Monte Carlo Based Statistical Power Analysis for Mediation Models: Methods and Software

Zhiyong Zhang

University of Notre Dame

Author Note

We thank Scott Maxwell and Ke-Hai Yuan for helpful discussion, and David Kenny and one anonymous reviewer for constructive suggestions that have significantly improved this research. Path diagrams used in the paper were generated using WebSEM (https://websem.psychstat.org).

Abstract

The existing literature on statistical power analysis for mediation models often assumes data normality and is based on a less powerful Sobel test instead of the more powerful bootstrap test. This study proposes to estimate statistical power to detect mediation effects based on the bootstrap method through Monte Carlo simulation. Non-normal data with excessive skewness and kurtosis are allowed in the proposed method. A free R package called bmem is developed to conduct the power analysis discussed in this study. Four examples, including a simple mediation model, a multiple-mediator model with a latent mediator, a multiple-group mediation model, and a longitudinal mediation model, are provided to illustrate the proposed method.

Keywords: Power analysis, mediation models, non-normal data, bootstrapping, R package bmem

Monte Carlo Based Statistical Power Analysis for Mediation Models: Methods and Software

Introduction

Mediation models are widely used in the social and behavioral sciences as demonstrated in recent books by Hayes (2013) and MacKinnon (2008). Mediation models are useful because they can be used to investigate the underlying mechanism related to why an input variable influences an output variable. In order to avoid under-powered research, statistical power analysis is always necessary before data collection. We are only aware of a few studies that have discussed how to conduct power analysis for mediation models. Most literature on power analysis for mediation models has focused on a simple mediation model (Fritz & MacKinnon, 2007; C. Wang & Xue, 2012; Vittinghoff et al., 2009; Beasley, 2012). Software is available in terms of R code (Kenny & Judd, 2013) and R package (Qiu, 2013) for conducting power analysis for certain types of mediation models. Thoemmes et al. (2010) proposed a general framework for power analysis for complex mediation models using Monte Carlo simulation in Mplus (Muthén & Muthén, 1998-2011). However, their method assumes that data are normally distributed and uses the Sobel test although it can be extended to non-normal data analysis.

Practically collected data are often non-normal. For example, Micceri (1989) reported that among 440 large-sample achievement and psychometric measures taken from journal articles, research projects, and tests, all were significantly non-normally distributed. Consequently, statistical tests developed for normal data often give inaccurate power estimation in the presence of non-normal data (e.g., L. Wang & Zhang, 2011; Zu & Yuan, 2010; Zhang & Wang, 2013b). Mediation analysis adds extra complexity to power analysis. For example, different methods are available for testing the mediation effect, and they can have different power for the same mediation effect (e.g., Cheung, 2007; Fritz & MacKinnon, 2007). Studies have shown that the bootstrap method achieves the highest power among many methods developed for detecting mediation in the literature.

This study extends Thoemmes et al. (2010) in several ways. First, it proposes a general method to conduct power analysis for mediation models based on the bootstrap method. The method is still based on Monte Carlo simulation but uses the bootstrap method to test mediation effects. Second, the method allows the specification of non-normal data in the Monte Carlo simulation and can thereby reflect more closely practical data collection. Third, a free, open-source R package, bmem, is developed to ease power analysis for mediation models using the proposed method.

In the following sections, we first present the proposed method for power analysis of mediation models. Then, we illustrate the use of the R package bmem for conducting power analysis. After that, we demonstrate the use of the proposed method through 4 examples including a simple mediation model, a multiple-mediator model with latent variables, a multiple group mediation model, and a longitudinal mediation model. Complete R code for the 4 examples is provided in the appendices.

Monte Carlo Based Statistical Power Analysis

In this section, we first present the proposed method. For better illustration, we focus our discussion on a simple mediation model even though the method applies to more complex models, as shown in our examples. Figure 1 displays the path diagram of the simple mediation model. In the figure, `x`, `m`, and `y` represent the independent or input variable, the mediation variable, and the dependent or outcome variable, respectively. In this model, the total effect of `x` on `y`, `c'+a*b`, consists of the direct effect `c'` and the mediation effect θ=`a*b`, the multiplication of the direct effect of `x` on `m` and the direct effect of `m` on `y`. The mediation effect is also called the indirect effect because it is the effect of `x` on `y` indirectly through `m`.

Statistical power analysis for mediation can be viewed as concerning a test whether the mediation effect (θ) is significantly different from 0. More specifically, we have the null

and alternative hypothesis

$$H_0 : \theta = 0 \text{ vs. } H_1 : \theta = \theta_1,$$

where θ_1 represents a given effect size. By its definition, the statistical power (π) is

$$\pi \quad = \quad \Pr(\text{reject } H_0 | H_1). \tag{1}$$

In addition to the use of null hypothesis testing, the power can be calculated using the confidence intervals. This is based on the equivalence of confidence intervals and hypothesis testing (e.g., Hoenig & Heisey, 2001; Meehl, 1997). That is, if a 1-α confidence interval does not include the null hypothesis value, one can infer a statistically significant result at the significant level α (e.g., Daly, 1991). More specifically, let $[l, u]$ denote the confidence interval of the mediation effect θ. The power is then

$$\pi = \Pr(0 \notin [l, u] | H_1). \tag{2}$$

In practice, the power π can be difficult to calculate analytically especially for complex mediation models. However, it can be estimated using the relative frequency of rejecting the null hypothesis in Monte Carlo simulation following Algorithm 1. The algorithm has been widely applied in the literature of statistical power analysis for both mediation analysis and other analysis (e.g., Cheung, 2007; Fritz & MacKinnon, 2007; Fritz et al., 2012; Hayes & Scharkow, 2013; MacKinnon et al., 2004; Muthén & Muthén, 2002; Thoemmes et al., 2010; Zhang & Wang, 2009, 2013b).

A critical component of such a Monte Carlo algorithm is the choice of the method for constructing the confidence interval of the mediation effect. In this study, we consider three types of confidence intervals: the normal confidence interval, the robust confidence interval, and the bootstrap confidence interval although we recommend the use of the bootstrap confidence interval.

Algorithm 1 Monte Carlo simulation algorithm for statistical power

1. Form a mediation model based on the hypothesized theory and set up the population parameters for the mediation model. The parameter values can be decided from previous studies in the literature or a pilot study.

2. Generate a data set with sample size n based on the model and its population parameter values.

3. Test the significance of a mediation effect by forming a confidence interval using the generated data.

4. Repeated Steps 2 and 3 for R times where R is the number of Monte Carlo replications.

5. Suppose among the R replications, the mediation effect is significant for r times. Then the power for detecting the mediation effect given the sample size n is r/R.

Normal Confidence Interval

In mediation analysis, model parameters and their covariance can be estimated using the maximum likelihood method. Under the normal data assumption, the estimated model parameters follow a multivariate normal distribution asymptotically. For example, for the simple mediation model, $\hat{a}$ and $\hat{b}$, estimates of a and b, have a bivariate normal distribution with the covariance matrix $\begin{pmatrix} \hat{\sigma}_a^2 & \hat{\sigma}_{ab} \\ \hat{\sigma}_{ab} & \hat{\sigma}_b^2 \end{pmatrix}$ where $\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$, and $\hat{\sigma}_{ab}$ are the estimated variances and covariance of $\hat{a}$ and $\hat{b}$. Using the delta method, $\hat{\theta} = \hat{a}\hat{b}$ is normally distributed with mean $\theta = ab$ and variance $\hat{b}^2\hat{\sigma}_a^2 + 2\hat{a}\hat{b}\hat{\sigma}_{ab} + \hat{a}^2\hat{\sigma}_b^2$ (p.298, Sobel, 1982). The $1-\alpha$ confidence interval for ab can be constructed as

$$[\hat{a}\hat{b} + \Phi^{-1}(\alpha/2) \times \widehat{se}(\hat{a}\hat{b}), \hat{a}\hat{b} + \Phi^{-1}(1-\alpha/2) \times \widehat{se}(\hat{a}\hat{b})], \tag{3}$$

where Φ is the standard normal cumulative distribution function and therefore $\Phi^{-1}(\alpha)$ gives the 100αth percentile of the standard normal distribution. For example, for the 95% confidence interval, $\Phi^{-1}(\alpha/2) = \Phi^{-1}(.05/2) = \Phi^{-1}(.025) \approx -1.96$ and $\Phi^{-1}(1-\alpha/2) = \Phi^{-1}(.975) \approx 1.96$. $\widehat{se}(\hat{a}\hat{b}) = \sqrt{\hat{b}^2\hat{\sigma}_a^2 + 2\hat{a}\hat{b}\hat{\sigma}_{ab} + \hat{a}^2\hat{\sigma}_b^2}$ is the standard error of

$\hat{a}\hat{b}$. We refer to this interval as the normal confidence interval. Note a power analysis based on the normal confidence interval is the same as the use of the Sobel test.

Robust Confidence Interval

When data are not normally distributed, the standard error estimates of the parameter estimates of the mediation models are not consistent. Therefore, the confidence interval in Equation (3) is problematic. However, if the fourth moments (or kurtosis) of the non-normal data still exist, the robust Sandwich-type standard errors are consistent and can be used (Zu & Yuan, 2010). Therefore, replacing the normal standard error with the Sandwich-type standard error in Equation (3), we obtain a robust confidence interval for the mediation effect.

Bootstrap Confidence Interval

Both the normal and robust confidence intervals are based on asymptotic theory and they might not perform well in finite sample experiments (e.g., MacKinnon et al., 2004; Zu & Yuan, 2010). In the literature, confidence intervals constructed using the bootstrap method have been shown to perform better under many studied conditions (e.g., Cheung, 2007; Fritz & MacKinnon, 2007; Fritz et al., 2012; Hayes & Scharkow, 2013; MacKinnon et al., 2004; Preacher & Hayes, 2004; Shrout & Bolger, 2002). Algorithm 2 can be followed to construct a bootstrap confidence interval.

Different bootstrap confidence intervals have been used for the bootstrap method in the literature of mediation analysis (e.g., Cheung, 2007; Fritz & MacKinnon, 2007; Fritz et al., 2012; Hayes & Scharkow, 2013; MacKinnon et al., 2004). Let θ denote a population mediation effect, $\hat{\theta}$ denote the estimate of θ from the original data, and $\hat{\theta}^b, b = 1, \ldots, B$ denote its estimate for the bth bootstrap sample. A $100(1-\alpha)\%$ bootstrap confidence interval is formed in the following ways. First, the percentile bootstrap confidence interval can be constructed by $[\hat{\theta}^b(\alpha/2), \hat{\theta}^b(1-\alpha/2)]$ for a parameter with $\hat{\theta}^b(\alpha)$ denoting the 100αth percentile of the B bootstrap estimates. Second, the bias-corrected bootstrap

Algorithm 2 Bootstrap confidence interval algorithm

1. Using the original data set (Sample size = n) as a population, draw a bootstrap sample of n persons randomly with replacement.

2. With the bootstrap sample, estimate model parameters and compute estimated mediation effects.

3. Repeat Steps 1 and 2 for a total of B times. B is the number of bootstrap samples.

4. The bootstrap confidence intervals of model parameters and mediation effects are constructed.

confidence interval can be constructed as $[\hat{\theta}^b(\tilde{\alpha}_l), \hat{\theta}^b(\tilde{\alpha}_u)]$ where $\tilde{\alpha}_l$ and $\tilde{\alpha}_u$ are used to get the quantiles and are calculated by

$$\tilde{\alpha}_l = \Phi[2z_0 + \Phi^{-1}(\alpha/2)] \tag{4}$$

and

$$\tilde{\alpha}_u = \Phi[2z_0 + \Phi^{-1}(1 - \alpha/2)] \tag{5}$$

with

$$z_0 = \Phi^{-1}\left[\frac{\text{number of times that } \hat{\theta}^b < \hat{\theta}}{B}\right]. \tag{6}$$

Remarks

Choice of Confidence Intervals. Simulation studies have been conducted to evaluate the normal, robust, and bootstrap confidence intervals (e.g., Cheung, 2007; Fritz et al., 2012; Hayes & Scharkow, 2013; MacKinnon et al., 2004; Zu & Yuan, 2010). Overall, the bootstrap confidence intervals, including the percentile and the bias-corrected ones, perform better than the non-bootstrap ones. It is found that the percentile bootstrap confidence interval has greater power than the normal one and, at the same time, maintains Type I error near its nominal level. The bias-corrected bootstrap confidence interval has even greater power than the percentile bootstrap confidence interval but at the cost of more liberal Type I error. In a recent study, Hayes & Scharkow (2013)

recommended that if power is at the forefront of concerns, the bias-corrected bootstrap confidence interval can be used and, in general, one can use the percentile bootstrap confidence interval as a good compromise test. In this study, by default, we adopt the percentile bootstrap confidence interval in our analysis. Our software allows the use of normal, robust, percentile bootstrap, and bias-corrected bootstrap confidence intervals.

Non-normal Data in Power Analysis. Typically in the Monte Carlo based power analysis, data are generated from a multivariate normal distribution assuming that data collected in the future study will be normally distributed (e.g., Zhang & Wang, 2009). In order to deal with non-normal data, one should allow for a power analysis based on non-normal data. In this study, continuous non-normal data with target skewness and kurtosis can be used. Specifically, the method developed by Vale & Maurelli (1983) is used to generate non-normal data with the same mean and variance as the normal data but with target skewness and kurtosis provided by a user. If the literature shows, or a researcher has reason to believe, that non-normality is a concern after data collection, power analysis should be conducted with simulated non-normal data. For non-normal data with excessive skewness and kurtosis, power based on the robust and bootstrap confidence intervals should be trusted more than the normal confidence interval. For studies with small sample size, the bootstrap method is expected to perform best.

Controlling Type I Error. The literature on mediation analysis has shown that the type I error for mediation tests is generally not well controlled (e.g., Fritz et al., 2012; MacKinnon et al., 2004; Zhang & Wang, 2013b). Through simulation, MacKinnon et al. (2004) showed that the normal method had too conservative empirical type I error and, on average, the bias-corrected bootstrap method had an acceptable empirical type I error. Recent studies further found that the bias-corrected bootstrap method had an inflated type I error rate when one of the direct effects, a or b, was not zero (e.g., Fritz et al., 2012; Hayes & Scharkow, 2013). The percentile bootstrap method, on the other hand, was found to have better controlled type I error. Therefore, if the power is calculated based on the

percentile bootstrap confidence interval, the type I error should not be a big concern.

If controlling type I error is the foremost concern in power analysis, we recommend the following strategy. Our Monte Carlo method can be conducted under the null hypothesis where the null values can be used as the population parameters. In this case, power becomes type I error. Suppose a researcher wants to control type I error at 0.05 level. The researcher can always starts with a significance level at 0.05. Then, the empirical type I error can be obtained. If the empirical type I error rate is smaller than .05, the researcher can increase the significance level. Otherwise, the researcher can decrease the significance level. By trial and error, the researcher can decide on a significance level to control the type I error rate at the desired value. Then, in power analysis, the significance level can be used to construct confidence intervals to test mediation effects.

Relevant Statistics from the Monte Carlo Method. Using the Monte Carlo method, the statistical power is estimated by r/R. The standard error of the power can be estimated by $\sqrt{r(1-r)/R^3}$. Note that with the increase of R, the power estimate becomes more accurate. In practice, we recommend $R \geq 1000$ because in the literature, 1000 replications are often used in evaluating power (e.g., Cheung, 2007; Thoemmes et al., 2010; Zhang & Wang, 2009).

In addition to power and its standard error, other statistics can be calculated. An important one is the empirical coverage probability. The empirical coverage probability is the rate that the constructed confidence interval covers the population value. For a well-performed confidence interval, the empirical coverage probability should be close to the confidence level $1-\alpha$. For example, Hayes & Scharkow (2013) showed that the percentile bootstrap confidence interval has better coverage than the bias-corrected one. With the R sets of parameter estimates and their standard errors, one can also calculate the mean and standard deviation of the parameter estimates and the mean of the standard errors.

R Package `bmem`

The proposed method in the above section is implemented in the free, open-source R package `bmem` (Zhang & Wang, 2013a). The package `bmem` uses the R package `lavaan` (Rosseel, 2012) for model estimation. The package can conduct power analysis based on the normal, robust, and bootstrap confidence intervals. We now illustrate the use of the package through a simple mediation model shown in Figure 2. The values in the figure are population parameters that can be decided from a pilot study or previous literature. In this example, we choose the parameter values to approximate a medium mediation effect. Some of the values are labeled using `a`, `b`, and `cp`. For demonstration, suppose we are interested in the power of the mediation effect `ab=a*b` and the total effect `abc=a*b+cp`.

To use `bmem`, one needs to specify the mediation model and the mediation effect. The package uses the `lavaan` model specification method but with some specific requirements. For example, for the simple mediation model, it is specified as below:

```
demo = "
     y ~ cp*x + start(0)*x + b*m + start(.39)*m
     m ~ a*x+start(.39)*x
     x ~~ start(1)*x
     m ~~ start(1)*m
     y ~~ start(1)*y
"
```

First, the name of the model is `demo` in R. Everything about the model is given in a pair of quotation marks. Each path in the model is described using a line of statement. For example, `m ~ a*x + start(.39)*x` means that `m` regresses on `x` with the coefficient `0.39` as in `start(0.39)`. Because the coefficient has a label `a`, it is also specified in the equation. The statement `x ~~ start(1)*x` means that the variance for `x` is 1. More generally, the regression relationships are specified using `~` and variance and covariance are specified

using `~~`. More about model specification can be found in Rosseel (2012). The use of bmem does not requires knowledge of R beyond what is discussed in this paper. However, users who are interested in learning R are directed to the webpage at `http://www.r-project.org/doc/bib/R-books.html` for a list of useful references.

With the model, we also need to tell the package the mediation effect to conduct power analysis on. In this example, the mediation effect `ab` and the total effect `abc` are of interest to us. They can be specified as

```
mediation = "
      ab := a*b
      abc:= a*b + cp
"
```

The notation `:=` means to calculate the indirect effect `ab` as the product of parameter `a` and `b`, where the labels on the right hand of "`:=`" should be consistent with those used in the model statement `demo`. Similarly, the total effect is calculated.

Only the labels for the parameters that will appear in the calculation of the mediation effect are necessary to use in the model specification part. For example, for the variance parameters, no labels are used. By default, the variance parameters will be set at 1. Therefore, in this example, the specifications of the three variance parameters are not required.

The package `bmem` conducts power analysis based on the percentile bootstrap method through the function `power.boot`. The code below presents an example to calculate power for the model in Figure 2 with a sample size 100.

```
power.result = power.boot(model=demo, indirect=mediation, nobs
   =100, nrep=1000, nboot=2000, parallel="multicore", skewness=
   c(0, 0, 1.3), kurtosis=c(0, 0, 10), ovnames=c("x", "m", "y")
   )
```

The function `power.boot` takes many arguments but only the `model` one is required. A model is provided using the argument `model`, for this example, `model=demo`. If a mediation effect is of interest, it should also be provided as in this example `indirect=mediation`. By default, the power is calculated for a sample size of 100, but one can change it by providing a different number to `nobs`. One can specify the number of Monte Carlo simulation replications in the calculation of power using option `nrep=` with a default 1000 and also the number of bootstrap using `nboot=` with a default 1000. To take the advantage of multi-core processors of modern computers, the package allows parallel computing by setting `parallel='snow'` which uses the R package snowfall (Knaus, 2013) for automatic parallelization. By default, all cores available on a computer are used to speed up calculation. If one suspects the data will be non-normal , the skewness and kurtosis for the observed variables can be provided. When specifying non-normal data, the observed variable names (`ovnames`) should also be provided to match the order of the skewness and kurtosis statistics.

The results of the power analysis can be summarized into a table using the function `summary(power.result)`. The results table (Output 1) shows several columns. First, the column `True` lists the population parameter values. Second, the column `Estimate` presents the average parameter estimates across all replications. Third, the column `MSE` is the average bootstrap standard error and the column `SD` is the standard deviation of the parameter estimates across all replications. Fourth, the column `Power` gives the power to detect whether a parameter is significant and the column `Power.se` provides the standard error of the estimated statistical power. Finally, the column `Coverage` presents the empirical coverage probability of the bias-corrected bootstrap confidence interval. The power for the mediation effect is listed at the end of the table entitled "Indirect/Mediation effects". The power to detect the mediation effect with a sample size 100 is about 0.935 using the percentile bootstrap confidence interval for the current example. During the phase of sample size planning, if a researcher targets a power of 0.8, he/she can reduce the

current sample size for another calculation.

Output 1. Output of bmem

```
                      True  Estimate      MSE      SD     Power Power.se Coverage
Regressions:
  math ~
    ME       (cp)    0.000    0.001    0.107    0.111    0.070    0.008    0.930
    HE        (b)    0.390    0.395    0.109    0.119    0.971    0.005    0.928
  HE ~
    ME        (a)    0.390    0.394    0.101    0.103    0.966    0.006    0.938

Variances:
    math             1.000    0.975    0.262    0.351    1.000    0.000    0.791
    HE               1.000    0.982    0.135    0.140    1.000    0.000    0.915

Indirect/Mediation effects:
    ab               0.152    0.156    0.061    0.064    0.935    0.008    0.933
    abc              0.152    0.158    0.108    0.111    0.346    0.015    0.927
```

In addition to the power for the mediation effect and the total effect, the results also include power for all parameters in the model. For variance parameters, since they are always larger than 0, its power is obtained based on the bootstrap standard error instead of the percentile confidence interval. Furthermore, if one adds the argument `ci="BC"` in the `power.boot` function, the power based on the bias-corrected confidence interval can be obtained. Although the purpose of this study focuses on the power using the bootstrap method, the package `bmem` also provides a function `power.basic` to conduct power analysis for mediation models using the normal and robust confidence intervals.

In estimating the power for the mediation effect and the total effect, we assume that the type I error is well controlled. If a researcher is concerned about the type I error, he/she can investigate it using bmem. For example, the code below specifies a possible model under the null hypothesis that assumes $a = b = 0$. Replacing the model `demo` using `demo.alpha`, we get the empirical type I error 0.003 for the mediation effect and 0.064 for the total effect. Furthermore, through trial-and-error, one can find that at the significance level 0.3, the empirical standard error is approximately 0.05 for $a = b = 0$ and at the same

time, the power for $a = b = 0.39$ boosts to 0.996. However, Fritz et al. (2012) showed that the type I error is related to the magnitude of a and b under null (also see Example 1). Therefore, a more serious investigation would evaluate the type I error according to the combination of different values of a and b.

```
demo.alpha = "
      y ~ cp*x + start(0)*x + b*m + start(0)*m
      m ~ a*x+start(0)*x
"
```

Examples

In this section, we present four examples to demonstrate how the proposed method can be applied in different scenarios. The first example is about a simple mediation analysis. The second example is on a multiple-mediator mediation model with a latent mediator. The third example involves mediation analysis in a multiple group analysis setting. The fourth examples shows the power analysis for longitudinal mediation models.

Example 1. Simple Mediation Analysis

In this example, the model with its population parameter values in Figure 2 is used to explore whether the relationship between mothers' education (ME) and children's mathematical achievement (math) is mediated by home environment (HE; Zhang & Wang, 2013b). Through this example, we demonstrate the difference in power for normal and non-normal data. In generating the non-normal data, the skewness is set at -0.3, -0.7, and 1.3 and the kurtosis is set at 1.5, 0, and 5 for ME, HE, and math, respectively. The skewness and kurtosis statistics are determined according to real data used in Zhang and Wang (2013). The sample size of 50 and 100 is investigated. The focus is the mediation effect `ab`. Complete R code for the analysis can be found in Appendix A.

The statistical power for detecting the mediation effect at the significance level

$\alpha = 0.05$ in this example is given in Table 1. Like a typical power analysis, power increased with sample size regardless of the normality of the data. It should be noted that non-normality may not necessarily reduce the power to detect significance of mediation effect. In this example, when data are non-normal, the power actually increased. The results are consistent with the previous literature on non-normal data with excessive kurtosis (e.g., Yuan et al., 2005).

Type I error is also investigated in this example. Because $ab = 0$ has different indications on the value of a and b, we evaluate the influence of different combinations of them. The results shows that when the magnitude of a or b is small, e.g., < 0.14, the empirical type I error is smaller than 0.05. On the other hand, when either a or b is large, the method based on the percentile bootstrap confidence interval tends to reject the null hypothesis more. The results here are consistent with Fritz et al. (2012).

Example 2. Mediation Analysis with a Latent Mediator (Power Curve)

A power curve is useful to graphically display how power changes with sample size (e.g., Zhang & Wang, 2009). Using the model shown in Figure 3, we show how to generate a power curve. The substantive idea of the model in Figure 3 is that the relationship between age and eduction and the performance on the everyday problem solving test (ept) is mediated by the memory ability measured by the Hopkins Verbal Learning Test (hvltt) and the reasoning ability measured by three reasoning tests including word series (ws), letter sets (lt), and letter series (ls) tests (see Zhang & Wang, 2013b). The population model parameters are also displayed in the figure. The R code in Appendix B generates the power curve in Figure 4. The power curve displays the power in detecting the effect of age and education on ept that is mediated by hvltt (`a*b+c*b`) for sample size from 100 to 1900 with an interval of 200. The plot shows that to get a power 0.8, a sample size about 1,500 is needed. Note that a power curve can be used to obtain power for a given sample size through interpolation, although the results might not be as accurate.

Example 3. Multiple Group Mediation Analysis (Moderated Mediation)

Thoemmes et al. (2010) considered a multiple group mediation model shown in Figure 5. Different from the simple mediation model in Figure 2, the mediator m is measured as a latent variable by three observed variables, m1, m2, and m3. Furthermore, two groups are considered with varying mediation effects. Specifically, the mediation effect for the first group is `a1*b1 = 0.26` and for the second group is `a2*b2 = 0.10`. This implies a moderated mediation because the mediation effects are different for the two groups. The moderated mediation can be evaluated using `a1*b1 - a2*b2`. The sample size for the first group is 400 and for the second group 200. The R code for this analysis is given in Appendix C.

Power for this example is given in Table 2. Comparing the power from Thommes et al. (2010), the power for `med2 = a2*b2` increased while the power for `diffmed = a1*b1 - a2*b2` decreased. The following reasons might explain the difference. First, x is binary in Thommes et al. but is continuous in the current study. Second, a close look at the bootstrap distributions revealed that the bootstrap distribution of `med2` was right-skewed and the bootstrap distribution of `diffmed` was left-skewed. Thommes et al. (2010) used the Sobel test that assumed the distribution of the indirect effects is normal while the bootstrap method does not require such an assumption.

Example 4. A Longitudinal Mediation Model

Maxwell & Cole (2007) has recommended the use of longitudinal mediation models in mediation analysis because of the involvement of causal process in mediation. Figure 6 is a longitudinal mediation model derived from Figure 3 of Maxwell and Cole (2007) with population parameter values calculated from Table 2 of Maxwell and Cole. In this example, each variable in the mediation model is measured three times repeatedly. The idea of longitudinal mediation is that the input variable at time 1 influences the mediator at time 2 which in turn affects the outcome variable at time 3. The mediation effect is then

measured by `a*b` as in the cross-sectional mediation models. The power for the mediation effect `a*b` is calculated with the code in Appendix 4. The power is 0.860 when the bootstrap method is utilized for a sample size of 50.

Discussion

In this study, we proposed to conduct power analysis for mediation models based on the bootstrap method. Specifically, the significance of the mediation effect is evaluated using the percentile bootstrap confidence interval. The proposed method is implemented in the free, open-source R package `bmem`. The use of the method is illustrated through four examples that cover a large variety of mediation models. The bootstrap method is recommended to use especially when data are not normally distributed, e.g., with excessive skewness and kurtosis.

The proposed method is computationally intensive because of the involvement of the Monte Carlo simulation and bootstrap. For example, for a power analysis with 1000 replications of Monte Carlo simulation and 1000 times of bootstrap, a total of one million models have to be estimated and evaluated. In order to take advantage of modern hardware such as multi-core processors, the package `bmem` implements automatic parallelization algorithms. Figure 7 displays the computing time along with the number of cores used on our desktop. Clearly, the computing time can be significantly reduced when multiple cores are used. Furthermore, the parallel method is very efficient because the computing time reduces almost linearly with the increase of the number of cores.

The bootstrap method requires about B times computing time of the normal or robust method, where B is the number of bootstraps. Furthermore, with the same sample size, the bootstrap method often has greater power. Therefore, in practice, one can first calculate power using the normal method to determine a rough sample size. Then, the bootstrap method can be carried out with a smaller sample size than the normal method. In this way, one can save a significant amount of computing time.

Although we have focused our discussion on the mediation models, the method and software in this study can be used to conduct power analysis for structural equation models as well. The calculation of the power using the Monte Carlo method needs the estimation of a mediation model. Therefore, after data collection, the same model can be estimated in R without the need of additional statistical software.

In the future, we will improve our method and software in the following ways. First, a better model estimation algorithm will be utilized to save computing time. For example, von Oertzen & Brick (2013) developed an efficient method that can significantly reduce the time of power analysis proposed in this study. Second, missing data are always a problem in practical power analysis (Zhang & Wang, 2009). The current method assumes the data collected will be complete. In the future, we will incorporate missing data in power calculation. Third, the algorithm for generating non-normal data will be improved. One limitation of the method developed by Vale & Maurelli (1983) is that it cannot generate non-normal data with all possible combinations of skewness and kurtosis. Fourth, the current study focuses on one type of non-normal data, namely, continuous data with excessive skewness and kurtosis. In the future, other types of non-normal data, for example, categorical data, count data, and survival data will be investigated.

References

Beasley, T. M. (2012). Power of product tests of mediation as a function of mediator collinearity. *Multiple Linear Regression Viewpoints*, *38*, 17–23.

Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling*, *14*(2), 227–246.

Daly, L. E. (1991). Confidence intervals and sample sizes: Don't throw out all your old sample size tables. *BMJ: British Medical Journal*, *302*, 333–336.

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

Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, *47*, 61–87.

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*, *in press*.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power. *The American Statistician*, *55*, 19–24.

Kenny, D. A., & Judd, C. M. (2013). Power anomalies in testing mediation. *Psychological Science*, *in press*. (R code at http://www.davidakenny.net/progs/PowMedR.txt)

Knaus, J. (2013). snowfall: Easier cluster computing (based on snow). [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=snowfall` (R package version 1.84-4)

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

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), 99–128.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*, 23–44.

Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393–425). Mahwah, NJ: Lawrence Erlbaum Associates.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156–166.

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus user's guide* (Sixth ed.). Los Angeles, CA: Muthén and Muthén. (http://www.statmodel.com)

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*(4), 599–620.

Preacher, K. J., & Hayes, A. F. (2004). Spss and sas procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731.

Qiu, W. (2013). powermediation: Power/sample size calculation for mediation analysis [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=powerMediation` (R package version 0.1.3)

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

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

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

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.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal disttributions. *Psychometrika*, *48*, 465–471.

Vittinghoff, E., Sen, S., & McCulloch, C. (2009). Sample size calculations for evaluating mediation. *Statistics In Medicine*, *28*, 541–557.

von Oertzen, T., & Brick, T. (2013). Efficient hessian computation using sparse matrix derivatives in ram notation. *Behavior Research Methods*, *in press*.

Wang, C., & Xue, X. (2012). Power and sample size calculations for evaluating mediation effects in longitudinal studies. *Statistical Methods in Medical Research*, *in press*.

Wang, L., & Zhang, Z. (2011). Estimating and testing mediation effects with censored data. *Structural Equation Modeling*, *18*(1), 18–34.

Yuan, K.-H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis. *Sociological Methods & Research*, *34*(2), 240–258. Retrieved from `http://dx.doi.org/10.1177/0049124105280200` doi: 10.1177/0049124105280200

Zhang, Z., & Wang, L. (2009). Statistical power analysis for growth curve models using sas. *Behavior Research Methods*, *41*, 1083–1094.

Zhang, Z., & Wang, L. (2013a). bmem: Mediation analysis with missing data using bootstrap [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=bmem` (R package version 1.5)

Zhang, Z., & Wang, L. (2013b). Methods for mediation analysis with missing data. *Psychometrika*, *78*, 154–184.

Zu, J., & Yuan, K. H. (2010). Local influence and robust procedures for mediation analysis. *Multivariate Behavioral Research*, *45*, 1–44.

Appendix A. R Code for Example 1

```
library(bmem)

ex1model<-'
math ~ c*ME+start(0)*ME + b*HE+start(.39)*HE
HE ~ a*ME+start(.39)*ME
'

indirect<-'ab:=a*b'

boot.normal<-power.boot(ex1model, indirect, 50, nrep=2000,
   nboot=2000, parallel='multicore')
summary(boot.normal)

boot.non.normal<-power.boot(ex1model, indirect, 100, nrep=2000,
    nboot=2000, parallel='multicore', skewness=c(-.3, -.7, 1.3)
   , kurtosis=c(1.5, 0, 5), ovnames=c('ME', 'HE', 'math'))
summary(boot.non.normal)
```

Appendix B. R Code for Example 2

```
library(bmem)

ex2model<-'
 ept ~ start(.4)*hvltt + b*hvltt + start(0)*age + start(0)*edu
    + start(2)*R
 hvltt ~ start(-.35)*age + a*age + c*edu + start(.5)*edu
```

```
 R ~ start(-.06)*age + start(.2)*edu
 R =~ 1*ws + start(.8)*ls + start(.5)*lt
 age ~~ start(30)*age
 edu ~~ start(8)*edu
 age ~~ start(-2.8)*edu
 hvltt ~~ start(23)*hvltt
 R ~~ start(14)*R
 ws ~~ start(3)*ws
 ls ~~ start(3)*ls
 lt ~~ start(3)*lt
 ept ~~ start(3)*ept
'

indirect<-'ind1 := a*b + c*b'

nobs <- seq(100, 2000, by=200)

power.curve(model=ex2model, indirect=indirect, nobs=nobs, type
   ='boot', parallel='multicore')
```

Appendix C. R Code for Example 3

```
library(bmem)

ex3model<-"
y ~ start(c(.283, .283))*x + c(c1,c2)*x + start(c(.36, .14))*m
   +c(b1,b2)*m
m ~ start(c(.721, .721))*x + c(a1,a2)*x
m =~ c(1,1)*m1 + start(c(.8,.8))*m2 + start(c(.8,.8))*m3
x ~~ start(c(.25, .25))*x
y ~~ start(c(.81, .95))*y
m ~~ start(c(.87, .87))*m
m1 ~~ start(c(.36, .36))*m1
m2 ~~ start(c(.36, .36))*m2
m3 ~~ start(c(.36, .36))*m3
"

indirect<-'
med1 := a1*b1
med2 := a2*b2
diffmed := a1*b1 - a2*b2
'

bootstrap<-power.boot(ex3model, indirect, nobs=c(400,200), nrep
   =2000, nboot=1000, parallel='multicore')
```

```
summary(bootstrap)
```

Appendix D. R Code for Example 4

```
ex4model<-'
x2 ~ start(.9)*x1 + x*x1
x3 ~ start(.9)*x2 + x*x2
m2 ~ start(.3)*x1 + a*x1 + start(.3)*m1 + m*m1
m3 ~ start(.3)*x2 + a*x2 + start(.3)*m2 + m*m2
y2 ~ start(.3)*m1 + b*m1 + start(.7)*y1 + y*y1
y3 ~ start(.3)*m2 + b*m2 + start(.7)*y2 + y*y2 + start(0)*x1 +
   c*x1
x1 ~~ start(.37)*m1
x1 ~~ start(.27)*y1
y1 ~~ start(.2278)*m1
x2 ~~ start(.19)*x2
x3 ~~ start(.19)*x3
m2 ~~ start(.7534)*m2
m3 ~~ start(.7534)*m3
y2 ~~ start(.3243)*y2
y3 ~~ start(.3243)*y3
'

indirect<-'ab:=a*b'

bootstrap<-power.boot(ex4model, indirect, nobs=50, nrep=1000,
   nboot=1000, parallel='multicore', ncore=60)
summary(bootstrap)
```

Table 1

Power ($a = b = .39$) and Type I error ($ab = 0$) in detecting the mediation effect in Example 1. The empirical type I error is different for different combination of a and b.

		N=50		N=100	
a	b	Normal Data	Non-normal data	Normal Data	Non-normal data
0.39	0.39	0.509	0.574	0.928	0.954
0	0	0.001	0.002	0	0.003
0	0.14	0.004	0.01	0.011	0.016
0	0.39	0.036	0.05	0.061	0.055
0	0.59	0.07	0.074	0.063	0.062
0.14	0	0.004	0.007	0.01	0.01
0.39	0	0.037	0.054	0.059	0.062
0.59	0	0.061	0.073	0.06	0.056

Table 2

Power from Thoemmes et al. (2010) and current analysis

	med1=a1*b1	med2=a2*b2	diffmed=a1*b1 - a2*b2
Thoemmes et al. 2010	0.99	0.37	0.53
Percentile bootstrap	1	0.411	0.475

Note. med1 and med2 are mediation effects of the first and second group, respectively. diffmed is the difference in the mediation effects of the two groups.

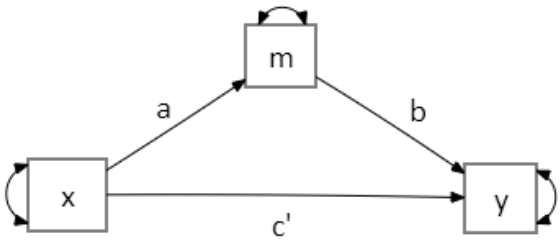

Figure 1. Path diagram of a simple mediation model.

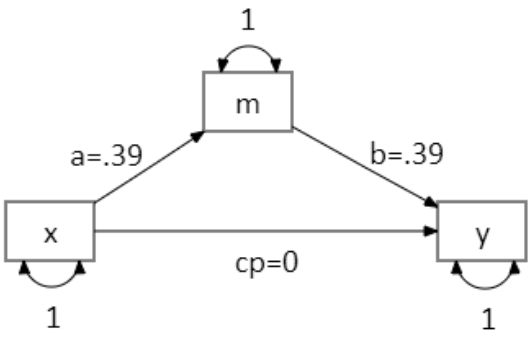

Figure 2. An example mediation model with population parameters

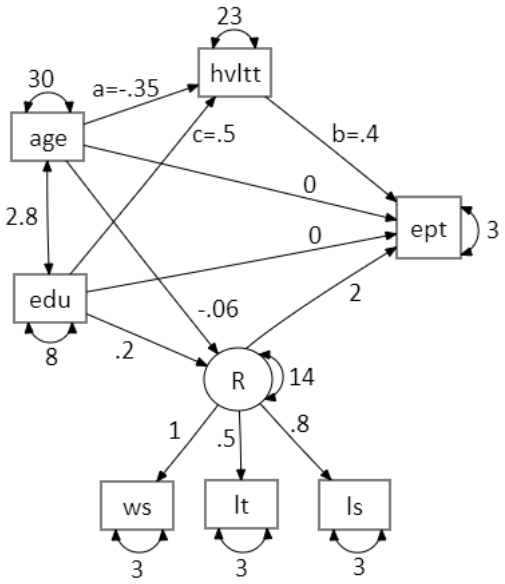

Figure 3. A multiple-mediator mediation model with population parameter values used in Example 2.

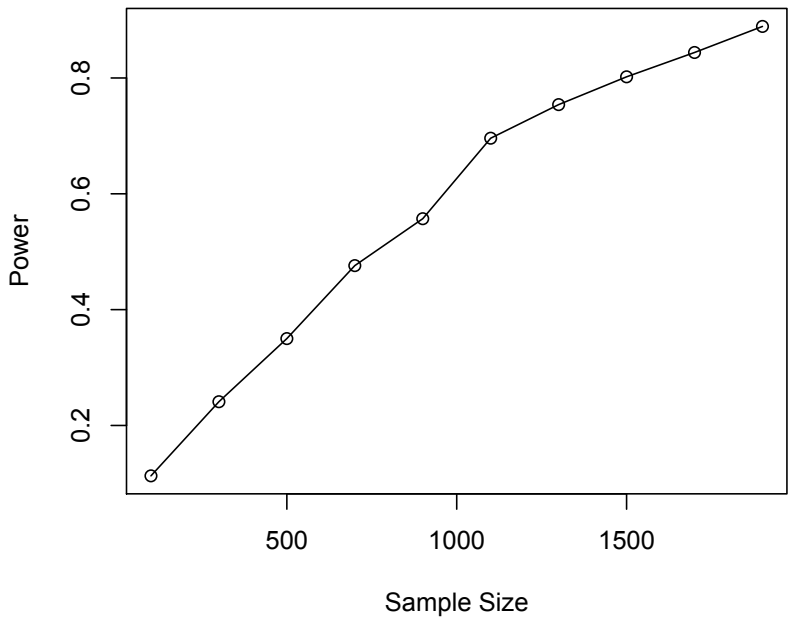

Figure 4. Power curve for testing the mediation effect in Example 2. To get a power 0.8, a sample size around 1500 is needed based on the power curve.

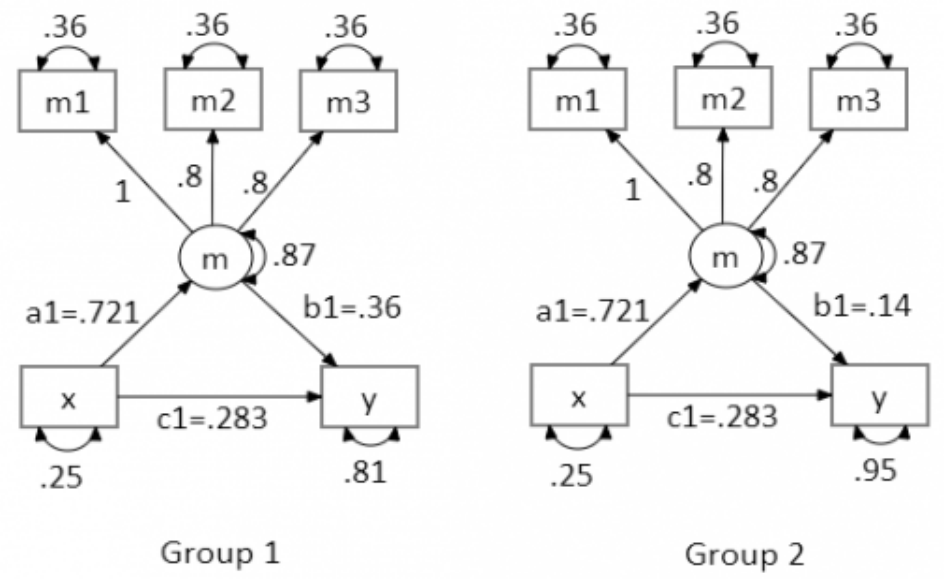

Figure 5. The path diagram of a multiple group mediation model with population parameter values.

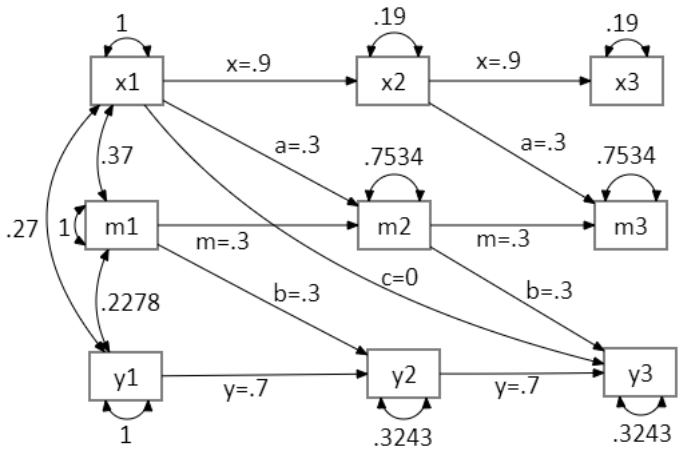

Figure 6. The path diagram for a longitudinal mediation model with population parameter values.

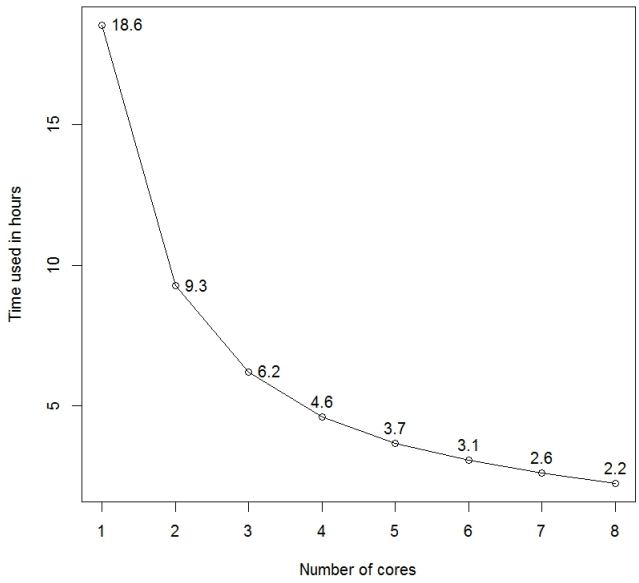

Figure 7. Computing time along with the number of CPU cores utilized for the mediation model in Example 1.