediation model as compared to the latent variable mediation model may pose a power disadvantage. In fact, I tested several conditions and demonstrated that using the auxiliary variable model (Method 2) and constraining the partial correlations between Y and the inexpensive mediators to zero (resulting in a reduction of the number of estimated parameters), the power to detect the mediated effect increases. In other words, when the true partial correlations of the auxiliary variables with Y are zero, removing these extra parameters increases power. Researchers will be at an advantage using the more parsimonious latent variable mediation model (Method 1) when possible. In other words, if the measurement model fits, when using the latent variable mediation model (Method 1) there is a slight boost in parsimony (i.e., fewer parameters), and thus, a resultant boost in power. If the measurement model is not tenable, the auxiliary variable model may be the better model because it ends up being no less parsimonious than a measurement model that is modified to fit (e.g., a model with post hoc correlated errors) and the auxiliary variable model avoids the need to conceptualize the mediators as indicators of a unitary construct.

The finding that the auxiliary variable model does not perform as well as the latent variable model is contrary to what might be expected. Previous research suggests that an inclusive strategy of including auxiliary variables is preferred due to improvements in efficiency (Collins et al., 2001; Rubin, 1996). In fact, Collins and colleagues (2001) noted that the inclusion of auxiliary variables is neutral in the worst case scenario and extremely beneficial in the best case scenario. Additional research has explicitly shown that, in terms of detecting mediation, auxiliary variables can increase the power to detect the mediated effect (Zhang & Wang, 2013). In the current study, the inclusive strategy of two auxiliary variables was beneficial compared to an analysis with only one auxiliary variable or maximum likelihood estimation with no additional mediators, but the inclusion of auxiliary variables was not beneficial as compared to the latent variable mediation model.

Illustrative Example of Analyzing Planned Missing Data in Mediation in Terms of Cost

As demonstrated by the results, with the exception of floor and ceiling effects (i.e., power approaching zero or 100%), in a mediation analysis, the use of planned missingness in the mediator will always decrease the power to detect the mediated effect. However, this decrease in power is not always as extreme as what may be expected. Furthermore, if a design is considered from the perspective of cost, it can be demonstrated that ***given a fixed budget*** there are scenarios where planned missing data designs can have ***more*** power than traditional complete data designs.

To better understand the potential for planned missing data to increase power to detect the mediated effect given a fixed budget, we can explore some artificial cost scenarios and compare power across different ways of allocating the same resources. These scenarios demonstrate that given a fixed budget, there are some scenarios in which a planned missing data design may perform better than a complete data design in terms of power to detect the mediated effect. To illustrate, consider a research design that costs \$10 per participant to enroll in the study (regardless of which variables are collected), an additional \$10 per participant per measure to collect data for measures X, Y, M_2 and M_3, and \$500 per participant to collect data on the expensive mediator, M_1. For the condition where the true mediated effect is composed of both small and medium paths (ie., $\alpha\beta = .03$ where α is small and β is medium **or** α is medium and β is small), using results for $\lambda = .8$, $N = 1000$, and 50% missing rate on M_1, the empirical power for the latent variable model (Method 1) is 82.85% and the empirical power for the model with two auxiliary variables (Method 2) is 79.40%. Both of these results would generally be considered adequate power using the convention of power = 80%. Alternatively, the researcher could use a complete data design with $N = 1000$ using only the expensive mediator, M_1, resulting in empirical power of 86.83%. The complete data design does provide a slight increase in power; however, this increase in power comes at considerable expense given a set of measures with costs as described above. For this cost scenario, the missing data designs described above (i.e., 50% missingness) would cost \$300,00, whereas the complete data design would cost and addition \$230,000 for a total cost of \$530,000. The calculations are below.

Missing Data Design (50% on M_1) with $N = 1000$ and empirical power = 82.85%

1. \$10 per participant for enrollment × 1000 participants = \$10,000
2. \$10 per participant to collect each of X, Y, M_2 and M_3 = \$40 per participant × 100 participants = \$40,000
3. \$500 for 50% of participants to collect M_1 = 500 participants × \$500 = \$250,000

Total: $300,000

<u>Complete Data Design with N = 1000 and empirical power = 86.83%</u>

1. $10 per participant for enrollment × 1000 participants = $10,000
2. $10 per participant to collect each of X and Y = $20 per participant × 1000 participants = $20,000 ($M_2$ and M_3 are not needed for complete data design)
3. $500 for 100% of participants to collect M_1 = 1000 participants × $500 = $500,000

Total: $530,000 ($550,000 if M_2 and M_3 are collected for the entire sample)

Given that the increase in power for the complete data condition is approximately four to seven percent higher as compared to the missing data models (Methods 1 and 2), the increase in power resulting from using a complete data design may not justify the increase in expense.

To further illustrate this point, Table 20 summarizes all the simulated conditions where the empirical power to detect the mediated effect is in the 75-87% range for $\alpha\beta$ = 0.03 (i.e., one small and one medium path comprising the mediated effect) and N = 1000. Cost is calculated using the method above based on the same hypothetical research design that costs $10 per participant included, $10 per participant to collect data for variables X, Y, M_2 and M_3, and $500 to collect data per participant on the expensive mediator, M_1. Results in Table 20 are ranked from the most expensive to the least expensive. The table shows that the latent variable model (Method 1) with 80% missing data produces an empirical power value of 79.45% at a cost of $150,000 compared to the complete data designs which produce slightly higher power (86.83%), but come at a substantially increased cost ($530,000 to $550,000 total cost).

The example above considered a scenario with 1,000 participants in the sample. With such a large sample size, concern about power loss from planned missingness may be somewhat mitigated. As another example, consider another scenario using the same costs associated with each measure as in the previous example, but using smaller sample sizes. If we consider the condition where both α and β are medium effect sizes ($\alpha = \beta =$ 0.30 and $\alpha\beta = 0.09$) for a medium correlation among the mediators ($\lambda = 0.6$), a complete data design with a sample of only 100 participants ($N = 100$) results in 70.30% empirical power to detect the mediated effect for a cost of $53,000 (assuming only one mediator, M_1, is collected; the cost increases for more mediators, but without an increase in power). On the other hand, a design with 200 participants ($N = 200$) with 80% missing data produces a power value of 81.35% at a reduced cost of $30,000. The calculation of costs is below.

Missing Data Design (80%) with *N* = 200 and empirical power = 81.35%

1. $10 per participant for enrollment $\times$ 200 participants = $2,000
2. $10 per participant to collect each of X, Y, M_2 and M_3 = $40 per participant $\times$ 200 participants = $8,000
3. $500 for 80% of participants to collect M_1 = 40 participants $\times$ $500 = $20,000

Total: $30,000

Complete Data Design with *N* = 100 and empirical power = 70.30%

1. $10 per participant for enrollment $\times$ 100 participants = $1,000
2. $10 per participant to collect each of X and Y = $20 per participant $\times$ 100 participants = $2,000
3. $500 for 100% of participants to collect M_1 = 100 participants $\times$ $500 = $50,000

Total: $53,000 ($55,000 if M_2 and M_3 are collected for the entire sample)

In this scenario, a design with 200 participants but 80% missing data on M_1 actually provides more power and costs $23,000 less than a design with complete data and 100 participants. This example demonstrates a condition where 80% missingness on M_1 for a larger sample, $N = 200$, can provide greater empirical power to detect the mediated effect than a complete data design on a smaller sample, $N = 100$. This is an example of a situation where a planned missing data design could have greater power at a potentially lower cost than a complete data design. These two hypothetical examples are supported by previous literature suggesting that given certain conditions, some missing data designs have more statistical power than the complete case designs that cost the same (Graham et al, 2001).

Recommendations

The illustrative example of analyzing planned missing data in mediation in terms of cost demonstrates that there are viable scenarios where missing data rates as high as 80% might actually be worthwhile in a planned missing data mediation analysis. However, the exact circumstances of when planned missing data can be used, and how high the missing data rate can be, depend largely on the anticipated effect sizes and cost of the measures. In situations where there isn't an extreme cost differential between the mediators, high planned missing rates may not be worth the reduction in power when considered from a cost perspective. The results also demonstrate that the specific planned missing design and analysis method a researcher uses may also depend on costs. For example, simulation results from the current study demonstrate the difference in power between the model with two auxiliary variables (Method 2) and the model with one auxiliary variable (Method 3) is often not very large. Depending on the cost of the inexpensive mediator, it may actually be most efficient to use the model with only one auxiliary model.

One major result from this study is that the latent variable mediation model had more power to detect the mediated effect than the mediation model with two auxiliary variables. As described previously, this difference can be attributed to the data generation model. However, because the latent variable mediation model estimates fewer parameters, this model is the recommended model for researchers evaluating planned missing data when there are multiple measures of the same mediation construct. Furthermore, the measurement portion of a latent variable model can help address issues of measurement error in the mediator and mitigate the potential detrimental impact of measurement error. The next logical inquiry is whether there are any conditions where a researcher should choose an auxiliary variable mediation model over a latent variable mediation model. There are three situations when a researcher may make this choice; the first two of these situations are based on theoretical considerations of the research. First, when the additional measures collected are not actually theoretically indicators of one latent mediation construct, but instead are highly correlated with the mediator of interest (and potentially X and/or M), it may make theoretically more sense to use an auxiliary variable model. In this case, only the desired mediator serves as the mediator. The researcher would then consider collecting highly correlated additional variables, but the burden of whether or not these additional variables represent the same construct of the mediator of interest need not be a concern. The auxiliary variable model eliminates the need to conceptualize the multiple mediators as indicators of a unitary construct. If researchers attempted to force a measurement model to fit data that does not represent a single construct, additional parameters would be required to make the mediators fit a measurement model (i.e., residual correlations between the M_i's and Y) and the resulting model would be statistically equivalent to the auxiliary variable model. Second, an auxiliary variable model may make more sense when X and Y are manifest variables and the study is only able to collect one additional inexpensive mediator. If X and Y are manifest variables, a latent variable model with only two indicators requires additional model constraints (e.g., both M_1 and M_2 correlate with another variable in the model or both loadings constrained to be equal). From a theoretical standpoint, the auxiliary variable model with one auxiliary variable may be more defensible than a latent variable model with the required constraint. Constraints aren't needed if X and Y are latent variables with multiple indicators. Finally, some researchers may be more comfortable avoiding the latent variable framework and prefer to use the MAR-based procedure of multiple imputation with a regression analysis that includes auxiliary variables. Unlike the latent variable model (Method 1), the auxiliary variable models (Methods 2 and 3) can be implemented in the OLS regression framework using multiple imputation.

The vital question that remains is, for a given effect size, how much missing data is recommended? Unfortunately, there is no simple answer to that question. The optimal planned missing data design depends on the effect size, the strength of the correlations of the auxiliary variables, the sample size, and the cost associated with each of the collected measures. Depending on the parameters listed above, a planned missing design with 80% missing data might be optimal design, but there are cases where a design with a lower missing rate or a traditional complete data might be optimal. Before implementing any design with purposeful missing data, researchers should conduct Monte Carlo power analyses to determine the best design strategy for the parameters of a given research scenario. A Monte Carlo analysis can take into account the cost of each measure and the anticipated effect sizes to determine which design might best optimize resources. There are several resources available to researchers describing power calculations for missing data designs (Enders, 2010, pp. 30–32; Mistler & Enders, 2012; L. K. Muthén & Muthén, 2002).

A power analysis may also want to take into consideration that research involving human participants may also include unplanned missing data. Little and Rhemtulla (2013) offer guidelines for preemptively accommodating unplanned missing data in a planned missing data design. First, power analyses of planned missing data designs should incorporate the potential for unplanned missing data. Rates of unplanned missing data from previous research may be used to estimate the amount of unplanned missingness that may arise above and beyond the planned missingness. Second, additional measures should be collected in order to ensure that unplanned missingness satisfies the MAR-mechanism. These measures might be chosen based on previous research and theoretical reasons unplanned missingness might arise (e.g., lack of conscientiousness, socioeconomic status). In some cases, the inexpensive mediators may serve both purposes (i.e., the inexpensive mediator chosen to help increase power on the mediator with missing data may also help satisfy the MAR-mechanism). In other research scenarios, additional auxiliary variables may be required to satisfy the MAR-mechanism when there are unplanned missing data.

Limitations and Future Directions

Planned missing data analysis in mediation is a research area that has many possibilities beyond the conditions explored in the current study. The current simulation has a number of limitations. The simulation is limited by the levels of the manipulated factors chosen for this study. For non-zero mediated effects, both α and β were positive values resulting in a positive mediated effect; the current study did not consider situations when the two paths that comprise the mediated effect are both negative or in opposite directions. The study also constrained τ' (the direct effect of *Y* on *X*) to zero and only evaluated conditions with complete mediation in the population. Although this choice is defensible given previous literature on the impact of partial mediation on power to detect the mediated effect (Fritz & MacKinnon, 2007; MacKinnon et al., 2004), methodologists have argued that complete mediation is unlikely with a single mediation construct (Baron & Kenny, 1986). The presence of inconsistent mediation (which only occurs in the presence of partial mediation) may also affect the results. In inconsistent mediation, τ' is in the opposite direction of $\alpha\beta$ (MacKinnon et al., 2000). Consequently, the mediator acts like a suppressor variable, and this suppression could potentially influence the performance of the evaluated analysis methods.

As another limitation, the current study only investigated missingness on the mediator, *M*. Planned missing designs may also be appropriate with planned missingness on the independent variable, *X*, and/or the dependent variable, *Y*. Furthermore, this research assumed all variables were normally distributed. Although the assumption of normally distributed variables is often tenable, there are many variables of research interest that are non-normal. Mediational planned missing designs need to be evaluated in the context of non-normal distributions on *X*, *M* and *Y* (e.g., skewed, binary outcomes, count variables). Because the distribution of the mediated effect, $\alpha\beta$, is not normally distributed, non-normal distributions may have unexpected effects on power to detect the mediated effect with planned missing data. Planned missingness with non-normal variables requires further investigation.

A major limitation of this study is the assumption that the expensive mediator, M_1, was perfectly reliable. Although many simulation studies make the assumption of perfect reliability, in reality, very few measures are perfectly reliable. As stated earlier in this document, the unreliability of *M* may cause an increased standard error in the α coefficient and may attenuate the β coefficient. Overall, the effect of unreliability of the mediator may decrease the power to detect the mediated effect. Thus, to the extent that the expensive mediator is unreliable, evaluated analysis methods may have decreased power to detect the mediated effect. On the other hand, we also found that a perfectly reliable mediator, M_1, may actually undermine the performance of the auxiliary variable analysis as compared to a latent variable model.

This study focused on maximum likelihood estimation in the analysis models (Methods 1 – 4). Multiple imputation is another viable MAR-based strategy that may be used to analyze a planned missing data design. Because multiple imputation is asymptotically equivalent to maximum likelihood estimation, there was no reason to expect major differences between the two approaches. In this study, maximum likelihood estimation was chosen because the method required reduced computing resources as compared to multiple imputation. However, at smaller sample sizes, there may be some differences between these two MAR-based approaches. Future research should investigate the use of planned missing data designs analyzed using multiple imputation. Jia et al. (2013) found evidence that at smaller sample sizes, maximum likelihood estimation performed better than multiple imputation.

The current research project considered planned missing data designs in terms of data that satisfy the MCAR mechanism. MAR-based designs may also be of interest and it is possible to consider the same analysis methods in the context of MAR-based designs. In a MAR-based design, it may be that the value of one collected measure predicts whether or not the researcher collects an additional measure. This may be particularly true in data collected in a clinical setting. It may be by design that one measure is only collected for people who score above or below a certain threshold on another measure. Schafer and Graham (2002) provide an example where only participants who score above a certain threshold on blood pressure return at a later date to provide a repeated blood pressure method. Examples exist in cross-sectional research as well. For example, perhaps participants visiting a clinic who score above a certain threshold for fasting blood sugar may have additional blood drawn for more extensive laboratory tests (e.g., A1C test for average blood glucose level).

Conclusion

There is no one-size fits all prescription for planned missing data. When designing a research study, researchers must carefully evaluate both the potential measures and cost of these measures relative to the study budget. Simulation studies are necessary to determine how to best allocate resources. It is clear that there are scenarios where planned missing data designs may be beneficial tools for optimizing resources to study questions of interest; there are also situations where planned missingness may not have much added benefit. However, in some scenarios, planned missing data designs do have the capacity to optimize resources. Rather than questioning why anyone would want to have missing data, perhaps we should always consider such a design as a possibility when designing a study. Researchers investigating mechanisms of change, particularly in situations where expensive measures are involved, would be prudent to consider a planned missing data design as a possibility. The results of this study support the rhetorical question posed by Graham et al. (2001), “why would anyone not want to consider a planned missing-data design?”

REFERENCES

Algina, J., & Penfield, R. D. (2009). Classical test theory. In *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 93–122). London: SAGE Publications.

Aroian, L. A. (1947). The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics, 18*(2), 265–271.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*(1), 5–37. http://doi.org/10.1016/j.jsp.2009.10.001

Baraldi, A. N., & MacKinnon, D. P. (2011, May). *Measurement error in mediation*. Presented at the Society for Prevention Research, Washington, DC.

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.

Beale, E. M. L., & Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society. Series B (Methodological), 37*(1), 129–145.

Beaton, A. E., Johnson, E. G., & Ferris, J. J. (1987). The assignment of exercises to students. In A. E. Beaton (Ed.), *Implementing the new design: The NAEP 1983–84 technical report*. Princeton, NJ: Educational Testing Services.

Berman, J. S., & Kenny, D. A. (1976). Correlational bias in observer ratings. *Journal of Personality and Social Psychology, 34*(2), 263–273.

Bollen, K. A. (1989). *Structural equations with latent variables* (1 edition). New York: Wiley-Interscience.

Braam, L. A. L. M., Ocké, M. C., Bueno-de-Mesquita, H. B., & Seidell, J. C. (1998). Determinants of obesity-related underreporting of energy intake. *American Journal of Epidemiology, 147*(11), 1081–1086.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 edition). Hillsdale, N.J: Routledge.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 558–577. http://doi.org/10.1037/0021-843X.112.4.558

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. http://doi.org/10.1037/1082-989X.6.4.330

Craig, C. C. (1936). On the frequency function of xy. *The Annals of Mathematical Statistics*, *7*(1), 1–15.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, Ohio: Wadsworth Pub Co.

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill Education.

Davey, A., & Savla, J. (2009). Estimating statistical power with incomplete data. *Organizational Research Methods*, *12*(2), 320–346. http://doi.org/10.1177/1094428107300366

Dempster, A. P., & Laird, N. M. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, *39*(1), 1–38.

Dodge, Y. (2006). *The Oxford dictionary of statistical terms*. Oxford University Press.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. http://doi.org/http://doi.org/10.1207/S15328007SEM0803_5

Enders, C. K., Fairchild, A. J., & MacKinnon, D. P. (2013). A Bayesian approach for estimating nediation effects with missing data. *Multivariate Behavioral Research*, *48*(3), 340–369. http://doi.org/10.1080/00273171.2013.784862

Enders, C. K., & Peugh, J. L. (2004). Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(1), 1–19. http://doi.org/10.1207/S15328007SEM1101_1

Fritz, M. S., Kenny, D. A., & MacKinnon, D. P. (2015). The opposing effects of simultaneously ignoring measurement error and omitting confounders in a single-mediator model. *Manuscript Submitted for Publication.*

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 233–239. http://doi.org/10.1111/j.1467-9280.2007.01882.x

Garnier-Villarreal, M., Rhemtulla, Mijke, & Little, Todd D. (2014). Two-method planned missing designs. *Manuscript Submitted for Publication.*

Gonzalez, J. M., & Eltinge, J. L. (2007). Multiple matrix sampling: A review. In *Proceedings of the Section on Survey Research Methods, American Statistical Association* (pp. 3069–3075).

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(1), 80–100. http://doi.org/10.1207/S15328007SEM1001_4

Graham, J. W. (2012). Planned missing data designs 1: The 3-form design. In J. W. Graham, *Missing Data* (pp. 279–294). New York, NY: Springer New York.

Graham, J. W., & Collins, N. L. (1991). Controlling correlational bias via confirmatory factor analysis of MTMM data. *Multivariate Behavioral Research*, *26*(4), 607–629. http://doi.org/10.1207/s15327906mbr2604_3

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research.* (pp. 325–366). Washington, DC, US: American Psychological Association. Retrieved from http://psycnet.apa.org/books/10222/010

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*(2), 197–218. http://doi.org/10.1207/s15327906mbr3102_3

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. *NIDA Research Monograph*, *142*, 13–63.

Graham, J. W., & Shevock, A. E. (2012). Planned missing data design 2: Two-method measurement. In *Missing Data* (pp. 295–323). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4614-4018-5_13

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. In *New Methods for the Analysis of Change* (pp. 335–353). Washington, DC, US: American Psychological Association.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. http://doi.org/10.1037/1082-989X.11.4.323

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. http://doi.org/10.1080/03637750903310360

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65*(4), 599–610.

Hoyle, R. H. (1991). Evaluating measurement models in clinical research: Covariance structure analysis of latent variable models of self-conception. *Journal of Consulting and Clinical Psychology*, *59*(1), 67–76.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: Sage.

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, *5*(1), 64–86. http://doi.org/10.1037/1082-989X.5.1.64

Iacobucci, D., Saldanha, N., & Deng, X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology*, *17*(2), 139–153. http://doi.org/10.1016/S1057-7408(07)70020-7

Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data designs with small sample sizes: How small is too small? *International Journal of Behavioral Development*, 0165025414531095. http://doi.org/10.1177/0165025414531095

Johnson, E. G. (1992). The design of the national assessment of educational progress. *Journal of Educational Measurement*, *29*(2), 95–110. http://doi.org/10.1111/j.1745-3984.1992.tb00369.x

Kaplan, D. (1995). The impact of BIB spiraling-induced missing data patterns on goodness-of-fit tests in factor analysis. *Journal of Educational and Behavioral Statistics*, *20*(1), 69–82. http://doi.org/10.3102/10769986020001069

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, *12*(3), 247–252. http://doi.org/10.1016/0022-1031(76)90055-X

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology, Vols. 1 and 2 (4th ed.)* (pp. 233–265). New York, NY, US: McGraw-Hill.

Kraemer, H. C., Kiernan, M., Essex, M., & Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology*, *27*(2, Suppl), S101–S108. http://doi.org/10.1037/0278-6133.27.2(Suppl.).S101

Lafay, L., Basdevant, A., Charles, M., Vray, M., Balkau, B., Borys, J., … Romon, M. (1997). Determinants and nature of dietary underreporting in a free-living population: The Fleurbaix Laventie Ville Sante (FLVS) Study. *International Journal of Obesity & Related Metabolic Disorders*, *21*(7), 567–73.

Le Cessie, S., Debeij, J., Rosendaal, F. R., Cannegieter, S. C., & Vandenbroucke, J. P. (2012). Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology*, *23*(4), 551–560. http://doi.org/10.1097/EDE.0b013e318254f5de

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, N.J.: Wiley.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199–204. http://doi.org/10.1111/cdep.12043

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley Publishing Company, Inc.

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In *Structural equation modeling: Concepts, issues, and applications* (pp. 16–36). Thousand Oaks, CA, US: Sage Publications, Inc.

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Lawrence Erlbaum Associates.

Mackinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*(2), 144–158. http://doi.org/10.1177/0193841X9301700202

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*(1), 593–614. http://doi.org/10.1146/annurev.psych.58.110405.085542

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*(3), 384–389. http://doi.org/10.3758/BF03193007

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181. http://doi.org/10.1023/A:1026595011371

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., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–99. http://doi.org/10.1207/s15327906mbr3901_4

Mallinckrodt, B., Todd, W., Wei, M., & Russell, D. W. (2006). Advances in testing the statistical significance of mediation effects. *Journal of Counseling Psychology*, *53*(3), 372–378. http://doi.org/10.1037/0022-0167.53.3.372

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod fata: Many problems and a few solutions. *Applied Psychological Measurement*, *13*(4), 335–361. http://doi.org/10.1177/014662168901300402

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142. http://doi.org/10.1037/0012-1649.38.1.115

McArdle, J. J., & Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research*, *18*(3-4), 145–166. http://doi.org/10.1080/03610739208253917

McArdle, J. J., & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods*, *2*(4), 403–435. http://doi.org/10.1037/1082-989X.2.4.403

Mistler, S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In Laursen, B., Little, T., & Card, N. (Eds.), *Handbook of Developmental Research Methods* (pp. 742–754). New York: Guilford.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*(3), 431–462. http://doi.org/10.1007/BF02294365

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. http://doi.org/10.1207/S15328007SEM0904_8

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. http://doi.org/10.3758/BRM.40.3.879

Raghunathan, T. E. (2004). What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health*, *25*(1), 99–117. http://doi.org/10.1146/annurev.publhealth.25.102802.124410

Rhemtulla, M., Jia, F., Wu, W., & Little, T. D. (2014). Planned missing designs to optimize the efficiency of latent growth parameter estimates. *International Journal of Behavioral Development*, 0165025413514324. http://doi.org/10.1177/0165025413514324

Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development*, *13*(4), 425–438. http://doi.org/10.1080/15248372.2012.717340

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. http://doi.org/10.1037/h0037350

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. http://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1978a). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, *6*(1), 34–58.

Rubin, D. B. (1978b). Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In *Proceedings of the Survey Research Methods Section of the American Statistical Association* (pp. 30–34).

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489. http://doi.org/10.1080/01621459.1996.10476908

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331.

Rubin, D. B., & Little, R. J. (2002). Statistical analysis with missing data. *Hoboken, NJ: J Wiley & Sons*.

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. http://doi.org/10.1111/j.1751-9004.2011.00355.x

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. http://doi.org/10.1037/1082-989X.7.2.147

Shoemaker, D. M. (1973). *Principles and procedures of multiple matrix sampling* (Vol. xviii). Oxford, England: Ballinger.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445. http://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*(1982), 290–312.

Stephenson, M. T., & Holbert, R. L. (2003). A Monte Carlo simulation of observable versus latent variable structural equation modeling techniques. *Communication Research*, *30*(3), 332–354. http://doi.org/10.1177/0093650203030003004

Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2007). Tests of the three-path mediated effect. *Organizational Research Methods*. http://doi.org/10.1177/1094428107300344

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*(3), 692–700. http://doi.org/10.3758/s13428-011-0076-x

Valeri, L., Lin, X., & VanderWeele, T. J. (2014). Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model. *Statistics in Medicine*, *33*(28), 4875–4890. http://doi.org/10.1002/sim.6295

VanderWeele, T. J., Asomaning, K., Tchetgen, E. J. T., Han, Y., Spitz, M. R., Shete, S., … Lin, X. (2012). Genetic variants on 15q25.1, smoking, and lung cancer: An assessment of mediation and interaction. *American Journal of Epidemiology*, *175*(10), 1013–1020. http://doi.org/10.1093/aje/kwr467

VanderWeele, T. J., Valeri, L., & Ogburn, E. L. (2012). The role of measurement error and misclassification in mediation analysis. *Epidemiology (Cambridge, Mass.)*, *23*(4), 561–564. http://doi.org/10.1097/EDE.0b013e318258f5e4

Wang, L., Zhang, Z., & Tong, X. (2014). Mediation analysis with missing data through multiple imputation and bootstrap. *arXiv Preprint arXiv:1401.2081*.

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, *15*(1), 18–37. http://doi.org/10.1037/a0015917

Wu, W., & Jia, F. (2013). A new procedure to test mediation with missing data through nonparametric bootstrapping and multiple imputation. *Multivariate Behavioral Research*, *48*(5), 663–691. http://doi.org/10.1080/00273171.2013.816235

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301–322. http://doi.org/10.1037/a0016972

Zhang, Z., & Wang, L. (2013). Methods for mediation analysis with missing data. *Psychometrika*, *78*(1), 154–184. http://doi.org/10.1007/s11336-012-9301-5

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

APPENDIX A

TABLES AND FIGURES

Table 1. 3-form Design

Form	Blocks			
	X	*A*	*B*	*C*
1	✓	✓	✓	⃠
2	✓	✓	⃠	✓
3	✓	⃠	✓	✓

Note. ✓ denotes included item sets. ⃠ denotes excluded item sets. Adapted from Graham et al. (2006) and Graham (2012).

Table 2. 3-Form Design as Described in Illustrative Example

Form	Items Included in Each Block				
	X	*A*	*B*	*C*	Total Items
1	1-25	26-50	51-75	-	**75**
2	1-25	26-50	-	76-100	**75**
3	1-25	-	51-75	76-100	**75**

Note. In this illustrative example, Block *X* contains items numbered 1-25, Block *A* contains 26-50, Block *B* contains 51-75, and Block *C* contains 76-100.

Table 3. Measurement Model Conditions

Factor Loading Condition	Standardized Factor Loadings			Residual Variances			Correlations between M_1, M_2, and M_3		
	λ_1	λ_2	λ_3	$\sigma^2_{\varepsilon_1}$	$\sigma^2_{\varepsilon_2}$	$\sigma^2_{\varepsilon_3}$	$r_{M_1M_2}$	$r_{M_1M_3}$	$r_{M_2M_3}$
Low	1	.4	.4	.00	.84	.84	.4	.4	.16
Medium	1	.6	.6	.00	.64	.64	.6	.6	.36
High	1	.8	.8	.00	.36	.36	.8	.8	.64

Table 4. Raw Bias by Missing Data Rate Averaged across All Other Factors

Missing Data Rate	Mean	Median	SD
Complete Data	-0.0001	-0.0002	0.0294
20% Missing on M_1	0.0001	-0.0002	0.0321
50% Missing on M_1	0.0004	-0.0003	0.0395
80% Missing on M_1	0.0009	-0.0006	0.0659
All Missing Rates	*0.0003*	*-0.0003*	*0.0442*

Table 5. Standardized Bias by Missing Data Rate and Method across All Other Factors

Method	*Mean*	*SD*	*Min*	*Max*
Complete Data				
1. Latent Variable Model	0.001	0.023	-0.065	0.050
2. Two Auxiliary Variable Model	0.000	0.023	-0.065	0.050
3. One Auxiliary Variable Model	0.000	0.023	-0.065	0.050
4. ML with No Additional Mediators	0.000	0.023	-0.065	0.050
5. Listwise Deletion	0.000	0.023	-0.065	0.050
20% Missing on M_1				
1. Latent Variable Model	0.002	0.024	-0.058	0.078
2. Two Auxiliary Variable Model	0.002	0.024	-0.060	0.077
3. One Auxiliary Variable Model	0.002	0.025	-0.066	0.079
4. ML with No Additional Mediators	0.003	0.027	-0.076	0.081
5. Listwise Deletion	0.000	0.026	-0.080	0.079
50% Missing on M_1				
1. Latent Variable Model	0.011	0.028	-0.071	0.079
2. Two Auxiliary Variable Model	0.012	0.030	-0.069	0.103
3. One Auxiliary Variable Model	0.014	0.031	-0.066	0.104
4. ML with No Additional Mediators	0.015	0.035	-0.082	0.104
5. Listwise Deletion	0.004	0.032	-0.078	0.081
80% Missing on M_1				
1. Latent Variable Model	0.025	0.053	-0.098	0.272
2. Two Auxiliary Variable Model	0.030	0.058	-0.135	0.225
3. One Auxiliary Variable Model	0.035	0.068	-0.170	0.268
4. ML with No Additional Mediators	0.041	0.087	-0.337	0.268
5. Listwise Deletion	0.001	0.068	-0.277	0.250

Table 6. Empirical Power (%) for each of the Analysis Methods Averaged across All Other Factors

Method	Power (%)
1. Latent Variable Model	61.47
2. Two Auxiliary Variable Model	58.72
3. One Auxiliary Variable Model	57.33
4. ML with No Additional Mediators	54.68
5. Listwise Deletion	53.52
All Methods	*57.14*

Table 7. Logistic Regression Results for Power Ranked by Wald χ^2 Combining Factors α and β into Mediated Effect Factor, $\alpha\beta$

Effect	Wald	df	NCP	*p*-value
Three-Way Interactions				
$\alpha\beta \times N \times$ Missing Rate	2168.729	30	2138.729	<.0001
Method $\times \lambda \times$ Missing Rate	1342.585	16	1326.585	<.0001
Method $\times \alpha\beta \times$ Missing Rate	1107.393	40	1067.393	<.0001
Method $\times \alpha\beta \times \lambda$	986.2488	40	946.2488	<.0001
Method $\times N \times$ Missing Rate	789.2546	24	765.2546	<.0001
Method $\times \alpha\beta \times N$	694.3005	60	634.3005	<.0001
Method $\times \lambda \times N$	389.1778	24	365.1778	<.0001
$\alpha\beta \times \lambda \times N$	200.1886	30	170.1886	<.0001
$\alpha\beta \times \lambda \times$ Missing Rate	178.0867	20	158.0867	<.0001
$\lambda \times N \times$ Missing Rate	139.9992	12	127.9992	<.0001
Conditional Two-Way Interactions				
$\lambda \times N$	110.8432	6	104.8432	<.0001
Method $\times N$	292.5393	12	280.5393	<.0001
$\alpha\beta \times \lambda$	418.2972	10	408.2972	<.0001
$\lambda \times$ Missing Rate	475.6014	4	471.6014	<.0001
N $\times$ Missing Rate	488.7682	6	482.7682	<.0001
$\alpha\beta \times$ Missing Rate	769.6084	10	759.6084	<.0001
Method $\times \alpha\beta$	846.6286	20	826.6286	<.0001
Method $\times$ Missing Rate	3171.798	8	3163.798	<.0001
Method $\times \lambda$	4203.292	8	4195.292	<.0001
$\alpha\beta \times N$	18806.91	15	18791.91	<.0001
Conditional Main Effects				
$\alpha\beta$	26524.68	5	26519.68	<.0001
N	3.5548	3	0.5548	0.3137
Missing Rate	0.8498	2	-1.1502	0.6538
Method	0.1974	4	-3.8026	0.9954
λ	0.0203	2	-1.9797	0.9899

Note. NCP is the non-centrality parameter calculated as Wald χ^2 – degrees of freedom.

Table 8. Empirical Power (%) for the Method×λ×Missing Rate Interaction Averaged across All Levels of Effect Size ($\alpha\beta$) and Sample Size (N)

Method	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$
Complete Data			
All Methods[a]	66.44	66.63	66.30
20% Missing on M_1			
1. Latent Variable Model	63.42	64.43	65.34
2. Two Auxiliary Variable Model	63.19	64.18	65.12
3. One Auxiliary Variable Model	62.76	63.51	64.39
4. ML with No Additional Mediators	62.35	62.27	61.99
5. Listwise Deletion	61.89	61.82	61.65
50% Missing on M_1			
1. Latent Variable Model	56.90	60.62	63.78
2. Two Auxiliary Variable Model	55.20	58.38	61.99
3. One Auxiliary Variable Model	54.06	56.37	59.70
4. ML with No Additional Mediators	52.77	53.12	53.04
5. Listwise Deletion	51.56	51.94	51.78
80% Missing on M_1			
1. Latent Variable Model	47.12	55.01	61.64
2. Two Auxiliary Variable Model	40.41	44.40	52.38
3. One Auxiliary Variable Model	38.93	41.36	47.46
4. ML with No Additional Mediators	37.24	37.04	37.05
5. Listwise Deletion	34.17	34.00	34.06

[a]For complete data conditions, percent power estimates are identical through the first decimal place for all five methods. The table presents percent power from the latent variable model (Method 1) to represent power for all five methods in the complete data condition.

Table 9. Empirical Power (%) for the Method × $\alpha\beta$ × Missing Rate Interaction

	Population $\alpha\beta$					
Method	0.01	0.03	0.05	0.09	0.15	0.25
Complete Data						
All Methods[a]	30.15	47.23	45.64	92.26	95.02	99.91
20% Missing on M_1						
1. Latent Variable Model	26.71	44.22	43.28	90.42	93.81	99.84
2. Two Auxiliary Variable Model	26.33	43.83	43.12	90.19	93.63	99.80
3. One Auxiliary Variable Model	25.53	42.94	42.32	89.63	93.27	99.80
4. ML with No Additional Mediators	23.07	41.05	40.94	88.16	92.46	99.71
5. Listwise Deletion	22.93	40.63	40.14	87.80	92.06	99.69
50% Missing on M_1						
1. Latent Variable Model	21.90	38.98	38.43	85.95	90.99	99.29
2. Two Auxiliary Variable Model	19.39	36.58	36.34	83.54	89.45	99.03
3. One Auxiliary Variable Model	16.75	34.33	34.22	81.41	88.23	98.69
4. ML with No Additional Mediators	12.36	29.39	30.23	76.68	85.41	97.70
5. Listwise Deletion	11.95	28.46	28.49	75.36	83.72	97.20
80% Missing on M_1						
1. Latent Variable Model	16.27	32.56	32.13	78.70	85.03	96.93
2. Two Auxiliary Variable Model	8.57	22.74	24.72	66.51	75.34	90.88
3. One Auxiliary Variable Model	6.22	19.54	21.63	62.42	71.73	88.82
4. ML with No Additional Mediators	3.41	14.04	17.41	53.46	64.91	84.38
5. Listwise Deletion	2.71	12.24	14.64	49.97	60.10	80.03

[a]For complete data conditions, percent power estimates are identical through the first decimal place for all five methods. The table presents percent power from the latent variable model (Method 1) to represent power for all five methods in the complete data condition.

Table 10. Empirical Power (%) for the Method $\times \alpha\beta \times \lambda$ Interaction

	Population $\alpha\beta$					
Method	0.01	0.03	0.05	0.09	0.15	0.25
λ = 0.40						
1. Latent Variable Model	16.89	34.00	33.66	80.00	86.43	97.25
2. Two Auxiliary Variable Model	14.64	30.90	31.61	75.79	82.97	94.98
3. One Auxiliary Variable Model	13.63	29.69	30.64	74.36	82.10	94.40
4. ML with No Additional Mediators	12.62	28.43	29.56	72.66	80.95	93.93
5. Listwise Deletion	12.16	27.37	27.71	70.91	78.63	92.35
λ = 0.60						
1. Latent Variable Model	21.68	38.58	38.10	85.44	90.32	99.06
2. Two Auxiliary Variable Model	17.56	33.71	34.17	79.54	85.85	96.33
3. One Auxiliary Variable Model	15.58	31.53	32.20	77.09	84.00	95.59
4. ML with No Additional Mediators	12.93	28.13	29.64	72.81	81.02	93.95
5. Listwise Deletion	12.55	27.06	27.94	71.05	78.65	92.40
λ = 0.80						
1. Latent Variable Model	26.30	43.18	42.08	89.62	93.07	99.74
2. Two Auxiliary Variable Model	22.09	38.53	38.40	84.91	89.59	98.41
3. One Auxiliary Variable Model	19.28	35.59	35.32	82.00	87.13	97.31
4. ML with No Additional Mediators	13.29	27.91	29.37	72.83	80.82	93.91
5. Listwise Deletion	12.88	26.91	27.63	71.17	78.60	92.17

Note. Methods averaged only over conditions where the missing data rate is non-zero.

Table 11. Logistic Regression Results for Type 1 Error Rates ranked by Wald χ^2

Effect	Wald	df	*p*-value
Three-Way Interactions			
Effect Size × *N* × Missing Data Rate	500.81	18	<.0001
Method × *N* × Missing Data Rate	117.03	24	<.0001
Effect Size × Method × Missing Data Rate	99.93	24	<.0001
Effect Size × Method × *N*	77.48	36	<.0001
Effect Size × Method × Lambda	38.12	24	0.03
N × Missing Data Rate × Lambda	29.39	12	<.01
Effect Size × Missing Data Rate × Lambda	20.20	12	0.06
Effect Size × *N* × Lambda	18.31	18	0.44
Method × *N* × Lambda	14.10	24	0.94
Conditional Two-Way Interactions			
Effect Size × *N*	2973.87	9	<.0001
Effect Size × Missing Data Rate	319.44	6	<.0001
N × Missing Data Rate	136.83	6	<.0001
Effect Size × Method	119.35	12	<.0001
Method × Missing Data Rate	57.29	8	<.0001
Method × *N*	43.74	12	<.0001
Effect Size × Lambda	22.32	6	<.01
Missing Data Rate × Lambda	10.30	4	0.04
Method × Lambda	8.66	8	0.37
N × Lambda	2.10	6	0.91
Conditional Main Effects			
Effect Size	21036.66	3	<.0001
Method	51.62	4	<.0001
N	51.08	3	<.0001
Missing Data Rate	20.83	2	<.0001
Lambda	11.17	2	<.01

Table 12. Type 1 Error Rates (%) for the Method × Effect Size × Missing Rate Interaction Averaged across All Levels of λ and Sample Size (N)

Method	$\alpha = \beta = 0$	α or $\beta = .1$	α or $\beta = .3$	α or $\beta = .5$
Complete Data				
All Methods[a]	0.16	1.99	**4.88**	**5.32**
20% Missing on M_1				
1. Latent Variable Model	0.13	1.87	**4.68**	**5.34**
2. Two Auxiliary Variable Model	0.12	1.85	**4.72**	**5.32**
3. One Auxiliary Variable Model	0.14	1.87	**4.72**	**5.36**
4. ML with No Additional Mediators	0.16	1.76	**4.72**	**5.45**
5. Listwise Deletion	0.15	1.73	**4.67**	**5.40**
50% Missing on M_1				
1. Latent Variable Model	0.19	1.66	**4.68**	**5.36**
2. Two Auxiliary Variable Model	0.19	1.58	**4.77**	**5.59**
3. One Auxiliary Variable Model	0.22	1.50	**4.69**	**5.62**
4. ML with No Additional Mediators	0.25	1.30	**4.63**	**5.64**
5. Listwise Deletion	0.20	1.19	**4.37**	**5.42**
80% Missing on M_1				
1. Latent Variable Model	0.18	1.44	**4.41**	**5.29**
2. Two Auxiliary Variable Model	0.48	1.33	**4.75**	6.40
3. One Auxiliary Variable Model	0.50	1.17	**4.79**	6.32
4. ML with No Additional Mediators	0.59	1.07	**4.52**	6.56
5. Listwise Deletion	0.29	0.65	3.48	**5.41**

Note. Values ranging from 4.04 to 5.96% are bold; a procedure with honest error rates would be expected to yield Type I errors between these values.

[a]Type 1 error rates were nearly identical across all methods for the complete data condition. The Type 1 error results from the latent variable model (Method 1) are presented to represent complete data results.

Table 13. Type 1 Error Rates (%) for the Method × N × Missing Data Rate Interaction

Method	N = 100	N = 200	N = 500	N = 1000
Complete Data				
All Methods[a]	3.03	3.32	3.61	**4.07**
20% Missing on M_1				
1. Latent Variable Model	3.02	3.24	3.49	3.91
2. Two Auxiliary Variable Model	3.00	3.23	3.52	3.91
3. One Auxiliary Variable Model	2.99	3.28	3.54	3.93
4. ML with No Additional Mediators	2.93	3.35	3.51	3.93
5. Listwise Deletion	2.87	3.33	3.46	3.90
50% Missing on M_1				
1. Latent Variable Model	2.80	3.14	3.57	3.98
2. Two Auxiliary Variable Model	3.06	3.21	3.56	3.91
3. One Auxiliary Variable Model	3.11	3.14	3.60	3.77
4. ML with No Additional Mediators	3.07	3.15	3.48	3.65
5. Listwise Deletion	2.66	2.92	3.44	3.65
80% Missing on M_1				
1. Latent Variable Model	2.73	2.90	3.46	3.74
2. Two Auxiliary Variable Model	**4.07**	3.37	3.39	3.71
3. One Auxiliary Variable Model	3.84	3.37	3.46	3.65
4. ML with No Additional Mediators	3.91	3.39	3.38	3.54
5. Listwise Deletion	2.17	2.49	3.03	3.37

Note. Values ranging from 4.04 to 5.96% are bold; a procedure with honest error rates would be expected to yield Type I errors between these values

[a]Type 1 error rates were nearly identical across all methods for the complete data condition. The Type 1 error results from the latent variable model (Method 1) are presented to represent complete data results.

.

Table 14. Asymmetric Confidence Interval Coverage (%) for each of the Analysis Methods Averaged across All Other Factors for Conditions where $\alpha\beta > 0$ and Missing Rate > 0

Method	Power (%)
1. Latent Variable Model	94.96
2. Two Auxiliary Variable Model	94.71
3. One Auxiliary Variable Model	94.78
4. ML with No Additional Mediators	94.89
5. Listwise Deletion	95.04
All Methods	***94.88***

Table 15. Logistic Regression Results for Confidence Interval Coverage Ranked by Wald χ^2 Combining Factors α and β into Mediated Effect Factor, $\alpha\beta$

Effect	Wald	df	p-value
Three-Way Interactions			
$\alpha\beta\times N\times$Missing Rate	741.65	30	<.0001
Method$\times\alpha\beta\times$Missing Rate	250.81	40	<.0001
Method$\times\alpha\beta\times N$	176.11	60	<.0001
$\alpha\beta\times\lambda\times N$	158.67	30	<.0001
Method$\times N\times$Missing Rate	87.88	24	<.0001
$\alpha\beta\times\lambda\times$Missing Rate	84.12	20	<.0001
$\lambda\times N\times$Missing Rate	45.59	12	<.0001
Method$\times\lambda\times$Missing Rate	43.83	16	<.001
Method$\times\alpha\beta\times\lambda$	34.02	40	.74
Method$\times\lambda\times N$	15.65	24	.90
Conditional Two-Way Interactions			
$\alpha\beta\times N$	3715.08	15	<.0001
$\alpha\beta\times$Missing Rate	612.92	10	<.0001
N$\times$Missing Rate	152.98	6	<.0001
Method$\times\alpha\beta$	133.92	20	<.0001
Method$\times$Missing Rate	121.13	8	<.0001
Method$\times N$	56.09	12	<.0001
$\lambda\times N$	33.16	6	<.0001
$\alpha\beta\times\lambda$	27.10	10	<.01
$\lambda\times$Missing Rate	18.09	4	<.01
Method$\times\lambda$	15.30	8	.05
Conditional Main Effects			
$\alpha\beta$	5084.28	5	<.0001
N	615.16	3	<.0001
Method	124.65	4	<.0001
Missing Rate	13.72	2	<.01
λ	9.41	2	.01

Table 16. Asymmetric Confidence Interval Coverage (%) for the Method × $\alpha\beta$ × Missing Rate Interaction Averaged Across λ and Sample Size (N)

	Population $\alpha\beta$					
Method	0.01	0.03	0.05	0.09	0.15	0.25
Complete Data						
All Methods[a]	96.57	**94.84**	**94.56**	**94.75**	**94.73**	**94.98**
20% Missing on M_1						
1. Latent Variable Model	96.56	**94.73**	**94.83**	**95.17**	**94.87**	**94.69**
2. Two Auxiliary Variable Model	96.60	**94.69**	**94.82**	**95.22**	**94.88**	**94.65**
3. One Auxiliary Variable Model	96.62	**94.72**	**94.85**	**95.12**	**94.80**	**94.72**
4. ML with No Additional Mediators	96.70	**94.76**	**94.86**	**95.22**	**94.86**	**94.80**
5. Listwise Deletion	96.67	**94.80**	**94.90**	**95.18**	**94.91**	**94.84**
50% Missing on M_1						
1. Latent Variable Model	96.77	**94.97**	**94.62**	**94.60**	**94.58**	**94.63**
2. Two Auxiliary Variable Model	96.80	**94.89**	**94.42**	**94.44**	**94.53**	**94.45**
3. One Auxiliary Variable Model	96.93	**94.99**	**94.43**	**94.49**	**94.48**	**94.60**
4. ML with No Additional Mediators	97.13	**95.06**	**94.35**	**94.56**	**94.54**	**94.68**
5. Listwise Deletion	97.10	**95.14**	**94.63**	**94.48**	**94.78**	**94.46**
80% Missing on M_1						
1. Latent Variable Model	97.12	**95.09**	**94.59**	**94.19**	**94.62**	**94.50**
2. Two Auxiliary Variable Model	97.28	**94.70**	93.45	93.70	93.60	93.98
3. One Auxiliary Variable Model	97.64	**95.04**	93.56	93.77	93.76	94.01
4. ML with No Additional Mediators	98.30	**95.32**	93.62	**94.32**	93.80	93.93
5. Listwise Deletion	98.79	**95.94**	**94.38**	**94.30**	93.83	93.75

Note. Values ranging from 94.05 to 95.96% are bold; a procedure with honest error rates would be expected to yield coverage values within this range.

[a]Coverage values were nearly identical across all methods for the complete data condition. The coverage results from the latent variable model (Method 1) are presented to represent complete data results.

Table 17. Confidence Interval Coverage (%) for the Method$\times\alpha\beta\times N$

	Population αβ					
Method	0.01	0.03	0.05	0.09	0.15	0.25
***N* = 100**						
1. Latent Variable Model	99.13	**95.05**	**94.24**	93.94	**94.09**	**94.12**
2. Two Auxiliary Variable Model	98.79	**94.87**	93.35	93.50	93.24	93.70
3. One Auxiliary Variable Model	98.81	**95.17**	93.38	93.76	93.34	93.82
4. ML with No Additional Mediators	98.82	**95.56**	93.37	**94.38**	93.43	93.93
5. Listwise Deletion	99.22	96.21	**94.43**	**94.28**	93.59	93.44
***N* = 200**						
1. Latent Variable Model	98.38	**94.85**	**94.48**	**94.97**	**94.67**	**94.74**
2. Two Auxiliary Variable Model	98.35	**94.71**	93.94	**94.54**	**94.56**	**94.49**
3. One Auxiliary Variable Model	98.48	**94.99**	94.02	**94.48**	**94.48**	**94.35**
4. ML with No Additional Mediators	98.69	**95.20**	**94.15**	**94.79**	**94.45**	**94.46**
5. Listwise Deletion	98.90	**95.47**	**94.38**	**94.64**	**94.56**	**94.32**
***N* = 500**						
1. Latent Variable Model	**94.88**	**94.88**	**94.96**	**94.66**	**95.16**	**94.54**
2. Two Auxiliary Variable Model	**95.68**	**94.63**	**94.80**	**94.73**	**94.81**	**94.33**
3. One Auxiliary Variable Model	**95.91**	**94.69**	**94.81**	**94.65**	**94.76**	**94.58**
4. ML with No Additional Mediators	96.39	**94.68**	**94.71**	**94.84**	**94.99**	**94.53**
5. Listwise Deletion	96.34	**94.73**	**94.77**	**94.90**	**95.07**	**94.59**
***N* = 1000**						
1. Latent Variable Model	**94.87**	**94.96**	**95.04**	**95.04**	**94.84**	**95.02**
2. Two Auxiliary Variable Model	**94.76**	**94.83**	**94.84**	**95.04**	**94.73**	**94.92**
3. One Auxiliary Variable Model	**95.06**	**94.81**	**94.91**	**94.94**	**94.82**	**95.01**
4. ML with No Additional Mediators	**95.61**	**94.75**	**94.87**	**94.79**	**94.74**	**94.97**
5. Listwise Deletion	**95.61**	**94.76**	**94.97**	**94.79**	**94.82**	**95.05**

Note. To be consistent with the logistic regression analysis, cells only represent replications where the missing data rate is non-zero. Values ranging from 94.05 to 95.96% are bold; a procedure with honest error rates would be expected to yield coverage values within this range.

Table 18. Confidence Interval Coverage (%) for the Method × N × Missing Rate Interaction

Method	N = 100	N = 200	N = 500	N = 1000
Complete Data				
All Methods[a]	**95.04**	**94.98**	**94.96**	**94.83**
20% Missing on M_1				
1. Latent Variable Model	**95.09**	**95.12**	**94.89**	**95.02**
2. Two Auxiliary Variable Model	**95.07**	**95.14**	**94.89**	**95.01**
3. One Auxiliary Variable Model	**95.09**	**95.09**	**94.88**	**95.03**
4. ML with No Additional Mediators	**95.07**	**95.22**	**95.01**	**94.99**
5. Listwise Deletion	**95.08**	**95.27**	**94.98**	**95.06**
50% Missing on M_1				
1. Latent Variable Model	**94.75**	**95.23**	**94.94**	**94.80**
2. Two Auxiliary Variable Model	**94.59**	**95.05**	**94.83**	**94.81**
3. One Auxiliary Variable Model	**94.61**	**95.09**	**94.92**	**94.86**
4. ML with No Additional Mediators	**94.74**	**95.14**	**95.01**	**94.78**
5. Listwise Deletion	**94.94**	**95.21**	**95.05**	**94.85**
80% Missing on M_1				
1. Latent Variable Model	**94.81**	**95.01**	**94.86**	**95.05**
2. Two Auxiliary Variable Model	93.31	**94.40**	**94.69**	**94.69**
3. One Auxiliary Variable Model	93.69	**94.59**	**94.75**	**94.80**
4. ML with No Additional Mediators	**94.13**	**94.82**	**94.81**	**94.92**
5. Listwise Deletion	**95.10**	**95.08**	**94.96**	**94.93**

Note. Values ranging from 94.05 to 95.96% are bold; a procedure with honest error rates would be expected to yield coverage values within this range.

[a]Coverage values were nearly identical across all methods for the complete data condition. The coverage results from the latent variable model (Method 1) are presented to represent complete data results.

Table 19. Complete Data Power (%) Results from Current Study compared to MacKinnon et al. (2002)

	Power (%)	
	Current Study	MacKinnon et al. (2002)
αβ = Small-Small		
N = 100	2.37	6.20
N = 200	6.58	27.40
N = 500	34.38	76.00
N = 1000	77.27	98.00
αβ =Medium-Medium		
N = 100	71.07	92.00
N = 200	97.97	100.00
N = 500	100.00	100.00
N = 1000	100.00	100.00
αβ = Large-Large		
N = 100	99.65	100.00
N = 200	100.00	100.00
N = 500	100.00	100.00
N = 1000	100.00	100.00

Note. The two studies defined effect sizes differently as described in the text.

Table 20. Empirical Power and Associated Cost for Conditions where $\alpha\beta = 0.03$, $N =$ 1000, and Power Values are in the 75 – 87% Range

Missing Rate	**Analysis Method**	**Power (%)**	**Cost**[a]
Complete Data	All Methods[b]	86.83	\$530,000 - \$550,000[b]
20% Missing	1. Latent Variable Model	85.23	\$450,000
20% Missing	2. Two Auxiliary Variable Model	85.00	\$450,000
20% Missing	3. One Auxiliary Variable Model	83.50	\$440,000
20% Missing	4. ML with No Additional Mediators	78.90	\$430,000
20% Missing	5. Listwise Deletion[c]	78.47	\$424,000[c]
50% Missing	1. Latent Variable Model	82.85	\$300,000
50% Missing	2. Two Auxiliary Variable Model	79.40	\$300,000
50% Missing	3. One Auxiliary Variable Model	75.50	\$290,000
80% Missing	1. Latent Variable Model	79.45	\$150,000

[a]Cost based on a hypothetical research design that costs \$10 per participant included, \$10 per participant to collect data for variables X, Y, M_2 and M_3, and \$500 to collect data per participant on the expensive mediator, M_1

[b]For complete data conditions, percent power estimates are identical through the first decimal points for all analysis methods. The cost will vary depending on whether or not additional mediators are collected. Methods 1 and 2 cost \$550,000 because all three mediators are collected for 1000 participants. Method 3 only collects two mediators, M_1 and M_2, and costs \$540,000. Methods 4 and 5 only collect one mediator, M_1, and costs \$530,000.

[c]This condition would be identical to a complete data condition where $N = 800$. The cost, \$424,000 reflects a situation where X, M_1, and Y are only collected for 800 participants. If additional data are collected on X and Y and thrown out, the cost would be \$430,000.

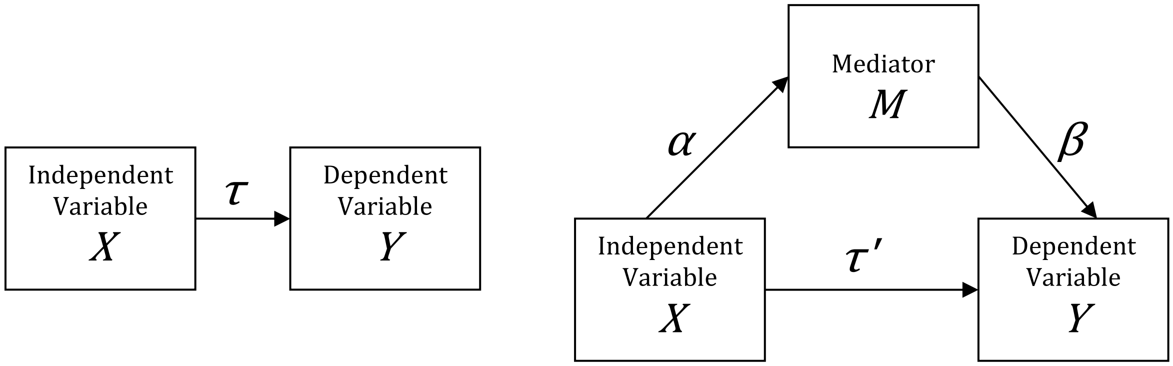


Figure 1. Illustration of the mediation model using path diagrams. Paths indicating error in prediction have been eliminated for simplicity. The path diagram on the left side is expressed by Equation 1.1. The path diagram on the right side is expressed by Equations 1.2 and 1.3.

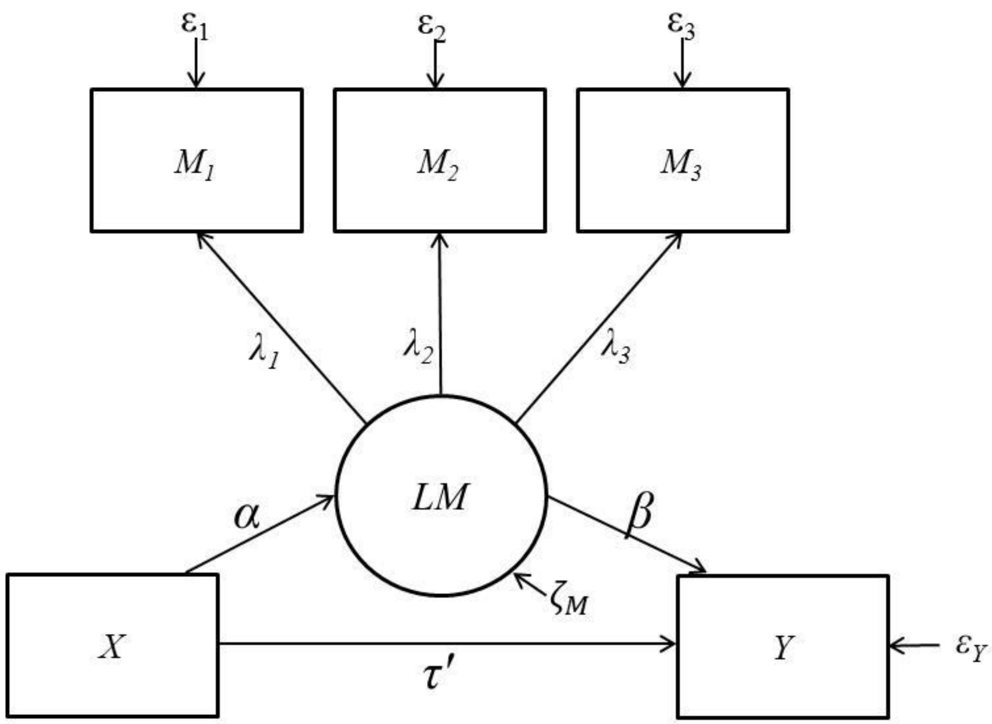


Figure 2. Path diagram of a mediation model with multiple indicators of the mediation construct. Similar to Figure 1, the mediated effect is assessed based on the α and β paths.

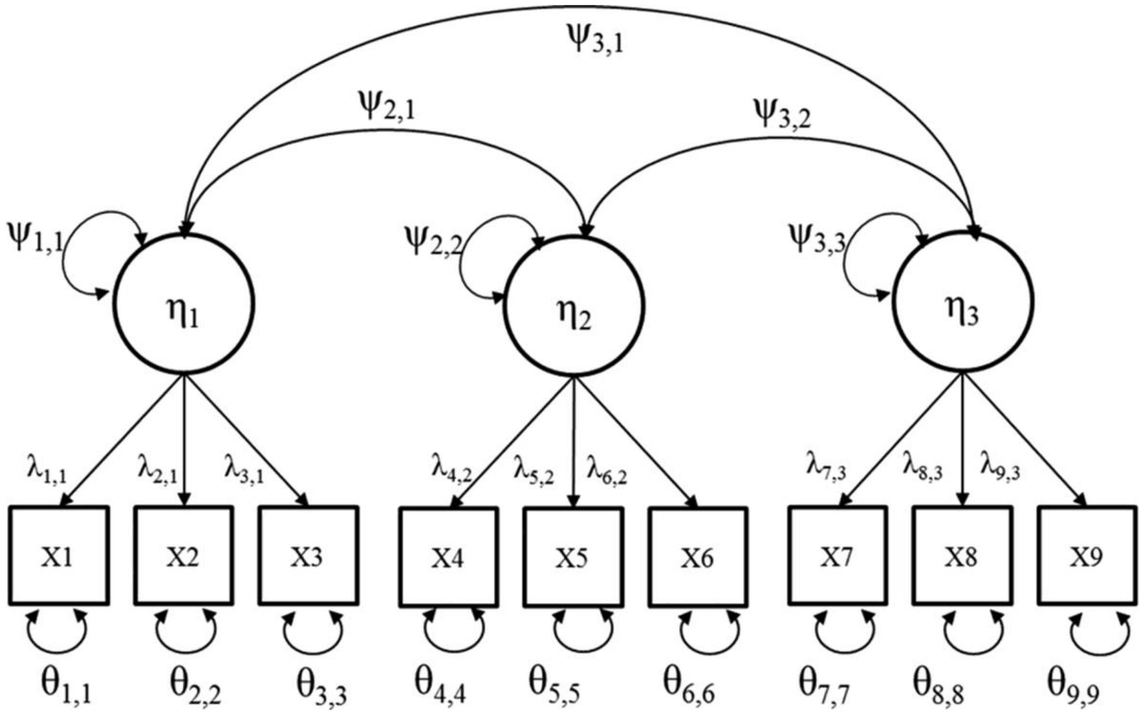


Figure 3. Reproduction of the cross-sectional confirmatory factor analysis model as described in Jia et al., 2014.

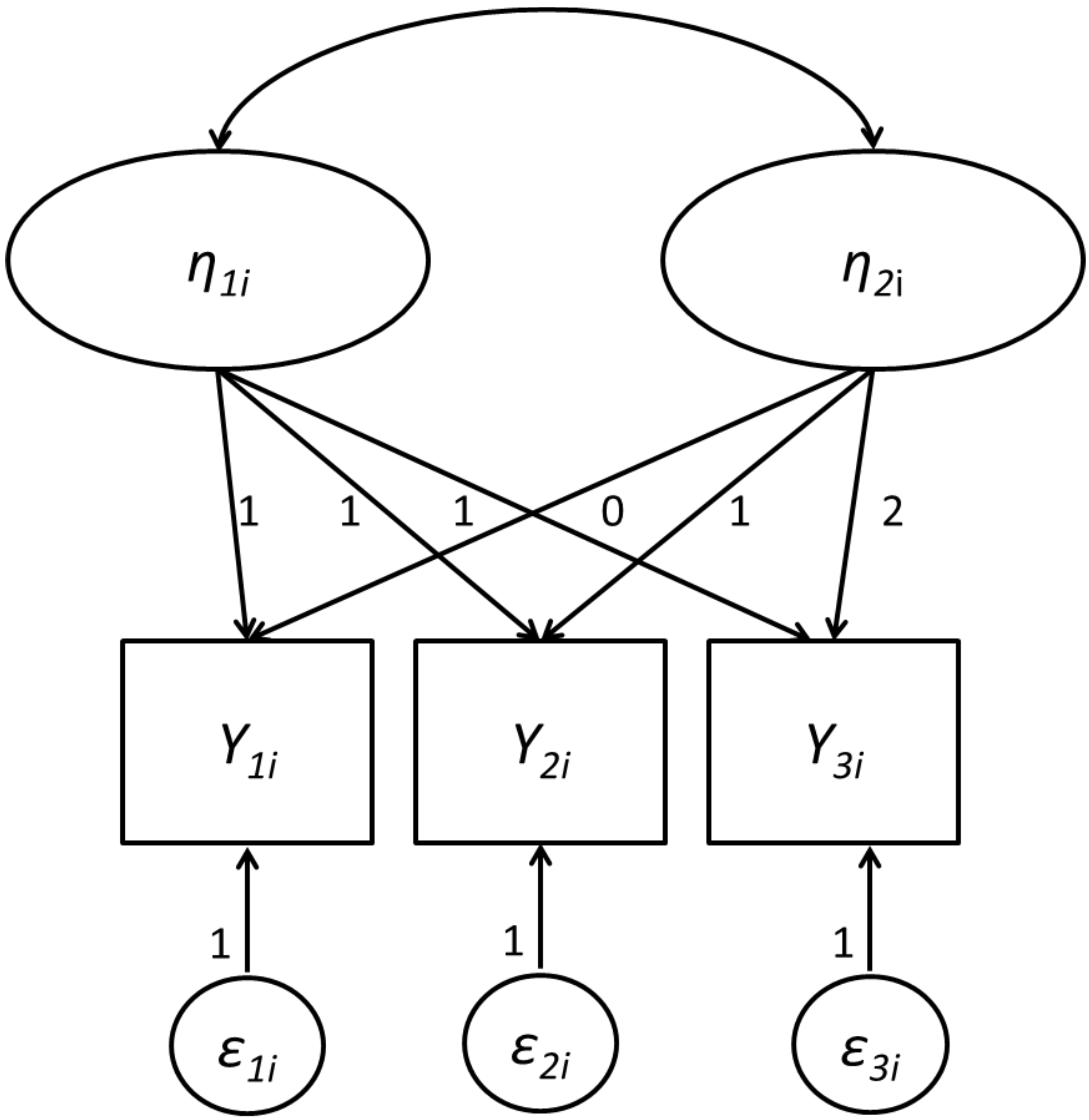


Figure 4. SEM measurement model.

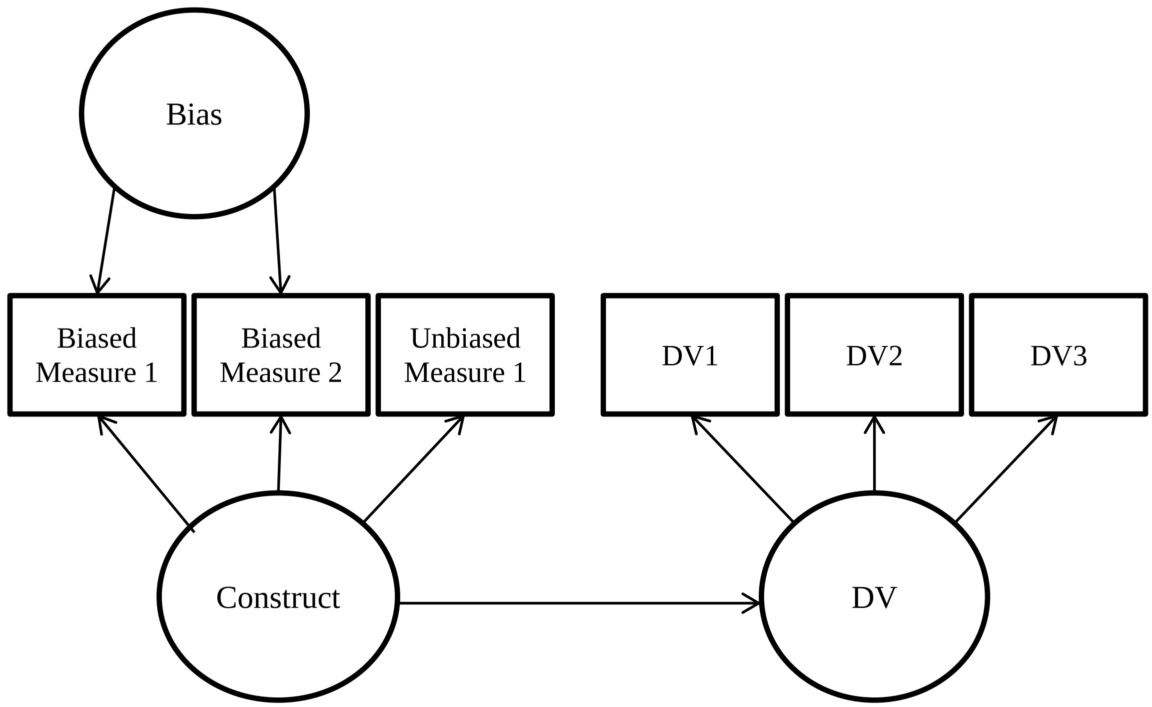


Figure 5. Bias response model for two-method measurement. DV = Dependent Variable.

There must be at least two biased measures to model bias.

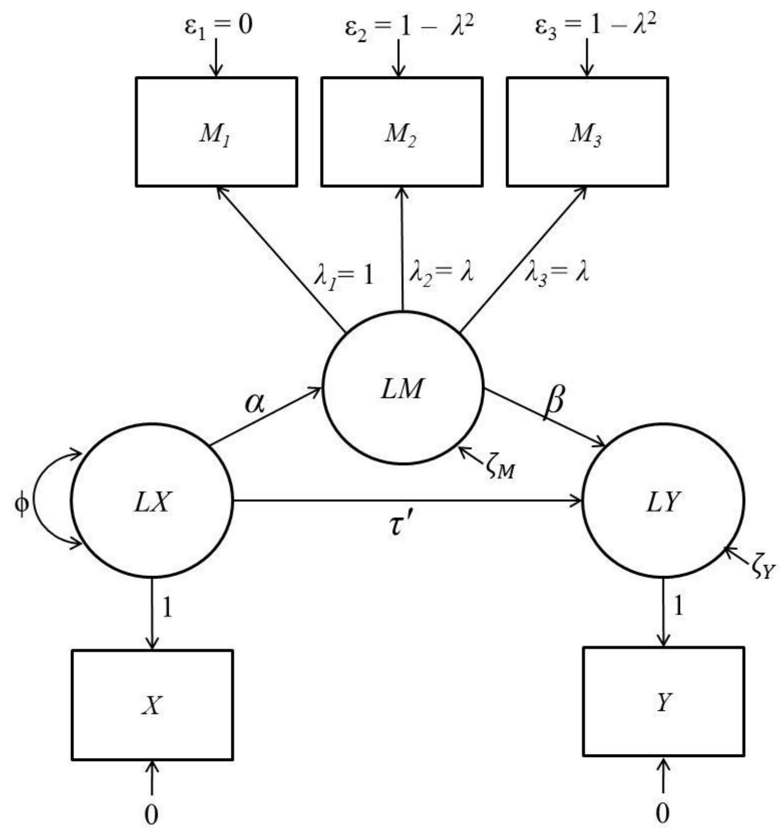
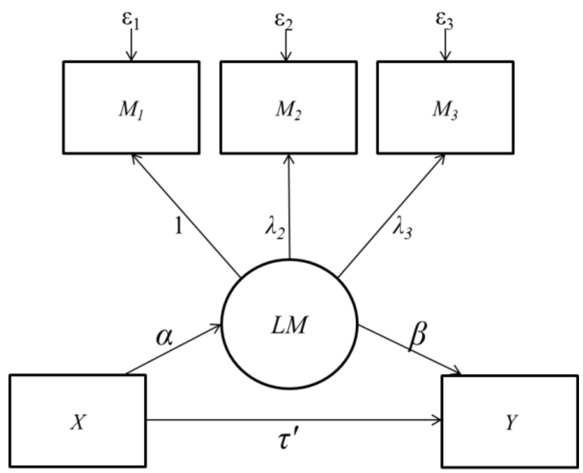
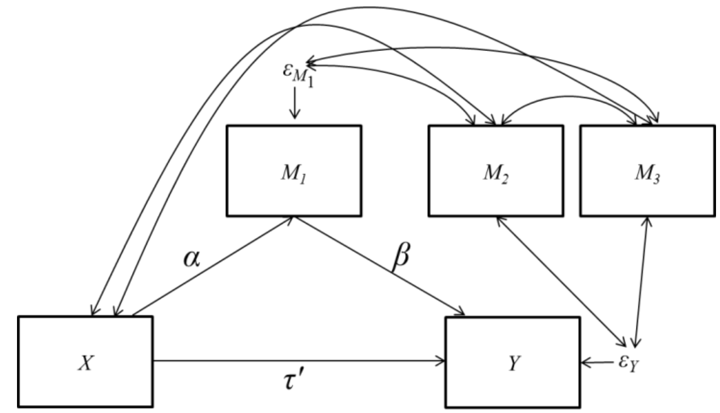


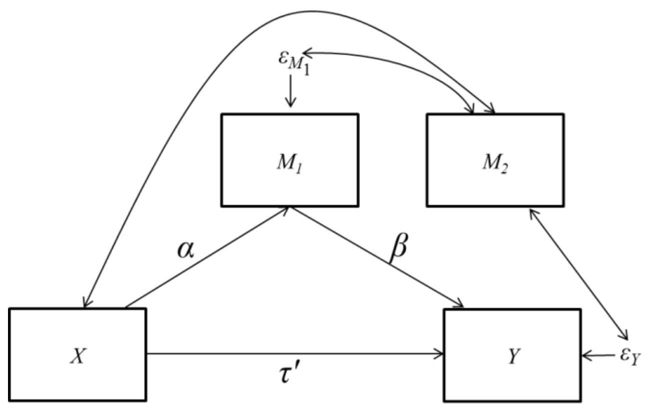
Figure 6. Data generation model. The observed variables are *X*, M_1, M_2, M_3 and *Y*. Variable M_1 is the "expensive" mediation variable and variables M_2 and M_3 are the "inexpensive" mediation variables. All variables are assumed to be multivariate normal. *X* and *Y* are manifest variables generated as single manifest indicator latent variables with loadings and residual variances of one and zero, respectively. The factor loading, λ_1 is constrained to one. The factor loadings for M_2 and M_3 are constrained such that $\lambda_2 = \lambda_3$ and varied to reflect low, medium and high loadings of 0.4, 0.6 and 0.8, respectively. The variances of *LX*, *LM* and *LY* are all constrained to one as described in the text. Parameters α and β are fully crossed with all levels of zero, small, medium and large effect sizes and τ' is constrained to zero with the exception of $\alpha = \beta = 0$, where τ' is constrained to small, medium and large.



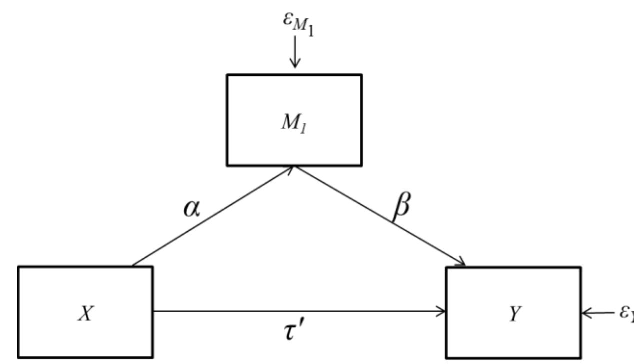
Method 1. Three indicator latent variable mediation model.



Method 2. Auxiliary variable mediation model with two auxiliary variables.



Method 3. Auxiliary variable mediation model with one auxiliary variable.



Methods 4 and 5. Inclusion of M_1 only using ML missing data methods (Method 4) or ML with listwise deletion (Method 5).

Figure 7. Analysis methods evaluated in simulation study.

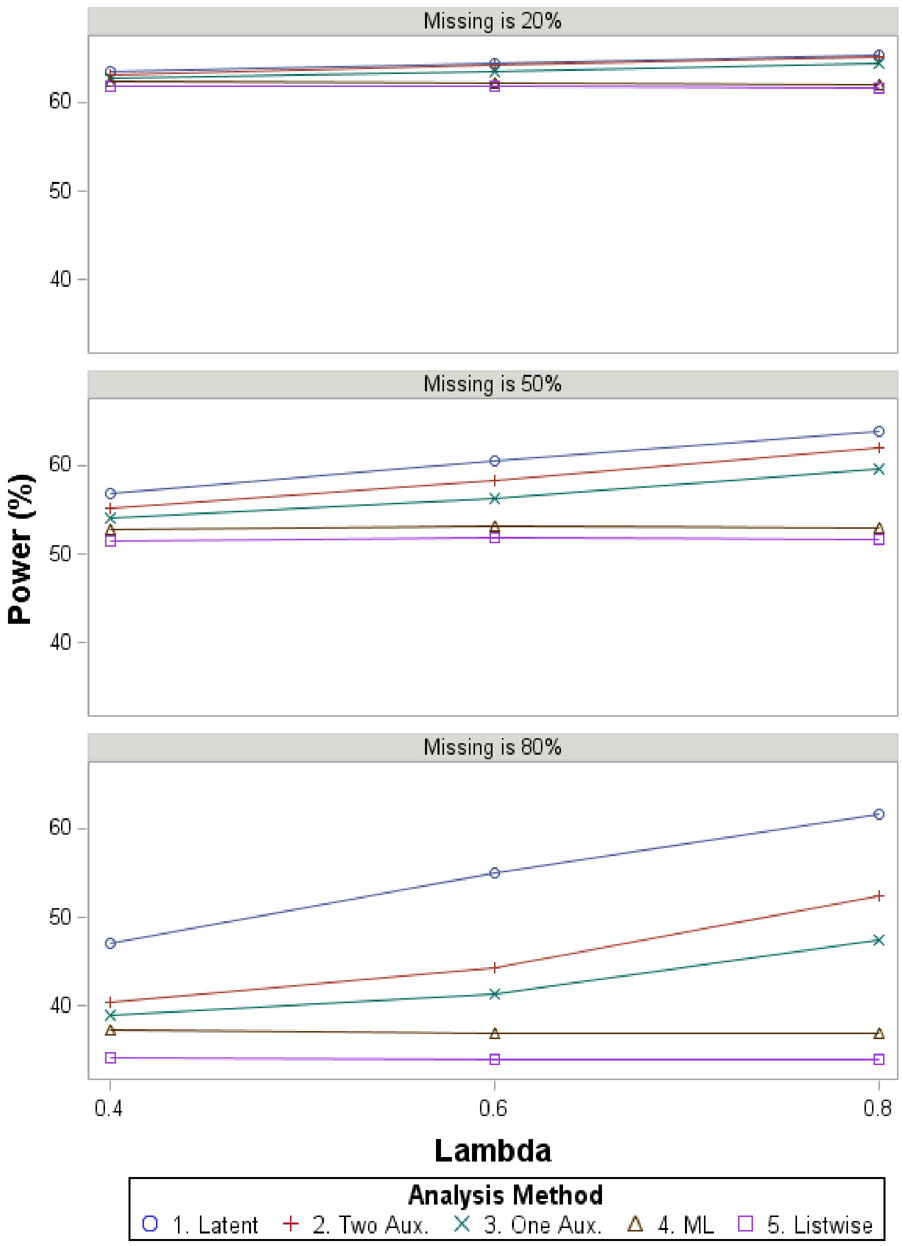


Figure 8. Percent power for the Method×λ×Missing Rate interaction. Graphs showing the relation between λ, method and power are paneled by missing rate on M_1. The horizontal axis shows λ and the lines represent analysis methods. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7. This graph depicts that the interaction between method and λ is modified by missing data rate.

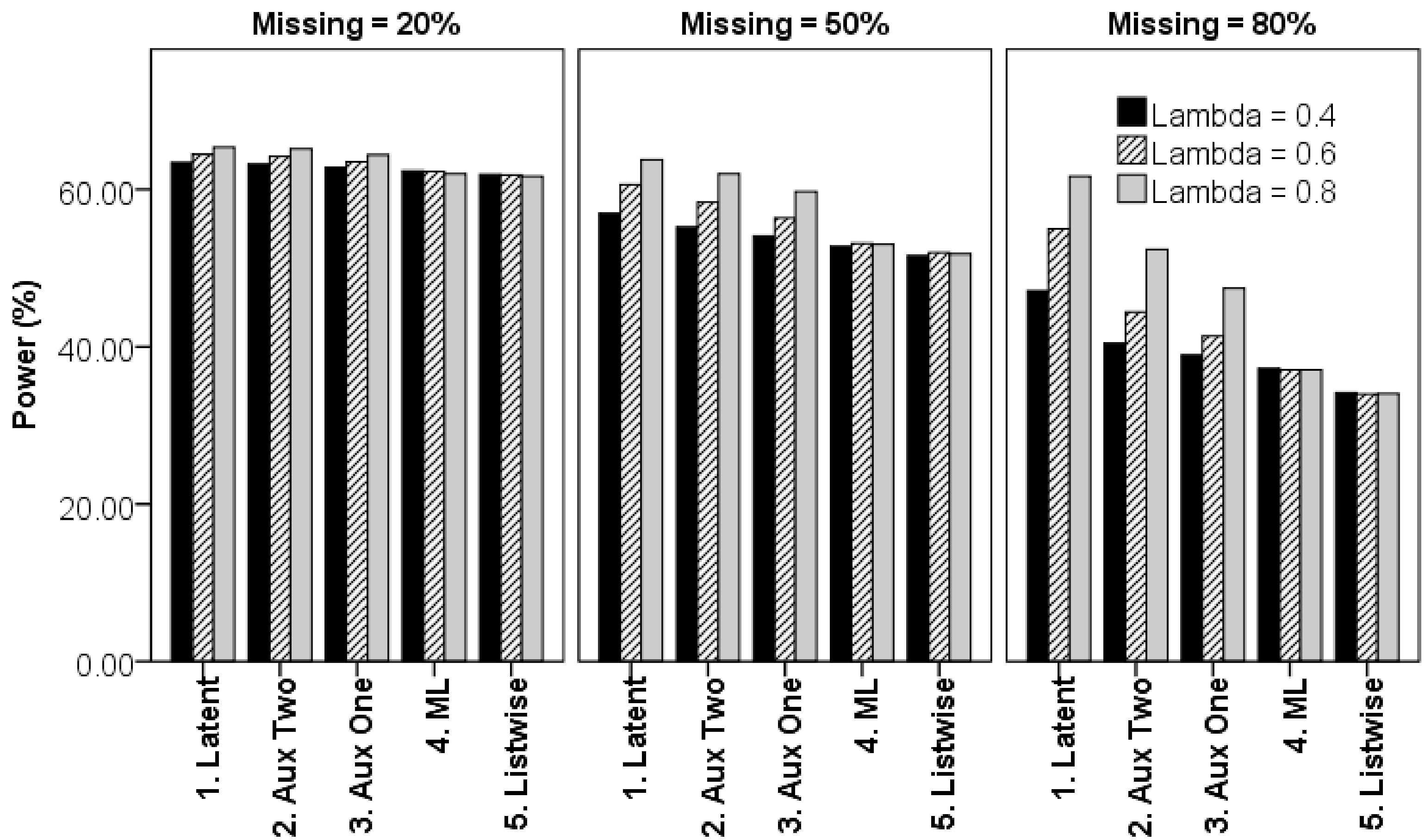


Figure 9. Percent power for each analysis method for the different values of λ paneled by the missing data rate of M_1. The solid bars denote power for $\lambda = 0.4$, the striped bars denote $\lambda = 0.6$, and the grey bars denote $\lambda = 0.8$. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7.

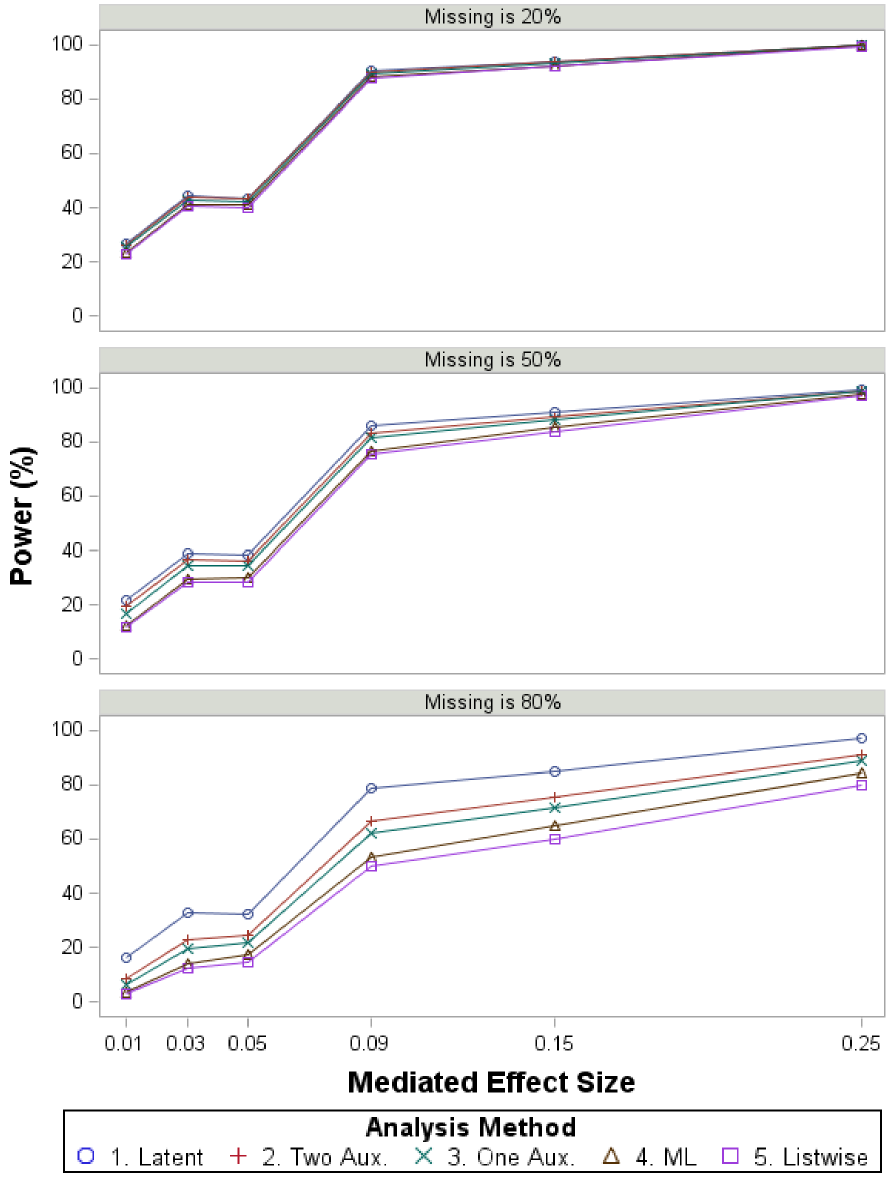


Figure 10. Percent power for the Method × $\alpha\beta$ × Missing Rate interaction. Graphs showing the relation between population $\alpha\beta$, method and power are paneled by missing data rate of M_1. The horizontal axis shows mediated effect size, $\alpha\beta$, and the lines represent analysis methods. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7. This graph depicts that the interaction between method and $\alpha\beta$ is modified by missing data rate.

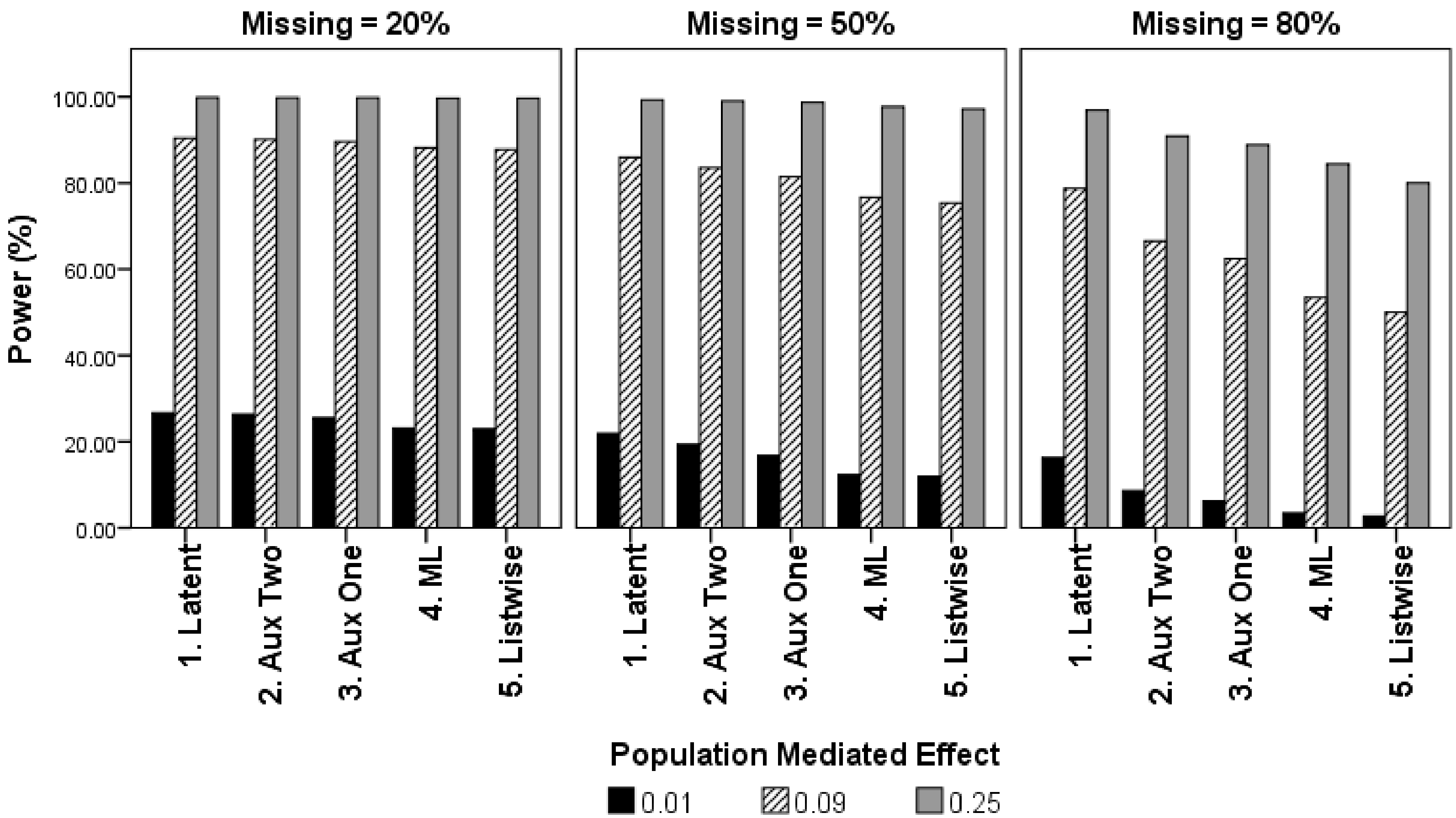


Figure 11. Percent power for each analysis method for three select mediated effect sizes ($\alpha\beta$ = 0.01, 0.09 and 0.25) paneled by missing data rate on M_1. The black bars denote power for $\alpha\beta$ = 0.01, the striped bars denote power for $\alpha\beta$ = 0.09, and the grey bars denote power for $\alpha\beta$ = 0.25. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7.

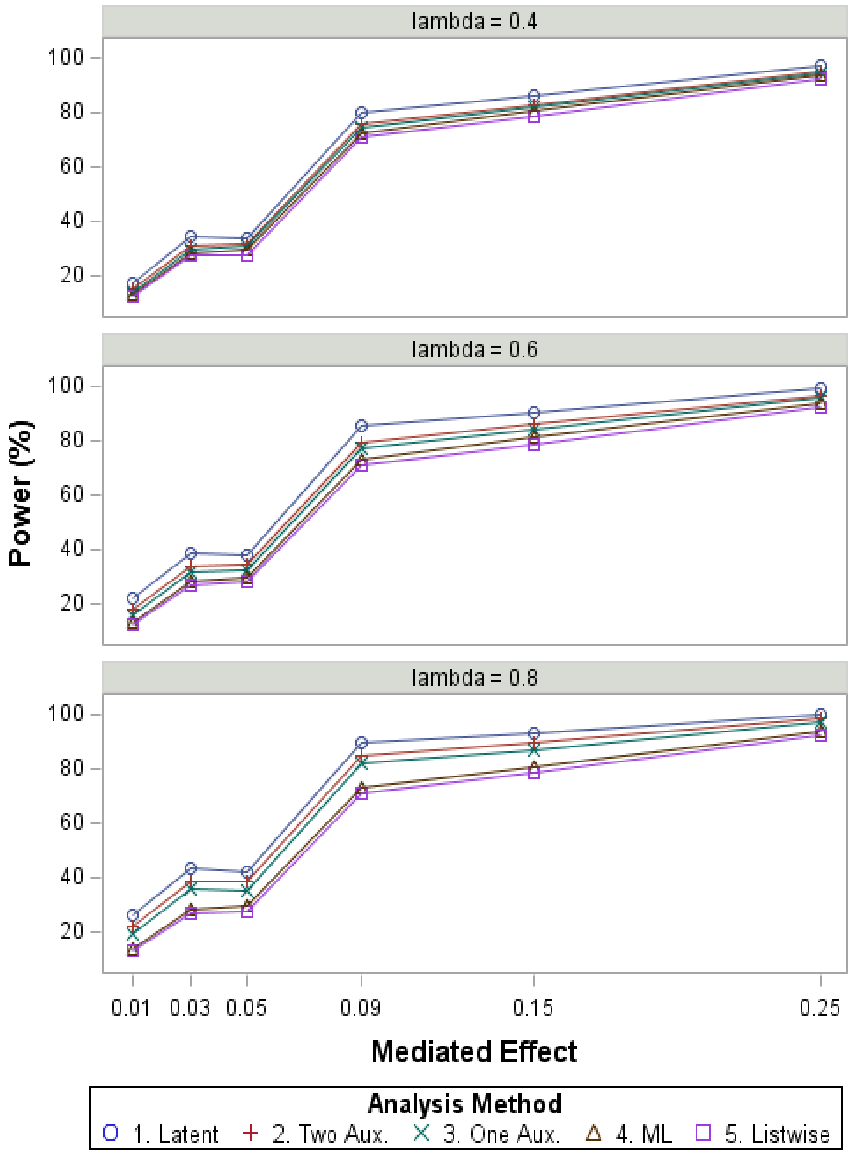


Figure 12. Percent power for the Method$\times\alpha\beta\times\lambda$ interaction. Graphs showing the relation between mediation effect size ($\alpha\beta$), method and power are paneled by missing data rate of M_1. The horizontal axis shows the value of $\alpha\beta$ and the lines represent analysis methods. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7. This graph depicts that the interaction between method and $\alpha\beta$ is moderated by λ.

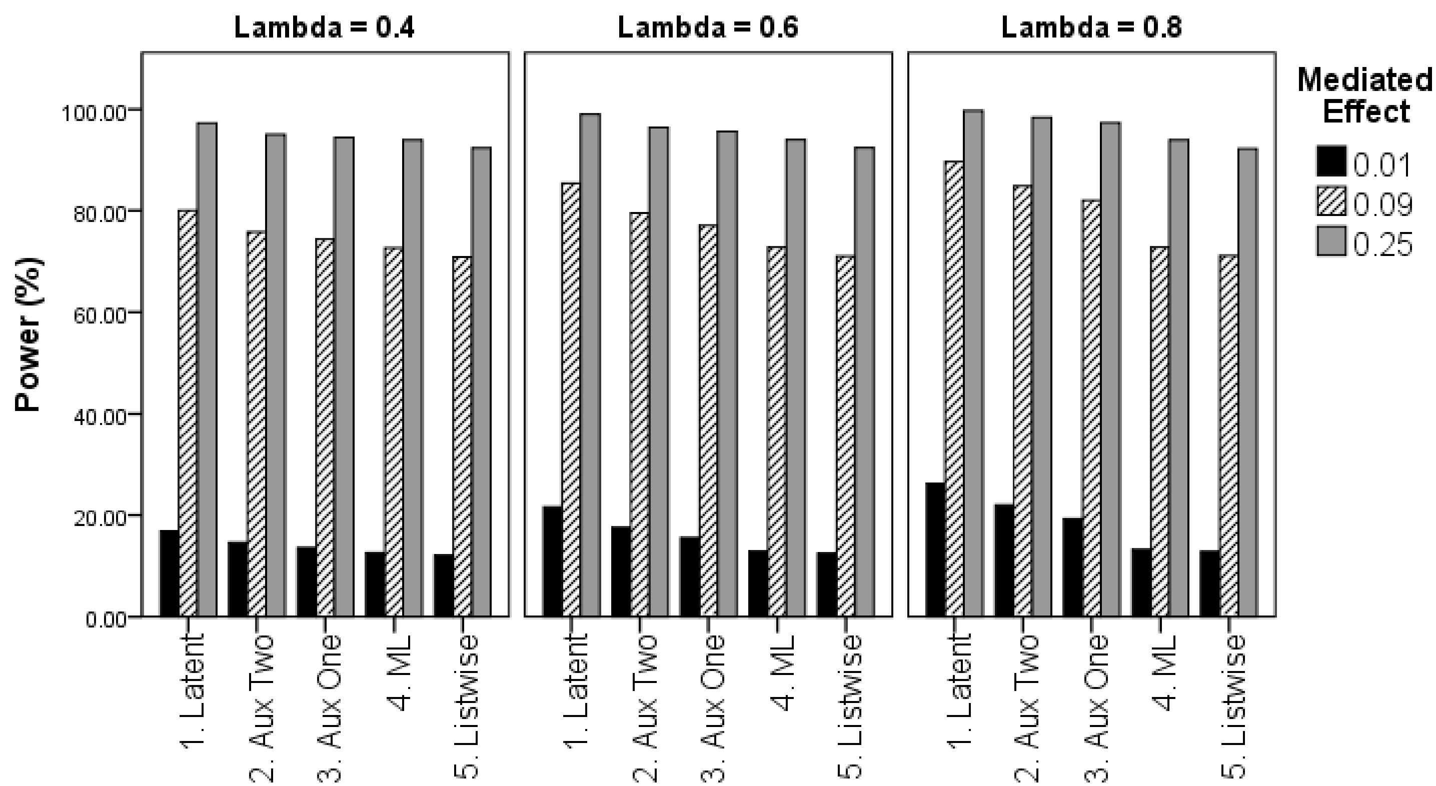


Figure 13. Percent power for each analysis for three select mediated effect sizes, $\alpha\beta$, paneled by values of λ . The solid bars denote power for $\alpha\beta$ = 0.01, the striped bars denote $\alpha\beta$ = 0.09, and the grey bars denote $\alpha\beta$ = 0.25. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7

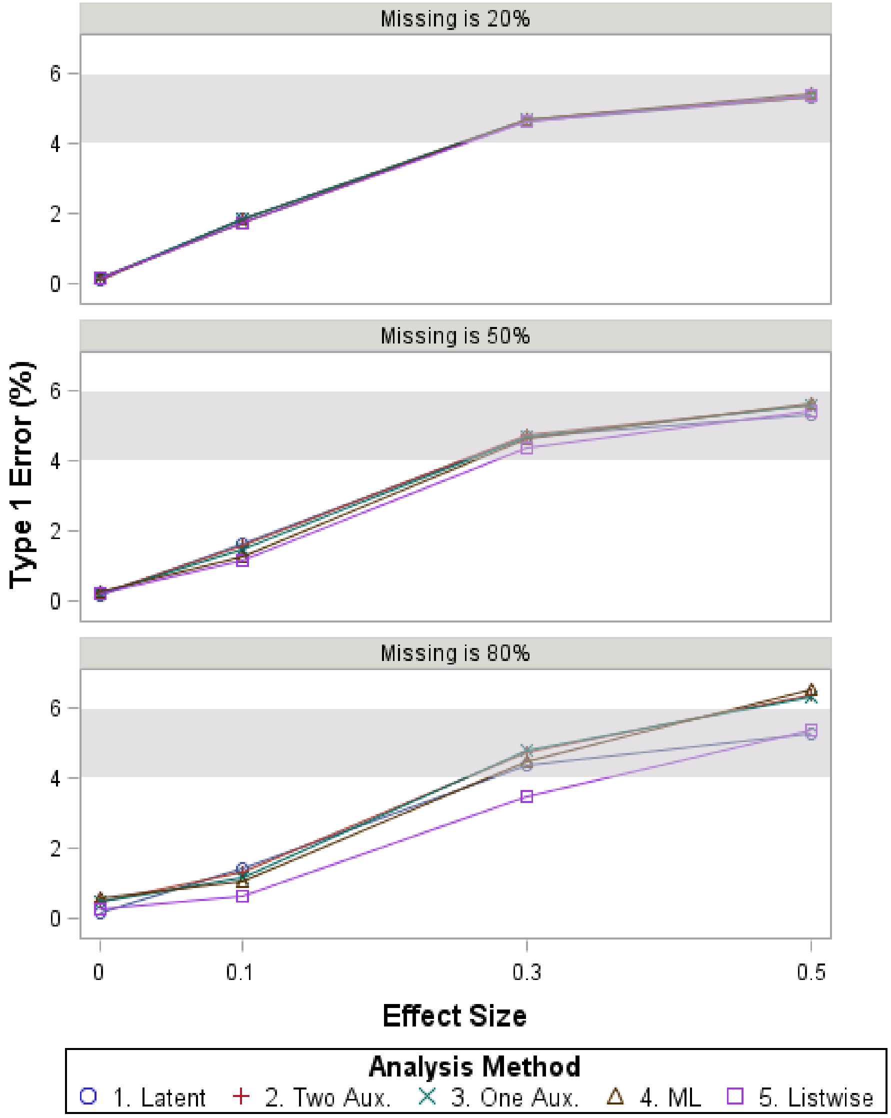


Figure 14. Type 1 error rates for Method×Effect Size×Missing Rate interaction. The relation between method and effect size are paneled by missing rate on M_1. Effect size indicates the size of non-zero α or β (except in the case of effect size = 0 where $\alpha = \beta = 0$). The lines represent analysis methods. Analysis methods are keyed with numbers corresponding to Analysis Models 1 – 5 in Figure 7. The grey band represents values ranging from 4.04 to 5.96%; a procedure with honest error rates would expect to yield Type I errors between these values.

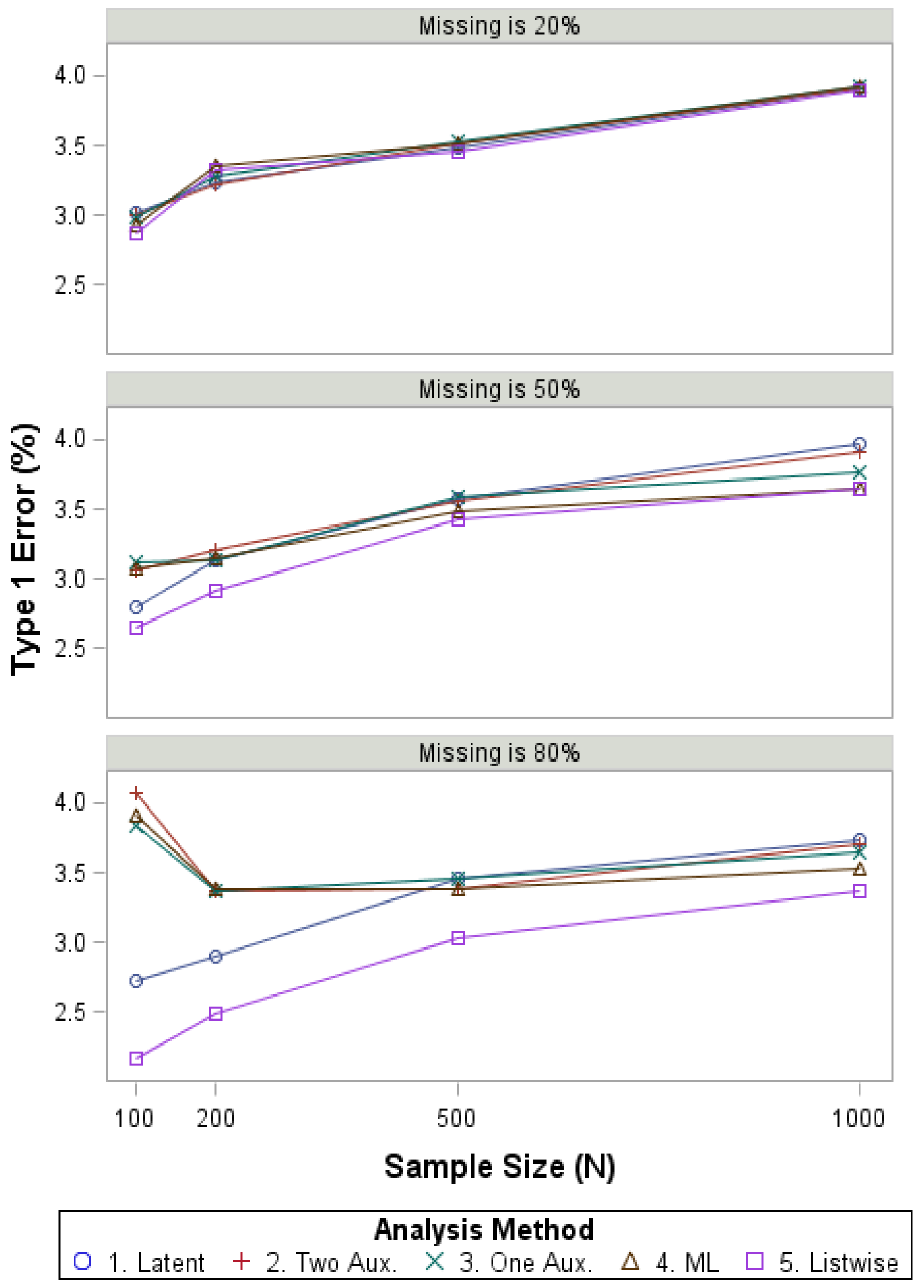


Figure 15. Type 1 error rates for the Method × Sample Size × Missing Rate interaction. The relation between method and ample size (*N*) on Type 1 error rates is paneled by missing rate on M_1. The lines represent analysis methods. Analysis methods are keyed with numbers corresponding to Analysis Models 1 – 5 in Figure 7.

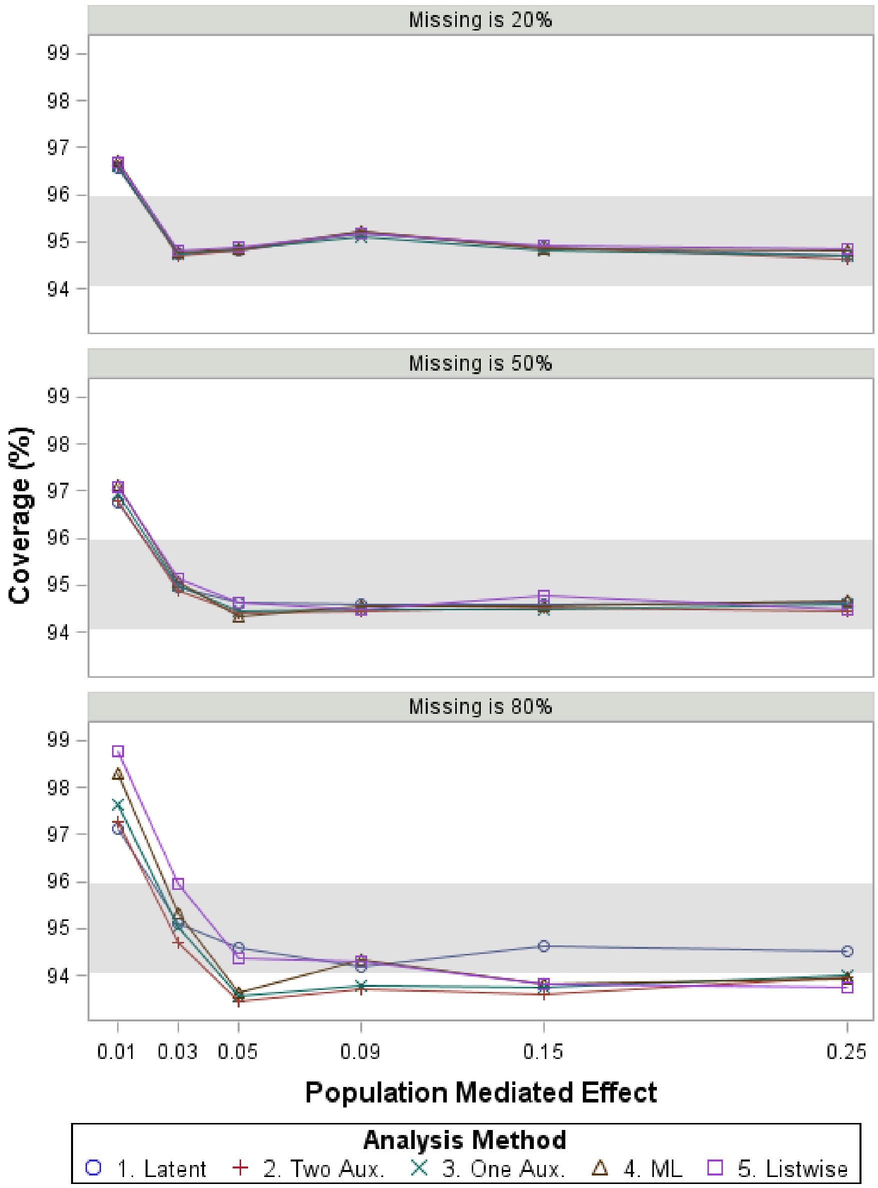


Figure 16. Confidence interval coverage for the Method × $\alpha\beta$ × Missing Rate interaction. The interaction between mediated effect ($\alpha\beta$) and method on coverage is paneled by missing rate. The lines represent analysis methods. Methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7. The grey band represents values ranging from 94.05 to 95.96%; a procedure with honest coverage would expect to yield coverage values within this range.

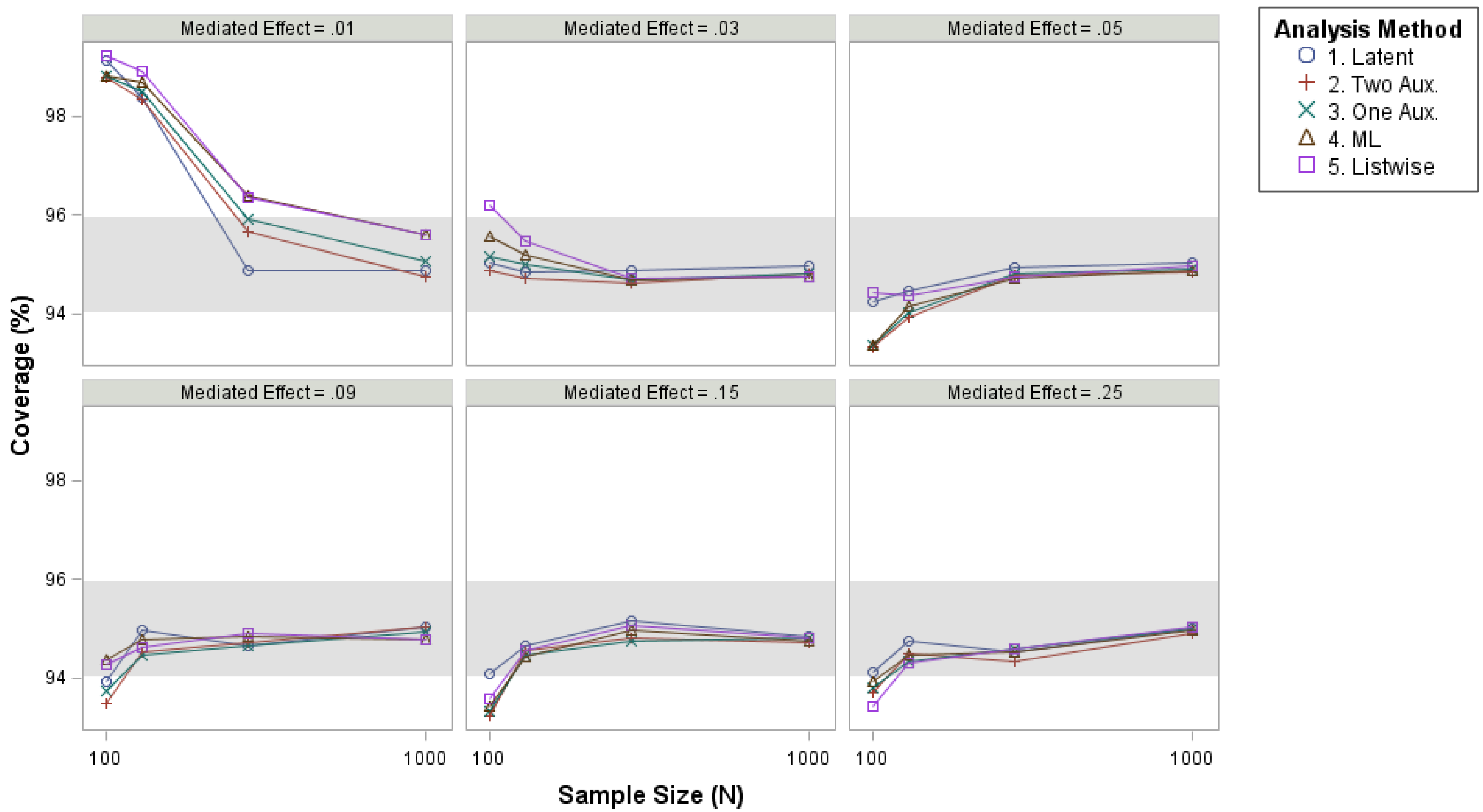


Figure 17. Confidence interval coverage for the Method $\times \alpha\beta \times N$ interaction. The interaction between method and sample size (N) on coverage is paneled by mediated effect, $\alpha\beta$. The lines represent analysis methods. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7. The grey band represents values ranging from 94.05 to 95.96%; a procedure with honest coverage would expect to yield coverage values within this range.

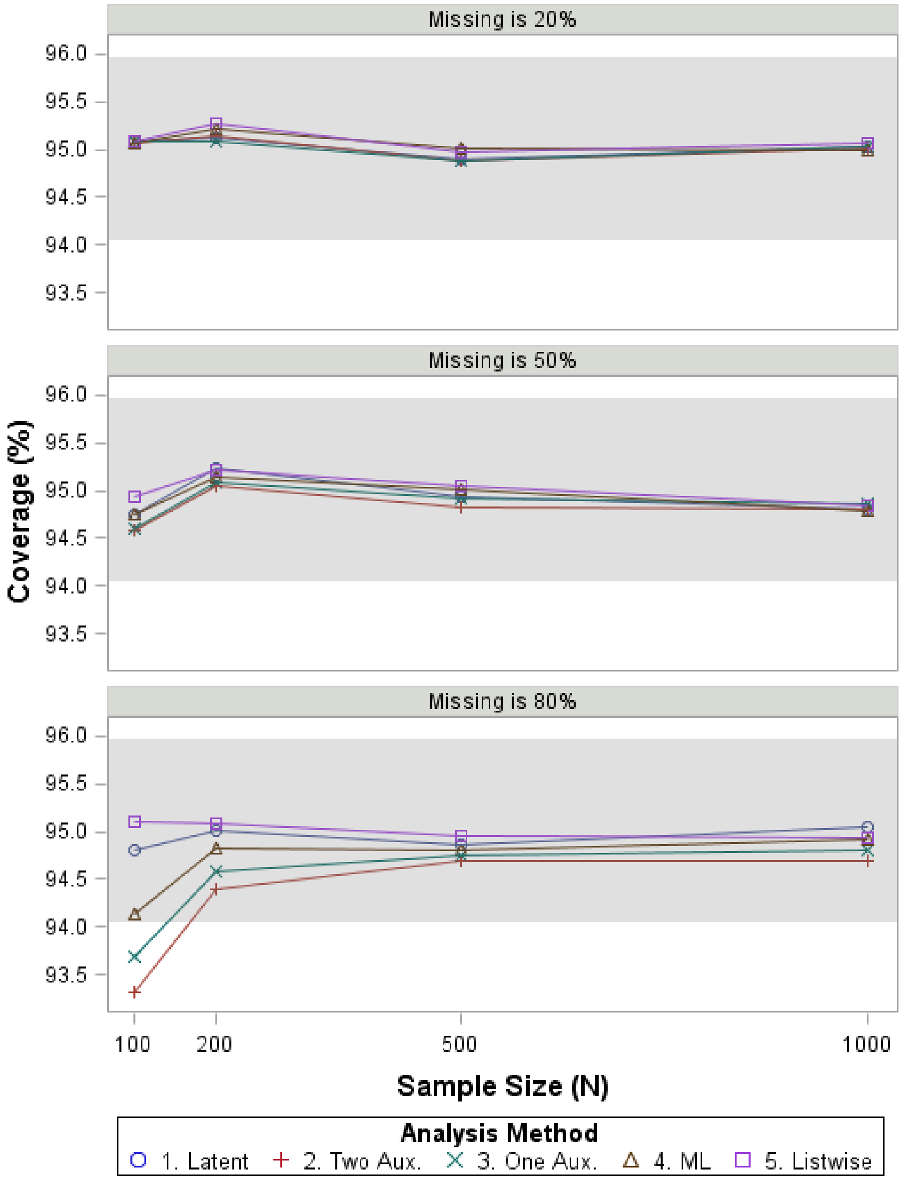


Figure 18. Confidence interval coverage for the Method$\times N \times$Missing Rate interaction. The interaction between method and sample size on coverage is paneled by missing rate. The lines represent analysis methods. Analysis methods are keyed with numbers corresponding to Analysis Methods 1 – 5 in Figure 7. The grey band represents values ranging from 94.05 to 95.96%; a procedure with honest coverage would expect to yield coverage values within this range.

APPENDIX B

MEASUREMENT AND STRUCTURAL MODELS OF THE POPULATION GENERATION MODEL

Equations 3.1 and 3.2 in the text provide the measurement model for the population generation model in matrix form. These matrix equations may be expanded into a set of equations to provide the measurement model as below.

$$X = 1LX + 0 = LX$$

$$M_1 = \lambda_1 LM + \varepsilon_1$$

$$M_2 = LM + \varepsilon_2$$

$$M_3 = \lambda_3 LM + \varepsilon_3$$

$$Y = 1LY + 0 = LY$$

Equation 3.7 in the text provides the structural model for the population generation model in matrix form and Equation 3.9 shows the matrix components. This matrix equation can be expanded into a set of equations that define the structural model as below.

$$LM = \alpha X + \zeta_M$$

$$LY = \beta LM + \tau' X + \zeta_Y$$

APPENDIX C

MODEL IMPLIED COVARIANCE MATRIX

Below assumes $\sigma_X^2 = \phi_X = 1$

		$\underline{X}$	$\underline{M_1}$	$\underline{M_2}$	$\underline{M_3}$
X	|	1			
M_1	|	$\alpha\boldsymbol{\lambda_1}$	$\boldsymbol{\lambda_1}^{\mathbf{2}}(\alpha^2 + \sigma_{\zeta_M}^2) + \sigma_{\varepsilon_1}^2$		
M_2	|	$\alpha\boldsymbol{\lambda_2}$	$\boldsymbol{\lambda_1}\boldsymbol{\lambda_2}(\alpha^2 + \sigma_{\zeta_M}^2)$	$\boldsymbol{\lambda_2}^{\mathbf{2}}(\alpha^2 + \sigma_{\zeta_M}^2) + \sigma_{\varepsilon_2}^2$	
M_3	|	$\alpha\boldsymbol{\lambda_3}$	$\boldsymbol{\lambda_1}\boldsymbol{\lambda_3}(\alpha^2 + \sigma_{\zeta_M}^2)$	$\boldsymbol{\lambda_2}\boldsymbol{\lambda_3}(\alpha^2 + \sigma_{\zeta_M}^2)$	$\boldsymbol{\lambda_3}^{\mathbf{2}}(\alpha^2 + \sigma_{\zeta_M}^2) + \sigma_{\varepsilon_3}^2$
Y	|	$\beta\alpha + \tau'$	$\boldsymbol{\lambda_1}(\beta\alpha^2 + \tau'\alpha + \beta\sigma_{\zeta_M}^2)$	$\boldsymbol{\lambda_2}(\beta\alpha^2 + \tau'\alpha + \beta\sigma_{\zeta_M}^2)$	$\boldsymbol{\lambda_3}(\beta\alpha^2 + \tau'\alpha + \beta\sigma_{\zeta_M}^2)$

Variance of $Y = \beta^2(\alpha^2\sigma_X^2 + \sigma_{\zeta_M}^2) + 2\beta\tau'\alpha + \tau'^2 + \sigma_{\zeta_Y}^2$

APPENDIX D

SYNTAX TO DERIVE POPULATION COVARIANCE MATRIX FROM MODEL

PARAMETERS

Note: Replace all parameter values $(\lambda_i, \sigma^2_{\varepsilon_i}, \alpha, \beta, \tau', \phi_X, \sigma^2_{\zeta_M}, \sigma^2_{\zeta_Y})$ with a numeric value.

Residual variance, $\sigma^2_{\zeta_M}$ and $\sigma^2_{\zeta_Y}$, are defined by values of α and β as described in the text and τ' is zero.

```
PROC IML;

*LambdaX;
LX = {1.00};
*LambdaY Matrix;
LY = {λ1 0, λ2 0, λ3 0, 0 1.00};
*Gamma Matrix;
GA = { α, τ'};
*Phi Covariance Matrix;
PH = {1};
*Psi Covariance Matrix;
PS = {σ²ζM 0.00, 0.00 σ²ζY};
*Theta Delta Matrix;
TD = {0};
*Theta Epsilon Matrix;
TE = {σ²ε1 0.00 0.00 0.00,
  0.00 σ²ε2 0.00 0.00,
  0.00 0.00 σ²ε3 0.00,
  0.00 0.00 0.00 0.00};
*Beta Matrix;
B = {0.00 0.00, β 0.00};
*Identity Matrix;
I = {1 0, 0 1};

COVY = LY*(INV(I-B))*(GA*PH*GA`+PS)*(INV(I-B`))*LY`+TE;
COVX = LX*PH*LX` + TD;
COVYX = LY*(INV(I-B))*GA*PH*LX`;
COVXY = LX*PH*GA`*(INV(I-B`))*LY`;

UPPER = COVX || COVXY;
LOWER = COVYX || COVY;
COV = UPPER // LOWER;
PRINT COV;

S=SQRT(DIAG(COV));  *** obtain the matrix with standard deviations on the diagonal;
S_INV=INV(S);    *** the inverse of S matrix;
R=S_INV*COV*S_INV;  *** obtain correlation matrix;

PRINT COV;
PRINT S;
PRINT R;
```

APPENDIX E

STANDARDIZED BIAS FOR MISSING RATE = 80% FOR ALL NON-ZERO MEDIATED EFFECT

Table E1. Standardized Bias for Missing Rate = 80% and N = 100 for all non-zero $\alpha\beta$

			Standardized Bias (for Missing Rate = 80% and *N* = 100				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	0.076	0.132	0.118	0.074	0.017
		0.6	0.014	-0.069	-0.109	-0.062	-0.106
		0.8	-0.015	-0.013	-0.079	0.047	-0.014
	0.3	**0.4**	0.030	0.001	0.036	0.019	-0.029
		0.6	0.045	0.009	0.004	0.048	-0.007
		0.8	-0.003	0.023	0.014	0.040	-0.009
	0.5	**0.4**	0.062	0.030	0.042	0.071	0.026
		0.6	0.002	0.009	0.026	0.023	0.008
		0.8	-0.031	0.010	0.036	0.036	-0.036
0.3	**0.1**	**0.4**	0.139	0.095	0.163	0.171	0.064
		0.6	0.013	0.051	0.068	-0.016	-0.064
		0.8	-0.020	0.019	0.007	-0.032	-0.101
	0.3	**0.4**	0.083	-0.037	-0.013	0.037	-0.055
		0.6	0.069	0.125	0.190	0.202	0.033
		0.8	0.021	0.046	0.066	0.142	0.011
	0.5	**0.4**	0.194	0.101	0.126	0.166	-0.022
		0.6	0.059	0.114	0.116	0.161	-0.081
		0.8	0.041	0.067	0.094	0.118	-0.038
0.5	**0.1**	**0.4**	0.035	-0.008	-0.013	0.007	-0.034
		0.6	0.012	0.032	0.042	0.110	0.055
		0.8	0.024	0.047	0.103	0.109	0.053
	0.3	**0.4**	0.188	0.059	0.107	0.122	-0.043
		0.6	0.063	0.009	0.070	0.125	-0.064
		0.8	0.006	0.013	0.010	-0.017	-0.149
	0.5	**0.4**	0.272	0.168	0.218	0.268	0.041
		0.6	0.177	0.148	0.217	0.247	0.007
		0.8	0.043	0.118	0.165	0.240	0.013

Table E2. Standardized Bias for Missing Rate = 80% and N = 200 for all non-zero $\alpha\beta$

			Standardized Bias for Missing Rate = 80% and *N* = 200				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	0.008	0.045	-0.010	-0.021	-0.056
		0.6	-0.009	-0.001	-0.016	0.009	0.019
		0.8	-0.005	0.010	0.024	0.073	0.032
	0.3	**0.4**	0.108	0.070	0.088	0.138	0.092
		0.6	0.017	0.057	0.048	0.147	0.064
		0.8	-0.021	-0.038	-0.026	-0.103	-0.143
	0.5	**0.4**	0.040	0.033	0.042	0.026	-0.014
		0.6	0.039	-0.022	-0.025	0.024	0.010
		0.8	0.019	0.014	0.030	0.068	0.001
0.3	**0.1**	**0.4**	0.079	0.083	0.100	0.096	0.038
		0.6	-0.008	0.040	0.028	0.013	0.017
		0.8	0.008	0.025	0.043	0.061	0.005
	0.3	**0.4**	0.152	0.149	0.143	0.143	0.062
		0.6	0.047	0.064	0.111	0.160	0.073
		0.8	0.028	0.072	0.118	0.106	0.005
	0.5	**0.4**	0.176	0.225	0.268	0.262	0.155
		0.6	0.056	0.108	0.112	0.141	0.003
		0.8	0.050	0.085	0.107	0.131	0.007
0.5	**0.1**	**0.4**	0.080	0.048	0.071	0.064	0.010
		0.6	0.030	0.032	0.031	0.031	-0.007
		0.8	0.041	0.095	0.091	0.102	0.060
	0.3	**0.4**	0.146	0.118	0.132	0.122	0.030
		0.6	0.063	0.060	0.098	0.128	0.021
		0.8	0.060	0.100	0.111	0.114	0.024
	0.5	**0.4**	0.198	0.168	0.184	0.205	0.002
		0.6	0.051	0.087	0.123	0.153	-0.044
		0.8	0.020	0.070	0.098	0.172	-0.032

Table E3. Standardized Bias for Missing Rate = 80% and N = 500 for all non-zero $\alpha\beta$

			Standardized Bias for Missing Rate = 80% and *N* = 500				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	0.052	0.100	0.099	0.073	0.063
		0.6	0.033	-0.012	0.000	-0.021	-0.033
		0.8	0.013	0.040	0.018	0.092	0.054
	0.3	**0.4**	0.122	0.084	0.105	0.115	0.077
		0.6	-0.022	-0.006	-0.025	-0.068	-0.101
		0.8	0.005	0.008	0.027	0.027	-0.006
	0.5	**0.4**	0.029	0.029	0.043	0.031	0.024
		0.6	-0.021	0.024	0.036	0.006	0.006
		0.8	0.012	0.018	0.001	0.015	-0.033
0.3	**0.1**	**0.4**	0.079	0.108	0.066	0.044	0.015
		0.6	0.020	0.088	0.082	0.062	0.035
		0.8	-0.019	0.010	0.035	0.029	0.010
	0.3	**0.4**	0.093	0.083	0.109	0.138	0.080
		0.6	0.025	0.025	0.041	0.067	-0.006
		0.8	0.016	0.092	0.114	0.096	0.035
	0.5	**0.4**	0.057	0.066	0.069	0.085	0.018
		0.6	0.047	0.127	0.114	0.142	0.036
		0.8	0.008	0.042	0.057	0.047	-0.050
0.5	**0.1**	**0.4**	0.074	0.088	0.080	0.065	0.034
		0.6	0.031	0.001	-0.003	0.002	-0.030
		0.8	0.001	0.054	0.042	-0.029	-0.059
	0.3	**0.4**	0.068	0.050	0.079	0.074	-0.013
		0.6	0.059	0.125	0.101	0.118	0.039
		0.8	-0.023	0.031	0.077	0.110	0.052
	0.5	**0.4**	0.158	0.169	0.178	0.182	0.067
		0.6	0.119	0.161	0.176	0.212	0.088
		0.8	-0.001	0.034	0.074	0.062	-0.033

Table E4. Standardized Bias for Missing Rate = 80% and N = 1000 for all non-zero $\alpha\beta$

			Standardized Bias for Missing Rate = 80% and *N* = 1000				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	0.004	-0.025	-0.013	-0.029	-0.046
		0.6	-0.005	0.016	0.042	0.057	0.041
		0.8	0.019	0.059	0.036	0.078	0.069
	0.3	**0.4**	-0.028	-0.003	-0.013	-0.009	-0.013
		0.6	-0.002	-0.025	-0.021	-0.043	-0.086
		0.8	-0.025	-0.004	-0.001	0.046	0.039
	0.5	**0.4**	-0.039	-0.037	-0.063	-0.053	-0.064
		0.6	0.021	0.047	0.049	0.040	0.031
		0.8	0.005	0.010	0.023	0.039	-0.015
0.3	**0.1**	**0.4**	0.056	0.031	0.025	0.045	0.024
		0.6	-0.035	-0.062	-0.032	-0.004	-0.021
		0.8	-0.022	-0.007	0.014	-0.049	-0.058
	0.3	**0.4**	0.029	0.013	0.015	0.022	0.003
		0.6	0.026	0.042	0.043	0.077	0.036
		0.8	0.046	0.084	0.093	0.053	0.007
	0.5	**0.4**	0.043	0.044	0.043	0.034	-0.055
		0.6	-0.004	0.039	0.051	0.050	0.000
		0.8	0.021	0.041	0.043	0.096	0.036
0.5	**0.1**	**0.4**	0.040	0.031	0.041	0.028	0.027
		0.6	-0.023	-0.046	-0.044	0.007	-0.010
		0.8	-0.004	0.034	0.055	0.026	0.015
	0.3	**0.4**	0.012	-0.019	-0.013	-0.001	-0.056
		0.6	0.047	0.096	0.120	0.150	0.086
		0.8	0.035	0.024	0.018	-0.026	-0.059
	0.5	**0.4**	0.073	0.055	0.069	0.094	0.006
		0.6	0.023	0.063	0.091	0.072	-0.026
		0.8	-0.017	-0.037	-0.036	-0.053	-0.160

APPENDIX F

POWER RESULTS FOR ALL CELLS

Table F1. Power (%) for Complete Data and N = 100

			Power (%) for Complete Data and N = 100				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	1.90	1.90	1.90	1.90	1.90
		0.6	2.70	2.70	2.70	2.70	2.70
		0.8	2.50	2.50	2.50	2.50	2.50
	0.3	**0.4**	13.35	13.35	13.35	13.35	13.35
		0.6	14.35	14.35	14.35	14.35	14.35
		0.8	14.05	14.15	14.10	14.10	14.10
	0.5	**0.4**	18.15	18.15	18.15	18.15	18.15
		0.6	16.95	16.95	16.95	16.95	16.95
		0.8	17.95	17.95	17.95	17.95	17.95
0.3	**0.1**	**0.4**	13.40	13.40	13.40	13.40	13.40
		0.6	12.75	12.75	12.75	12.75	12.75
		0.8	14.45	14.45	14.45	14.45	14.45
	0.3	**0.4**	70.85	70.85	70.85	70.85	70.85
		0.6	70.40	70.40	70.40	70.35	70.35
		0.8	71.95	71.95	71.95	71.95	71.95
	0.5	**0.4**	86.40	86.40	86.40	86.40	86.40
		0.6	88.15	88.15	88.15	88.15	88.15
		0.8	86.05	86.05	86.05	86.05	86.05
0.5	**0.1**	**0.4**	14.80	14.85	14.85	14.80	14.80
		0.6	16.25	16.25	16.25	16.25	16.25
		0.8	15.65	15.70	15.70	15.70	15.70
	0.3	**0.4**	77.75	77.75	77.70	77.75	77.75
		0.6	77.05	77.05	77.05	77.05	77.05
		0.8	77.75	77.75	77.75	77.70	77.70
	0.5	**0.4**	99.70	99.70	99.70	99.70	99.70
		0.6	99.75	99.75	99.75	99.75	99.75
		0.8	99.50	99.50	99.50	99.50	99.50

Table F2. Power (%) for Complete Data and N = 200

			Power (%) for Complete Data and *N* = 200				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
α	***β***	***λ***					
0.1	**0.1**	**0.4**	6.95	6.95	6.95	6.95	6.95
		0.6	6.50	6.50	6.50	6.50	6.50
		0.8	6.30	6.30	6.30	6.30	6.30
	0.3	**0.4**	31.05	31.05	31.05	31.05	31.05
		0.6	29.60	29.60	29.60	29.60	29.60
		0.8	29.55	29.55	29.60	29.55	29.55
	0.5	**0.4**	29.55	29.55	29.55	29.55	29.55
		0.6	30.00	30.00	30.00	30.00	30.00
		0.8	28.15	28.15	28.15	28.15	28.15
0.3	**0.1**	**0.4**	27.10	27.10	27.10	27.10	27.10
		0.6	28.45	28.45	28.45	28.45	28.45
		0.8	25.10	25.10	25.10	25.10	25.10
	0.3	**0.4**	98.10	98.10	98.10	98.10	98.10
		0.6	97.80	97.80	97.80	97.80	97.80
		0.8	98.00	98.00	98.00	98.00	98.00
	0.5	**0.4**	99.65	99.65	99.65	99.65	99.65
		0.6	99.15	99.15	99.15	99.15	99.15
		0.8	99.45	99.45	99.45	99.45	99.45
0.5	**0.1**	**0.4**	24.95	24.95	24.95	24.85	24.85
		0.6	25.35	25.35	25.35	25.30	25.30
		0.8	22.80	22.80	22.80	22.80	22.80
	0.3	**0.4**	96.80	96.80	96.80	96.80	96.80
		0.6	95.85	95.85	95.85	95.85	95.85
		0.8	96.50	96.50	96.50	96.50	96.50
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F3. Power (%) for Complete Data and N = 500

			Power (%) for Complete Data and N = 500				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	33.80	33.80	33.75	33.75	33.75
		0.6	35.95	35.95	35.95	35.95	35.95
		0.8	33.45	33.45	33.45	33.45	33.45
	0.3	**0.4**	61.15	61.10	61.10	61.10	61.10
		0.6	61.95	61.95	61.95	61.95	61.95
		0.8	60.50	60.50	60.50	60.50	60.50
	0.5	**0.4**	60.50	60.50	60.50	60.50	60.50
		0.6	62.70	62.70	62.70	62.70	62.70
		0.8	59.20	59.20	59.20	59.20	59.20
0.3	**0.1**	**0.4**	58.00	58.00	58.00	58.00	58.00
		0.6	58.00	58.00	58.00	58.00	58.00
		0.8	57.35	57.35	57.35	57.35	57.35
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
0.5	**0.1**	**0.4**	50.40	50.45	50.45	50.35	50.35
		0.6	49.45	49.45	49.45	49.45	49.45
		0.8	51.25	51.25	51.20	51.20	51.20
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F4. Power (%) for Complete Data and N = 1000

			Power (%) for Complete Data and *N* = 1000				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
α	β	λ					
0.1	**0.1**	**0.4**	77.10	77.10	77.10	77.10	77.10
		0.6	77.00	77.00	77.00	77.00	77.00
		0.8	77.65	77.70	77.70	77.70	77.70
	0.3	**0.4**	88.85	88.85	88.85	88.85	88.85
		0.6	88.50	88.50	88.50	88.50	88.50
		0.8	89.05	89.05	89.05	89.05	89.05
	0.5	**0.4**	88.00	88.00	88.00	88.00	88.00
		0.6	89.00	89.00	89.00	89.00	89.00
		0.8	88.85	88.85	88.85	88.85	88.85
0.3	**0.1**	**0.4**	85.45	85.45	85.45	85.45	85.45
		0.6	86.85	86.85	86.85	86.85	86.85
		0.8	84.60	84.60	84.60	84.60	84.60
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
0.5	**0.1**	**0.4**	78.05	78.00	78.00	78.00	78.00
		0.6	78.10	78.10	78.10	78.05	78.05
		0.8	79.25	79.25	79.25	79.20	79.20
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F5. Power (%) for Missing Rate = 20% and N = 100

			Power (%) for Missing Rate = 20% and N = 100				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
α	β	λ					
0.1	**0.1**	**0.4**	2.15	2.00	1.75	1.75	1.65
		0.6	1.25	1.60	1.40	1.20	1.15
		0.8	1.90	1.95	1.90	1.50	1.35
	0.3	**0.4**	11.95	12.05	12.45	11.30	11.40
		0.6	12.20	12.60	11.95	11.95	11.55
		0.8	13.75	13.60	13.45	11.70	10.95
	0.5	**0.4**	16.20	15.85	16.00	15.75	14.80
		0.6	17.35	17.05	16.00	16.80	16.05
		0.8	17.40	17.60	16.65	14.80	14.45
0.3	**0.1**	**0.4**	12.25	12.40	11.45	11.35	11.25
		0.6	11.65	11.55	11.45	11.30	10.75
		0.8	13.55	13.60	12.55	10.50	10.10
	0.3	**0.4**	62.65	61.60	60.30	59.05	57.80
		0.6	64.90	63.55	61.65	57.60	56.35
		0.8	68.30	68.00	66.05	58.60	57.45
	0.5	**0.4**	83.65	82.95	82.80	81.90	80.05
		0.6	83.50	83.15	82.00	79.70	77.25
		0.8	85.65	84.55	83.45	80.10	77.70
0.5	**0.1**	**0.4**	14.80	14.15	14.50	14.90	14.45
		0.6	14.30	14.50	13.85	13.75	13.45
		0.8	15.40	15.05	14.25	14.35	13.90
	0.3	**0.4**	70.05	70.05	68.90	67.85	67.20
		0.6	72.75	72.45	71.50	68.90	68.20
		0.8	74.70	73.35	72.10	68.35	67.65
	0.5	**0.4**	99.20	99.00	98.90	98.80	98.65
		0.6	99.55	99.40	99.30	99.20	99.10
		0.8	99.30	99.25	99.35	98.50	98.55

Table F6. Power (%) for Missing Rate = 20% and N = 200

			Power (%) for Missing Rate = 20% and *N* = 200				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
α	β	λ					
0.1	**0.1**	**0.4**	5.30	4.80	4.35	4.20	4.10
		0.6	6.15	6.25	5.80	4.95	5.00
		0.8	6.80	6.70	6.65	5.80	5.65
	0.3	**0.4**	25.20	25.05	24.40	24.00	23.25
		0.6	26.10	25.40	25.25	23.35	22.30
		0.8	28.70	27.75	26.25	23.70	23.35
	0.5	**0.4**	26.75	27.75	27.55	27.10	26.35
		0.6	29.60	29.85	29.15	28.15	26.55
		0.8	28.00	27.80	26.75	25.55	24.80
0.3	**0.1**	**0.4**	22.45	23.35	22.80	21.90	21.70
		0.6	24.60	23.95	23.10	21.80	21.70
		0.8	26.05	25.60	24.30	22.20	22.15
	0.3	**0.4**	95.30	95.10	94.75	93.75	93.55
		0.6	96.55	96.75	95.55	94.55	94.30
		0.8	97.35	97.25	97.20	94.40	94.15
	0.5	**0.4**	98.20	98.40	98.25	97.75	97.70
		0.6	99.00	98.90	98.75	97.90	97.70
		0.8	98.80	98.75	98.40	97.80	97.50
0.5	**0.1**	**0.4**	21.55	21.70	22.20	21.50	21.35
		0.6	22.20	21.70	21.20	20.25	20.10
		0.8	22.50	22.10	21.35	18.95	18.90
	0.3	**0.4**	94.10	93.50	93.15	93.05	92.90
		0.6	95.75	95.80	94.75	93.85	93.65
		0.8	95.25	95.25	94.40	92.20	92.05
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F7. Power (%) for Missing Rate = 20% and N = 500

			Power (%) for Missing Rate = 20% and *N* = 500				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
α	β	λ					
0.1	**0.1**	**0.4**	24.60	23.40	23.00	21.95	21.80
		0.6	29.35	28.60	27.00	23.65	23.55
		0.8	31.35	30.90	29.60	23.50	23.55
	0.3	**0.4**	55.20	54.30	53.15	53.40	51.85
		0.6	56.65	55.50	54.90	52.05	51.60
		0.8	58.45	58.30	57.35	52.60	51.95
	0.5	**0.4**	56.45	55.95	55.75	55.30	52.15
		0.6	57.75	57.45	56.00	54.75	52.35
		0.8	59.50	59.50	57.90	53.30	52.70
0.3	**0.1**	**0.4**	50.70	49.55	47.90	47.70	47.60
		0.6	54.65	53.40	52.00	49.40	49.25
		0.8	54.00	53.20	52.15	47.20	47.15
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
0.5	**0.1**	**0.4**	44.90	44.60	43.50	43.00	42.95
		0.6	46.10	45.90	45.05	42.15	42.15
		0.8	45.25	45.20	43.70	41.30	41.15
	0.3	**0.4**	100.00	100.00	100.00	99.90	99.90
		0.6	100.00	100.00	100.00	99.95	99.95
		0.8	100.00	100.00	100.00	99.95	99.95
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F8. Power (%) for Missing Rate = 20% and *N* = 1000

			Power (%) for Missing Rate = 20% and *N* = 1000				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
α	***β***	***λ***					
0.1	**0.1**	**0.4**	68.40	67.75	65.40	63.65	63.25
		0.6	69.30	68.30	67.00	62.00	61.45
		0.8	73.95	73.70	72.45	62.75	62.65
	0.3	**0.4**	83.95	83.65	83.30	82.50	81.70
		0.6	86.20	85.50	84.55	82.15	81.45
		0.8	87.15	87.00	85.50	80.30	79.50
	0.5	**0.4**	83.35	83.00	82.40	82.45	79.90
		0.6	85.55	85.85	84.85	83.10	81.25
		0.8	89.90	89.45	88.15	85.10	83.55
0.3	**0.1**	**0.4**	80.45	79.60	78.35	77.40	77.30
		0.6	82.15	82.00	80.40	77.85	77.85
		0.8	83.30	83.00	81.50	77.50	77.45
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
0.5	**0.1**	**0.4**	73.45	73.20	72.05	71.35	71.35
		0.6	74.30	73.60	72.10	69.60	69.55
		0.8	76.15	76.05	74.75	69.25	69.25
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F9. Power (%) for Missing Rate = 50% and N = 100

			Power (%) for Missing Rate = 50% and *N* = 100				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	1.50	1.80	1.70	1.65	1.00
		0.6	1.75	1.60	1.75	1.75	1.20
		0.8	2.20	2.50	1.75	1.05	.75
	0.3	**0.4**	8.90	8.35	7.35	6.70	5.55
		0.6	9.50	8.40	8.15	6.60	5.80
		0.8	12.65	11.50	9.95	6.65	5.85
	0.5	**0.4**	14.30	14.05	14.20	13.05	12.55
		0.6	16.45	13.90	12.50	11.45	11.05
		0.8	16.70	15.85	15.50	13.35	12.15
0.3	**0.1**	**0.4**	8.60	7.95	7.25	7.30	6.45
		0.6	9.90	9.20	7.90	7.60	6.80
		0.8	12.40	11.55	10.35	7.90	6.40
	0.3	**0.4**	42.95	37.95	35.20	32.50	29.65
		0.6	52.60	45.85	41.10	33.35	30.50
		0.8	62.95	56.95	49.95	32.70	29.30
	0.5	**0.4**	71.95	67.80	66.45	63.80	58.10
		0.6	77.55	72.80	69.25	63.25	57.75
		0.8	81.35	79.00	75.10	62.30	56.30
0.5	**0.1**	**0.4**	12.95	13.05	12.10	12.05	11.05
		0.6	13.15	12.75	12.55	11.65	9.95
		0.8	14.20	14.00	13.30	11.25	9.95
	0.3	**0.4**	57.10	52.75	52.35	50.35	47.25
		0.6	64.40	59.90	56.00	50.80	48.00
		0.8	71.55	67.85	63.30	52.80	50.65
	0.5	**0.4**	94.50	93.65	92.10	90.35	88.50
		0.6	97.75	96.40	95.40	91.80	90.20
		0.8	99.35	98.60	97.10	91.05	88.75

Table F10. Power (%) for Missing Rate = 50% and N = 200

			Power (%) for Missing Rate = 50% and *N* = 200				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\lambda}$					
0.1	**0.1**	**0.4**	3.15	3.15	2.75	2.45	2.35
		0.6	4.50	4.45	3.75	3.10	2.75
		0.8	5.80	5.45	3.80	2.75	2.35
	0.3	**0.4**	19.60	18.40	17.85	16.60	15.70
		0.6	22.35	20.15	18.45	14.85	14.20
		0.8	26.70	24.90	22.75	15.15	14.80
	0.5	**0.4**	22.45	21.55	19.55	19.30	17.35
		0.6	25.00	23.35	20.70	19.45	17.80
		0.8	26.00	24.90	23.60	18.50	18.10
0.3	**0.1**	**0.4**	18.70	17.35	16.20	14.75	13.95
		0.6	19.70	18.00	15.95	13.70	12.80
		0.8	23.30	21.40	19.45	14.05	13.45
	0.3	**0.4**	84.50	80.35	77.20	73.55	70.95
		0.6	92.30	88.55	84.20	74.80	72.65
		0.8	96.20	93.20	89.80	75.10	73.05
	0.5	**0.4**	94.55	93.20	92.20	91.15	86.95
		0.6	97.80	96.40	94.70	91.75	88.25
		0.8	98.70	97.65	96.35	91.35	87.35
0.5	**0.1**	**0.4**	16.60	16.15	16.00	15.25	14.10
		0.6	18.75	18.10	17.70	15.40	14.55
		0.8	21.15	20.50	19.00	15.00	14.60
	0.3	**0.4**	84.15	81.15	79.95	78.15	77.40
		0.6	90.20	86.65	83.80	78.40	77.20
		0.8	94.60	92.20	89.30	78.90	77.65
	0.5	**0.4**	99.85	99.75	99.70	99.65	99.60
		0.6	100.00	100.00	100.00	99.80	99.75
		0.8	100.00	100.00	99.95	99.70	99.65

Table F11. Power (%) for Missing Rate = 50% and N = 500

			Power (%) for Missing Rate = 50% and *N* = 500				
α	β	λ	**Method 1** Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
0.1	**0.1**	**0.4**	14.85	12.25	10.70	10.30	9.60
		0.6	22.85	17.95	14.30	11.50	11.05
		0.8	26.90	23.10	19.55	11.10	10.75
	0.3	**0.4**	43.75	41.95	39.65	36.55	35.55
		0.6	50.60	46.35	43.55	36.85	35.40
		0.8	55.00	52.15	47.75	36.50	34.80
	0.5	**0.4**	45.25	43.25	41.95	39.25	35.45
		0.6	53.25	49.95	46.00	41.10	37.80
		0.8	57.30	53.35	49.45	40.00	35.20
0.3	**0.1**	**0.4**	39.75	36.60	35.70	33.65	33.25
		0.6	45.75	41.55	39.95	34.90	34.30
		0.8	52.30	48.30	45.05	34.50	33.80
	0.3	**0.4**	99.90	99.70	99.65	99.35	99.40
		0.6	100.00	99.90	99.80	99.45	99.50
		0.8	100.00	100.00	100.00	99.40	99.35
	0.5	**0.4**	100.00	100.00	99.95	99.95	99.90
		0.6	100.00	100.00	100.00	99.95	99.90
		0.8	100.00	100.00	100.00	99.95	99.85
0.5	**0.1**	**0.4**	32.15	29.90	27.30	26.20	25.90
		0.6	38.85	35.70	34.00	28.85	28.35
		0.8	44.30	42.00	37.65	28.40	27.95
	0.3	**0.4**	99.75	99.45	99.15	98.75	98.70
		0.6	100.00	99.90	99.70	99.10	99.05
		0.8	100.00	100.00	99.95	99.10	99.10
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F12. Power (%) for Missing Rate = 50% and N = 1000

			Power (%) for Missing Rate = 50% and *N* = 1000				
			Method 1 Latent Variable Model	**Method 2** Two Auxiliary Variables	**Method 3** One Auxiliary Variable	**Method 4** ML with M_1 only	**Method 5** Listwise Deletion
α	***β***	***λ***					
0.1	**0.1**	**0.4**	47.95	41.85	37.55	33.00	32.65
		0.6	60.35	52.60	44.65	33.70	33.45
		0.8	70.95	65.95	58.75	36.00	35.45
	0.3	**0.4**	71.10	67.95	64.85	62.35	60.60
		0.6	78.30	74.35	69.70	62.25	60.75
		0.8	85.65	82.85	79.25	64.25	62.05
	0.5	**0.4**	73.65	70.25	68.20	65.80	61.25
		0.6	80.75	77.00	72.80	67.40	62.15
		0.8	84.90	81.85	77.40	66.20	61.10
0.3	**0.1**	**0.4**	66.75	62.40	60.10	57.00	56.60
		0.6	74.20	70.40	65.05	58.45	58.15
		0.8	80.05	75.95	71.75	56.30	56.10
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
0.5	**0.1**	**0.4**	57.35	53.15	51.45	49.10	48.85
		0.6	63.85	59.60	55.80	49.20	48.95
		0.8	73.00	68.00	62.50	48.20	47.65
	0.3	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00

Table F13. Power (%) for Missing Rate = 80% and N = 100

			Power (%) for Missing Rate = 80% and *N* = 100				
α	***β***	***λ***	1. Latent Variable Model	2. Two Auxiliary Variable Model	3. One Auxiliary Variable Model	4. ML with No Additional Mediators	5. Listwise Deletion
0.1	**0.1**	**0.4**	1.05	2.25	2.10	1.90	.75
		0.6	1.05	1.80	1.85	1.85	.95
		0.8	1.30	2.05	1.50	2.30	.80
	0.3	**0.4**	4.65	4.80	5.00	4.50	3.05
		0.6	7.30	6.25	5.55	4.10	1.95
		0.8	9.95	7.35	6.30	3.95	2.50
	0.5	**0.4**	9.80	9.45	9.40	8.35	5.50
		0.6	11.95	9.45	9.15	7.60	5.55
		0.8	14.90	13.25	11.90	8.20	5.65
0.3	**0.1**	**0.4**	4.90	5.85	5.45	5.35	3.15
		0.6	6.80	6.70	6.15	4.55	2.05
		0.8	9.55	7.05	5.80	4.60	2.00
	0.3	**0.4**	19.15	14.35	13.80	12.00	8.00
		0.6	37.65	19.55	16.90	12.15	8.15
		0.8	56.60	31.60	23.40	13.25	8.45
	0.5	**0.4**	45.85	31.05	28.80	26.35	19.30
		0.6	64.30	40.60	34.90	27.95	18.75
		0.8	79.55	56.50	45.35	25.60	18.05
0.5	**0.1**	**0.4**	9.00	12.00	11.15	10.20	6.25
		0.6	9.70	12.55	11.40	10.65	7.25
		0.8	12.95	14.20	13.40	11.35	7.25
	0.3	**0.4**	34.50	28.15	26.40	24.90	17.90
		0.6	48.85	31.10	27.00	22.95	15.80
		0.8	64.45	44.10	35.95	23.10	16.30
	0.5	**0.4**	75.40	57.15	54.65	52.95	40.95
		0.6	91.65	65.55	61.00	53.05	40.95
		0.8	98.25	84.05	74.50	53.50	40.90

Table F14. Power (%) for Missing Rate = 80% and *N* = 200

			Power (%) for Missing Rate = 80% and *N* = 200				
α	***β***	***λ***	1. Latent Variable Model	2. Two Auxiliary Variable Model	3. One Auxiliary Variable Model	4. ML with No Additional Mediators	5. Listwise Deletion
0.1	**0.1**	**0.4**	1.30	1.95	1.75	1.80	1.10
		0.6	2.90	1.75	1.50	.95	.60
		0.8	4.85	2.75	2.30	1.55	1.00
	0.3	**0.4**	11.60	8.45	7.65	7.40	5.80
		0.6	17.10	9.90	7.60	7.15	4.85
		0.8	22.20	15.00	9.90	5.10	3.95
	0.5	**0.4**	15.00	12.85	12.45	11.50	10.05
		0.6	20.45	15.15	13.90	11.50	9.80
		0.8	24.35	20.85	17.25	12.00	9.55
0.3	**0.1**	**0.4**	10.05	8.55	7.60	6.65	4.30
		0.6	15.05	10.65	8.65	6.95	5.30
		0.8	22.60	14.95	11.85	6.70	4.65
	0.3	**0.4**	57.35	35.80	32.00	28.05	21.05
		0.6	81.35	46.10	37.20	27.95	21.75
		0.8	94.10	72.65	59.80	26.50	21.65
	0.5	**0.4**	81.80	65.45	61.00	57.90	48.35
		0.6	92.05	75.15	67.95	55.70	46.40
		0.8	97.75	88.45	80.05	53.60	44.85
0.5	**0.1**	**0.4**	13.10	13.00	12.15	11.40	8.60
		0.6	15.75	13.40	12.35	11.80	8.75
		0.8	19.90	15.80	14.25	11.50	9.00
	0.3	**0.4**	64.10	49.40	47.35	43.45	36.90
		0.6	81.85	57.10	52.30	45.55	38.80
		0.8	91.35	74.60	62.95	45.60	38.75
	0.5	**0.4**	98.10	90.25	87.55	85.35	80.70
		0.6	99.75	94.60	91.35	83.90	79.20
		0.8	100.00	99.00	96.85	84.30	78.40

Table F15. Power (%) for Missing Rate = 50% and *N* = 500

			Power (%) for Missing Rate = 80% and *N* = 500				
α	***β***	***λ***	1. Latent Variable Model	2. Two Auxiliary Variable Model	3. One Auxiliary Variable Model	4. ML with No Additional Mediators	5. Listwise Deletion
0.1	**0.1**	**0.4**	8.20	3.95	3.45	2.95	2.35
		0.6	15.65	6.15	4.55	3.00	2.30
		0.8	24.90	10.95	7.70	2.85	2.60
	0.3	**0.4**	34.15	20.50	18.70	16.85	15.30
		0.6	41.30	26.55	21.40	14.80	13.55
		0.8	52.85	38.25	32.10	15.85	14.30
	0.5	**0.4**	32.70	26.35	24.00	22.40	17.80
		0.6	40.90	28.90	25.15	19.55	18.55
		0.8	54.15	40.05	32.20	21.05	18.60
0.3	**0.1**	**0.4**	28.45	21.00	19.15	16.70	14.75
		0.6	37.65	24.60	20.55	15.40	13.55
		0.8	47.20	32.05	26.55	14.30	13.10
	0.3	**0.4**	98.20	85.40	80.70	75.95	72.90
		0.6	99.95	94.30	88.90	75.55	71.40
		0.8	100.00	99.25	97.80	75.45	72.30
	0.5	**0.4**	99.30	96.00	95.10	93.20	87.10
		0.6	100.00	98.55	96.70	93.10	87.85
		0.8	100.00	99.85	99.40	92.65	88.50
0.5	**0.1**	**0.4**	23.00	18.60	17.60	16.25	14.70
		0.6	31.55	20.75	17.95	15.80	14.20
		0.8	41.00	29.30	22.70	16.35	15.10
	0.3	**0.4**	95.55	83.80	81.30	78.20	76.35
		0.6	99.65	92.50	87.70	79.05	77.10
		0.8	99.90	98.10	95.50	79.75	78.05
	0.5	**0.4**	100.00	100.00	99.95	100.00	99.85
		0.6	100.00	100.00	100.00	99.70	99.65
		0.8	100.00	100.00	100.00	99.85	99.80

Table F16. Power (%) for Missing Rate = 80% and *N* = 1000

			Power (%) for Missing Rate = 80% and *N* = 1000				
α	***β***	***λ***	1. Latent Variable Model	2. Two Auxiliary Variable Model	3. One Auxiliary Variable Model	4. ML with No Additional Mediators	5. Listwise Deletion
0.1	**0.1**	**0.4**	24.25	10.55	9.10	5.85	5.35
		0.6	45.10	19.65	13.45	7.55	7.10
		0.8	64.70	39.10	25.40	8.35	7.65
	0.3	**0.4**	52.95	38.20	35.05	32.05	29.25
		0.6	72.05	47.75	40.65	31.10	28.75
		0.8	80.95	62.35	54.05	31.40	29.15
	0.5	**0.4**	54.90	40.95	37.15	33.90	28.95
		0.6	72.65	49.90	44.10	36.75	30.50
		0.8	82.55	66.35	54.90	35.95	28.70
0.3	**0.1**	**0.4**	50.05	33.45	31.25	28.40	27.55
		0.6	64.15	38.45	33.90	26.05	24.70
		0.8	77.95	57.10	48.15	27.00	26.30
	0.3	**0.4**	100.00	99.20	98.75	97.70	97.60
		0.6	100.00	99.90	99.80	98.35	98.05
		0.8	100.00	100.00	100.00	98.60	98.30
	0.5	**0.4**	100.00	99.95	99.85	99.70	98.95
		0.6	100.00	100.00	99.95	99.70	99.50
		0.8	100.00	100.00	100.00	99.90	99.60
0.5	**0.1**	**0.4**	38.15	27.95	26.70	24.05	23.45
		0.6	54.35	33.65	28.60	24.70	23.85
		0.8	68.45	48.65	39.20	25.00	23.80
	0.3	**0.4**	99.80	98.30	97.35	96.45	96.20
		0.6	100.00	99.50	98.95	96.90	96.50
		0.8	100.00	100.00	99.65	96.65	96.55
	0.5	**0.4**	100.00	100.00	100.00	100.00	100.00
		0.6	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00