Sample size determination for mediation analysis of longitudinal data

Haitao Pan1, Suyu Liu2, Danmin Miao3* and Ying Yuan2*

Abstract

Background: Sample size planning for longitudinal data is crucial when designing mediation studies because sufficient statistical power is not only required in grant applications and peer-reviewed publications, but is essential to reliable research results. However, sample size determination is not straightforward for mediation analysis of longitudinal design.

Methods: To facilitate planning the sample size for longitudinal mediation studies with a multilevel mediation model, this article provides the sample size required to achieve 80% power by simulations under various sizes of the mediation effect, within-subject correlations and numbers of repeated measures. The sample size calculation is based on three commonly used mediation tests: Sobel’s method, distribution of product method and the bootstrap method.

Results: Among the three methods of testing the mediation effects, Sobel’s method required the largest sample size to achieve 80% power. Bootstrapping and the distribution of the product method performed similarly and were more powerful than Sobel’s method, as reflected by the relatively smaller sample sizes. For all three methods, the sample size required to achieve 80% power depended on the value of the ICC (i.e., within-subject correlation). A larger value of ICC typically required a larger sample size to achieve 80% power. Simulation results also illustrated the advantage of the longitudinal study design. The sample size tables for most encountered scenarios in practice have also been published for convenient use.

Conclusions: Extensive simulations study showed that the distribution of the product method and bootstrapping method have superior performance to the Sobel’s method, but the product method was recommended to use in practice in terms of less computation time load compared to the bootstrapping method. A R package has been developed for the product method of sample size determination in mediation longitudinal study design.

Keywords: Sample size determination, Mediation analysis, Longitudinal study

Background

Mediation analysis is a statistical method that helps researchers to understand the mechanisms underlying the phenomena they study. It has broad application in psychology, prevention research, and other social sciences. A simple mediation framework (see Fig. 1) involves three variables: the independent variable, dependent variable and mediating variable [4, 27]. The aim of mediation analysis is to determine whether the relation between the independent and dependent variables is due, wholly or in part, to the mediating variables. Since the seminal work of Baron and Kenney [4], extensive research has been conducted in mediation analysis, including that of [7, 22, 25]; [34]; and [18], among others. A comprehensive review of mediation analysis can be found in the book by [27].

When planning a mediation study, the investigator commonly determines the required sample size. An appropriately chosen sample size is critical for the success of the study. If the sample size is too small, the study may lack adequate statistical power to detect an effect size of practical importance, which leads the investigator to incorrectly conclude that an efficacious intervention is inefficacious. Reviews of the
psychological literature suggest that insufficient statistical power is a common problem in psychological studies [1, 29, 30]. On the other hand, an unnecessarily large sample size is wasteful and increases the duration of the study. Because of the importance of sample size, funding agencies such as the National Institutes of Health routinely require investigators to justify the sample size for funded projects.

Unfortunately, sample size determination is not straightforward for mediation analysis. No simple formula is available to carry out this task. Using Monte Carlo simulations, Fritz and MacKinnon [14] investigated power calculations for the simple mediation model and provided guidance in choosing sample sizes for mediation studies with independent data. Their results, however, are not applicable to longitudinal studies, in which data are correlated.

A longitudinal study design is common in psychological and social research [13]. Compared with a cross-sectional study design, the longitudinal design requires fewer subjects and allows investigators to study the trajectory of each subject. In longitudinal studies, repeated measures are collected from each subject over time. Since measures collected from the same subject are more likely to be similar when compared to those collected from other subjects, data from the same subject tend to be correlated. Analyzing such correlated data requires special statistical methods, such as the multilevel model [33]. In this article, assuming a multilevel mediation model and using Monte Carlo simulation, we investigate sample size determination for longitudinal mediation studies. Our objective is to provide practical guidance and easy-to-use R software to help researchers determine the sample size when designing longitudinal mediation studies.

**Methods**

This section starts by formulating single-level mediation model, then multilevel mediation model for longitudinal data is described. We focus on lower-level multilevel mediation model and relevant model assumptions are discussed.

**Simple single-level mediation model**

Let $Y$ denote the dependent (or outcome) variable, $X$ denote the independent variable, and $M$ denote the mediating variable (or mediator). A single-level mediation model (Fig. 1) can be expressed in the form of three regression equations:

\[
Y = \beta_0 + \beta_c X + \epsilon_1 \quad (1)
\]

\[
Y = \beta_0 + \beta_c X + \beta_M M + \epsilon_2 \quad (2)
\]

\[
M = \beta_0 + \beta_c X + \epsilon_3 \quad (3)
\]

where $\beta_c$ quantifies the relation between the independent variable and dependent variable (i.e., the total effect of $X$ on $Y$); $\beta_M$ quantifies the relation between the independent variable and dependent variable after adjusting for the effect of the mediating variable (i.e., the direct effect of $X$ on $Y$ adjusted for $M$); $\beta_c$ quantifies the relation between the mediating variable and dependent variable after adjusting for the effects of the independent variable; $\beta_M$ measures the relation between the independent variable and mediating variable; $\beta_{01}$, $\beta_{02}$, and $\beta_{03}$ are intercepts; and $\epsilon_1$, $\epsilon_2$, and $\epsilon_3$ are error terms that follow normal distributions with mean 0 and respective variances of $\sigma_1^2$, $\sigma_2^2$, and $\sigma_3^2$.

The mediation effect can be defined by two ways: $\beta_c - \beta_M$ and $\beta_M \beta_c$ [16, 17, 27]. For the single-level mediation model, the two definitions of the mediation effect are equivalent [28], but they are generally different in the multilevel mediation models we will describe.

**Multilevel mediation model for longitudinal data**

For correlated longitudinal data, the simple mediation model, which assumes independence of observations, is not appropriate. Using the single-level mediation model for longitudinal data leads to biased estimates of standard errors and confidence intervals [3].

Multilevel mediation modeling is a powerful technique for analyzing mediation effects in longitudinal data. Multilevel models assume that there are at least two levels in the data, an upper level and a lower level. The lower-level units (e.g., repeated measures) are often nested within the upper-level units (e.g., subjects). Assuming that the lower-level units are random, also known as random effects, multilevel models appropriately account for correlations.
among the observations from the same subject, and yield valid statistical inference. For a comprehensive coverage of multilevel modeling techniques, see the book by Raudenbush & Bryk [33].

The multilevel mediation model is much more complex than the single-level model because mediation effects can occur at the different model levels. Two kinds of mediation, upper-level mediation and lower-level mediation, can be distinguished in the context of multilevel mediation models [5]. In upper-level mediation, the initial causal variable for which the effect is mediated is an upper-level variable. In lower-level mediation, the mediator is a lower-level variable. Krull [21] and MacKinnon [22] offered examples of upper-level mediation, while [18] studied lower-level mediation, in which the mediation links varied randomly across the upper-level units. In this study, we focus on a specific type of lower-level mediation model (Fig. 2) that is appropriate for analyzing longitudinal studies. In this model, an initial variable X is mediated in the lower level (i.e., measurement level), but the mediator M and outcome Y are affected by upper-level (i.e., subject level) variations. A simple scenario for this model is a longitudinal experimental study in which subjects are randomly assigned to a treatment (time-invariant) or the multiple treatments can be assigned to a same subject in cross-over design (i.e., initial variable X, in this paper, variable X is treated as time-varying), and mediating variable M, such as a psychosocial measure, is believed to change individual behavior (i.e., dependent variable Y) over time.

The lower-level mediation model
Let \( X_{ij} \), \( Y_{ij} \) and \( M_{ij} \) denote the independent variable, dependent variable, and mediating variable, respectively, for the \( i \)th observation from the \( j \)th subject. The lower-level mediation model in Fig. 2 can be expressed in the form of the following two-level regression equations,

Lower : \( Y_{ij} = \beta_{01j} + \beta_c X_{ij} + \epsilon_{1ij} \)  \hspace{1cm} (4)

Upper : \( \beta_{01j} = y_1 + u_{1j} \)  \hspace{1cm} (5)

Lower : \( Y_{ij} = \beta_{02j} + \beta_c X_{ij} + \beta_h M_{ij} + \epsilon_{2ij} \)  \hspace{1cm} (6)

Upper : \( \beta_{02j} = y_2 + u_{2j} \)  \hspace{1cm} (7)

Lower : \( M_{ij} = \beta_{03j} + \beta_a X_{ij} + \epsilon_{3ij} \)  \hspace{1cm} (8)

Upper : \( \beta_{03j} = \gamma_3 + u_{3j} \)  \hspace{1cm} (9)

where at the lower (or within-subject) level, similar to the simple single-level mediation model, \( \beta_c \) measures the total effect of the independent variable on the dependent variable; \( \beta_h \) measures the direct effect of the independent variable on the dependent variable, adjusted for the mediating variable; \( \beta_a \) measures the effect of the mediating variable on the dependent variable, adjusted for the independent variable; and \( \beta_{01j}, \beta_{02j}, \) and \( \beta_{03j} \) are subject-specific intercepts that differ from subject to subject, as reflected by the subscript \( j \) in these parameters. These subject-specific intercepts are also known as random intercepts. The terms \( \epsilon_{1ij}, \epsilon_{2ij}, \) and \( \epsilon_{3ij} \) are lower-level (or within-subject) error terms that follow normal distributions with a mean of zero and respective variances \( \sigma_1^2, \sigma_2^2, \) and \( \sigma_3^2 \). At the upper (or between-subject) level \( y_1, y_2, \) and \( y_3 \) are overall or population average intercepts; and \( u_{1j}, u_{2j}, \) and \( u_{3j} \) are upper-level (between-subject) error terms that follow normal distributions with a mean of zero and respective variances \( \tau_1^2, \tau_2^2, \) and \( \tau_3^2 \).

In the multilevel model, the upper-level errors induce within-subject correlations. Let \( y_i \) and \( y_j \) denote the i-
th and i-th measures for the same subject j, then y_{ij} and y_{ij}' are correlated as
\[
\text{cov}(y_{ij}, y_{ij}') = \text{cov}(\beta_{00} + \beta_1 x_i + \beta_2 M_i + \epsilon_{0j}, \beta_{00} + \beta_1 x_i + \beta_2 M_i + \epsilon_{0j}) = \text{cov}(\beta_{00}, \beta_{00}) = r^2_j.
\]

Such within-subject correlation is often measured by the intraclass correlation coefficient (ICC), which is defined as
\[
\text{ICC} = \frac{\text{within-subject covariance}}{\text{overall variance}}
\]
Under the above two-level mediation model, the value of ICC for Y is given by
\[
\text{ICC} = \frac{r^2_j}{\sigma^2 + r^2_j}. \tag{10}
\]
Larger values of ICC represent strong within-subject correlations, i.e., measures from the same subject are more similar. When ICC = 0, measures from the same subject are independent.

Due to the within-subject correlation, the two definitions of the mediation effects, $\beta_c - \beta_c'$ and $\beta_a \beta_{b0}$, are generally not equivalent in multilevel models [21], although they are equivalent in the single-level mediation model. The different behaviors of multilevel and single-level models are caused by the fact that the weighting matrix used to estimate the multilevel model is typically not identical to single-level equations. The non-equivalence between $\beta_c - \beta_c'$ and $\beta_a \beta_{b0}$, however, is unlikely to be problematic because the difference between the two estimates is typically small and unsystematic and tends to vanish at large sample sizes [21]. In this article, we focus on $\beta_a \beta_{b0}$ as the measure of the mediation effect.

Test of the mediation effect
As the independence assumption is violated, conventional statistical methods, such as the ordinary least squares method, are not appropriate for estimating the multilevel mediation model. Instead, maximum likelihood methods and/or empirical Bayes methods are typically used. Let $\hat{\beta}_a$ and $\hat{\beta}_b$ denote the maximum likelihood estimates of $\beta_a$ and $\beta_b$, respectively. The maximum likelihood estimate of the mediation effect is given by $\hat{\beta}_a \hat{\beta}_b$. To test whether the mediation effect $\beta_a \beta_{b0}$ equals zero, three approaches can be taken.

Sobel’s method
Sobel’s method is a widely used test of the mediation effect, based on the first-order multivariate delta method [35, 36]. In this approach, assuming $\hat{\beta}_a$ and $\hat{\beta}_b$ are independent, the standard deviation of $\hat{\beta}_a \hat{\beta}_b$ is estimated by
\[
\hat{s}_{\beta_a \beta_b} = \sqrt{s^2_{\hat{\beta}_a} \hat{\beta}_b^2 + \hat{s}^2_{\hat{\beta}_a} \beta_{b0}^2}, \tag{11}
\]
where $\hat{s}^2_{\hat{\beta}_a}$ and $\hat{s}^2_{\beta_{b0}}$ are the squared standard errors of $\hat{\beta}_a$ and $\hat{\beta}_b$, respectively. The 100(1-\(\alpha\))% confidence interval (CI) of the mediation effect is given by
\[
\left(\hat{\beta}_a \hat{\beta}_b - z_{1-\alpha/2} \hat{s}_{\beta_a \beta_{b0}} \beta_{b0}, \hat{\beta}_a \hat{\beta}_b + z_{1-\alpha/2} \hat{s}_{\beta_a \beta_{b0}} \beta_{b0}\right), \tag{12}
\]
where $z_{1-\alpha/2}$ is the (1 - \(\alpha/2\))th quantile of the standard normal distribution. If \(\alpha = 0.05\), the familiar 95% CI results. If this CI does not contain zero, we reject the null hypothesis and conclude that the mediation effect is statistically significant.

Sobel’s method relies on the assumption that $\hat{\beta}_a \hat{\beta}_b$, the product of two normal random variables $\hat{\beta}_a$ and $\hat{\beta}_b$, is normally distributed. However, several studies have shown that the distribution of the product of two normal random variables is not actually normal, but skewed [23]. The violation of the normality assumption compromises the performance of Sobel’s method and leads to invalid CIs [26]. To address this problem, [26] discussed several improved CIs that account for the fact that $\hat{\beta}_a \hat{\beta}_b$ is not normally distributed, including the CI based on the distribution of the product of two normal random variables and the CI based on the bootstrap method [6, 34].

Distribution of the product method
Instead of assuming the normality of $\hat{\beta}_a \hat{\beta}_b$, the distribution of the product method proposed by MacKinnon and Lockwood (2001) constructs the CI of the mediation effect based on the distribution of the product of two normal random variables. Although such a distribution does not take a simple closed form, Meeker et al. [31] provided tables of critical values for this distribution that can be used to construct the CI. Alternatively, the critical values can also be obtained based on the empirical distribution of the product of two normal random variables through Monte Carlo simulations. Let $\delta_{\text{lower}}$ and $\delta_{\text{upper}}$ denote critical values that correspond to the lower and upper bounds of the CI, then the CI of the mediation effect is given by
\[
\left(\hat{\beta}_a \hat{\beta}_b - \delta_{\text{lower}} \times \hat{s}_{\beta_a \beta_{b0}}, \hat{\beta}_a \hat{\beta}_b + \delta_{\text{upper}} \times \hat{s}_{\beta_a \beta_{b0}}\right). \tag{13}
\]

Bootstrap method
Another approach for constructing the CI without imposing a normal assumption on $\hat{\beta}_a \hat{\beta}_b$ is the bootstrap method [11]. The bootstrap method, based on resampling, is useful for finding the standard error and forming CIs for estimates when their sampling distributions
are unknown. In this study, we use the percentile bootstrap [6] to construct the CI for the mediation effect. We repeatedly resample the original data with replacement, obtaining the so-called bootstrap samples. For each of the bootstrap samples, we estimate the mediation effect using the maximum likelihood method. These estimates form the empirical distribution of the mediation effect. Let \( q_{a/2} \) and \( q_{1-a/2} \) denote the \( (a/2) \)th and \( (1-a/2) \)th percentiles of this empirical distribution; then the \( 100(1-a)\% \) CI of the mediation effect is given by

\[
\left( q_{a/2}, q_{1-a/2} \right).
\]

When conducting bootstrap resampling for the multilevel mediation model, in principle, we should resample both the upper-level (subjects) and lower-level (measures) units. However, in a multilevel context, we should be careful of not breaking the structure of the dataset, therefore, a resampling scheme for multilevel models must take into account the hierarchical data structure. There are three approaches can be applied to bootstrap two-level models: the parametric bootstrap, the residual bootstrap, and the cases bootstrap. We chose the cases bootstrap since it requires minimal assumptions of hierarchical dependency in the data being assumed to be specified correctly. de Leeuw & Meijer [9] suggest that when the number of lower-level units (measures) is small, the approach of resampling only the upper level when the number of lower-level units (measures) is specified correctly. de Leeuw & Meijer [9] suggest that the bootstrap is as follows:

1. Draw a sample of size \( J \) with replacement from the upper level; that is, draw a sample \( \{ j_k, k = 1, \ldots, J \} \) (with replacement) of upper level numbers.
2. For each \( k \), draw a sample of entire cases, with replacement, from (the original) upper level unit \( j = j_k \). This sample has the same size \( n_k = n_{j_k} = n_j \) as the original unit from which the cases are drawn. Then, for each \( k \), we have a set of data \( \{ Y_{ik}, X_{ik}, M_{ik} \}; i = 1, \ldots, n_k \} \).
3. Compute estimates for all parameters of the two-level model.
4. Repeat steps 1–3 \( B \) times.

**Simulation study**

We conducted a simulation study to determine the sample size that is needed to achieve 80% power when using Sobel’s method, the distribution of the product method, and the bootstrap method for longitudinal mediation studies. In our simulation, we varied three factors. The first one is the effect size of the mediation effect \( \hat{\beta}_a \hat{\beta}_b \).

We considered four values of \( \beta_a \) and \( \beta_b \): 0.14, 0.26, 0.39 and 0.59, respectively corresponding to smaller, medium, halfway (between medium and large), and large effect sizes. These values yielded 16 combinations of effect sizes of the mediation effect. Another factor is the ICC. We considered five values of ICC, 0.1, 0.3, 0.5, 0.7 and 0.9, to cover various within-subject correlations from low to high. The last factor is the number of repeated measures. We considered 2, 3, 4 and 5 repeated measures for each subject. For other parameters, we set the overall intercepts \( \gamma_2 \) and \( \gamma_3 \) as zero. Since there were no repeated measurements in Fritz et al. [14] and the samples were all drawn from a standard normal distribution, for fair comparisons, we set marginal variances of \( Y_{ij} \) and \( M_{ip} \) that is, \( \sigma_2^2 + r_2^2 \) and \( \sigma_3^2 + r_3^2 \), as 1. Based on the definition of ICC, we have \( r_2^2 = r_3^2 = ICC \).

To simulate data, we first simulated the independent variable \( X \) from the standard normal distribution, then generated random intercepts \( \beta_{02j} \) and \( \beta_{03j} \) according to eqs. (7) and (9). Conditional on the values of \( \beta_{02j} \) and \( \beta_{03j} \), we generated the dependent variable \( Y \) and mediating variable \( M \) according to eqs. (6) and (8).

To determine the power of the three test methods, under each of the parameter settings, we generated 1000 simulated datasets, and applied the methods to each of the datasets to test the mediation effect. We calculated the power of the methods as the proportion of tests that rejected the null hypothesis of no mediation effects, i.e., the CI excluded zero. For the bootstrap method, we based the construction of the CI on 500 bootstrap samples.

To determine the sample size that yields 80% power, we started with an initial guess of the sample size. If we found the power achieved with that sample size to be too low, we increased the sample size; and if we found the power to be too high, we decreased the sample size. We repeated this procedure until the sample size allowed us to reach the level of power nearest to 80%.

**Results**

Tables 1, 2, 3, 4 and 5 show the sample sizes necessary to achieve 80% power under five different ICCs (ICC = 0.1, 0.2, 0.4, 0.6, 0.9). For completeness, results with other ICCs, say, 0.3, 0.5, 0.7, and 0.8, are also shown, which can be found in the Additional file 1: Tables S1-S4, respectively. In each table, the 16 mediation effect sizes are denoted by two letters, with the first one referring to the size of \( \beta_a \), and the second letter referring to the size of \( \beta_b \). We use S for small (0.14), M for medium (0.39), L for large (0.59) and H for halfway (0.26) between large and medium effect sizes, e.g., the effect size ML indicates \( \beta_a = 0.39 \) and \( \beta_b = 0.59 \). Among the three methods of testing the mediation effects, Sobel’s method required the largest sample size to
achieve 80% power. Bootstrapping and the distribution of the product method performed similarly and were more powerful than Sobel’s method, as reflected by the relatively smaller sample sizes. For instance, when the mediation effect size was medium (i.e., SM) and the ICC was 0.2, with 4 repeated measures, Sobel’s method required 191 subjects to achieve 80% power, whereas the distribution of the product and bootstrap methods required 188 and 185 subjects, respectively, to achieve the same power.

For all three methods, the sample size required to achieve 80% power depended on the value of the ICC (i.e., within-subject correlation). A larger value of ICC typically required a larger sample size to achieve 80% power. For example, under the design with two repeated measures and using the distribution of the product method, to detect a small effect size of SS, a sample size of 299 was needed when ICC = 0.1, while a sample size of 420 was needed when ICC = 0.4.

Simulation results also illustrated the advantage of the longitudinal study design. Compared with the results reported by Fritz and MacKinnon [14] for the cross-sectional study, the required sample size under the longitudinal design was substantially smaller. When the ICC was low, such as 0.1, the required sample size under the longitudinal study design was a fraction of that under the cross-sectional design, and was approximately equal to the sample size of the cross-sectional study divided by the number of repeated measures. For example, under the longitudinal design with three repeated measures and using the distribution of the product method, the sample size under the longitudinal design was 215 to detect a small effect size of SS, which was approximately one-third of that required under the cross-sectional design (667). Even when the ICC was relatively high, we still observed dramatic sample size savings. For example, when ICC = 0.6 and using the bootstrap method, to detect the mediation effect size SM, the cross-sectional design required 422 subjects, while the longitudinal design with 4 repeated measures only required 351 subjects. This observation is in accordance to findings in literatures [19].

Figure 3 shows the type I error rates for the sample sizes corresponding to 5 examples of zero mediation effects when ICC = 0.3 for three repeated measures. A parameter combination of zero/zero (ZZ) had error rates around zero for all numbers of observations and sample sizes across the mediation tests. The distribution of the product method had the most precise rates; whereas Sobel’s method had less type I error probability and bootstrapping inflated the error rates in the case of a zero/0.59 (ZL) parameter, as with small sample sizes. However, the rates approached 0.05 when the number of sample sizes increased. Other scenarios taking various ICCs and repeated measures showed results similar to those in Fig. 3 and they are not shown in the paper.

Table 1 Estimated numbers of required subjects for 2, 3, 4 and 5 observations with ICC = 0.1

| Effect size | Observations |
|-------------|--------------|
|             | 2            | 3            | 4            | 5            |
| SS          | 365          | 299          | 304          | 257          | 215          | 215          | 207          | 163          | 169          |
| SH          | 272          | 235          | 237          | 194          | 172          | 180          | 151          | 140          | 150          |
| SM          | 248          | 226          | 230          | 188          | 188          | 200          | 146          | 144          | 145          |
| SL          | 238          | 248          | 251          | 176          | 176          | 176          | 148          | 147          | 149          |
| HS          | 238          | 201          | 209          | 161          | 138          | 143          | 123          | 104          | 108          |
| HH          | 109          | 88           | 94           | 78           | 65           | 69           | 61           | 50           | 57           |
| HM          | 87           | 77           | 99           | 58           | 55           | 57           | 49           | 42           | 52           |
| HL          | 74           | 72           | 73           | 57           | 51           | 25           | 47           | 46           | 20           |
| MS          | 215          | 200          | 209          | 138          | 134          | 140          | 110          | 102          | 105          |
| MH          | 79           | 65           | 69           | 54           | 46           | 46           | 42           | 35           | 36           |
| MM          | 51           | 41           | 44           | 38           | 30           | 33           | 29           | 23           | 28           |
| ML          | 40           | 36           | 38           | 29           | 27           | 28           | 24           | 22           | 25           |
| LS          | 204          | 206          | 204          | 139          | 132          | 140          | 105          | 101          | 103          |
| LH          | 65           | 60           | 69           | 44           | 41           | 40           | 34           | 31           | 32           |
| LM          | 36           | 31           | 33           | 25           | 22           | 24           | 19           | 17           | 18           |
| LL          | 24           | 20           | 22           | 17           | 16           | 15           | 14           | 13           | 12           |

*Effect size: The first letter is the size of $\beta_a$, the second letter is the size of $\beta_b$. S is small (0.14), M is medium (0.39), L is large (0.59) and H is halfway (0.26) between large and medium effect sizes.
Discussion
Assuming a two-level mediation model and using Monte Carlo simulations, we determined the sample sizes required to achieve 80% power for longitudinal mediation studies under various practical settings. The simulation results provide guidance for researchers when choosing appropriate sample sizes in the design of longitudinal mediation studies. Our simulations also show that the distribution of the product and bootstrap methods are more powerful than Sobel’s method for testing the mediation effect. In addition, the required sample size is closely related to the ICC. A high ICC generally requires

Table 2 Estimated numbers of required subjects for 2, 3, 4 and 5 observations with ICC = 0.2

| Observations |
|--------------|
| Effect size  |
| 2            |
| Sobel       |
| Product     |
| Bootstrap   |
| 3            |
| Sobel       |
| Product     |
| Bootstrap   |
| 4            |
| Sobel       |
| Product     |
| Bootstrap   |
| 5            |
| Sobel       |
| Product     |
| Bootstrap   |
| SS          | 408 | 330 | 341 |
| SH          | 301 | 282 | 283 |
| SM          | 294 | 287 | 279 |
| SL          | 282 | 267 | 278 |
| HS          | 240 | 206 | 224 |
| HH          | 120 | 97  | 104 |
| HM          | 95  | 87  | 90  |
| HL          | 88  | 85  | 85  |
| MS          | 213 | 194 | 202 |
| MH          | 81  | 68  | 73  |
| MM          | 56  | 46  | 50  |
| ML          | 47  | 38  | 37  |
| LS          | 215 | 189 | 204 |
| LH          | 66  | 60  | 65  |
| LM          | 38  | 32  | 33  |
| LL          | 25  | 22  | 22  |

Table 3 Estimated numbers of required subjects for 2, 3, 4 and 5 observations with ICC = 0.4

| Observations |
|--------------|
| Effect size  |
| 2            |
| Sobel       |
| Product     |
| Bootstrap   |
| 3            |
| Sobel       |
| Product     |
| Bootstrap   |
| 4            |
| Sobel       |
| Product     |
| Bootstrap   |
| 5            |
| Sobel       |
| Product     |
| Bootstrap   |
| SS          | 479 | 394 | 395 |
| SH          | 379 | 351 | 564 |
| SM          | 376 | 350 | 351 |
| SL          | 360 | 361 | 361 |
| HS          | 258 | 215 | 219 |
| HH          | 143 | 117 | 123 |
| HM          | 116 | 109 | 111 |
| HL          | 112 | 109 | 110 |
| MS          | 217 | 210 | 212 |
| MH          | 88  | 72  | 81  |
| MM          | 64  | 54  | 55  |
| ML          | 54  | 49  | 50  |
| LS          | 209 | 198 | 204 |
| LH          | 68  | 61  | 63  |
| LM          | 41  | 34  | 36  |
| LL          | 29  | 24  | 26  |
a larger sample size to detect a given effect size. The simulation results show that when the ICC is high, above 0.8 for instance, the required sample sizes in these scenarios are close to the values provided in Fritz et al. [14], suggesting that we should choose cross-sectional studies instead of longitudinal studies since the former is relatively easy to conduct but does not lose power. However, in real studies, especially in psychotherapy clinical trial studies, a meta-analysis of ICCs found that ICCs varied widely, ranging from 0 to 0.729, with an average around 0.08 [8]. Similar results have been found in clinical trial data [12, 20] and clinical practice data.

Table 4 Estimated numbers of required subjects for 2, 3, 4 and 5 observations with ICC = 0.6

| Effect size | Observations | 2 | 3 | 4 | 5 |
|-------------|--------------|---|---|---|---|
|             |              | Sobel | Product | Bootstrap | Sobel | Product | Bootstrap | Sobel | Product | Bootstrap | Sobel | Product | Bootstrap |
| SS          | 551          | 467 | 477 | 438 | 387 | 385 | 400 | 326 | 333 | 351 | 326 | 326 |
| SH          | 451          | 426 | 453 | 376 | 369 | 370 | 363 | 349 | 350 | 332 | 326 | 334 |
| SM          | 451          | 438 | 451 | 380 | 377 | 380 | 357 | 346 | 351 | 326 | 323 | 333 |
| SL          | 454          | 444 | 445 | 385 | 376 | 380 | 344 | 326 | 340 | 326 | 313 | 324 |
| HS          | 289          | 234 | 239 | 200 | 171 | 179 | 171 | 136 | 157 | 145 | 120 | 127 |
| HH          | 156          | 132 | 130 | 127 | 111 | 122 | 111 | 101 | 108 | 104 | 93  | 96  |
| HM          | 142          | 125 | 132 | 115 | 111 | 113 | 107 | 102 | 105 | 100 | 98  | 98  |
| HL          | 133          | 131 | 127 | 115 | 111 | 112 | 102 | 97  | 97  | 100 | 97  | 99  |
| MS          | 226          | 194 | 202 | 157 | 145 | 148 | 129 | 108 | 118 | 106 | 88  | 85  |
| MH          | 99           | 80  | 91  | 76  | 62  | 73  | 63  | 52  | 60  | 57  | 48  | 52  |
| MM          | 72           | 62  | 70  | 61  | 51  | 53  | 54  | 47  | 48  | 49  | 46  | 48  |
| ML          | 63           | 59  | 61  | 53  | 52  | 52  | 49  | 49  | 49  | 47  | 46  | 46  |
| LS          | 211          | 201 | 204 | 138 | 132 | 135 | 117 | 105 | 108 | 91  | 79  | 87  |
| LH          | 72           | 62  | 65  | 52  | 45  | 49  | 43  | 36  | 38  | 37  | 31  | 32  |
| LM          | 45           | 38  | 39  | 35  | 29  | 34  | 30  | 24  | 28  | 26  | 23  | 24  |
| LL          | 33           | 31  | 32  | 28  | 24  | 25  | 25  | 24  | 25  | 23  | 21  | 23  |

Table 5 Estimated numbers of required subjects for 2, 3, 4 and 5 observations with ICC = 0.9

| Effect size | Observations | 2 | 3 | 4 | 5 |
|-------------|--------------|---|---|---|---|
|             |              | Sobel | Product | Bootstrap | Sobel | Product | Bootstrap | Sobel | Product | Bootstrap | Sobel | Product | Bootstrap |
| SS          | 638          | 576 | 581 | 544 | 495 | 523 | 499 | 465 | 454 | 476 | 426 | 455 |
| SH          | 551          | 545 | 550 | 499 | 468 | 473 | 462 | 463 | 464 | 458 | 440 | 465 |
| SM          | 551          | 550 | 550 | 507 | 485 | 490 | 476 | 479 | 478 | 444 | 447 | 445 |
| SL          | 568          | 576 | 577 | 512 | 499 | 500 | 461 | 466 | 466 | 447 | 425 | 435 |
| HS          | 308          | 246 | 251 | 236 | 189 | 204 | 193 | 163 | 168 | 175 | 148 | 162 |
| HH          | 195          | 165 | 170 | 159 | 144 | 149 | 144 | 139 | 142 | 143 | 129 | 134 |
| HM          | 183          | 171 | 170 | 143 | 148 | 149 | 139 | 133 | 132 | 134 | 128 | 129 |
| HL          | 167          | 164 | 166 | 154 | 145 | 148 | 134 | 134 | 132 | 134 | 129 | 131 |
| MS          | 232          | 208 | 212 | 171 | 145 | 153 | 136 | 113 | 111 | 115 | 95  | 102 |
| MH          | 109          | 92  | 102 | 90  | 74  | 85  | 79  | 66  | 70  | 71  | 65  | 67  |
| MM          | 86           | 76  | 82  | 73  | 66  | 65  | 65  | 64  | 64  | 63  | 60  | 62  |
| ML          | 82           | 74  | 77  | 70  | 65  | 66  | 63  | 63  | 63  | 65  | 61  | 62  |
| LS          | 213          | 207 | 216 | 144 | 133 | 137 | 111 | 108 | 109 | 91  | 84  | 84  |
| LH          | 77           | 64  | 76  | 56  | 48  | 54  | 49  | 40  | 48  | 43  | 36  | 40  |
| LM          | 50           | 42  | 46  | 41  | 36  | 34  | 35  | 31  | 32  | 32  | 27  | 30  |
| LL          | 41           | 34  | 37  | 33  | 31  | 33  | 31  | 29  | 30  | 30  | 28  | 30  |
In studies in the field of education, small ICCs are also common [15], with 0.20 as a median value. Another interesting finding for multilevel mediation is that the power of testing the mediation effect depends on not only the overall value of the mediation effects $\beta_a \beta_b$, but also the values of the individual regression coefficients $\beta_a$ and $\beta_b$. For instance, the sample size required to detect the effect size of LS is different from that required to detect the effect size SL. In other words, the sample size depends on the position of the effect sizes. Such a “positioning” effect for testing the mediation effect in multilevel mediation depends on the ICC. A high ICC leads to a stronger positioning effect. For example, in Table 5, when ICC = 0.9, detecting the effect size SL requires 568 subjects, while detecting the effect size LS only requires 213 subjects. The positioning effect does not appear in the single-level mediation model, which can be viewed as an extreme case of the multilevel model with ICC = 0. In the single-level mediation model, the required sample size (or power) only depends on the position of the effect sizes.

Fig. 3 Type I error rates of Sobel’s (black line), distribution of the product (red line), and bootstrap (blue line) methods under various sample sizes with 3 repeated measures and ICC of 0.3.
value of $\beta_a \beta_b$, but not the individual values of $\beta_a$ and $\beta_b$ [14]. For example, the number of subjects needed to detect the effect size $LS$ was equal to that required to detect the effect size $SL$. The different behavior of multilevel mediation compared to single-level mediation is due to the within-cluster correlation in the multilevel model. Therefore, when conducting power calculations for longitudinal mediation studies, in addition to the mediation effect $\beta_a \beta_b$, it is equally important to report the effect size of $\beta_a$ and $\beta_b$.

Our simulation studies showed that the bootstrap and the distribution of the product methods have similar performance in testing the mediation effect. However, as the bootstrap is much more computer-intensive and time-consuming, we recommend using the distribution of the product method in practice. One limitation is that in the paper, coefficients $\beta_a$, $\beta_b$, $\beta_a$ and $\beta_c$ in the model were treated as fixed-effects coefficients only. More flexible model by treating these as random-effects variables and two-level random-slopes model can also be considered. Another limitation is that in practice, effects size estimates are just estimates, not the true values, so uncertainty needs to be considered in the effect size estimates for sample size planning. Interested readers can consult the papers by [2, 10] for more information. There is a recent paper [38] discusses power and sample size for mediation model in longitudinal studies, however, in their model, the mediator was assumed to be time-invariant instead of time-variant in our research.

**Conclusion**

Mediation analysis using longitudinal data allows researchers to investigate biological pathways and identifies their direct and indirect contribution to interested outcome variable. However, though this method is common in psychological and social research, sample size determination is still a challenging problem. This paper gives a way of using multilevel model for longitudinal data to provide the sample size under various sizes of the mediation effect, within-subject correlations and numbers of repeated measures via simulations by using three methods, Sobel, distribution of product and bootstrap. We found that the bootstrap and distribution of the product methods had comparable results and were more powerful than the Sobel’s method in terms of relatively smaller sample sizes. We recommend to use the distribution of product method due to its less computational load. For the mediation model of longitudinal data, the sample size depended on the ICC (i.e., the intra-subject correlation), number of repeated measurements, “position” of $\beta_a$ and $\beta_b$. Sample size tables for commonly encountered scenarios in practice were also provided for researchers’ convenient use.
8. Crits-Christoph P, Mintz J. Implication of therapist effects for the design and analysis of comparative studies of psychotherapies. J Consult Clin Psychol. 1991;59:20–6.
9. de Leeuw J, Meije R. Handboek van multivariate analyse. New York Springer; 2008.
10. Du H, Wang L. A Bayesian power analysis procedure considering uncertainty in effect size estimates from a meta-analysis. Multivar Behav Res. 2016;51(5):589–605.
11. Elkin B. Bootstrap methods: another look at the jackknife. Ann Stat. 1979;7:1–26.
12. Elkin I, Falconnier L, Martinovich Z, Mahoney C. Therapist effects in the NIMH Treatment of Depression Collaborative Research Program. Psychother Res. 2006;16:144–60.
13. Frees EW. Longitudinal and panel data: analysis and applications in the social sciences. Cambridge: Cambridge University Press; 2004.
14. Fritz MS, MacKinnon DP. Required sample size to detect the mediated effect. Psychol Sci. 2007;18:233–9.
15. Hox JJ. Multilevel analysis: techniques and applications. Mahwah, NJ: Erlbaum; 2002.
16. Judd CM, Kenny DA. Estimating the effects of social interventions. Cambridge: Cambridge University Press; 1981a.
17. Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. Eval Rev. 1981b;5:602–19.
18. Kenny DA, Korchmaros JD, Bolger N. Lower level mediation in multilevel models. Psychol Methods. 2003;8:115–28.
19. Killip S, Mahfoud Z, Pearce K. What is an Intraclass correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med. 2004;2(3):204–8. https://doi.org/10.1370/afm.141
20. Kim D-M, Wampold BE, Bolt DM. Therapist effects in psychotherapy: a random effects modeling of the NIMH TDCRP data. Psychother Res. 2006;16:161–72.
21. Krull JL, MacKinnon DP. Multilevel mediation modeling in group-based intervention studies. Eval Rev. 1999;23:418–44.
22. Krull JL, MacKinnon DP. Multilevel modeling of individual and group level mediated effects. Multivar Behav Res. 2001;36:249–77.
23. Lomnicky ZA. On the distribution of product of random variables. J R Stat Soc. 1967;70:513–24.
24. Lutz, Wolfgang; Leon, Scott C.; Martinovich, Zoran; Lyons, John S.; Stiles, William B. Therapist effects in outpatient psychotherapy: a three-level growth curve approach. J Couns Psychol, Vol 54(1), Jan 2007, 32–39.
25. MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. Eval Rev. 1999;17:144–58.
26. MacKinnon DP, Lockwood CM, Williams J. Confi dence limits for the indirect effect: distribution of the product and resampling methods. Multivar Behav Res. 2004;39:99–128.
27. MacKinnon DP. Introduction to statistical mediation analysis. New York: Lawrence Erlbaum Associates; 2008.
28. MacKinnon DP, Wards G, Dwyer JH. A simulation study of mediated effect measures. Multivar Behav Res. 1999;34:1–62.
29. Maxwell SE. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. Psychol Methods. 2000;5(2):147.
30. Maxwell SE, Kelley K, Rausch J. Sample size planning for statistical power and accuracy in parameter estimation. Annu Rev Psychol. 2008;59:537–63.
31. Meeker WQ Jr, Cornell LW, Aroian LA. The product of two normally distributed random variables. In: Kennedy WJ, Odeh RE, editors. Selected tables in mathematical statistics, vol. VII. Providence, RI: American Mathematical Society; 1981.
32. Okiishi J, Lambert MJ, Nielsen SL, Ogles BM. Waiting for Supershink: an empirical analysis of therapist effects. Clinical Psychology and Psychotherapy. 2003;10:361–73.
33. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods. 2nd ed. Newbury Park, CA: Sage; 2003.
34. Shrum PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychology Methods. 2002;7:422–45.
35. Sobel ME. Asymptotic conﬁdence intervals for indirect effects in structural equation models. Sociol Methodol. 1982;2:300–312.
36. Sobel ME. Direct and indirect effects in linear structural equation models. Sociol Methods and Research. 1987;16:155–67.
37. Wampold BE, Brown GS. Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. J Consult Clin Psychol. 2005;73:914.
38. Wang C, Xue X. Power and sample size calculations for evaluating mediation effects in longitudinal studies. Stat Methods Med Res. 2016 Apr;25(2):686–705.