Published in final edited form as:

Struct Equ Modeling. 2019 ; 26(4): 636–645. doi:10.1080/10705511.2018.1540934.

Empirical Sample Size Guidelines for Use of Latent Difference Score Mediation

Melissa Simone, Ph.D.^{a,b} and Ginger Lockhart, Ph.D.^{a,c}

^aUtah State University, Logan, Utah; Department of Psychology, 2810 Old Main Hill, Logan, UT, 84322-2810, USA.

^bDepartment of Psychiatry, 2450 Riverside Avenue, F227, Minneapolis, MN, 55454, USA. simon996@umn.edu; Phone: +1 (612) 273-0912; Fax: +1 (612) 273-9779

^cginger.lockhart@usu.edu

Abstract

Mediation models are commonly used to identify the mechanisms through which one variable influences another. Among longitudinal mediation methods, latent difference score mediation stands out due to its unique ability to capture non-linear change over time. However, there is limited information regarding sample size demands to achieve adequate power with this method, resulting in few applications of latent difference score mediation. To address this limitation, the current study presents empirically supported sample size guidelines for 10 common latent difference score mediation structural models and 9 unique population models. The results of this study offer researchers with a set of representative sample estimates that may be used when designing studies or seeking funding.

Keywords

Longitudinal Mediation; Monte Carlo; Latent Difference Score; Sample Requirements

Mediation analysis is a powerful statistical tool that allows researchers to answer substantive questions about *how* one variable influences another (MacKinnon, 2008; MacKinnon, Krull & Lockwood, 2000; MacKinnon & Dwyer, 1993). Mediation analyses examine two main pathways (see Figure 1): the direct effect of a predictor (X) on an outcome (Y), and the indirect effect of X on Y through a mediator (M) ($X \rightarrow M$ and $M \rightarrow Y$; MacKinnon et al., 2000). For example, family environment may mediate the relationship between disordered eating and alcohol misuse among girls. In this example, girls who engage in disordered eating behaviors may experience increased parent-child conflict in response to their eating patterns, which in turn may result in alcohol misuse.

Address correspondence to: Melissa Simone, Ph.D., Department of Psychiatry, 2450 Riverside Avenue, F227, Minneapolis, MN, 55454, USA. simon996@umn.edu; Phone: +1 (612) 273-0912; Fax: +1 (612) 273-9779.

Declarations: An earlier version of this manuscript was submitted in partial fulfillment of the doctoral degree requirements at Utah State University and is available on ProQuest.

Traditional mediation methods examine all tested effects simultaneously, which prevents researchers from examining how processes unfold over time. Yet, often times researchers are interested in understanding the temporal order of events (MacKinnon & Fairchild, 2009). Recent advances in the field of quantitative methodology offer a range of longitudinal mediational tools that can better describe these changes as they occur over time (Selig & Preacher, 2009). Longitudinal mediation is superior to traditional mediation methods as longitudinal analyses allow researchers to examine how developmental processes unfold over time (Selig & Preacher, 2009). The ability to model change as it occurs is important because it allows researchers to evaluate whether an effect, or outcome, is dependent upon the amount of time elapsed (Selig & Preacher, 2009), and allows researchers to describe the temporal order of effects (MacKinnon & Fairchild, 2009).

Among existing longitudinal mediation methods, latent difference score mediation stand out as one of the more flexible methods, as it does not assume systematic change across measurement occasions. Specifically, latent difference score mediation represents change at the variable level as the difference between adjacent observations, which allows researchers to explain the temporal order of modeled variables and relaxes the assumption of systematic change over time (Selig & Preacher, 2009; McArdle, 2009; MacKinnon, 2008; McArdle & Hamagami, 2001). Thus, latent difference score mediation models are useful when change is expected to vary across measurement occasions (MacKinnon, 2008). Moreover, latent difference score mediation directly models both within and between individual variability (Selig & Preacher, 2009), thus offering greater utility to answer questions related to development over time.

When planning to use mediational methods, researchers are often interested in how many participants are needed to achieve adequate statistical power (Fritz & MacKinnon, 2007). Yet, latent difference score mediation methods lack empirically supported sample size guidelines, and sample requirements have been shown to vary greatly across models (Muthén & Muthén, 2002). Thus the current study aimed to address this gap in the literature by developing a set of empirically supported sample size guidelines for a wide range of latent difference score mediation structural and population models by conducting a Monte Carlo simulation study.

Latent Difference Score Mediation

Latent difference score mediation models examine change over time among predictors, mediators, and/or outcomes. In these models, change is represented as the difference in scores in adjacent observation periods (Selig & Preacher, 2009). The adjacent observations that make up latent difference scores are included in the model as lower-level latent factors. For example, when a predictor (X) is modeled as a difference variable (X_{1-2}), X_1 and X_2 are modeled as unobserved lower-level latent factors with multiple indicators (see Figure 2 for an example). Lower-level latent factors are then used to form the latent difference score variables. The latent difference scores represent the change between adjacent measurement occasions as well as the effect of the variable at the first time point on the change between occasions. See Figure 3 for an example.

In a latent difference score mediation model for which the predictor (X), mediator (M), and outcome (Y) variables are each represented by a difference score (see Figure 3), changes in X from time 1 to time 2 (ΔX_{1-2}) are modeled to predict changes in the mediator from time 2 to time 3 (ΔM_{2-3}), which in turn predict changes in the outcome from time 3 to time 4 (ΔY_{3-4}). In such model, X at time 2 (X_2) is defined by Equation 1, where $\beta_{a1}X_1$ is the effect of X at time 1 (X_1) on X_2 , $\beta_{a2}\Delta X_{1-2}$ is the effect of ΔX_{1-2} on X_2 , and e_{x2} is the residual term. Both $\beta_{a1}X_1$ and $\beta_{a2}\Delta X_{1-2}$ are fixed to 1, which is reflected in Figure 3 (Selig & Preacher, 2009):

$$X_2 = \beta_{a1}X_1 + \beta_{a2}\Delta X_{1-2} + e_{x2} \quad (1)$$

The change in X from time 1 to time 2, ΔX_{1-2} , is defined by Equation 2. In Equation 2, $\beta_{x1}X_1$ reflects the effect of X_1 on ΔX_{1-2} (Selig & Preacher, 2009) and e_{x12} is the residual term:

$$\Delta X_{1-2} = \beta_{x1}X_1 + e_{x12} \quad (2)$$

Further, M_2 is defined by Equation 3, where $\alpha_1\Delta X_{1-2}$ accounts for the effect of ΔX_{1-2} on M_2 , α_2X_1 accounts for the effect of X_1 on M_2 , and e_{m2} is the residual term.

$$M_2 = \alpha_1\Delta X_{1-2} + \alpha_2X_1 + e_{m2} \quad (3)$$

Similarly, M_3 is defined in Equation 4, where $\beta_{a3}M_2$ accounts for the effect of M_2 on M_3 , $\beta_{a4}\Delta M_{2-3}$ accounts for the effect of ΔM_{2-3} on M_3 and e_{m3} is the residual term. Similar to Equation 1, both $\beta_{a3}M_2$ and $\beta_{a4}\Delta M_{2-3}$ are fixed to 1.

$$M_3 = \beta_{a3}M_2 + \beta_{a4}\Delta M_{2-3} + e_{m3} \quad (4)$$

Where ΔM_{2-3} is defined in Equation 5. In Equation 5, $\alpha_3\Delta X_{1-2}$ accounts for the effect of ΔX_{1-2} on ΔM_{2-3} , α_4X_1 accounts for the effect of X_1 on ΔM_{2-3} , $\beta_{m1}M_2$ accounts for the effect of M_2 on ΔM_{2-3} , and e_{m23} is the residual term.

$$\Delta M_{2-3} = \alpha_3\Delta X_{1-2} + \alpha_4X_1 + \beta_{m1}M_2 + e_{m23} \quad (5)$$

Further, Y at time 3 (Y_3) is defined by Equation 6, where $c'_{1}\Delta X_{1-2}$ is the effect of ΔX_{1-2} on Y_3 after controlling for the mediator, $c'_{2}X_1$ is the effect of X_1 on Y_3 after controlling for the mediator, $b_1\Delta M_{2-3}$ is the effect of ΔM_{2-3} on Y_3 , b_2M_2 is the effect of M_2 on Y_3 and e_{y3} is the residual term:

$$Y_3 = c'_1 \Delta X_{1-2} + c'_2 X_1 + b_1 \Delta M_{2-3} + b_2 M_2 + e_{y3} \quad (6)$$

Finally, Y_4 is defined in Equation 7, where $\beta_{a5} Y_3$ accounts for the effect of Y_3 on Y_4 , $\beta_{a6} Y_{3-4}$ accounts for the effect of Y_{3-4} on Y_4 and e_{y4} is the residual term. Similar to Equations 1 and 2, both $\beta_{a5} Y_3$ and $\beta_{a6} Y_{3-4}$ are fixed to 1.

$$Y_4 = \beta_{a5} Y_3 + \beta_{a6} \Delta Y_{3-4} + e_{y4} \quad (7)$$

Where the change in Y from time 3 to time 4 (Y_{3-4}) is defined by Equation 8. In Equation 8, $c'_3 X_{1-2}$ is the effect of X_{1-2} on Y_{3-4} , $c'_4 X_1$ is the effect of X_1 on Y_{3-4} , $b_3 M_{2-3}$ is the effect of M_{2-3} on Y_{3-4} , $b_4 M_2$ is the effect of M_2 on Y_{3-4} , $\beta_{y1} Y_3$ is the effect of Y_3 on Y_{3-4} and e_{y34} is the residual term:

$$\Delta Y_{3-4} = c'_3 \Delta X_{1-2} + c'_4 X_1 + b_3 \Delta M_{2-3} + b_4 M_2 + \beta_{y1} Y_3 + e_{y34} \quad (8)$$

Latent difference score mediation models examine change in model predictors, mediators, and outcomes dynamically over time, without assuming systematic change within or across measurement occasions. However, there are several important considerations to address in order to capture the mediated effect of interest.

Considerations with Latent Difference Score Mediation

Due to the complex nature of latent difference score mediation models, a wide range of factors such as the role of time, span of the study, the developmental period of interest, the model assumptions and distribution qualities must be considered. Specifically, the role of time is a critical component in latent difference score mediation because researchers need to collect data at the appropriate measurement occasions to capture the mediated effect over the course of the study (Selig & Preacher, 2009). For example, a researcher interested in detecting the mediated effect of adolescent social environment on the relationship between late childhood eating patterns and adult alcohol misuse, must carefully select the amount of time between measurement occasions in order to make sure enough time elapses for the expected effects to take place. That is to say, while the value of disordered eating at the first measurement occasion may predict the family environment scores at the second measurement occasion, latent difference score mediation captures whether the change in a predictor across measurement occasions predicts change in a mediator. Thus, failure to measure study variables at the appropriate developmental period, or with insufficient time between measurement occasions, may result in insignificant findings, or change, across occasions. Therefore, when considering the role of time, the choice of developmental period, space between data collection, and the span of a study must be determined on a case-by-case basis (Selig & Preacher, 2009), as differences in these components directly impact the interpretation of the change in each construct.

Mediation analyses contain several assumptions (MacKinnon, 2008), which should be checked prior to examining any data. Specifically, mediation models assume that there is no interaction between the X and M variables, and that the causal ordering and direction of effects are accurate (e.g. $X \rightarrow M \rightarrow Y$, not $Y \rightarrow M \rightarrow X$; Lockhart, MacKinnon & Ulrich, 2011; MacKinnon, Fairchild & Fritz, 2007). Further, it is assumed that the analyses use valid and reliable measures (Lockhart et al., 2011), that there is no misspecification due to unmeasured variables (Holland, 1988; James & Brett, 1984), and that residual variables are not correlated with the mediated effect (McDonald, 1997).

Mediation analysis involves obtaining at least one product between two coefficients and the underlying distribution of this product is asymmetrical, which violates the standard regression assumption of normality (MacKinnon, Lockwood & Williams, 2004). Thus, resampling methods (e.g. Jackknife, Percentile Bootstrap and Bias-corrected Bootstrap; MacKinnon et al., 2004; Efron & Tibshirani, 1993; Mosteller & Tukey, 1977) are typically used to adjust for bias. Resampling techniques generate a variety of datasets by resampling observations from the original dataset to test distributional hypotheses and generate confidence intervals (Williams & MacKinnon, 2008). Of current resampling techniques, the bias-corrected bootstrap method provides the most accurate confidence intervals and more power than normal theory and non-corrected methods (MacKinnon et al., 2004).

Types of Indirect Effects

In order to capture change in X , M and Y , the most complex version of the latent difference score mediation model with a single mediator includes four measurement occasions ($X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$). In this model, there are several different indirect effects that may be calculated. For example, the indirect effect of X_{1-2} on M_{2-3} and M_{2-3} on Y_{3-4} may be examined, as can of X_1 on M_{2-3} and M_{2-3} on Y_3 , or of X_{1-2} on M_2 and M_2 on Y_{2-3} . Because each indirect effect includes different forms of the predictor, mediator and outcome variables, resulting in different substantive meanings, the calculation of a total indirect effect is not recommended (Selig & Preacher, 2009). For example, while it is possible to calculate the total indirect effect of $(a_3 X_{1-2})(b_3 M_{2-3})$ and $(a_2 X_1)(b_2 M_2)$, the total indirect effect would be uninterpretable. Specifically, $(a_3 X_{1-2})(b_3 M_{2-3})$ explains the total indirect effect of the change in X from time 1 to time 2 on the change in Y from time 3 to time 4, as mediated by the change in M from time 2 to time 3. Here, the indirect effect explains a mediated effect by examining changes between measurements. Whereas $(a_2 X_1)(b_2 M_2)$ explains the extent to which the status of X at time 1 influences the status of Y at time 3, as mediated by the status of M at time 2. In this example, the indirect effect explains the mediated effect by examining individual differences on each construct, rather than changes across measurements. Thus, the meaning of a total indirect effect that combines static effects ($X_1 \rightarrow M_2 \rightarrow Y_3$) with changes across measurement occasions ($X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$) cannot be interpreted.

Utility of Power Analysis

Though the ideal latent difference score mediation structural model includes four measurement occasions, other forms of the model may be used to examine change over time.

Specifically, it is not required that all variables included in the study be represented by a latent difference score. For example, a randomized controlled trial with 3 measurement occasions and a dichotomous X variable (e.g. type of treatment) a research may use a model where $X_d \rightarrow M_{2-3} \rightarrow Y_{3-4}$. To date, no known power analyses have been conducted to determine the sample size requirements for the complete latent difference score mediation model ($X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$) or its variants. Thus, the current study seeks to conduct a comprehensive Monte Carlo simulation study to detect the required sample size for a wide range of latent difference score mediation structural models, based on several commonly used study designs.

Monte Carlo simulations offer a powerful tool to evaluate the performance of statistical models under various conditions (Muthén & Muthén, 2002). Monte Carlo studies offer an empirical method for evaluating statistical methods and power (Paxton, Curran, Bollen, Kirby & Chen, 2001) and are often used to determine the required sample size to detect effects within a wide range of statistical models. In Monte Carlo studies, data are simulated from a predetermined population model with known population values (Muthén & Muthén, 2002; Paxton et al., 2001). Population values are determined either through a review of the literature that has applied the statistical model of interest, or through covariance algebra calculations based on effect size selections (MacKinnon, 2008; Muthén & Muthén, 2002). A large number of samples are then drawn from the population model, for which all model parameters are estimated. Parameter values and standard errors reflect the average value over the samples drawn from the population model (Muthén & Muthén, 2002). These estimates can then be compared to the predefined population model to evaluate statistical power, bias, and coverage.

In order to determine the required sample size for any given effect, a model must meet four criteria: (1) parameter and standard error biases do not exceed 10% for any parameter in the model; (2) the standard error bias for the parameter being examined, in this case the mediated path, does not exceed 5%; (3) coverage remains between .91 - .98; and (4) statistical power is close to .80 (Muthén & Muthén, 2002). The method to calculate parameter and standard error bias are displayed in Equations 9 and 10, respectively. Parameter bias captures how close the average parameter estimate across all replications for any given parameter is to the specified population parameter. Similarly, standard error bias captures how close the average standard error estimate across all replications for any given parameter is to the population standard error. The average standard deviation of each parameter estimate over a large number of replications in a Monte Carlo study is considered to be the population standard error (Muthén & Muthén, 2002), and is provided in the simulation output when conducted with the Mplus statistical software (Muthén & Muthén, 1998–2011).

$$\frac{Parameter_{avg} - Parameter_{pop}}{Parameter_{pop}} \quad (9)$$

$$\frac{\text{Standard Error}_{avg} - \text{Standard Error}_{pop}}{\text{Standard Error}_{pop}} \quad (10)$$

Coverage accounts for the proportion of replications for which the 95% confidence interval contains the true parameter value. Finally, statistical power is defined as the probability of accurately rejecting or accepting the null hypothesis (Wilson VanVoorhis & Morgan, 2007; Sedlmaier & Gigerenzer, 1989). A power of .80 has been widely accepted as adequate (Murphy, Myors & Wolach, 2014; Muthén & Muthén, 2002), meaning that rejecting the null is four times as likely as failure (Murphy et al., 2014).

The required sample size to detect any given effect varies greatly by the specific model of interest and expected effect size (Murphy et al., 2014; Muthén & Muthén, 2002). Monte Carlo power analyses identify the required sample size for specific population models, which researchers can then refer to in order to ensure that they collect a large enough sample to detect their hypothesized effects (Thoemmes, MacKinnon & Reiser, 2010; Muthén & Muthén, 2004; Paxton et al., 2001).

Purpose

Latent difference score mediation allows researchers to examine change over time, variability in change within and between individuals, and offers flexibility in how change is modeled over time. The lack of empirically supported sample size requirements for latent difference score mediation has resulted in few applications of the model. To address this gap in the literature, the current study includes a Monte Carlo simulation study to develop an empirically supported set of sample size guidelines for ten possible latent difference score mediation structural models with nine unique population models. The structural and population models were selected to offer a generalizable set of sample size requirements.

Statistical Analysis

A Monte Carlo simulation was conducted to examine the required sample size to detect mediation among several variations of the latent difference score mediation model using Mplus statistical software (Muthén & Muthén, 1998–2011). More specifically, ten unique structural models were evaluated and nine unique population models, resulting in 90 unique models. The models were selected to offer a generalizable set of sample size guidelines that can be applied to answer a wide range of research questions. Sample Mplus input is provided as an Appendix.

Tested Models Included in the Monte Carlo Simulation

Structural models.—Each structural model included in the Monte Carlo simulation was selected to reflect a possible single mediator latent difference score mediation model. To this end the structural models varied by the structure of the X , M , and Y variables across each model. Specifically, X was modeled as a dichotomous variable (X_{d1}), a latent variable (X_1), or a latent difference score (X_{1-2}), whereas the M and Y variables were modeled as either

a latent variable (M_2 or Y_3 , respectively) or as latent difference scores (M_{2-3} or Y_{3-4} , respectively) across structural models. All structural models included in the simulation are presented in Table 1.

Population models.—Population models were selected to offer a set of generalizable guidelines across a wide range of possible effect sizes, however the limited number of applications of latent difference score mediation offer limited information regarding expected effects, thus population models were set to vary by the magnitude of the effect of X on M (α) and the effect of M on Y (b). Specifically, the α and b paths were set to have small, medium, or large effect sizes, resulting in nine possible effect size pairings. Effect size estimates used to develop the population model parameters followed Cohen's (1988) guidelines for R^2 estimates, which describes the amount of explained variance in the outcome. Cohen (1988) defined R^2 effect sizes as small (0.02), medium (0.13), and large (0.26) effects. The R^2 effect size statistic was selected to provide empirical sample estimates that compare to those from other Monte Carlo studies examining power among complex mediation models (e.g., Thoemmes, MacKinnon, & Reiser, 2010; Ma & Zeng, 2014; Fritz & MacKinnon, 2007). Corresponding unstandardized path coefficients were calculated using covariance algebra from a set of predefined effect size estimates.

The value of the unstandardized α paths of interest were derived from the following equation:

$$\text{Var}(M) = \alpha^2 \text{Var}(X) + \text{Var}(e_m)$$

Consider a model with a dichotomous X and a small effect size for the α path ($R^2 = .02$) where the variance of M [$\text{Var}(M)$] was fixed to 1 and the residual term [$\text{Var}(e_m)$] was calculated as $1 - R^2$. Because X is modeled as a dichotomous variable, the variance of X [$\text{Var}(X)$] is set to .25, which reflects an even split of participants across the conditions of dichotomous X variable. When these values are plugged into the equation, the resulting unstandardized α path estimate is .28:

$$1 = \alpha^2(.25) + (1 - .02)$$

$$\frac{1 - .98}{.25} = \alpha^2 = .08$$

$$\alpha = \sqrt{.08} = .28$$

In this same example, where the magnitude of the b path was also set to reflect a small R^2 effect size (.02) and the total variance of Y was set to 1, the unstandardized b path value was solved with the following equation:

$$b^2 = .02 \text{ and therefore } b = \sqrt{.02} = .14$$

Further, the c' path can be solved for with the following equation:

$$c' = \sqrt{\frac{R^2}{\text{Var}(X)}}$$

Thus, in the running example, c' is solved as:

$$c' = \sqrt{\frac{.02}{.25}} = .28$$

Complete details regarding the covariance algebra used to calculate these values have been outlined in MacKinnon (2008).

Unstandardized path coefficients for each model are presented in Table 1. Specifically, unstandardized path coefficients of .28, .72, and 1.02 were used to represent small, medium, and large effect sizes for the a path when X was dichotomous. Whereas, path values of .14, .36, and .51 represent small, medium, and large effect sizes, respectively, for the a and b paths when X was a latent or latent difference score variable. The c' pathway, or the path representing the direct effect of X on Y while controlling for M , was fixed to a small effect size, resulting in a path coefficient of .14 when X was a latent or latent difference score, and .28 when dichotomous.

Parameter and Model Specifications

Consistent with other Monte Carlo studies examining the required sample size to detect mediation among complex mediation models (e.g., Thoemmes et al., 2010), all continuous variables were generated to have a mean of 0 and a variance of 1. Distributions were assumed to be normal, and the data were generated with no missing data. Factor loadings were set to 0.90 to retain strong loadings without assuming perfect reliability (e.g., by setting all loadings equal to 1). Residual variances for latent variables were fixed so that the total variance equaled 1. Each of the latent variables were identified by three observed variable indicators. For simplicity purposes, the residual variances for the observed indicators were set as equal within their respective factor. Latent variables are comprised of two components: the true score and error. Because the indicators were not assumed to have perfect reliability, the residual variances were fixed to a non-zero value (Brown, 2006), and thus, were calculated with the following equation, where λ represents the factor loading and e represents the indicator residual variance:

$$1 = (\lambda)^2 + e$$

The equation to calculate the observed residual variance values was solved as follows:

$$1 = (.90)^2 + e = (.81) + e = 1 - (.81) = .19$$

Moreover, preliminary analyses were conducted to determine the most appropriate beta value for a latent X , M , and Y variable onto its respective difference score variable. These preliminary analyses revealed an average unstandardized path value of $-.30$, and thus when applicable, the X , M , and Y loadings onto their respective difference score variables were fixed to $-.30$.

Given the complex nature of the indirect effects tested within the full latent difference score mediation model, only the indirect path of interest for each model was tested. For example, in the model $X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$ only the $X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$ indirect path was tested. Each model was tested with 1,000 Monte Carlo replications with a random seed value of 11787 across models. Bias-corrected bootstrap tests were used to adjust for the asymmetrical distribution inherent in mediation analysis, as this method provides the most accurate confidence intervals of current resampling techniques (MacKinnon et al., 2004). Each model included 500 bootstrap draws.

Model Selection

The primary goal of this study was to generate a set of guidelines for the required sample size to detect mediation among several latent difference score mediation model variants. Muthén and Muthén's (2002) four criteria for sample size selection as previously described were followed to determine the appropriate sample estimate for each model included in the simulation. Thus, the final sample size estimates represent models with adequate statistical power and coverage, as well as minimal parameter and standard error bias.

Sample sizes were adjusted in an iterative process, accordingly, until the empirical power estimate is within 0.001 of 0.80, coverage ranged between .91 - .98, and both parameter and standard error bias were at an acceptable level (e.g., less than 10%). For example, if power was greater than 0.80, the sample size was decreased, and if the power estimate was smaller than 0.80, the sample size was increased (Fritz & MacKinnon, 2007). This iterative process was conducted until the smallest sample size for which all criteria were met for each of the tested models was identified.

Results

The final empirical sample estimates for each structural and population model are shown in Table 2. Empirical sample estimates ranged from 270 – 780, with sample requirements varying greatly across each structural model and population models. In general, the complexity of each structural model had a large impact on the empirical sample size estimates. To this end, more simplistic structural models required smaller samples than more complex structures. For example, structural models in which X , M , or Y variables were modeled as latent variables required fewer participants than models in which X , M , or Y variables were modeled as latent difference score variables. Similarly, structural models in

which X was modeled as a dichotomous variable required smaller samples than models in which X was modeled as a continuous or latent difference score variable.

Further, population models with larger effect sizes tended to require smaller sample sizes than those with smaller effect size pairings. Moreover, the magnitude of effect for the b path tended to influence the sample size estimates more strongly than the effect size of the a path. Specifically, population models in which the b path was specified as small required larger samples than any other models. Sample size requirements were highest among the population model with a LS effect size pairing ranging from 600 – 780.

Discussion

The purpose of this paper was to develop a set of empirically supported sample size guidelines for ten unique latent difference score mediation structural models and nine population models. There are three key findings from the current Monte Carlo analysis: (1) simple structural models require fewer participants than more complex structural models; (2) population models with larger effect size pairings generally require smaller sample sizes than those with smaller effect size pairings; and (3) the effect size of the b path has a greater impact on the required sample size than that of the a path. Each of the main findings are discussed in more detail below.

The structural models played a large role in determining the size of the empirical sample estimates. To this end, less complicated models, such as those with a dichotomous X , required fewer participants across population models than those with multiple difference score variables. This finding highlights the benefits of simplicity, and suggests that researchers should only include difference scores for the variables in which modeling change across adjacent measurement occasions is necessary to answer the question of interest. For example, consider a researcher who is interested in whether disordered eating among adolescents predicts changes in family environment, which ultimately predicts alcohol misuse among adult women. In this example, family environment would be modeled as a latent difference score, but disordered eating and alcohol misuse can be modeled as latent variables.

The magnitude of effect for both the a and b paths strongly influenced the empirical sample estimates across population models. More specifically, population models with larger effect size pairings generally required smaller samples than population models with smaller effect size pairings. This finding is consistent with findings from other Monte Carlo simulations with mediation analysis (e.g., Fritz & Mackinnon, 2007).

Moreover, the magnitude of effect of the b path has a greater impact on the required sample size than that of the a path. Models with a small effect size for the b path required larger sample estimates than models with small, medium, or large effect sizes the a path, and models with medium or large effect sizes for the b path. This finding suggests that the b path is more important in determining the required sample size to detect mediation than the a path. While these findings are consistent with other mediational power analyses (e.g., Fritz & Mackinnon, 2007), the discrepancies in the sample size requirements among models with

a small b path effect size were much larger than those seen in previous power analyses. For example, the results of the current study identified a required sample size of 435 with a LS effect size pairing and 655 with a SL effect size pairing in the $X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$ structural model. In this case, the population model with a LS effect size pairing requires a sample size with more than 200 more participants than the SL effect size pairing. Findings from a previous Monte Carlo simulation study on mediation report a required sample size of 396 for a LS effect size pairing and 385 for a SL effect size pairing with bias-corrected bootstrapping (Fritz & MacKinnon, 2007), which is a much smaller discrepancy than those seen in the current study.

On the basis of the results from the empirical sample estimates, there are four key recommendations for researchers interested in using latent difference score mediation models: (1) simplicity is considered a virtue and thus, difference scores should be included as necessary within any given model; (2) researchers must thoroughly evaluate the literature to determine the expected effect sizes before choosing their sample size; (3) sample estimates should be viewed as a lower bound; and (4) the number of mediators and kinds of indirect effects to be tested must be determined before selecting a sample size. Each of the key recommendations are discussed in detail below.

First, because simplicity is considered a virtue of latent difference score mediation models, researchers should strongly consider which variables will be modeled as latent difference scores. This consideration is important because sample size requirements drastically increase with the addition of each difference score variable. Moreover, if change across measurement occasions is not a key component of the research question at hand, modeling a variable as such may result in no change over time and thus the findings will be rendered insignificant.

Second, given the impact of effect size pairings on the required sample size to detect mediation, it is suggested that researchers thoroughly review the literature in their research area of interest to determine their hypothesized effect size. Because the b pathway more strongly influences the required sample size to detect mediation, researchers should seek out the results of other mediational models in their substantive area to inform their decision on the effect size estimates for the b pathway.

Third, because the simulated data included in this study were normally distributed and contained no missing data, researchers should view the empirical sample estimates from the current study as a lower limit of the number of participants needed to achieve .8 power. Consider a researcher who is interested in testing whether changes in self-esteem mediate the relationship between changes in disordered eating and changes in alcohol abuse. This researcher hypothesizes that the effect of changes in disordered eating on self-esteem will be of medium size and the effect of changes in self-esteem on changes in alcohol abuse to be of small size. Using the results from the current power analysis, the researcher can identify a lower sample size bound of 750 for their study to achieve .80 power.

Finally, researchers must consider how many mediators they would like to include and which specific indirect paths they hope to test prior to selecting a sample size. The current study examined the required sample size to detect mediation among ten structural latent

difference score mediation models and nine population models, all with a single mediator and one indirect effect. Researchers looking to include more than one mediator or test more than one indirect path will likely require more participants to achieve adequate power than those identified in the empirical sample size estimates within the current study.

Power estimates that offer empirical estimates for complex statistical models are becoming a common requirement when applying for grants and other funding mechanisms (Fritz & MacKinnon, 2007). The current study aimed to offer a generalizable set of sample size guidelines for researchers interested in applying latent difference score mediation models. There is no general rule of thumb for determining the required sample size for any given model, thus the sample size needed for a study depends on many factors (e.g., effect size; Muthén & Muthén, 2002). Monte Carlo simulations are common empirical method to evaluate structural equation models (Paxton et al., 2001), thus Monte Carlo models were used to general a set of empirical sample estimates across ten structural and nine population models. As a result, the current study offers future researchers a reference for determining sample size requirements for a wide range of single-mediator latent difference score mediational models.

APPENDIX:: SAMPLE MPLUS CODE

TITLE: MC WITH DICHOTOMOUS X AND SMALL M (CHANGE) AND SMALL Y (CHANGE)

ANALYSIS: Estimator = ML;

Bootstrap is 500;

MONTECARLO:

names are x m1A m1B m1C

M2A M2B M2C

y2a y2b y2c

y3a y3b y3c;

cutpoints = x (0);

nobs = 500;

nreps = 1000;

seed = 17721;

MODEL POPULATION:

[X @ 0];

X @ .25;

M1 BY M1a@.9 M1b@.9 M1c@.9;

[M1a - M1c @ 0];

M1a@.19;

M1b@.19;

M1c@.19;

M2 BY M2a@.9 M2b@.9 M2c@.9;

[M2a - M2c @ 0];

M2a@.19;

M2b@.19;

M2c@.19;

[M1 - M2@0];

M1@.98;

M2@0;

Y2 BY Y2A@.9 Y2B@.9 Y2C@.9;

[Y2a - Y2C @ 0];

Y2A@.19;

Y2B@.19;

Y2C@.19;

Y3 BY Y3A@.9 Y3B@.9 Y3C@.9;

[Y3a - Y3C @ 0];

Y3A@.19;

Y3B@.19;

Y3C@.19;

[Y2 - Y3@0];

Y2@.86;

Y3@0;

M2 ON M1 @1;

MDIFF BY M2 @1;

MDIFF ON M1*-.3;

MDIFF@.89;

Y3 ON Y2 @1;

YDIFF BY Y3 @1;

YDIFF ON Y2*-.3;

YDIFF@.77;

MDIFF ON X*.28;

YDIFF ON MDIFF*.14 X*.28;

MODEL INDIRECT:

YDIFF IND X;

MODEL:

M1 BY M1a@.9 M1b*.9 M1c*.9;

M1a*.19;

M1b*.19;

M1c*.19;

M2 BY M2a@.9 M2b*.9 M2c*.9;

M2a*.19;

M2b*.19;

M2c*.19;

[M1 - M2@0];

M1*.98;

M2@0;

Y2 BY Y2A@.9 Y2B*.9 Y2C*.9;

Y3A*.19;

Y3B*.19;
 Y3C*.19;
 Y3 BY Y3A@.9 Y3B*.9 Y3C*.9;
 Y2A*.19;
 Y2B*.19;
 Y2C*.19;
 [Y3 - Y3@0];
 Y2*.86;
 Y3@0;
 M2 ON M1 @1;
 MDIFF BY M2 @1;
 MDIFF ON M1*-.3;
 MDIFF*.89;
 Y3 ON Y2 @1;
 YDIFF BY Y3 @1;
 YDIFF ON Y2*-.3;
 YDIFF*.77;
 MDIFF ON X*.28;
 YDIFF ON MDIFF*.14 X*.28;
 MODEL INDIRECT:
 YDIFF IND X;
 output: tech9 CINTERVAL(BCBOOTSTRAP);

References

- Brown TA (2006). *Confirmatory Factor Analysis*. New York, NY: The Guilford Press
- Bryan CJ, Clemans TA, Hernandez AM, Mintz J, Peterson AL, Yarvis JS, Resick PA & The STRONG STAR Consortium. (2016). Evaluating potential iatrogenic suicide risk in trauma-focused group cognitive behavioral therapy for the treatment of PTSD in active duty military personnel. *Depression and Anxiety*, 33, 549–557. DOI: 10.1002/da.22456 [PubMed: 26636426]
- Bryan CJ, Gonzalez J, Rudd MD, Bryan AO, Clemans TA, Ray-Sannerud B, Wertenberger E, Leeson B, Heron EA, Morrow CE & Etieene N (2015). Depression mediates the relation of insomnia

- severity with suicide risk in three clinical samples of U.S. military personnel. *Depression and Anxiety*, 32, 647–655. DOI: 10.1002/da.22383 [PubMed: 26047362]
- Cohen J (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum
- Efron B & Tibshirani RJ, (1993). *An introduction to the bootstrap*. New York: Chapman & Hall
- Fritz MS, & MacKinnon DP (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18(3), 233–239. DOI: 10.1111/j.1467-9280.2007.01882.x [PubMed: 17444920]
- Holland PW (1988). Causal inference, path analysis, and recursive structural equation models. *Sociological Methodology*, 18, 449–84.
- Infurna FJ & Mayer A (2015). The effects of constraints and mastery on mental and physical health: Conceptual and methodological considerations. *Psychology of Aging*, 30(2), 432–448. Doi: 10.1037/a0039050
- James LR & Brett JM (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, 69, 307–321
- Lockhart G, MacKinnon DP, & Ulrich V (2011). Mediation analysis in psychosomatic medicine research. *Psychosomatic Medicine*, 73(1), 29–43. DOI: 10.1097/PSY.0b013e318200a54b [PubMed: 21148809]
- Ma Z, & Zeng W (2014). A multiple mediator model: Power analysis based on Monte Carlo simulation. *American Journal of Applied Psychology*, 3(3), 72–79. DOI: 10.11648/j.ajap.20140303.15
- MacKinnon DP, & Fairchild AJ (2009). Current directions in mediation analysis. *Current Directions in Psychological Science*, 18(1), 16–20. DOI: 10.1111/j.1467-8721.2009.01598.x [PubMed: 20157637]
- MacKinnon DP (2008). *Introduction to Statistical Mediation Analysis*, New York: Erlbaum and Taylor Francis Group.
- MacKinnon DP, Fairchild AJ, & Fritz MS (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593. DOI: 10.1146/annurev.psych.58.110405.085542
- MacKinnon DP, Lockwood CM & Williams J (2004) Confidence limits for the indirect effects: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99–128. DOI: 10.1207/s15327906mbr3901_4 [PubMed: 20157642]
- MacKinnon DP, Krull JL & Lockwood CM (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, 1(4), 173–186. [PubMed: 11523746]
- MacKinnon DP & Dwyer JH (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2), 144–158. DOI: 10.1177/0193841X9301700202
- McArdle JJ (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605. DOI: 10.1146/annurev.psych.60.110707.163612
- McArdle JJ & Hamagami F (2001). A latent difference score approach to structural models for linear dynamic analyses with incomplete longitudinal data In Cudeck R, du Toit S, & Sörbom D (Eds.) *Structural equation modeling: Present and future. A festschrift in honor of Karl Jöreskog* (pp. 341–380). Lincolnwood, IL: Scientific Software International.
- McDonald RP (1997). Haldane's lungs: a case study in path analysis. *Multivariate Behavioral Research*, 32, 1–38. [PubMed: 26751104]
- Mosteller R, & Tukey JW (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley
- Murphy KR, Myers B, & Wolach A (2014). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. New York: Taylor and Francis Group
- Muthén LK, & Muthén BO (2002). How to use Monte Carlo study to decide on sample size and detect power. *Structural Equation Modeling*, 9(4), 599–620. DOI: 10.1207/S15328007sem0904_8
- Muthén LK, & Muthén BO (1998–2011). *Mplus User's Guide*. Sixth Edition Los Angeles, CA: Muthén & Muthén.
- Paxton P, Curran PJ, Bollen KA, Kirby J, & Chen F (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8(2), 287–312.
- Reeve RA, Paul JM & Butterworth B (2015). Longitudinal changes in young children's 0–100 to 0–1000 number-line error signatures. *Frontiers Psychology*, 6, 647 DOI: 10.3389/fpsyg.2015.00647

- Sedlmaier R, & Gigerenzer G (1989). Do studies of statistical power have an effect on the power of the studies? *Psychological Bulletin*, 105, 309–316.
- Selig JP & Preacher KJ (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, 6(2–3), 166–164. DOI: 10.1080/15427600902911247
- Thoemmes F, MacKinnon DP, & Reiser MR (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, 17(3), 510–534. DOI: 10.1080/10705511.2010.489379 [PubMed: 23935262]
- Williams J, & MacKinnon DP (2008). Resampling and distribution of the product methods for testing indirect effects in complex models, *Structural Equation Modeling*, 15(1), 23–51. DOI: 10.1080/10705510701758166 [PubMed: 20179778]
- Wilson VanVoorhis CR, & Morgan BL (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Psychology*, 3(2), 43–50. DOI: 10.20982/tmq.03.2.p043

Figure 1:

Example of a simple mediation model with a single predictor (X), mediator (M) and outcome (Y) variable

Figure 2:
Example of a single lower-level latent factor with three observed indicators

Figure 3:

Latent difference mediation model with a single predictor, mediator, and outcome

Table 1:

Unstandardized Parameter Values for Each Structural Model Representing Small, Medium, and Large R^2 Values

Structural Models	Unstandardized paths	
	α	b
$X_{d1} \rightarrow M_{1-2} \rightarrow Y_{2-3}$	0.28, 0.72, 1.02	0.14, 0.36 and 0.51
$X_{d1} \rightarrow M_{1-2} \rightarrow Y_3$	0.28, 0.72, 1.02	0.14, 0.36 and 0.51
$X_{d1} \rightarrow M_2 \rightarrow Y_{2-3}$	0.28, 0.71, 1.02	0.14, 0.36 and 0.51
$X_1 \rightarrow M_{1-2} \rightarrow Y_{2-3}$	0.14, 0.36, 0.51	0.14, 0.36 and 0.51
$X_1 \rightarrow M_{1-2} \rightarrow Y_3$	0.14, 0.36, 0.51	0.14, 0.36 and 0.51
$X_1 \rightarrow M_2 \rightarrow Y_{2-3}$	0.14, 0.36, 0.51	0.14, 0.36 and 0.51
$X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$	0.14, 0.36, 0.51	0.14, 0.36 and 0.51
$X_{1-2} \rightarrow M_{2-3} \rightarrow Y_3$	0.14, 0.36, 0.51	0.14, 0.36 and 0.51
$X_{1-2} \rightarrow M_2 \rightarrow Y_{2-3}$	0.14, 0.36, 0.51	0.14, 0.36 and 0.51
$X_{1-2} \rightarrow M_2 \rightarrow Y_3$	0.14, 0.36, 0.51	0.14, 0.36 and 0.51

Note. α = path connecting the predictor to the mediator, b = path connecting the mediator to the outcome, X_{d1} = dichotomous predictor variable, = latent difference score variable, M = mediator, Y = outcome.

Table 2:

Empirical Sample Estimates for Each Model

Models	Effect Size Combinations								
	SS	SM	SL	MS	MM	ML	LS	LM	LL
$X_{d1} \rightarrow M_{1-2} \rightarrow Y_{2-3}$	500	465	440	492	400	325	600	490	400
$X_{d1} \rightarrow M_{1-2} \rightarrow Y_3$	480	380	390	550	425	365	660	515	430
$X_{d1} \rightarrow M_2 \rightarrow Y_{2-3}$	520	420	400	500	285	270	615	365	320
$X_1 \rightarrow M_{1-2} \rightarrow Y_{2-3}$	510	420	465	520	425	355	630	510	445
$X_1 \rightarrow M_{1-2} \rightarrow Y_3$	510	430	445	545	420	370	690	510	455
$X_1 \rightarrow M_2 \rightarrow Y_{2-3}$	515	455	450	530	460	410	650	560	475
$X_{1-2} \rightarrow M_{2-3} \rightarrow Y_{3-4}$	515	435	435	530	415	325	655	510	430
$X_{1-2} \rightarrow M_{2-3} \rightarrow Y_3$	525	430	420	590	440	380	780	585	480
$X_{1-2} \rightarrow M_2 \rightarrow Y_{2-3}$	535	460	455	535	455	385	660	640	440
$X_{1-2} \rightarrow M_2 \rightarrow Y_3$	545	460	450	565	445	415	720	550	500

Note. X = predictor, d = dichotomous variable, M = mediator, Y= outcome. S = small effect, M = medium effect, L = large effect