Published in final edited form as:

Stat Methods Med Res. 2016 April ; 25(2): 686–705. doi:10.1177/0962280212465163.

Power and sample size calculations for evaluating mediation effects in longitudinal studies

Cuiling Wang and Xiaonan Xue

Department of Epidemiology and Population Health, Albert Einstein College of Medicine of Yeshiva University, 1300 Morris Park Ave, NY 10461, U.S.A.

Abstract

Current methods of power and sample size calculations for the design of longitudinal studies to evaluate mediation effects are mostly based on simulation studies and do not provide closed form formulae. A further challenge due to the longitudinal study design is the consideration of missing data, which almost always occur in longitudinal studies due to staggered entry or drop out. In this paper, we consider the product of coefficients as a measure for the longitudinal mediation effect and evaluate three methods for testing the hypothesis on the longitudinal mediation effect: the joint significant test, the normal approximation and the test of b methods. Formulae for power and sample size calculations are provided under each method while taking into account missing data. Performance of the three methods under limited sample size are examined using simulation studies. An example from the Einstein Aging Study (EAS) is provided to illustrate the methods.

Keywords

drop out; joint significance test; linear mixed effects model; missing data; power analysis; product of coefficients

1 Introduction

Mediation effects are often of great interest in longitudinal studies. For example, in the Einstein Aging Study (EAS) [1] program project, one primary goal is to investigate the role of clinical cardiovascular disease (CAD) and cerebral vascular reactivity (CVR) with CO₂ challenge on the rate of cognitive decline. CVR assessed by transcranial doppler ultrasound (TCD) with CO₂ challenge is a measure of cerebrovascular function. Decreased reactivity may reflect microvascular damage. One possible pathway of the effect of CAD on cognitive decline is through cerebral microvascular damage. Hence it is hypothesized that the effect of CAD on cognitive decline is partly mediated by CVR.

There are currently limited research on the power analysis for mediation effects on the rate of change in a continuous outcome, i.e., the longitudinal mediation effect. Current work on power estimation for mediation effects is mostly empirical, calculated based on simulations.

Address for correspondence: Cuiling Wang, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461. cuiling.wang@einstein.yu.edu.

Conflict of interest: None declared.

For example, to evaluate mediation effects in cross-sectional studies, MacKinnon et al. [1] examined empirical power based on simulation studies for some common sample size values, and Fritz and MacKinnon [3] reported sample size needed with 80% power obtained from simulation studies. For more complicated situations including longitudinal studies with repeatedly measured mediator, Thoemmers et al [4] also used simulation studies to calculate power. The empirical approach through simulations is useful for power calculations as it can handle various statistical models and complications such as missing data. However, it can be computationally intensive and is cumbersome to use when sample size estimation is of primary interest.

In this paper, we develop closed form formulae to calculate sample size and power for detecting a mediation effect with consideration of missing data. In particular, a linear mixed effects model is used for evaluating the rate of change, or slope, in the longitudinal outcome. We adopt the product of coefficients as a measure of the longitudinal mediation effect, and evaluate three methods for testing the hypothesis on the longitudinal mediation effect: the joint significance test, the normal approximation and the test of b methods. Section 2 provides the description of the product of coefficients as the measure of the longitudinal mediation effect. In Section 3, three methods of testing the hypothesis are evaluated. The type I error rate and power of each method are examined, and power and sample size calculation formulae are provided. Simulation studies to examine the performance of the three methods under various situations are presented in Section 4. The methods are applied to an example from EAS in Section 5. We conclude the paper in Section 6.

2 Measure of Longitudinal Mediation Effect

2.1 Background and Notation

Consider a longitudinal study of n subjects indexed by i , with k planned visits. Due to staggered entry and drop out, K_i measurements, $K_i = 1, \dots, k$, are observed for subject i . Here k is considered fixed so that the asymptotic properties of the estimators are applicable for large n . In presence of missing data, it is assumed that the data are missing at random (MAR) [5]. A monotone missing data pattern is also assumed, i.e., once a subject misses a visit, he/she will not return to the study. This pattern is common in longitudinal studies as missing data are usually caused by staggered entry and drop out. We assume that the observations from different subjects are independent. For subject i , $i = 1, \dots, n$, let X_i and M_i denote the risk factor and the mediator, respectively. Both are time-invariant (e.g., baseline measures). Denote $\mathbf{Z}_i$ as the set of covariates for subject i . Let Y_{ij} and t_{ij} denote the outcome and time, respectively, at the j th measurement ($t_{ij} = 0$ corresponds to the baseline), and $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iK_i})^T$, where superscript T denotes matrix transpose. Here the variables M and Y are both continuous. Although in reality t_{ij} might vary between subjects, in the design stage it is typical to assume t_{ij} as fixed with value t_j for all subjects.

Suppose the effect of X on M can be modelled as follows:

$$M_i = \alpha_0 + aX_i + \zeta^T \mathbf{Z}_i + e_i, \quad (1)$$

where ζ is a vector of coefficients for the vector of covariates $\mathbf{Z}$; e_i is the random error assumed normally distributed with zero mean.

For the longitudinal trajectory of the outcome $\mathbf{Y}$, suppose that, given X , M and $\mathbf{Z}$, the following linear mixed effects model [6] holds:

$$Y_{ij} = \beta_0 + \beta_x^c X_i + \beta_m^c M_i + \beta_t t_{ij} + \beta_x^l X_i t_{ij} + b M_i t_{ij} + \phi^T \mathbf{Z}_{it} + e_{ij}, \quad (2)$$

where $\mathbf{Z}_{it}$ is a vector consisting of $\mathbf{Z}_i$ and possibly some interactions with time t_{ij} ; ϕ is the vector of coefficients for $\mathbf{Z}_{it}$, $\mathbf{e}_i = (e_{i1}, \dots, e_{iK_i})^T$ has variance covariance matrix V_i . For example, in a random intercept model, $e_{ij} = u_i + \varepsilon_{ij}$, where u_i is the subject-specific random effect which is normally distributed with mean 0 and variance σ_u^2 , and ε_{ij} is the normally distributed random error term with variance σ_e^2 and assumed independent of u_i . The parameter b measures how one unit difference in the mediator M affects the rate of change in the outcome $\mathbf{Y}$ adjusting for the risk factor X and the covariate $\mathbf{Z}$, i.e., it is a measure of the longitudinal effect of M on the outcome $\mathbf{Y}$ adjusting for X and $\mathbf{Z}$. Similarly, β_x^l measures the longitudinal effect of X on $\mathbf{Y}$ adjusting for M and $\mathbf{Z}$. The parameters β_x^c and β_m^c measure the association between X and the outcome at baseline, and between M and the outcome at baseline, respectively, and therefore are referred as the baseline effects.

2.2 Measure of longitudinal mediation effect

According to (2), the expectation of Y_{ij} given X_i , M_i and $\mathbf{Z}_i$ is

$$E(Y_{ij} | X_i, M_i, \mathbf{Z}_i) = \beta_0 + \beta_x^c X_i + \beta_m^c M_i + \beta_t t_{ij} + \beta_x^l X_i t_{ij} + b M_i t_{ij} + \phi^T \mathbf{Z}_{it}. \quad (3)$$

From (1), this implies that the expectation of Y_{ij} given X_i and $\mathbf{Z}_i$ is

$$\begin{aligned} & E(Y_{ij} | X_i, \mathbf{Z}_i) \\ &= \beta_0 + \beta_x^c X_i + \beta_m^c E(M_i | X_i, \mathbf{Z}_i) + \beta_t t_{ij} + \beta_x^l X_i t_{ij} + b E(M_i | X_i, \mathbf{Z}_i) t_{ij} + \phi^T \mathbf{Z}_{it} \quad (4) \\ &= (\beta_0 + \alpha_0 \beta_m^c) + (\beta_x^c + a \beta_m^c) X_i + (\beta_t + \alpha_0 b) t_{ij} + (\beta_x^l + ab) X_i t_{ij} + \gamma^T \mathbf{Z}_{it}, \end{aligned}$$

where $\gamma^T \mathbf{Z}_{it} = \phi^T \mathbf{Z}_{it} + \beta_m^c \zeta^T \mathbf{Z}_i + b \zeta^T \mathbf{Z}_i t_{ij}$. The difference in the coefficients of $X_i t_{ij}$ between (4) and (3), which measures the effect of X on the rate of change in the outcome, is ab . Hence, the product of a and b , denoted as ab , measures the difference in the longitudinal effects of the risk factor X on the rate of change in the outcome Y , with and without the presence of the mediator M . In other words, it measures the indirect longitudinal effect, through the mediator M , of the risk factor X on the rate of change in the outcome Y . Thus is a measure of the longitudinal mediation effect. The inference based on ab is called the product of coefficients method ([7]-[9]).

Suppose a random intercept model is used for model (2), then the residual variance

$\text{Var}(\mathbf{Y}_i | X_i, M_i, \mathbf{Z}_i) = \sigma^2 \rho \mathbf{J}_{K_i} + \sigma^2 (1 - \rho) \mathbf{I}_{K_i}$, where $\sigma^2 = \sigma_u^2 + \sigma_e^2$, $\rho = \sigma_u^2 / \sigma^2$, $\mathbf{J}_{K_i}$ is a $K_i \times K_i$ matrix with all elements equal to 1, and $\mathbf{I}_{K_i}$ is the identity matrix with dimension K_i , is

compound symmetry with variance σ^2 and intraclass correlation coefficient ρ . The variance given X and Z is

$$\begin{aligned}\text{Var}(\mathbf{Y}_i|X_i, \mathbf{Z}_i) &= E\{\text{Var}(\mathbf{Y}_i|X_i, M_i, \mathbf{Z}_i)|X_i, \mathbf{Z}_i\} + \text{Var}\{E(\mathbf{Y}_i|X_i, M_i, \mathbf{Z}_i)|X_i, \mathbf{Z}_i\} \\ &= \sigma^2 \rho \mathbf{J}_{K_i} + \sigma^2(1 - \rho) \mathbf{I}_{K_i} + \mathbf{\Gamma},\end{aligned}\quad (5)$$

where the expectation and variance are taken over the distribution of the mediator M given X and Z . The $(j, k)^{th}$ entry of the variance-covariance matrix $\mathbf{\Gamma}$ is $(\beta_m^c + bt_{ij})(\beta_m^c + bt_{ik})\sigma_{m|x,z}^2$, a quadratic function of time unless $b = 0$; here $\sigma_{m|x,z}^2 = \text{Var}(M|X, Z)$. This shows that in general the structure of the residual variance matrix in the model without the mediator does not have the same structure as that in the model with the mediator. As shown in the above example, when model (2) is a random intercept model, $\text{Var}(\mathbf{Y}_i|X_i, \mathbf{Z}_i)$ is no longer compound symmetry as $\text{Var}(\mathbf{Y}_i|X_i, M_i, \mathbf{Z}_i)$ is. In practice, the difference in coefficients method is sometimes used, in which model (2), and the following linear mixed effects model without M is fit:

$$Y_{ij} = \beta_0^* + \beta_x^{c*} X_i + \beta_t^* t_{ij} + \beta_x^{l*} X_i t_{ij} + \phi^{*T} \mathbf{Z}_{it} + e_{ij}^*. \quad (6)$$

The difference in the coefficients of $X_i t_{ij}$ in model (2) and (6), $\beta_x^{l*} - \beta_x^l$, is then used as the measure of the longitudinal the mediation effect. However, model (6) needs to be fit with caution as the residual $\mathbf{e}_i^* = (e_{i1}^*, \dots, e_{iK_i}^*)^T$ usually has a different and more complicated variance structure compared to model (2). Furthermore, the calculation of the covariance between β_x^{l*} and β_x^l is not easy. In the cross sectional mediation analysis, the difference in coefficients method does not perform better than the product of coefficients method (e.g. [2]). The product of coefficients measure is thus used and serves as the basis of the hypothesis test.

2.3 Estimation

Denote the estimates of a from model (1) and b from model (2) as $\hat{a}$ and $\hat{b}$, respectively. Because $\hat{a}$ and $\hat{b}$ are consistent estimate of a and b , the longitudinal mediation effect can be consistently estimated by

$$\hat{\Delta} = \hat{a}\hat{b}.$$

Similar to [10], it can be shown that $\hat{a}$ and $\hat{b}$ are asymptotically independent. The exact variance of $\hat{\Delta}$ can thus be expressed as

$$V_{\Delta}^e = a^2 \text{Var}(\hat{b}) + b^2 \text{Var}(\hat{a}) + \text{Var}(\hat{b}) \text{Var}(\hat{a}).$$

However, the distribution of $\hat{c}$ is not normal so the inference on the mediation effect can not be made simply by using the point estimate and its variance. Theoretically, the exact form of the distribution of the product of two random variables has been studied (e.g., [11]-[17]). For example, if $a = 0$ and $b = 0$, it is a Bessel function of the second kind [14, 18]. However, the implementation of the distribution is not easy. Although tables of the distribution of the product of two normal variables are available [14, 15, 19], and some algorithms on calculating the distribution function have been developed [16, 17], in general it is not available in commonly used commercial statistical softwares. In practice, the bootstrap technique [20] has been recommended to evaluate the distribution of the product of coefficients and obtain confidence intervals [21]. Although easily applicable at the data analysis stage, this method is not practical at the design stage. In the cross sectional setting, a table of the empirical distribution of the estimate of the product of coefficients based on simulation studies is available [22]. However, this is not particularly useful for longitudinal studies because, in addition to the sample size and other parameters, the variance of $\hat{b}$ further depends on several factors specific to longitudinal studies including number of repeated measures, time at each measurement, the missing data distribution, and correlations among the repeated measures. A simple and commonly used method is to use a normal distribution to approximate the distribution of $\hat{c}$ [12, 23, 24]. Due to the skewness of the distribution of $\hat{c}$, this method can be problematic, especially when the sample size is not large [2, 21, 25, 26]. However, it has been shown that when a or b is large in magnitude, the product of two normal variables approaches normal distribution [13], in which case the normal approximation method might work well.

In the next section, we examine three methods for testing the mediation effect, some of which do not require assumptions on the distribution of the mediation effect estimate. We also provide formulae for calculating power and sample size under each method.

3 Hypothesis test and power analysis

The hypothesis on the mediation effect, $H_0: c = ab = 0$ versus $H_1: c = a_0b_0$, where $a_0 \neq 0$ and $b_0 \neq 0$, is complicated by the fact that it involves two parameters a and b : values of both parameters have to be specified for either the null or the alternative hypothesis. The null hypothesis $H_0: c = ab = 0$ corresponds to either both $a = 0$ and $b = 0$, or one of a and b is zero. We examine the performance of the following three methods for testing the mediation effect.

3.1 Joint significance test

Because $ab \neq 0$ is equivalent to $a \neq 0$ and $b \neq 0$, the hypothesis H_0 versus H_1 on the mediation effect can be tested by jointly testing the two hypotheses on a and b . Specifically, the hypothesis on the coefficient a , $H_{0a}: a = 0$ versus $H_{1a}: a = a_0$, and the hypothesis on the coefficient b , $H_{0b}: b = 0$ versus $H_{1b}: b = b_0$, are tested jointly, and the null hypothesis on the mediation effect, $H_0: c = ab = 0$ versus $H_1: c = ab, a = a_0, b = b_0$, is rejected when both H_{0a} and H_{0b} are rejected [2, 27]. This method addresses the problem of testing the mediation effect without dealing with the complicated distribution of $\hat{c}$.

The distributions of the estimates of both components, $\hat{a}$ and $\hat{b}$, are asymptotically normal. Without covariate Z , the variance of $\hat{a}$ can be expressed as

$$\text{Var}(\hat{a}) = \frac{\sigma_m^2 - a^2 \sigma_x^2}{n \sigma_x^2},$$

where σ_x^2 and σ_m^2 are the marginal variance of X and M , respectively. In the presence of covariate Z , using the method of Hsieh (1998) [28], the variance of $\hat{a}$ can be conservatively approximated by

$$\text{Var}(\hat{a}) = \frac{\sigma_m^2 - a^2 \sigma_x^2}{n \sigma_x^2 (1 - R_{x|z}^2)},$$

where $R_{x|z}^2$ is the proportion of the variance of X explained by Z .

The variance of the estimate of the vector of parameters from model (2) is

$$\{\mathbf{E}(\mathbf{W}_i^T \mathbf{V}_i^{-1} \mathbf{W}_i)\}^{-1} / n, \quad (7)$$

where $\mathbf{W}_i$ is the design matrix for the fixed effects, and $\mathbf{V}_i = \text{Var}(\mathbf{Y}_i | \mathbf{X}_i, \mathbf{M}_i, \mathbf{Z}_i)$ is the residual variance. When $\mathbf{V}_i$ is compound symmetry, the closed form of the variance of $\hat{b}$ without missing data is well known (e.g. [29]). It has also been studied extensively in the presence of missing data ([30] - [38]). For a general structure of $\mathbf{V}_i$ denote ϕ_{lj} as the $(l, j)^{th}$ element of $\mathbf{V}_i^{-1}$, $l, j = 1, \dots, K_i$, $\mathbf{t}^{(\mathbf{K}_i)} = (1, \dots, t_{K_i})^T$, similar to [38], it can be shown that the variance of b can be expressed as

$$\text{Var}(\hat{b}) = \frac{1}{n A_t \sigma_m^2 (1 - R_{m|x,z}^2)}, \quad (8)$$

where $A_t = \mathbf{E}(c_i) - \{\mathbf{E}(b_i)\}^2 / \mathbf{E}(a_i)$, $a_i = \mathbf{1}_{K_i}^T \mathbf{V}_i^{-1} \mathbf{1}_{K_i} = \sum_{l=1}^{K_i} \sum_{j=1}^{K_i} \phi_{lj}$, $b_i = \mathbf{1}_{K_i}^T \mathbf{V}_i^{-1} \mathbf{t}^{(\mathbf{K}_i)} = \sum_{l=1}^{K_i} \sum_{j=1}^{K_i} (\phi_{lj} t_l)$, $c_i = \mathbf{t}^{(\mathbf{K}_i)T} \mathbf{V}_i^{-1} \mathbf{t}^{(\mathbf{K}_i)} = \sum_{l=1}^{K_i} \sum_{j=1}^{K_i} (\phi_{lj} t_l t_j)$, and $R_{m|x,z}^2$ is the proportion of the variance of M explained by X and Z . If there is no covariate Z included in the model, then $R_{m|x,z}^2$ is replaced by $R_{m|x}^2 = a^2 \sigma_x^2 / \sigma_m^2$.

If $\mathbf{V}_i$ is compound symmetry with common correlation coefficient ρ and variance σ^2 , then

$$A_t = \frac{\mathbf{E}(S_{K_i}^2) + (1 - \rho)\delta}{(1 - \rho)\sigma^2}, \text{ where } S_{K_i}^2 = \sum_{j=1}^{K_i} t_j^2 - (\sum_{j=1}^{K_i} t_j)^2 / K_i,$$

$$\delta = \mathbf{E}\left\{\frac{(\sum_{j=1}^{K_i} t_j)^2}{K_i(1 - \rho + K_i\rho)}\right\} - \left\{\mathbf{E}\left(\frac{\sum_{j=1}^{K_i} t_j}{1 - \rho + K_i\rho}\right)\right\}^2 / \mathbf{E}\left(\frac{K_i}{1 - \rho + K_i\rho}\right). \text{ Note that } A_t \text{ increases as } \rho$$

or the planned number of measurements k increases, and decreases as the drop out rate increases. When there is no missing data, i.e., $K_i \equiv k$, then $\delta = 0$ and $E(S_{K_i}^2) = S_k^2$, thus

$A_t = \frac{S_k^2}{(1-\rho)\sigma^2}$, and (3.1) reduces to the well-known full data formula

$$\text{Var}(\hat{b}) = \frac{(1-\rho)\sigma^2}{nS_k^2\sigma_m^2(1-R_{m|x,z}^2)},$$

which is the basis for formula (2.4.1) in chapter 2 of [29].

If the structure of $\mathbf{V}_i$ is first order autoregressive (AR1), i.e., the $(l, j)^{th}$ element of $\mathbf{V}_i$ is $\sigma^2\rho^{|j-l|}$, in which the correlation weakens as the time lag between observations increases, then it can be shown that the components a_i , b_i and c_i of A_i in the formula can be expressed

as follows: $a_i = \frac{K_i - K_i\rho + 2\rho}{(1+\rho)\sigma^2}$, $b_i = \frac{t_1}{\sigma^2}$ if $K_i = 1$, $\frac{t_1+t_2}{(1+\rho)\sigma^2}$ if $K_i = 2$, and

$\frac{t_1+t_{K_i}+(1-\rho)\sum_{j=2}^{K_i-1}t_j}{(1+\rho)\sigma^2}$ if $K_i \geq 3$; and $c_i = \frac{t_1^2}{\sigma^2}$ if $K_i = 1$, $\frac{t_1^2+t_2^2-2\rho t_1 t_2}{(1-\rho^2)\sigma^2}$ if $K_i = 2$, and

$\frac{t_1^2+t_{K_i}^2+(1+\rho^2)\sum_{j=2}^{K_i-1}t_j^2-2\rho\sum_{j=1}^{K_i-1}t_j t_{j+1}}{(1-\rho^2)\sigma^2}$ if $K_i \geq 3$. Similar as in the compound

symmetry case, A_i increases as ρ increases or as the drop out rate decreases. When $k > 2$, due to the reduced correlation coefficient as time lag increases in the AR1 structure, with the same parameter ρ , using an AR1 residual variance matrix will result in a lower power than a compound symmetry one. It is therefore important to use the correct variance structure for the longitudinal outcome in power calculations. Information of the variance structure can usually be obtained from preliminary data.

Although a monotone missing data pattern is assumed in the formula for $\text{Var}(\hat{b})$, when the missing data pattern is not monotone, $\text{Var}(\hat{b})$ can be obtained using the general formula (7). To approximate $\text{Var}(\hat{b})$ under a non-monotone missing data pattern using the closed-form formula obtained from monotone missing data patterns, data from the subsequent visits after a subject missed a visit are either discarded or moved backward in time as if they were from a monotone missing data pattern. Since a subject with measurements at later time points contributes more information to the estimation of slopes, the monotone approximation is conservative. If the proportion of missed visits is low, it will be similar to the exact value calculated using the general formula (7).

Suppose significance levels α_1 and α_2 are used for the tests of a and b , respectively. Denote Z_p as the p^{th} percentile of the standard normal distribution. To test the hypothesis on a , we

use the test statistics $T_a = \frac{\hat{a}}{\sqrt{\text{Var}(\hat{a})}}$, and reject H_{0a} if $|T_a| > Z_{1-\alpha_1/2}$. Similarly, to test the

hypothesis on b , we use the test statistics $T_b = \frac{\hat{b}}{\sqrt{\text{Var}(\hat{b})}}$, and reject H_{0b} if $|T_b| > Z_{1-\alpha_2/2}$. The null hypothesis H_0 on the mediation effect is rejected when both H_{0a} and H_{0b} are rejected.

We examine the type I error rate and power of the joint significance test. Denote P_a and P_b as the powers for detecting the values of a_0 and b_0 , respectively, under the alternative hypothesis for testing a and b , respectively. There are three possible situations under H_0 : (1) $a = 0, b = 0$; (2) $a = a_0, b = 0$; and (3) $a = 0, b = b_0$. The type I error rate, α_J , of the joint significance test under each situation can be expressed as follows because of the asymptotic independence of $\hat{a}$ and $\hat{b}$. Under situation (1),

$$\alpha_J \doteq P(\text{reject } H_{0a}|H_{0a} \text{ is true})P(\text{reject } H_{0b}|H_{0b} \text{ is true}) = \alpha_1 \alpha_2;$$

under situation (2),

$$\alpha_J \doteq P(\text{reject } H_{0a}|H_{1a} \text{ is true})P(\text{reject } H_{0b}|H_{0b} \text{ is true}) = P_a \alpha_2;$$

and under situation (3),

$$\alpha_J \doteq P(\text{reject } H_{0a}|H_{0a} \text{ is true})P(\text{reject } H_{0b}|H_{1b} \text{ is true}) = \alpha_1 P_b,$$

where $\doteq$ denotes asymptotic equivalence. Suppose the target significance level for the test of H_0 is α . If the tests of a and b are both set at the significance level α , i.e., $\alpha_1 = \alpha_2 = \alpha$, then in any of the above situations $\alpha_J = \alpha$. Hence, the joint significance test is conservative in the sense that the type I error rate is no more than the target level. Under situation (1) of the null hypothesis H_0 in which $a = 0$ and $b = 0$, the joint significance test is especially conservative as $\alpha_J = \alpha^2$. Under situations (2) and (3) of the null hypothesis H_0 in which only one of a or b equals zero, the actual type I error rate is the product of the power for detecting the non-zero coefficient and the significance level for the test of the zero coefficient. It is close to the target value α if the power for testing the non-zero coefficient is close to 1. In order to yield the target significance level α , one needs to inflate either one or both significance levels for the tests of a and b . For example, under situation (1), α_1 and α_2 can be set at value $\sqrt{\alpha}$. Under situation (2), if $\alpha_1 = \alpha$ and the power P_a can be determined, then the value of α/P_a can be chosen for α_2 . But in reality the true value of a and b are unknown, it is common to set both α_1 and α_2 at α because it guarantees a type I error rate no more than α . We adopt this setting of $\alpha_1 = \alpha_2 = \alpha$ here.

Given a sample size n , the power of detecting a_0 for the test of a is

$$P_a = \Phi \left(\sqrt{\frac{a_0^2}{\text{Var}(\hat{a})}} - Z_{1-\alpha/2} \right) = \Phi \left(\sqrt{\frac{na_0^2\sigma_x^2(1-R_{x|z}^2)}{\sigma_m^2 - a_0^2\sigma_x^2}} - Z_{1-\alpha/2} \right), \quad (9)$$

where Φ is the cumulative distribution function of a standard normal random variable, and the power of detecting b_0 for the test of b is

$$P_b = \Phi \left(\sqrt{\frac{b_0^2}{\text{Var}(\hat{b})}} - Z_{1-\alpha/2} \right) = \Phi \left(\sqrt{nb_0^2 A_t \sigma_m^2 (1 - R_{m|x,z}^2)} - Z_{1-\alpha/2} \right), \quad (10)$$

The power from the joint significance test is thus

$$P_J = P_a P_b = \Phi \left(\sqrt{\frac{na_0^2 \sigma_x^2 (1 - R_{x|z}^2)}{\sigma_m^2 - a_0^2 \sigma_x^2}} - Z_{1-\alpha/2} \right) \Phi \left(\sqrt{nb_0^2 A_t \sigma_m^2 (1 - R_{m|x,z}^2)} - Z_{1-\alpha/2} \right), \quad (11)$$

which reduces to

$$P_J = \Phi \left(\sqrt{\frac{na_0^2 \sigma_x^2}{\sigma_m^2 - a_0^2 \sigma_x^2}} - Z_{1-\alpha/2} \right) \Phi \left(\sqrt{nb_0^2 A_t (\sigma_m^2 - a_0^2 \sigma_x^2)} - Z_{1-\alpha/2} \right)$$

when there is no covariate Z .

It shows clearly from (11) that the power from the joint significance test increases as the sample size n or the value b_0 increases. From the formulae of A_t , the power also increases as the drop out rate decreases, or as σ^2 decreases in the case of common residual variance at each visit, or as the correlation among repeated measures increases for the compound symmetry or AR1 residual variance structure. However, it is not monotone with a_0 , σ_x^2 and σ_m^2 .

For a given power level P for testing $H_0: \theta = 0$ versus $H_1: \theta = ab$, $a = a_0$, $b = b_0$, the sample size needed, n_J , is the solution of n to the equation $P_J = P$. A closed form formula for the sample size n is not directly available, but it can be easily obtained through numerical iteration.

3.2 Normal approximation method

Since $\hat{a}$ and $\hat{b}$ are the maximum likelihood estimates (MLE) of a and b , respectively, based on asymptotic theory, the asymptotic distribution of $\hat{\theta}$ is normal with mean ab and variance V_{Δ}^a . Sobel ([23]) derived the asymptotic variance of $\hat{\theta}$ using the multivariate delta method ([39]) based on a first order Taylor series approximation, which can be expressed as

$$V_{\Delta}^a = a^2 \text{Var}(\hat{b}) + b^2 \text{Var}(\hat{a}).$$

Combining with the expression of $\text{Var}(\hat{a})$ and $\text{Var}(\hat{b})$, we have

$$V_{\Delta}^a = \frac{a^2}{n A_t \sigma_m^2 (1 - R_{m|x,z}^2)} + \frac{b^2 (\sigma_m^2 - a^2 \sigma_x^2)}{n \sigma_x^2 (1 - R_{x|z}^2)}.$$

If a second order Taylor series approximation is used, the asymptotic variance expression is the same as the exact variance V_{Δ}^e ([12]). The difference between V_{Δ}^a and V_{Δ}^e is usually very small because the additional component, $\text{Var}(\hat{b})\text{Var}(\hat{a})$, in V_{Δ}^e is usually ignorable relative to V_{Δ}^a . Other variance formulae are also available (e.g., [24]), but they are all close to V_{Δ}^a . In practice, V_{Δ}^a is the most frequently used.

The $(1 - \alpha)$ confidence interval based on the normal approximation can be obtained as

$$\hat{\Delta} \pm Z_{1-\alpha/2} \sqrt{V_{\Delta}^a}.$$

To test the hypothesis on the mediation effect at significance level α based on the asymptotic normal approximation, we use the test statistic

$$T_{\Delta} = \frac{\hat{\Delta}}{\sqrt{V_{\Delta}^a}},$$

in which the null hypothesis that $\Delta = 0$ is rejected if $|T| > Z_{1-\alpha/2}$. This test is sometimes referred as Sobel's test.

However, the actual type I error rate is at or below the target level α . This can be seen by comparing it with the joint significance test. Since $|T| > Z_{1-\alpha/2}$ is equivalent to

$$\frac{1}{T_{\Delta}^2} < \frac{1}{Z_{1-\alpha/2}^2}, \text{ and}$$

$$\frac{1}{T_{\Delta}^2} = \frac{\hat{a}^2 \text{Var}(\hat{b}) + \hat{b}^2 \text{Var}(\hat{a})}{\hat{a}^2 \hat{b}^2} = \frac{\text{Var}(\hat{a})}{\hat{a}^2} + \frac{\text{Var}(\hat{b})}{\hat{b}^2} = \frac{1}{T_a^2} + \frac{1}{T_b^2},$$

it follows that

$$P(|T_{\Delta}| > Z_{1-\alpha/2}) = P\left(\frac{1}{T_a^2} + \frac{1}{T_b^2} < \frac{1}{Z_{1-\alpha/2}^2}\right).$$

Because the set $\left\{\frac{1}{T_a^2} + \frac{1}{T_b^2} < \frac{1}{Z_{1-\alpha/2}^2}\right\}$ is a subset of $\left\{\frac{1}{T_a^2} < \frac{1}{Z_{1-\alpha/2}^2}, \frac{1}{T_b^2} < \frac{1}{Z_{1-\alpha/2}^2}\right\}$, we have

$$P(|T_{\Delta}| > Z_{1-\alpha/2}) \leq P(|T_a| > Z_{1-\alpha/2})P(|T_b| > Z_{1-\alpha/2}).$$

The left hand of the above inequality is the probability of rejecting H_0 using the normal approximation method, while the right hand side is the probability of rejecting H_0 using the

joint significant test method. The equality would hold when either $|T_a| > Z_{1-\alpha/2}$ or $|T_b| > Z_{1-\alpha/2}$ occurs with probability 1. This means that the null hypothesis on the mediation effect is less likely to be rejected in the normal approximation approach than the joint significance test approach, hence the type I error rate and power from the normal approximation approach are lower than or equal to that from the joint significance approach. As the type I error rate from the joint significance approach is already no more than α , the normal approximation approach also has a type I error rate lower than or equal to the target level α .

Consistent with our theoretical finding, in the cross-sectional setting, the joint significance test has been shown to be more powerful and has better type I error rate than the normal approximation method through simulation [2, 27].

The power from the normal approximation method for a given sample size n can be expressed as

$$\begin{aligned} P_s &= \Phi \left(\sqrt{\frac{a_0^2 b_0^2}{V_\Delta}} - Z_{1-\alpha/2} \right) \\ &= \Phi \left(\sqrt{\frac{n}{\frac{1}{b_0^2 A_t \sigma_m^2 (1-R_{m|x,z}^2)} + \frac{\sigma_m^2 - a_0^2 \sigma_x^2}{a_0^2 \sigma_x^2 (1-R_{x|z}^2)}}} - Z_{1-\alpha/2} \right), \end{aligned}$$

which reduces to

$$P_s = \Phi \left(\sqrt{\frac{n}{\frac{1}{b_0^2 A_t (\sigma_m^2 - a_0^2 \sigma_x^2)} + \frac{\sigma_m^2 - a_0^2 \sigma_x^2}{a_0^2 \sigma_x^2}}} - Z_{1-\alpha/2} \right)$$

when there is no covariate Z .

To calculate the sample size n needed to achieve power P for testing $H_0: \gamma = 0$ versus $H_1: \gamma = ab$, with $a = a_0$, $b = b_0$, where $a_0 \neq 0$ and $b_0 \neq 0$, the equation $(a_0 b_0)^2 = V (Z_{1-\alpha/2} + Z_P)^2$ is solved, where Z_P is the P^{th} percentile of the standard normal distribution. The sample size needed can then be expressed as

$$n_s = \left\{ \frac{1}{b_0^2 A_t \sigma_m^2 (1 - R_{m|x,z}^2)} + \frac{\sigma_m^2 - a_0^2 \sigma_x^2}{a_0^2 \sigma_x^2 (1 - R_{x|z}^2)} \right\} (Z_{1-\alpha/2} + Z_P)^2. \quad (12)$$

When there is no covariate, it is simplified as

$$n_s = \left\{ \frac{1}{b_0^2 A_t (\sigma_m^2 - a_0^2 \sigma_x^2)} + \frac{\sigma_m^2 - a_0^2 \sigma_x^2}{a_0^2 \sigma_x^2} \right\} (Z_{1-\alpha/2} + Z_P)^2.$$

Note that the power or sample size estimates using the normal approximation method can be misleading if the true distribution of $\hat{\gamma}$ deviates severely from the normal distribution.

3.3 Test of b

In practice, there might be situations that the investigators are quite certain that one of the component of γ is not zero. In this case, test of the other parameter would be almost the same as that of the test of the mediation effect. Since the estimation of a only requires cross-sectional data of X and M while the estimation of b requires the collection of all data on X , M and the longitudinal outcome Y , most likely the investigators have more knowledge of a than b at the design stage. If there are evidences that the mediator has some effect on the outcome, usually this effect is established without adjustment of X . Furthermore, it is straightforward to test a with existing softwares. Therefore, we focus on developing power and sample size estimates for the test of b . Using the test of b approach in the power and sample size calculations for evaluating mediation effect has been proposed in other settings [40].

The power for testing b under H_1 : $b = b_0$, is P_b as expressed in (10). And the sample size needed for achieving power P from this method is

$$n_b = \frac{(Z_{1-\alpha/2} + Z_P)^2}{b_0^2 A_t \sigma_m^2 (1 - R_{m|x,z}^2)}, \quad (13)$$

which reduces to

$$n_b = \frac{(Z_{1-\alpha/2} + Z_P)^2}{b_0^2 A_t (\sigma_m^2 - a_0^2 \sigma_x^2)}$$

when there is no covariate Z .

However, in general, the test of b and the test of γ is not the same. The test of γ can be approximated by the test of b only when P_a , the power for testing a , is close to 1. From the joint significance test, the power for testing the mediation effect P_J is the product of P_a and P_b , so $P_b \geq P_J$, and the two are close when $P_a \approx 1$. For example, suppose $\alpha_1 = 0.05$ and there is no covariates, and the value of a_0 and the sample size n are large enough so that

$\frac{n}{1/r^2 - 1} \geq 18.373$, where $r^2 = \frac{a_0^2 \sigma_x^2}{\sigma_m^2}$ is the correlation coefficient between X and M , then $P_a \geq 0.99$ so P_J is no more than 1% less than P_b . For $n = 100, 200, 300, 400, 500$, this occurs when $r = 0.394, 0.290, 0.240, 0.210, 0.188$, respectively.

The fact that the test of b is more liberal than the test of γ can be also seen from the comparison with the normal approximation approach. In the normal approximation approach, the test statistics, T , can be expressed in the forms of T_a and T_b , the test statistics for testing a and b , respectively, as follows

$$T_{\Delta}^2 = \frac{a^2 b^2}{a^2 \text{Var}(\hat{b}) + b^2 \text{Var}(\hat{a})} = \frac{1}{\frac{\text{Var}(\hat{b})}{b^2} + \frac{\text{Var}(\hat{a})}{a^2}} = \frac{1}{\frac{1}{T_b^2} + \frac{1}{T_a^2}}.$$

Hence $|T| \leq |T_b|$ and $|T| \leq |T_a|$. This also shows that the test of the mediation effect has lower power than the test of either a or b . The test of mediation and the test of b are equivalent when $|T_a|$ is infinitely large. When all the other conditions are fixed, this occurs when $|a|$ and n are large.

A comparison of the three methods discussed in this section is summarized in Table 1.

4 Simulation studies

Simulation studies were performed to examine the performance of the three methods for testing the longitudinal mediation effect, in particular, their Type I error rate and power, under limited sample size. We considered longitudinal studies with 2 and 5 visits, representing short and moderate length of follow-up, respectively. For $k = 2$, the planned follow-up times $t = (0, 1)$, representing baseline and follow-up visits; for $k = 5$, $t = (0, 1, 2, 3, 4)$, representing baseline and 4 follow-up visits with equal time intervals (e.g. annual follow-up). Values of 100, 200, ..., 1000 were considered for the sample size n . The risk factor X was generated from Bernoulli distribution with $P(X = 1) = 0.5$ for the case of $k = 2$, and standard normal for the case of $k = 5$. The mediator M was generated as a normal random variable with marginal variance $\sigma_m^2 = 1$ and depends on X through model (1), with $\alpha_0 = 0$, and a took values 0.25, 0.5 and 0.75 for the case of $k = 2$, and 0.10, 0.20 and 0.30 for the case of $k = 5$. Here a is the correlation coefficient between X and M for continuous X ($k = 5$), and the expected difference in M in SD unit between $X = 1$ and $X = 0$ for the binary X case ($k = 2$). The repeatedly measured outcome Y was generated from a random intercept model for model (2), with $\beta_0 = \beta_x^c = \beta_m^c = \beta_t = \beta_x^l = 0$, $\sigma^2 = 1$, $\sigma_u^2 = \rho\sigma^2$, $\sigma_e^2 = (1 - \rho)\sigma^2$, and b was set as 0.1, 0.2 and 0.3 for $k = 2$, and 0.02, 0.03, 0.04 and 0.05 for the case of $k = 5$. Here b is the expected difference in the change in the outcome at follow up from baseline for $k = 2$, and the expected difference in the slope for $k = 5$, corresponding to 1 SD unit difference in M adjusting for X . We considered values of 0.25, 0.50 and 0.75 for the correlation coefficient ρ , representing low, medium and high levels of correlation, respectively. The first (or baseline) measure Y_1 was observed for every subject. For the missing data distribution, a constant dropout rate p_d , defined as the probability of missing Y_{ij} given Y_{ij-1} was observed, $j = 2, \dots, k$, was considered. Values of 0, 0.1, 0.2 and 0.3 were considered for the dropout rate p_d . All simulations were repeated 2000 times. The significance level for all tests was set at $\alpha = 0.05$.

To examine the empirical type I error rates of the methods, the data were generated with either $a = b = 0$, or one of a or b was zero, and the proportions of rejecting the null hypothesis H_0 were reported. To examine the performance of the power calculation formulae, the data were generated from non-zero values of a and b . The asymptotic power estimates (Asym) and empirical power estimates from 2000 simulations (Emp) were obtained for the separate tests of a and b , the joint significance test (Joint Sig) and the

normal approximation (Norm Appr) methods for testing the mediation effect. The bias of the point estimate of the longitudinal mediation effect (Bias) was calculated as the difference between the average of the estimates of from 2000 simulations and the true value. The 95% confidence intervals were obtained from the normal approximation method and the proportion that the true values of were covered out of 2000 simulations (Cover) were reported.

Results on the empirical type I error rates for dropout rates of $p_d=0, 0.3, \rho=0.25, 0.75$ and sample size $n=100, 200, 300, 500$ and 1000 were reported in Table 2a and Table 2b for $k=2$ and $k=5$, respectively. In the case of $b=0$, only the cases with $\rho=0.25$ were reported. As expected, when both $a=0$ and $b=0$, the empirical Type I error rates of the joint significance test were close to the expected level of $\alpha^2=0.0025$; when one of a or b is not zero, the empirical type I error rate of the join significance test is approximately 0.05 multiplied by the proportion of rejecting the hypothesis on the non-zero component, and hence is lower than 0.05, but is close to 0.05 when the power for testing the non-zero parameter is close to 1. Also consistent with the theoretical results, the empirical type I error rate of the normal approximation method is lower than that of the joint significance test method, and the differences are smaller when either a or b is not zero and the power for detecting the non-zero component is close to 1.

Results on the empirical power for $k=2$, with $a_0=0.25, 0.75$ and $b_0=0.1, 0.3$, and for $k=5$, with $a_0=0.1, 0.3$ and $b_0=0.02, 0.05$, were shown in Table 3a and Table 3b, respectively, for sample size $n=100, 200, 500$ and 1000, $\rho=0.25, 0.75$, and dropout rates $p_d=0, 0.1$ and 0.3. In all scenarios the bias of the estimate of the mediation effect is ignorable, showing that $\hat{\gamma}$ is a consistent estimate. The empirical power estimates from the joint significance test are close to the asymptotic power estimates in all situations. The test of b method has empirical power similar to that of the joint significance test only when the power for detecting a is close to 1. The asymptotic power for testing b is close to the empirical power in all situations considered, which increases as ρ increases or the dropout rate decreases. The asymptotic power estimate of the normal approximation method deviates from its empirical power estimate unless the sample size or at least one of the values of a and b is large. The coverage proportion of the 95% confidence interval from the normal approximation method is either within or below the nominal level.

5 Example

We applied the power analysis methods to an example from the Einstein Aging Study (EAS) [1]. EAS is a prospective cohort study aimed to identify risk factors for dementia which was initiated in 1994 and last renewed in 2011. One aim in this study is to examine whether baseline cerebral vascular reactivity (CVR) from Transcranial Doppler ultrasonography (TCD) with carbon dioxide (CO₂) challenge, denoted by M , will mediate the effect of baseline clinical cardiovascular disease (CAD) status, denoted by X , on the rate of decline in cognitive performance (denoted by Y) measured by the free and cued reminding test (FCSRT) [41]. The rationale behind this hypothesis is that cerebral microvascular damage might be a possible pathway for the effect of CAD on cognitive decline, where CVR is a measure of cerebromicrovascular function and decreased reactivity may reflect

microvascular damage. There are 400 subjects available for annual follow-up for up to 5 years. The sample consists of a mix of current EAS participants and new subjects that will be enrolled. It is projected that 30%, 8%, 24%, 19% and 19% of the sample will have 1, 2, 3, 4 and 5 repeated measures of FCSRT.

Based on results from previous studies, the residual variance is compound symmetry, with a correlation coefficient ρ conservatively estimated as 0.50 and the residual standard deviation σ approximately 6.4. The prevalence of clinical CAD at baseline is $p_x = 14\%$ so that $\sigma_x^2 = p_x(1 - p_x) = 0.12$. Suppose the CVR measure M is standardized, i.e., $\sigma_m^2 = 1$, and the difference in CVR between subjects with and without CAD, the value of a_0 , is 0.5. The value of b_0 , i.e., the change in the rate of decline in FCSRT corresponding to 1 SD unit increase in CVR adjusting for CAD, is expected to be at least 0.6. The joint significance test method is proposed to test the hypothesis on the mediation effect, and the estimate of power to detect the mediation effect $= a_0 b_0 = 0.30$ is 94%. If the normal approximation method is used, the power estimate is 81%. As expected, it underestimates power as compared to the joint significance test. If the test of b method is used, the power estimate is 99.8%. Because the power for detecting a_0 is 94%, compared to the joint significance test, it overestimates power by 6%. The difference is not big here because the power for detecting a_0 is not too far from 1. The empirical power estimates from 2000 simulations based on the joint significance test, the normal approximation and the test of b methods are 94.1%, 90.2% and 99.95%, respectively; similar to their corresponding asymptotic estimates except for the normal approximation method whose asymptotic power estimate is lower than the empirical power estimate.

6 Discussion

We have shown that the product of coefficients a , the parameter measures the effect of the risk factor on the mediator, and b , the parameter measures the independent effect of the mediator on the change in outcome adjusting for the risk factor, is an appropriate measure of the mediation effect. Currently available methods for estimating power and sample size for detecting a mediation effect mostly rely on simulation studies and do not incorporate the presence of missing data. We have evaluated three methods to test the presence of longitudinal mediation effect: the joint significance test, the normal approximation and the test of b methods. Closed form formulae for power and sample size estimation are developed while allowing the existence of missing data, which is commonly seen in longitudinal studies. Such closed form not only allows easy computation for power and/or sample size, it clearly demonstrates how each parameter affects the power for testing the mediation effect so that it can provide guidance on how to design an optimal longitudinal study, taking into consideration of the extent of missing data, number of possible repeated measurements and level of correlations among repeated measured outcome, etc.

Simulation studies show that the joint significance test has better empirical type I error rate and empirical power than the normal approximation method. The asymptotic power estimate from the joint significance test is close to the empirical power estimate under limited sample sizes. The joint significance test is conservative because the type I error rate is lower than the target value, but it is close to the target value when the null hypothesis corresponds to only

one zero component and the power for detecting the nonzero component is close to 1. For the normal approximation method, the asymptotic power estimate may deviate from the empirical power unless the sample size or at least one of values of a and b is large. It can provide confidence intervals for the mediation effect, but the coverage proportion can be lower than expected. The test of b is not the same as the test of the mediation effect unless the null hypothesis of the mediation effect corresponds to nonzero a and the power for detecting it is close to 1, in which case it is a reasonable shortcut for testing the mediation effect. In general, the joint significance test performs the best for testing mediation effects and is recommended. Although the joint significance test is only for hypothesis testing and can not provide confidence intervals of the mediation effects, it is sufficient for the power analysis purpose at the design stage. Since the bootstrap method can provide a reliable estimate of the distribution of the mediation effect estimate and thus can serve for both the purposes of hypothesis testing and estimation of confidence intervals, it is usually recommended at the data analysis stage. However, it is hard to implement at the design stage without actual data unless simulation data are used. Nevertheless, values calculated from our formulae can serve as a starting point for the bootstrap or other computation intensive power analysis approaches.

When the mediator is binary, the linear model (1) is not appropriate for the association between the risk factor and the mediator, so the product of coefficients ab is no longer the difference in the coefficients of X_{itij} in model (2) and (6). Approaches that either model the mediator directly or treat it as the indicator of a latent continuous variable can be applied [42, 43, 44]. Recent years also saw growing interests in the mediation model that includes the interaction between the risk factor and the mediator on the outcome (e.g., [45, 46, 47]). In our case, this means that the interaction term between X and M , X_iM_i measures the interaction effect at baseline, as well as its product with time, $X_iM_it_{ij}$ measures the interaction effect in the rate of change in the outcome, will be added to model (2). The general formula (7) can be used to calculate the variance of the parameter estimates for the model with interactions between X and M . Power analysis under these situations will be considered in future research.

Acknowledgments

The authors are grateful to the referees for helpful suggestions which greatly improved the manuscript.

Funding: This work was supported by National Institutes of Health [P01-AG03949 and R01-AG02511903].

References

1. Lipton RB, Katz MJ, Kuslansky G, Sliwinski M, Stewart W, Verghese J, Crystal H, Buschke H. Screening for dementia by telephone using the memory impairment screen. *Journal of the American Geriatrics Society*. 2003; 51:1382–1390. [PubMed: 14511157]
2. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*. 2002; 7:83–104. [PubMed: 11928892]
3. Fritz MS, MacKinnon DP. Required sample size to detect the mediated effect. *Psychological Science*. 2007; 18:233–239. [PubMed: 17444920]

4. Thoemmes F, MacKinnon DP, Reiser MR. Power analysis for complex mediational designs using monte carlo methods. *Structural Equation Modeling: A Multidisciplinary Journal*. 2010; 17:510–534. [PubMed: 23935262]
5. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–592.
6. Laird NM, Ware JH. Random effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
7. MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Evaluation Review*. 1993; 17:144–158.
8. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annual review of Psychology*. 2007; 58:593–614.
9. MacKinnon, DP. Introduction to statistical mediation analysis. Lawrence Erlbaum; New York: 2008.
10. MacKinnon DP, Lockwood CM, Brown CH, Wang W, Hoffman JM. The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*. 2007; 4:499–513. [PubMed: 17942466]
11. Craig CC. On the frequency function of xy . *Annals of Mathematical Statistics*. 1936; 7:1–15.
12. Aroian LA. The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics*. 1947; 18:265–271.
13. Aroian LA, Taneja VS, Cornwell LW. Mathematical forms of the distribution of the product of two normal variables. *Communications in Statistics: Theory and Methods*. 1978; 7:165–172.
14. Springer MD, Thompson WE. The distribution of product of independent random variables. *SIAM Journal on Applied Mathematics*. 1966; 14:511–526.
15. Meeker, WQ.; Cornwell, 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. p. 1-256.
16. Meeker WQ, Escobar LA. An algorithm to compute the cdf of the product of two normal random variables. *Communications in Statistics: Simulation and Computation*. 1994; 23:271–280.
17. Glen AG, Leemis LM, Drew JH. Computing the distribution of the product of two continuous random variables. *Computational Statistics & Data Analysis*. 2004; 44:451–464.
18. Epstein B. Some applications of the Mellin Transform in Statistics. *Annals of Mathematical Statistics*. 1948; 19:370–379.
19. MacKinnon DP, Fritz MS, Williams J, Lockwood CM. Distribution of the product confidence limits for the indirect effects: Program PRODCLIN. *Behavior Research Methods*. 2007; 39:384–389. [PubMed: 17958149]
20. Efron, B.; Tibshirani, R. An introduction to the bootstrap. Chapman & Hall/CRC; New York: 1993.
21. Bollen KA, Stine RA. Direct and indirect effects: classical and bootstrap estimates of variability. *Sociological Methodology*. 1990; 20:115–140.
22. MacKinnon, DP.; Lockwood, C.; Hoffman, J. A new method to test for mediation. Paper presented at the annual meeting of the Society for Prevention Research; Park City, UT. 1998.
23. Sobel, ME. Asymptotic confidence intervals for indirect effects in structural equation models. In: Leinhardt, S., editor. *Sociological Methodology*. Washington DC: American Sociological Association; 1982. p. 290-312.
24. Goodman LA. On the exact variance of products. *Journal of the American Statistical Association*. 1960; 55:708–713.
25. Stone CA, Sobel ME. The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika*. 1990; 55:337–352.
26. MacKinnon DP, Warsi DP, Dwyer JH. A simulation study of mediated effect measures. *Multivariate Behavioral Research*. 1995; 30:41–62. [PubMed: 20157641]
27. Mallinckrodt B, Abraham WT, Wei M, Russell DW. Advances in testing the statistical significance of mediation effects. *Journal of Counseling Psychology*. 2006; 53:372–378.
28. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*. 1998; 17:1623–1634. [PubMed: 9699234]
29. Diggle, PD.; Heagerty, P.; Liang, KY.; Zeger, SL. Analysis of longitudinal data. 2nd. Oxford University Press; New York: 2002.

30. Wu MC. Sample size for comparison of changes in the presence of right censoring caused by death, withdrawal, and staggered entry. *Controlled Clinical Trials*. 1988; 9:32–46. [PubMed: 3356151]
31. Jung S, Ahn C. Sample size estimation for GEE method for comparing slopes in repeated measurements data. *Statistics in Medicine*. 2003; 22:1305–1315. [PubMed: 12687656]
32. Ahn C, Jung SH. Effect of dropouts on sample size estimates for test on trends across repeated measurements. *Journal of Biopharmaceutical Statistics*. 2005; 15:33–41. [PubMed: 15702603]
33. Lefante JJ. The power to detect differences in average rates of change in longitudinal studies. *Statistics in Medicine*. 1990; 9:437–446. [PubMed: 2362980]
34. Dawson JD. Sample size calculations based on slopes and other summary statistics. *Biometrics*. 1998; 54:323–330. [PubMed: 9544525]
35. Tu XM, Zhang J, Kowalski J, Shults J, Feng C, Sun W, Tang W. Power analyses for longitudinal study designs with missing data. *Statistics in Medicine*. 2007; 26:2958–2981. [PubMed: 17154250]
36. Liu G, Liang KY. Sample size calculations for studies with correlated observations. *Biometrics*. 1997; 53:937–947. [PubMed: 9290224]
37. Liu A, Shih WJ, Gehan E. Sample size and power determination for clustered repeated measurements. *Statistics in Medicine*. 2002; 21:1787–1801. [PubMed: 12111912]
38. Wang C, Hall C, Kim M. A comparison of power analysis methods for evaluating effects of a predictor on slopes in longitudinal designs with missing data. *Statistical Methods in Medical Research*. 2012 in press.
39. Bishop, YMM.; Fienberg, SE.; Holland, PW. *Discrete multivariate analysis: theory and practice*. MIT Press; Cambridge, MA: 1975.
40. Vittinghoff E, Sen S, McCulloch CE. Sample size calculations for evaluating mediation. *Statistics in Medicine*. 2009; 28:541–557. [PubMed: 19065627]
41. Buschke H. Cued recall in amnesia. *Journal of Clinical and Experimental Neuropsychology*. 1984; 6:433–440.
42. Stolzenberg RM. The measurement and decomposition of causal effects in nonlinear and nonadditive models. *Sociological Methodology*. 1980; 11:459–488.
43. Winship C, Mare RD. Structural equations and path analysis for discrete data. *American Journal of Sociology*. 1983; 89:54–110.
44. Li Y, Schneider JA, Bennett DA. Estimation of the mediation effect with a binary mediator. *Statistics in Medicine*. 2007; 26:3398–3414. [PubMed: 17066450]
45. Judd, CM.; Kenny, DA. *Estimating the effects of social intervention*. Cambridge, England: Cambridge University Press; 1981.
46. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*. 2002; 59:877–883. [PubMed: 12365874]
47. Albert JM. Mediation analysis via potential outcomes. *Statistics in Medicine*. 2008; 27:1282–1304. [PubMed: 17691077]

Table 1
Comparison of the three hypothesis testing methods

	Test of b^*	Joint Significance test	Normal Approximation
Power	P_b	$P_a P_b$	$\Phi\left(\sqrt{\frac{a_0^2 b_0^2}{V_{\Delta}^a}} - Z_{1-\alpha/2}\right)$
Type I error			
$a = b = 0$	NA	α^2	α^2
$a = 0, b = b_0$	NA	αP_b	αP_b
$a = a_0, b = 0$	α	αP_a	αP_a

Note

* Applicable only when $P_a = 1$

Table 2a

Empirical Type 1 error rates for 2-visit studies for a_0 (true value of a), b_0 (true value of b), sample size n and drop out rate, with joint significance test (Joint Sig) and normal approximation (Norm Appr) methods used for the test of $= ab$

		no drop out				30% drop out			
		Test of a	Test of b	Test of $= ab$		Test of a	Test of b	Test of $= ab$	
a_0	b_0	n	Joint Sig	Norm Appr		Joint Sig	Norm Appr		
0	0	100	0.0595	0.0505	0.0030	0.0455	0.0540	0.0015	0.0000
		200	0.0540	0.0600	0.0050	0.0500	0.0525	0.0025	0.0010
		300	0.0490	0.0510	0.0025	0.0540	0.0475	0.0015	0.0000
		500	0.0535	0.0455	0.0035	0.0545	0.0415	0.0015	0.0000
		1000	0.0490	0.0480	0.0025	0.0540	0.0595	0.0030	0.0000
0	0.10	100	0.0520	0.1295	0.0070	0.0500	0.1240	0.0050	0.0000
		200	0.0515	0.2200	0.0110	0.0610	0.1840	0.0090	0.0000
		300	0.0570	0.2985	0.0160	0.0465	0.2505	0.0095	0.0000
		500	0.0510	0.4315	0.0240	0.0535	0.3845	0.0215	0.0030
		1000	0.0520	0.7295	0.0380	0.0490	0.6380	0.0315	0.0050
	0.20	100	0.0495	0.3820	0.0185	0.0495	0.3065	0.0155	0.0010
		200	0.0485	0.6325	0.0290	0.0470	0.5045	0.0180	0.0035
		300	0.0515	0.7840	0.0395	0.0510	0.6880	0.0315	0.0080
		500	0.0485	0.9505	0.0440	0.0535	0.9010	0.0485	0.0140
		1000	0.0515	0.9995	0.0515	0.0500	0.9970	0.0500	0.0265
	0.30	100	0.0455	0.6745	0.0300	0.0535	0.5720	0.0260	0.0020
		200	0.0525	0.9260	0.0475	0.0550	0.8505	0.0450	0.0120
		300	0.0430	0.9845	0.0430	0.0475	0.9620	0.0460	0.0150
		500	0.0505	1.0000	0.0505	0.0420	0.9975	0.0420	0.0260
		1000	0.0490	0.0495	0.0145	0.2480	0.0515	0.0135	0.0015
0.25	0	200	0.4330	0.0520	0.0170	0.4370	0.0445	0.0200	0.0035
		300	0.6030	0.0545	0.0310	0.5675	0.0525	0.0335	0.0035
		500	0.8055	0.0530	0.0435	0.8110	0.0515	0.0435	0.0095
		1000	0.9795	0.0470	0.0460	0.9805	0.0485	0.0470	0.0225

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

30% drop out				no drop out						
Test of $= ab$		Test of b	Test of a	Test of $= ab$		Test of b	Test of a			
Joint Sig	Norm Appr			Joint Sig	Norm Appr			n	b_0	a_0
0.0505	0.0605	0.7430	0.0045	0.0240	0.0390	0.7355	100	0	0.50	
0.0540	0.0555	0.9595	0.0160	0.0410	0.0440	0.9455	200			
0.0470	0.0470	0.9930	0.0255	0.0490	0.0495	0.9930	300			
0.0490	0.0490	0.9995	0.0420	0.0560	0.0560	1.0000	500			
0.0560	0.0570	0.9820	0.0205	0.0470	0.0480	0.9790	100	0	0.75	
0.0495	0.0495	1.0000	0.0380	0.0575	0.0575	1.0000	200			
0.0555	0.0555	1.0000	0.0375	0.0500	0.0500	1.0000	300			
0.0545	0.0545	1.0000	0.0420	0.0495	0.0495	1.0000	500			

Table 2b
Empirical Type 1 error rates for 5-visit studies for a_0 (true value of a), b_0 (true value of b), sample size n and drop out rate, with joint significance test (Joint Sig) and normal approximation (Norm Appr) methods used for the test of $= ab$

		no drop out				30% drop out			
		Test of a	Test of b	Test of $= ab$		Test of a	Test of b	Test of $= ab$	
a_0	b_0	n	Joint Sig	Norm Appr		Joint Sig	Norm Appr	Joint Sig	Norm Appr
0	0	100	0.0540	0.0435	0.0025	0.0610	0.0525	0.0010	0.0000
		200	0.0530	0.0425	0.0010	0.0535	0.0515	0.0030	0.0000
		300	0.0575	0.0525	0.0025	0.0555	0.0450	0.0035	0.0000
		500	0.0490	0.0615	0.0025	0.0490	0.0485	0.0020	0.0000
		1000	0.0450	0.0490	0.0015	0.0505	0.0420	0.0020	0.0000
0	0.02	100	0.0495	0.1260	0.0065	0.0505	0.0760	0.0045	0.0010
		200	0.0445	0.1710	0.0080	0.0500	0.1060	0.0055	0.0005
		300	0.0500	0.2315	0.0110	0.0555	0.1380	0.0065	0.0000
		500	0.0535	0.3720	0.0195	0.0500	0.1960	0.0115	0.0010
		1000	0.0500	0.6215	0.0320	0.0425	0.3245	0.0155	0.0015
	0.03	100	0.0550	0.1840	0.0065	0.0550	0.1120	0.0060	0.0000
		200	0.0520	0.3195	0.0180	0.0465	0.1635	0.0100	0.0005
		300	0.0550	0.4755	0.0245	0.0555	0.2320	0.0155	0.0010
		500	0.0480	0.6885	0.0340	0.0395	0.3540	0.0140	0.0020
		1000	0.0485	0.9350	0.0445	0.0505	0.5990	0.0330	0.0040
	0.05	100	0.0545	0.4415	0.0210	0.0495	0.2035	0.0080	0.0010
		200	0.0540	0.7415	0.0395	0.0465	0.3830	0.0170	0.0015
		300	0.0540	0.8775	0.0470	0.0530	0.5130	0.0270	0.0040
		500	0.0440	0.9795	0.0425	0.0440	0.7455	0.0305	0.0055
		1000	0.0465	1.0000	0.0465	0.0570	0.9540	0.0555	0.0240
0.10	0	100	0.1625	0.0590	0.0125	0.1615	0.0480	0.0080	0.0005
		200	0.2830	0.0475	0.0155	0.2890	0.0395	0.0125	0.0015
		300	0.4100	0.0590	0.0195	0.3900	0.0505	0.0210	0.0005
		500	0.6215	0.0500	0.0355	0.5905	0.0545	0.0325	0.0075

		no drop out				30% drop out								
		Test of a		Test of b		Test of $= ab$		Test of a		Test of b		Test of $= ab$		
a_0	b_0	n	Joint Sig		Norm Appr		Joint Sig		Norm Appr		Joint Sig		Norm Appr	
0.20	0	1000	0.8900	0.0495	0.0445	0.0095	0.8860	0.0525	0.0460	0.0120				
		100	0.5400	0.0525	0.0285	0.0050	0.5250	0.0500	0.0265	0.0070				
		200	0.8225	0.0495	0.0425	0.0120	0.8150	0.0585	0.0490	0.0120				
		300	0.9345	0.0515	0.0470	0.0140	0.9430	0.0570	0.0560	0.0225				
		500	0.9970	0.0470	0.0465	0.0280	0.9950	0.0505	0.0500	0.0290				
0.30	0	1000	1.0000	0.0455	0.0455	0.0385	1.0000	0.0485	0.0485	0.0345				
		100	0.8745	0.0500	0.0425	0.0120	0.8805	0.0435	0.0380	0.0170				
		200	0.9915	0.0500	0.0495	0.0235	0.9930	0.0485	0.0485	0.0225				
		300	0.9995	0.0515	0.0515	0.0355	0.9995	0.0500	0.0500	0.0300				
		500	1.0000	0.0400	0.0400	0.0330	1.0000	0.0520	0.0520	0.0430				

Table 3a

Asymptotic (Asym) and empirical (Emp) powers for 2-visit studies from 2000 simulations under various values for a_0 (true value of a), b_0 (true value of b), sample size n and drop out rate p_d with joint significance test (Joint Sig) and normal approximation (Norm Appr) methods used for the test of $\hat{a} = \hat{b}$. For normal approximation method, percentage that 95% confidence interval covers the true value (Cover) is also presented. Bias is the simulation bias of $\hat{a} = \hat{b}$.

				Test of a			Test of b			Test of $\hat{a} = \hat{b}$		
a	b	n	p	p_d	Bias	Asym	Emp	Asym	Emp	Asym	Emp	Norm Appr
0.25	0.10	100	0.25	0	-0.0004	0.242	0.231	0.125	0.125	0.030	0.026	0.951
				0.10	0.0005	0.242	0.247	0.120	0.136	0.029	0.030	0.945
				0.30	0.0024	0.242	0.245	0.107	0.139	0.026	0.036	0.961
				0	0.0001	0.242	0.234	0.289	0.283	0.070	0.077	0.919
	0.75	200	0.25	0.10	-0.0003	0.242	0.239	0.268	0.279	0.065	0.071	0.915
				0.30	-0.0007	0.242	0.258	0.223	0.212	0.054	0.050	0.934
				0	0.0008	0.429	0.426	0.208	0.214	0.089	0.095	0.923
				0.10	-0.0003	0.429	0.424	0.197	0.193	0.085	0.084	0.919
	0.30	500	0.25	0.30	0.0005	0.429	0.429	0.173	0.178	0.074	0.081	0.928
				0	0.0002	0.429	0.442	0.510	0.498	0.219	0.222	0.911
				0.10	0.0000	0.429	0.432	0.474	0.462	0.203	0.197	0.912
				0.30	-0.0005	0.429	0.417	0.395	0.381	0.169	0.151	0.901
0.30	0.10	100	0.25	0	0.0002	0.804	0.798	0.441	0.442	0.355	0.351	0.924
				0.10	-0.0002	0.804	0.800	0.418	0.409	0.336	0.332	0.918
				0.30	0.0005	0.804	0.794	0.363	0.385	0.292	0.301	0.928
				0	-0.0002	0.979	0.979	0.726	0.714	0.711	0.699	0.943
	0.75	200	0.25	0.10	0.0000	0.979	0.982	0.698	0.702	0.683	0.691	0.935
				0.30	-0.0001	0.979	0.981	0.624	0.613	0.610	0.603	0.935
				0	0.0012	0.242	0.250	0.681	0.674	0.165	0.166	0.926
				0.10	0.0011	0.242	0.245	0.652	0.624	0.158	0.150	0.913
	0.30	500	0.25	0.30	0.0000	0.242	0.240	0.579	0.564	0.140	0.136	0.903
				0	0.0000	0.242	0.240	0.579	0.564	0.140	0.136	0.903
				0.10	0.0000	0.242	0.240	0.579	0.564	0.140	0.136	0.903
				0.30	0.0000	0.242	0.240	0.579	0.564	0.140	0.136	0.903

|--|

Asymptotic (Asym) and empirical (Emp) powers for 5-visit studies from 2000 simulations under various values for z_0 (true value of a), b_0 (true value of b), sample size n and drop out rate p_d with joint significance test (Joint Sig) and normal approximation (Norm Appr) methods used for the test of $\gamma = ab$. For normal approximation method, percentage that 95% confidence interval covers the true value (Cover) is also presented. Bias is the simulation bias of $\gamma = ab$.

									Test of a			Test of b			Test of $\gamma = ab$		
a	b	n	p	p _d	Bias	Asym	Emp	Asym	Emp	Asym	Emp	Cover	Asym	Emp			
0.10	0.02	100	0.25	0	0.0001	0.170	0.174	0.109	0.124	0.019	0.022	0.970	0.085	0.002			
				0.10	-0.0001	0.170	0.169	0.093	0.099	0.016	0.019	0.973	0.078	0.002			
				0.30	0.0001	0.170	0.181	0.067	0.075	0.011	0.018	0.982	0.062	0.003			
			0.75	0	0.0001	0.170	0.176	0.242	0.244	0.041	0.045	0.938	0.120	0.006			
				0.10	0.0000	0.170	0.171	0.188	0.180	0.032	0.029	0.942	0.110	0.003			
				0.30	0.0000	0.170	0.172	0.112	0.111	0.019	0.017	0.970	0.087	0.002			
		200	0.25	0	-0.0001	0.295	0.292	0.176	0.173	0.052	0.048	0.928	0.130	0.009			
				0.10	0.0001	0.295	0.310	0.145	0.164	0.043	0.049	0.940	0.115	0.010			
				0.30	0.0001	0.295	0.304	0.096	0.105	0.028	0.026	0.971	0.086	0.003			
			0.75	0	0.0001	0.295	0.315	0.429	0.439	0.127	0.126	0.909	0.198	0.031			
				0.10	0.0000	0.295	0.300	0.329	0.324	0.097	0.094	0.907	0.178	0.017			
				0.30	0.0001	0.295	0.287	0.182	0.203	0.054	0.059	0.925	0.132	0.008			
		500	0.25	0	0.0001	0.613	0.607	0.369	0.382	0.226	0.225	0.919	0.260	0.074			
				0.10	0.0000	0.613	0.620	0.296	0.330	0.181	0.204	0.911	0.225	0.057			
				0.30	0.0001	0.613	0.627	0.178	0.184	0.109	0.115	0.932	0.154	0.025			
			0.75	0	0.0000	0.613	0.608	0.804	0.808	0.493	0.492	0.921	0.419	0.268			
				0.10	0.0000	0.613	0.620	0.670	0.657	0.411	0.405	0.915	0.375	0.194			
				0.30	0.0000	0.613	0.606	0.383	0.381	0.235	0.230	0.908	0.266	0.072			
		1000	0.25	0	0.0000	0.888	0.897	0.632	0.660	0.562	0.589	0.927	0.461	0.359			
				0.10	0.0000	0.888	0.901	0.521	0.530	0.463	0.479	0.916	0.398	0.263			
				0.30	0.0000	0.888	0.890	0.311	0.317	0.276	0.286	0.934	0.265	0.141			

						Test of a		Test of b		Test of $\chi^2 = ab$	
								</			

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</