Social Psychological and
Personality Science
2017, Vol. 8(4) 379-386

DOI: 10.1177/1948550617715068
journals.sagepub.com/home/spp
SAGE

Determining Power and Sample Size for Simple and Complex Mediation Models

Alexander M. Schoemann[1], Aaron J. Boulton[2], and Stephen D. Short[3]

Abstract

Mediation analyses abound in social and personality psychology. Current recommendations for assessing power and sample size in mediation models include using a Monte Carlo power analysis simulation and testing the indirect effect with a bootstrapped confidence interval. Unfortunately, these methods have rarely been adopted by researchers due to limited software options and the computational time needed. We propose a new method and convenient tools for determining sample size and power in mediation models. We demonstrate our new method through an easy-to-use application that implements the method. These developments will allow researchers to quickly and easily determine power and sample size for simple and complex mediation models.

Keywords

power, mediation, sample size, *R*

Mediation analysis has been one of the most popular statistical methods utilized by social psychologists for decades. For example, a search of articles published in *Social Psychological & Personality Science* (*SPPS*) from January 2010 to September 2016 revealed 208 articles with mediation mentioned within the text of the article. Simply put, if social psychology and personality researchers are not conducting mediation analyses, they are very likely encountering the technique in the literature.

Several authors have provided detailed reviews of mediation analysis (Gunzler, Chen, Wu, & Zhang, 2013; Hayes, 2009, 2013; MacKinnon, 2008; Preacher, 2015; Rucker, Preacher, Tormala, & Petty, 2011), but discussions on power analysis and sample size calculations for these models are relatively sparse. Current best practice recommendations for assessing power and sample size in mediation models are to use a Monte Carlo power analysis (Muthen & Muthen, 2002; Thoemmes, MacKinnon, & Reiser, 2010) and, preferably, to test the indirect effect with a bootstrapped confidence interval (e.g., Zhang, 2014). However, this practice may rarely be adopted by researchers due to limited software options and the long computational time required. In an editorial, Vazire (2016) highlighted the need for adequately powered studies to be published in *SPPS*. With recent increased focus on study replication (Open Science Collaboration, 2015) and research practices (John, Loewenstein, & Prelec, 2012) in the social sciences, we find it important to highlight advances in power analysis and sample size determination for mediation analysis and provide researchers with a new easy-to-use tool to determine power and sample size for simple and complex mediation models. We begin this article with a brief review of mediation models and statistical power. Next, we describe our newly developed application for power and sample size calculations that utilizes the free statistical software R (R Core Team, 2016) and provide a brief tutorial for new users. Finally, we discuss planned extensions to and limitations of our app.

Overview of Mediation Analysis

The simple mediation model involves three measured variables (i.e., X, M, and Y) and examines if the relation between a predictor variable, X, and an outcome variable, Y, is carried through one mediating variable M. First, recall from simple regression (Equation 1) that the outcome variable Y is regressed on the predictor variable X.

$$Y = cX + e_{Y1}. \tag{1}$$

Here, following mediation analysis labeling conventions,[1] the slope for X is labeled c and is the total effect of X on Y. If a researcher is interested in addressing questions of "why" or "how" X affects Y, then a third variable, M, may be examined

[1] Department of Psychology, East Carolina University, Greenville, NC, USA
[2] University of Delaware, Newark, DE, USA
[3] College of Charleston, Charleston, SC, USA

Corresponding Author:
Alexander M. Schoemann, Department of Psychology, East Carolina University, 104 Rawl Building, Greenville, NC 27858, USA.
Email: schoemanna@ecu.edu

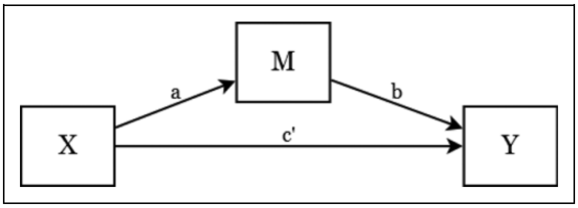

Figure 1. Simple mediation model.

as a potential mediator. Figure 1 displays a simple mediation model which can be represented by the regression Equations 2 and 3:

$$M = aX + e_M, \quad (2)$$

$$Y = c^{'}X + bM + e_Y. \quad (3)$$

The direct effect of X on Y is now labeled $c^{'}$. The indirect effect of X on Y through M is quantified as the product of a, the effect of X on M, and b, the effect of M on Y controlling for X. The total effect, c, is equal to the sum of the direct effect c' and indirect effect ab.

Historically, mediation was examined with the "causal steps" approach by satisfying four criteria described by Baron and Kenny (1986) through a series of regression models. Although the causal steps approach has been widely popular, research suggests that it is low in power to detect mediation (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) and is no longer considered best practice (Hayes, 2009). Researchers are instead encouraged to examine the indirect effect, ab.

The indirect effect can be tested for significance through a variety of methods. Sobel (1982) proposed a formula for calculating the standard error of ab, SE_{ab}, which permits calculation of a z-score statistic (ab/SE_{ab}) as well as a confidence interval for ab. This method of testing assumes that the product ab is normally distributed and has been referred to as the "Sobel test" or "normal theory approach." However, Bollen and Stine (1990) have noted the distribution of ab can deviate from normality, particularly in smaller samples, and methods that do not assume normality of the indirect effect are preferred. Many such approaches have been proposed (see, e.g., MacKinnon, 2008). In this article, we focus on two methods that do not make the normality assumption for ab and are considered best practice for testing indirect effects: bootstrap confidence intervals and Monte Carlo confidence intervals.

Bootstrap confidence intervals do not assume a normal distribution for ab and instead allow the researcher to empirically generate a sampling distribution for the indirect effect. Bootstrapping begins with the researcher assuming their collected sample represents the population that one wishes to make inferences about. A new sample of size N is then collected by resampling observations with replacement from the original sample. Analyses are conducted on this new sample to estimate the indirect effect ab. The value of ab is saved, and this process is repeated several times (e.g., 5,000) to create a sampling distribution of ab. A confidence interval for this bootstrapped sampling distribution is calculated and used for statistical inference. For example, if a researcher were interested in testing whether the indirect effect was significantly different from zero (i.e., H_0: $ab = 0$) using a 95% confidence interval, then the 2.5th and 97.5th percentiles of the bootstrapped sampling distribution would represent the lower and upper bounds of the confidence interval for ab. If this confidence interval did not include 0, then the researcher would reject the null hypothesis. Bootstrap confidence intervals have been a popular approach to testing indirect effects; however, they can be computationally intensive, especially for power analyses. In contrast, Monte Carlo confidence intervals provide a powerful, accurate test of the indirect effect and are significantly less computationally intensive.

Monte Carlo confidence intervals, also known as parametric bootstrap confidence intervals, assume normality of the regression coefficients a and b but not the product of the two terms[2] (Preacher & Selig, 2012). To form a Monte Carlo confidence interval, one obtains estimates of a, b, the variance of each coefficient (the square of a coefficient's standard error), and, if possible, the covariance between the coefficients. The regression coefficients a and b are assumed to be normally distributed with means corresponding to the parameter estimates of each coefficient and standard deviations corresponding to the standard error of each coefficient. Values of a and b are randomly drawn from these distributions, multiplied together to form an estimate of ab, and the process is repeated many times. The results from these random draws form an empirical sampling distribution of the indirect effect, and much the same as in bootstrapping, the percentiles of the distribution can be used to form a confidence interval.

Monte Carlo confidence intervals have been shown to perform as well or better than bootstrap confidence intervals in a variety of situations and models. For example, in a simple mediation model, Hayes and Scharkow (2013) note that the bias-corrected bootstrap confidence interval was more powerful than a Monte Carlo confidence interval when $N < 200$, but this power advantage was due to an increased Type I error rate. Thus, researchers interested in balancing power and Type I error rate to test the indirect effect should consider the Monte Carlo confidence interval (Hayes & Scharkow, 2013). Tofighi and MacKinnon (2016) examined power and Type I error in a more complex mediation model (i.e., $X \rightarrow M1 \rightarrow M2 \rightarrow M3 \rightarrow Y$) and note that Monte Carlo and percentile bootstrap methods did not differ in Type I error rates or power when $N = 200$, but the Monte Carlo confidence interval demonstrated more power when sample sizes were $N = 50$ and $N = 100$.

Multiple mediator models. The methods discussed for a simple trivariate mediation model can be easily extended to a wide range of models with more than one mediating variable. Simple, two mediator examples are shown in Figure 2 where multiple mediators can operate in parallel, Figure 2A, or in sequence, Figure 2B. With multiple mediators, multiple indirect effects exist in each model, for example, $a1b1$ and $a2b2$ in Figure 2A, and more complex functions of indirect effects

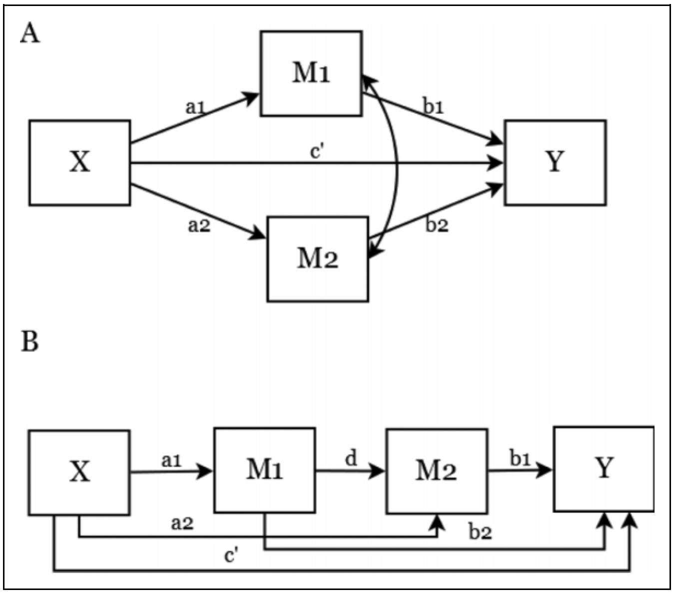

Figure 2. Multiple mediator models.

are possible, for example, *a1db1* in Figure 2B or the difference in indirect effects, *a1b1–a2b2*, in Figure 2A. Multiple mediator models can be estimated through a series of regression equations or through path analysis/structural equation models. All indirect effects or functions of indirect effects can be tested using bootstrap or Monte Carlo confidence intervals. For further details on multiple mediator models, the reader is referred to Hayes (2013).

Limitations of mediation. The discussion of mediation thus far has assumed cross-sectional data. Mediation analyses with cross-sectional data have severe limitations. Mediation analyses imply a causal model with X causing changes in M and, in turn, M causing changes in Y. With cross-sectional data, even if X is experimentally manipulated, determining the causal ordering of variables is difficult or impossible. Longitudinal data, or application of additional model assumptions, can lead to stronger causal claims about the relationships between variables (Cole & Maxwell, 2003; Preacher, 2015). Authors of *SPPS* articles employing mediation analyses are encouraged to carefully evaluate causal language used in their article and note possible causal limitations (McConnell, 2013). However, one could argue that without sufficient statistical power to detect mediation effects, concerns about the causal interpretation of effects are misplaced.

Overview of Statistical Power

Statistical power is defined as the probability of rejecting the null hypothesis (H_0), given that H_0 is false in the population. Ensuring a study has adequate power is critical for drawing conclusions from data. If a study is lacking in power, the conclusions that can be drawn if H_0 is not rejected are limited. Specifically, in a low-powered study, failing to reject H_0 may be due to the absence of an effect (H_0 is true in the population) or it may be due to the lack of power (the alternative hypothesis, H_1, is true in the population). The power of a given study is primarily affected by three components: effect size (ES), Type I error rate (α), and sample size (N). Power analysis and sample size determination are based on the fact that if three of the four quantities (power, ES, α, and N) are known, the fourth can be computed. For example, if ES, α, and N are known, power can be computed, a procedure often used for post hoc power analysis. Sample size can be determined by specifying α, ES, and power.

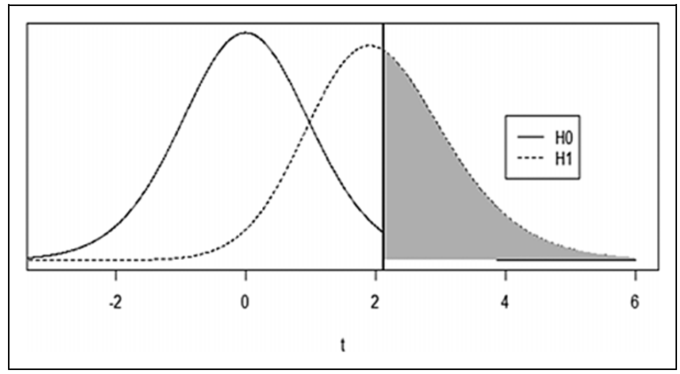

Figure 3. *t* Distribution under H_0 and H_1.

Methods of power analysis. Traditionally, power analyses and sample size determination have been based on the analytic methods. To determine power using analytic methods, values of ES, α, and N are used to construct distributions of the test statistic of interest (e.g., t statistics) consistent with H_0 and consistent with H_1. Power is the proportion of the distribution consistent with H_1 that exceeds the critical value under H_0. In Figure 3, t distributions consistent with H_0 and consistent with H_1 (where under H_1 $d = 1.0$) with 15 degrees of freedom are shown. The vertical line represents the two-tailed critical value under H_0 and the shaded portion of the H_1 distribution is the proportion of the distribution that is greater than the critical value. In this example, power, the proportion of the H_1 distribution above the critical value, is .46. To determine the sample size needed to achieve a desired level of power (e.g., .80), researchers would repeat the power analysis, varying only sample size until the desired level of power is achieved.

Power analysis methods using analytic methods have been applied to many types of statistical models and research designs including linear regression, the generalized linear model (Faul, Erdfelder, Buchner, & Lang, 2009), randomized trials and cluster randomized trails (Spybrook et al., 2011), and structural equation models (Satorra & Saris, 1985). In addition, many of these methods have been implemented in user-friendly software such as *G*Power* (Faul et al., 2009) and *Optimal Design* (Raudenbush et al., 2011). Power analyses using analytic methods provide accurate estimates of power and, for many simple research designs, are easily and quickly implemented. However, power analyses using analytic methods only cover a small portion of possible analyses (e.g., t tests, *analysis of* *variance*, correlation, regression), and even in these cases, researchers are often forced to make possibly unrealistic assumptions (e.g., equal group sizes, no missing data). When study designs or analyses are complex (e.g., mediation models with bootstrapping), analytic methods are often not available and a Monte Carlo simulation approach to power analysis is preferred.[3]

Power analysis based on Monte Carlo simulations. The idea behind the Monte Carlo simulation approach to power analysis is straightforward. Because power is the probability of rejecting H_0 given H_1 is true, if one can draw a large number (e.g., 5,000) of random samples (replications) from the population defined by H_1 and fit the hypothesized model (e.g., a regression equation) on the samples, power can be estimated as r/R, the number of samples that reject H_0 (r) divided by the total number of samples (R). Monte Carlo simulations have several advantages over traditional power analysis methods based on analytic methods. First, they allow researchers to specify the values of all parameters in a statistical model, thereby equating the power analysis and data analysis models for a more specific assessment of power. Second, power estimates can be obtained for multiple parameters in a single model. Third, greater flexibility in the specification of model assumptions (e.g., missing data) is permitted, which ideally are matched to the conditions under which a study is expected to take place. Finally, the number and types of models for which power simulations can be conducted are practically limitless. Indeed, for complex models such as mediation models, Monte Carlo power analysis may be the only method available to estimate statistical power.

To determine an appropriate sample size for a proposed study using Monte Carlo power analyses, a researcher needs to draw many random samples under the population model (defined by the researcher) with different sample sizes until he or she finds the sample size that yields the desired level of power. This process can become extremely tedious and time consuming, especially for models that are computationally intensive (e.g., mediation models using bootstrapping). Fortunately, a new method of power analysis based on varying sample size across replications can alleviate some of these limitations (Schoemann, Miller, Pornprasermanit, & Wu, 2014).

In a traditional Monte Carlo power simulation, all simulation parameters (e.g., N) are static across all replications (e.g., all replications have the same N). Power is estimated by the proportion of significant replications to the total number of replications and can only be computed for a single sample size at a time. In other words, one has to run the simulation again to know the power associated with a different sample size. Conversely, with a varying parameters approach, the design parameters (e.g., N) can take on a different set of values for each replication, and these parameters can either vary randomly or increase by small increments over a range of specified values. Power from the simulation is then analyzed with a regression model. Specifically, the significance of a parameter (coded as $0 =$ *not significant*, $1 =$ *significant*) computed from each replication serves as the outcome variable in a logistic regression analysis in which it is predicted by N. The estimated logistic regression equation can then be used to predict power from any sample size (within the specified range) without rerunning the simulation. This general approach allows researchers to run a single Monte Carlo simulation (albeit one with many replications) and compute power for a specific sample size, compute power for several sample sizes, or plot power curves over a range of sample sizes.

Power for mediation models. Despite the popularity of mediation models, determining appropriate power or sample size for one or more indirect effects is not straightforward. Guidelines for sample size in mediation models exist (e.g., Fritz & MacKinnon, 2007), but they provide guidance for a limited range of models and analytic conditions. Alternatively, researchers could attempt to determine power or sample size for each component of an indirect effect (e.g., a and b) and use the smallest power or largest sample size from these analyses when planning a study. This approach would entail using traditional power software such as *G*Power* (Faul et al., 2009) to determine the required sample size for a and b and using the larger of the two sample sizes. However, this approach will systematically underestimate the sample size needed to test the indirect effect and does not generalize to quantities from complex mediation models (e.g., $a1b1–a2b2$). The online application *WebPower* (Zhang & Yuan, 2015) can be used to determine power based on the Sobel test for simple mediation models and a path diagram-based method for assessing power for complex mediation models. However, this application does not include the ability to assess power via Monte Carlo confidence intervals or bootstrapping.

Monte Carlo power analyses are best practice for determining power and sample size in mediation models, but currently available software, though extremely flexible, has several limitations. Implementing a Monte Carlo Power analysis for mediation models requires knowledge of specific software (e.g., R, or Mplus), is computationally intensive (Zhang, 2014), and can prove difficult as users must specify all population parameters for a specific model of interest. We offer an application which is user-friendly, requires no specific programming knowledge, estimates power for the indirect effect(s) quickly, and provides an easy interface for specifying population parameters.

Application for Monte Carlo Power Analysis for Mediation Models

To facilitate use of the power analysis method based on Monte Carlo confidence intervals described above, we created a freely available application written in the R statistical computing language (R Core Team, 2016). In this section, we provide a tutorial on how to use the app to conduct a power analysis for a simple mediation model. The app employs an easy-to-use graphical user interface. Users may access the app by visiting https://schoemanna.shinyapps.io/mc_power_med/ or downloading it from within R. We recommend the latter as it will run faster on one's local machine. To download the app, users must have R as well as the shiny and MASS add-on packages

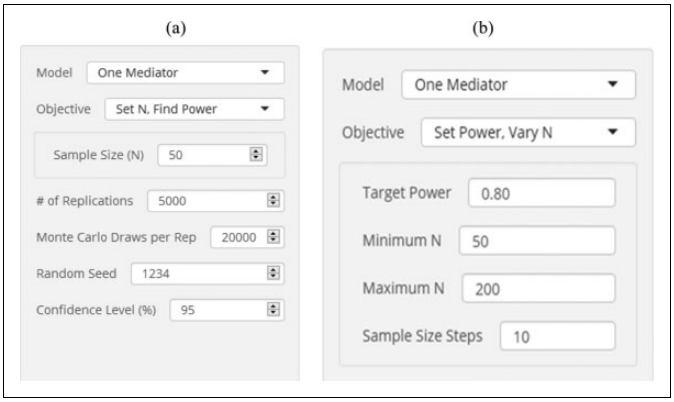

Figure 4. Application options.

installed.[4] Once installed, the user opens an R session and runs the following command:

```
library(shiny)
runGitHub("mc_power_med", "schoam4")
```

The first command loads routines from the shiny add-on package that are needed to run the app in the user's current R session. The second command downloads the app and opens it in a user's default web browser program.[5]

Once the app is running, the user is presented with a variety of program options as shown in Figure 4a. Starting at the top of the options menu, the user must first select the mediation model to be used in the power analysis. At the time of this writing, only two models—one mediator and two parallel mediators—are available, although more will be made available in future releases. For this tutorial, we will be calculating power for the default option, the simple trivariate mediation model shown in Equations 1–3. Note that a path diagram of the selected model will appear to the right of the options menu when selected (Figure 5). Next, the user must select the objective of their power analysis. Currently, two options are offered: (a) "Set *N*, Find Power," which calculates the statistical power for an indirect effect(s) under the chosen model specification and target sample size, or (b) "Set Power, Vary *N*," which uses the varying sample size approach to calculate the sample size required to achieve a specific level of power—designated by the user along with a range of sample sizes—for a target indirect effect(s). If option (a) is selected, only the target sample size is required to be entered by the user (Figure 4a). If option (b) is selected, the user is presented with a submenu of additional options, shown in Figure 4b.

For the present example, we will select option (b). In the submenu, the user must set the target power level, the minimum and maximum *N* for the range of sample sizes considered, and increments of *N* to calculate power estimates for within the specified range. For this tutorial, we have selected the conventional power level of .80, a minimum sample size of 50, a maximum sample size of 200, and a step size of 10. Note that smaller step sizes combined with wider ranges will require more computation time; thus, the user might opt to specify a large range of sample sizes with a large step size in a preliminary analysis and subsequently narrow the range and decrease the step size in additional runs for more precise sample size estimates.

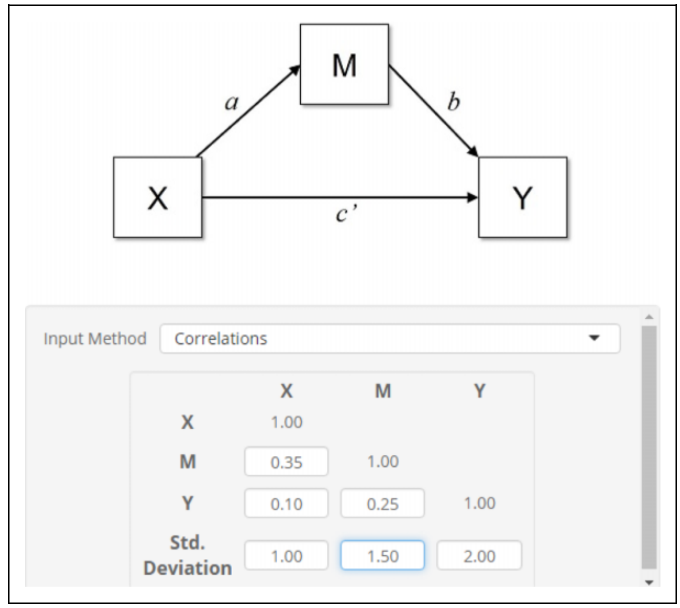

Figure 5. App model input section.

The remaining options shown in Figure 4a are typical parameters that need to be set for Monte Carlo power analyses (Muthén & Muthén, 2002). First, the total number of replications need to be selected. This number is typically 1,000 or greater, although little published guidance on the number of replications needed in simulation studies exists. Mundform et al. (2011) provided empirically based recommendations, suggesting that 5,000 may be enough for many applications. Ultimately, the number of replications should be sufficient to ensure stable power or sample size estimates and that number will depend on the modeling context. Therefore, it is recommended that the user run the power analysis at least twice with differing numbers of total replications (e.g., 5,000, 10,000) to ensure that the final estimate(s) has converged to a stable value(s). The next option, "Monte Carlo Draws per Rep," refers to the number of times each target coefficient is sampled from its sampling distribution within each power analysis replication to calculate the Monte Carlo confidence interval(s). Once again, published recommendations are scarce: We note only that several thousand draws are likely needed, such as 20,000, which was chosen for the empirical examples discussed in Preacher and Selig (2012). The logic presented above for the total number of replications also applies here such that conducting many runs with increasing values can reassure the user that estimates are stable. In our running example, we chose values of 5,000 and 20,000 for the total number of power analysis replications and the number of coefficient draws per replication, respectively. The final two options are the random number generator seed and the confidence interval width. The seed should be a positive integer and ensures results from a run of the app are replicable. A researcher using the same seed and parameter values will replicate another researcher's results, whereas a different seed may lead to slightly different results. The default seed in our application, 1,234, was used for this example. The confidence interval width [$100(1 - \alpha)\%$] sets the width of the confidence intervals for all indirect effects calculated within each replication. In our example, the width is set to 95% (corresponding to $\alpha = .05$).

Once all options for the power analysis are set, the user must input population parameters for the model, akin to choosing an ES in a traditional power analysis. Specifically, the information entered, in one form or another, must be sufficient for calculation of the hypothesized indirect effect and its associated confidence interval; at a minimum, this implies the hypothesized *a* coefficient(s), *b* coefficient(s), and the coefficient standard errors for the simple mediation model. There are a few different quantities that will meet this criterion, including model parameter estimates (Selig & Preacher, 2008; Zhang, 2014), measures of variance explained (Thoemmes et al., 2010), and correlation or covariance matrices. The default option in the app is to enter a correlation matrix and, if applicable, the standard deviations of the variables, which are used to transform the correlation matrix to a covariance matrix. In the running example, suppose we have found in previous studies or meta-analyses that our focal predictor *X* correlates with the mediator *M* at approximately .35, *M* correlates with the outcome variable *Y* at approximately .25, and the *X* and *Y* variables correlate at approximately .10. Additionally, prior research has found the standard deviations of *X*, *M*, and *Y* to be 1.00, 1.50, and 2.00, respectively. In the middle column of the app (Figure 5), we enter this information in the appropriate boxes that change responsively to the model and input method selected.

Now that the program options and hypothesized model have been fully specified in our example, we click the "Calculate Power" button on the right side of the app. If any errors were made in the previous steps, the program will terminate and an error message will appear below the button. If this occurs, users should change the relevant input and press the button again. Once the app begins to run, a progress bar will appear. If the power analysis calculations terminate successfully, output will appear below the button. In our running example, the app took approximately 52 s to run. Using the continuously varying sample size approach to Monte Carlo power analysis, approximately 150 individuals are required to ensure statistical power is at least 80% for detecting the hypothesized indirect effect.

The application remains in development and we foresee several extensions in new releases, detailed next.

Current Limitations and Potential Extensions

Many extensions to the existing app are possible. Foremost, including a larger number of models than those currently offered would considerably improve the flexibility of the app. Due to the large number of models circulating in social psychology journals, including all possible models is difficult. Implementing models that are the most common in and relevant to social psychology research, such as dyadic mediation models (Ledermann, Macho, & Kenny, 2011), longitudinal mediation models (Selig & Preacher, 2009), or models combining mediation and moderation (Hayes, 2013), is our priority. Moreover, missing data are pervasive in psychological research and reduce statistical power in addition to other potentially harmful consequences (Enders, 2010). Permitting missing data in the calculation of power would promote more accurate sample size estimates. Finally, nonnormal variables have to be potential to provide inaccurate power estimates (Zhang, 2014). Extending the app to allow for the specification of nonnormal variables, which tend to be normative in psychology (Micceri, 1989), would also enhance the accuracy of the app. Extensions and news about the development of the app will be posted on the following webpage: http://marlab.org/power_mediation/

Conclusion

Mediation analysis is a popular tool for social and personality psychologists, but for mediation to be an effective tool, researchers must plan studies with sufficient statistical power. Accurately determining statistical power for mediation models can be tricky for applied researchers, especially when using bootstrapping or Monte Carlo confidence intervals to test the indirect effect. We have developed an application which makes determining power or sample size for mediation models relatively straightforward. By utilizing a simple interface, population parameters expressed as correlations, and varying sample sizes within a power analysis, our application provides social and personality psychologists with a powerful, easy-to-use tool to aid in study planning when mediation is of interest.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Following the conventions in Hayes (2013), all regression equations omit the intercept. The inclusion of an intercept will not affect any tests of the indirect effect.
2. The assumption of the normality of regression coefficients in a Monte Carlo confidence interval is the same assumption made when using Wald tests with *z* values to determine statistical significance of a single regression coefficient. Thus, Monte Carlo confidence intervals are appropriate in any situation where interpretation of the Wald tests statistic is warranted. Furthermore, simulation studies (reported in the technical appendix for the app at http://marlab.org/technical_appendix/) have demonstrated that Monte Carlo confidence intervals perform as well as bootstrap confidence intervals when estimating power under normal and nonnormally distributed variables.

3. Two procedures used in this article contain the term Monte Carlo. To clarify between them, Monte Carlo confidence intervals are used to test model parameters (e.g., indirect effects) and Monte Carlo power analyses are used to determine power to reject H0 for a parameter(s) in a statistical model.
4. R can be downloaded from https://cran.r-project.org/. R includes some basic functionality, but the majority of routines used in R are included in add-on software packages developed by independent contributors. For instance, the shiny package, which is required to download and run the app, provides functionality that allow users to easily deploy R-based applications to the Internet. Expertise in R programming is not required to be able to download and run the app as detailed in this article; numerous webpages and books are easily located online for readers seeking additional information on R.
5. Although the app opens in an Internet browser, it is technically "off-line"—that is, the app is running locally on one's machine from files downloaded via the runGitHub command as opposed to running on a web server. For the best experience using the app, we recommend maximizing the web browser to fill the screen. Smaller browser sizes may make input boxes difficult to read.

References

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

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

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*, 558–577.

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analysis using G*Power 3.1: Tests for correlation and regression analyses. *Behavioral Research Methods*, *41*, 49–60.

Fritz, M. S., & McKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*, 233–239.

Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry*, *25*, 390–395.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*, 408–420.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression based approach*. New York, NY: Guilford Press.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*, 1918–1927.

John, L. K., Lowenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.

Ledermann, T., Macho, S., & Kenny, D. A. (2011). Assessing mediation in dyadic data using the actor-partner interdependence model. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 595–612.

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

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *Psychological Methods*, *7*, 83–104.

McConnell, A. R. (2013). Editorial. *Social Psychological and Personality Science*, *4*, 3–5.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166.

Mundform, D. J., Schaffer, J., Kim, M. J., Shaw, D., Thongteeraparp, A., & Supawan, P. (2011). Number of replications required in Monte Carlo simulation studies: A synthesis of four studies. *Journal of Modern Applied Statistical Methods*, *10*, 4.

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*, *9*, 599–620.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 943.

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, *66*, 825–852.

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*, 77–98.

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X., Martinez, A., Bloom, H., & Hill, C. (2011). *Optimal design software for multi-level and longitudinal research* (Version 3.01) [Software]. Retrieved from https://www.wtgrantfoundation.org

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*, 359–371.

R Core Team. (2016). *R: A language and environment for statistical computing* (version 3.3.0) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from the comprehensive R archive network (CRAN): https://www.R-project.org/

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.

Selig, J. P., & Preacher, K. J. (2008, June). *Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects* [Computer software]. Retrieved from http://quantpsy.org/

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*, 144–164.

Schoemann, A. M., Miller, P. M., Pornprasermanit, S., & Wu, W. (2014). Using Monte Carlo simulations to determine power and sample size for planned missing designs. *International Journal of Behavioral Development*, *38*, 471–479.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp. 290–312). San Francisco, CA: Jossey-Bass.

Spybrook, J., Bloom, H., Congdon, R., Hill, Cl., Martinez, A., & Raudenbush, S. (2011). *Optimal design for longitudinal and multilevel research: Documentation for the optimal design software* (Version 3.0). Retrieved from www.wtgrantfoundation.org

Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, *17*, 510–534.

Tofighi, D., & MacKinnon, D. P. (2016). Monte Carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 194–205.

Vazire, S. (2016). Editorial. *Social Psychological and Personality Science*, *7*, 3–7.

Zhang, Z. (2014). Monte Carlo based statistical power analysis for mediation models: Methods and software. *Behavior Research Methods*, *46*, 1184–1198.

Zhang, Z., & Yuan, K. H. (2015). *WebPower: Statistical power analysis online*. Retrieved from http://webpower.psychstat.org

Author Biographies

Alexander M. Schoemann is an assistant professor of psychology at East Carolina University. His research interests include missing data, latent variable modeling, meta-analysis, power analysis, and developing user friendly statistical software.

Aaron J. Boulton is a senior biostatistician in the Center for Health Research Assessment and Translation at the University of Delaware. His broad research interests include longitudinal data analysis, multilevel modeling, latent variable modeling, and the psychological and behavioral determinants of health across the life span.

Stephen D. Short is an assistant professor of psychology at the College of Charleston. His research interests include scale development, measurement invariance, latent variable modeling, and individual differences in attitudes toward and knowledge of science.

Handling Editor: Joseph Simmons