Title
Sample Size for Joint Testing of Indirect Effects

Journal
Prevention Science, 16(8)

ISSN
1389-4986

Authors
Vittinghoff, Eric
Neilands, Torsten B

Publication Date
2015-11-01

DOI
10.1007/s11121-014-0528-5

Peer reviewed

HHS Public Access

Author manuscript

Prev Sci. Author manuscript; available in PMC 2016 May 25.

Sample Size for Joint Testing of Indirect Effects

Eric Vittinghoff and

Department of Epidemiology and Biostatistics, University of California San Francisco, 550 16th Street, 2nd Floor, San Francisco, CA 94158, USA

Torsten B. Neilands

Department of Medicine, Center for AIDS Prevention Studies, University of California San Francisco, 550 16th Street, 3rd Floor, San Francisco, CA 94158, USA

Eric Vittinghoff: eric.vittinghoff@ucsf.edu

Abstract

This paper presents methods to calculate sample size for evaluating mediation by joint testing of both links in an indirect pathway from exposure to mediator to outcome. Calculations rely on simulations of the underlying data structure, with testing of the two links performed under the simplifying assumption that the two test statistics are asymptotically independent. Simulations show that the proposed methods are accurate. Continuous and binary exposures and mediators, as well as continuous, binary, count, and survival outcomes are accommodated, along with over-dispersion of count outcomes, design effects, and confounding of the exposure-mediator and mediator-outcome relationships. An illustrative example is provided, and a documented R program implementing the calculations is available online.

Keywords

Mediation; Indirect pathway; Sample size; Power; Generalized linear models

Scientific interest, especially in the social sciences, has long focused on mediating pathways through which an exposure or treatment affects an outcome. Recent developments in causal analysis have extended the reach of these analyses, including cases where the exposure and mediator interact (MacKinnon et al. 2007; VanderWeele 2009; Pearl 2011; Breen et al. 2013), deepened our understanding of the assumptions required for valid causal inference (Pearl 2001; Cole and Hernán 2002; Petersen et al. 2006; VanderWeele 2009; Imai et al. 2010; Pearl 2012), and provided new and convenient analytic tools (Hicks and Tingley 2011; Kohler et al. 2011; Muthén 2011; Valeri and Vander-Weele 2013). While some methods are available for sample size and power calculations (Freedman and Schatzkin 1992; Vittinghoff et al. 2008; Muthén 2011; Wang and Xue 2012; Kenny 2013), convenient

Correspondence to: Eric Vittinghoff, eric.vittinghoff@ucsf.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s11121-014-0528-5) contains supplementary material, which is available to authorized users.

Conflict of interests The authors declare that they have no conflict of interest.

tools with broader reach are needed to ensure adequate power at the design stage for mediation analyses.

This paper presents sample size calculations for joint testing of both links in an indirect pathway from an exposure through a mediator to an outcome. We follow the mediation literature (e.g., MacKinnon et al. 2002) in defining the joint test as a simultaneous test of the composite null hypothesis that the coefficients for exposure and mediator—in models for the mediator and outcome, respectively—are both zero. In earlier work, we focused on testing the second link only, under the assumption that the first link was known to exist (Vittinghoff et al. 2008). Subsequently, Wang and Xue (2012) showed that this approach could substantially underestimate sample size if both links in the indirect pathway must be established, especially when the first link in the indirect pathway is more difficult to detect than the second. In this work, our earlier methods using Monte Carlo integration are extended to joint testing in cases where exposure and mediator are continuous or binary, and outcomes are continuous, binary, counts, or survival times. The performance of the new methods is investigated using simulations and demonstrated in a detailed illustrative example. We first review assumptions and data-generating models, describe power estimators for linear, logistic, Poisson, and Cox models, and provide examples for each. We then present a simulation study validating the proposed methods. The proposed methods are implemented in an R (R Development Core Team 2014) program, available online.

Assumptions and Data-Generating Models

Suppose the exposure X_1 is normal with standard deviation (SD) σ_1, or binary with prevalence f_1 and $\mathrm{SD}\ \sqrt{f_1(1-f_1)}$. The mediator X_2, with marginal SD σ_2, is assumed to arise from the generalized linear model (GLM)

$$h_1[\mathrm{E}(X_2|X_1)]=\gamma_0+\gamma_1 X_1. \quad (1)$$

For continuous X_2, h_1 is the identity link and the errors are assumed normal with mean zero and variance depending on γ_1, σ_1, and σ_2; for binary X_2, the logit link and a Bernoulli distribution with marginal mean E(X_2) are assumed. Continuous, binary, and count outcomes are also assumed to arise from the GLM

$$h_2[\mathrm{E}(Y|X_1,X_2)]=\beta_0+\beta_1 X_1+\beta_2 X_2, \quad (2)$$

using identity, logit, and log links, respectively, with normal errors with mean zero and SD σ_e in the first case, and Bernoulli and Poisson distributions with marginal mean E(Y) in the second and third. For count outcomes, we allow for over-dispersion by a scale factor φ = Var(Y)/E(Y), with φ = 1 as the default. Finally, for survival outcomes, the proportional hazards model

$$\lambda(t,X_1,X_2)=\lambda_0(t)\exp(\beta_1 X_1+\beta_2 X_2) \quad (3)$$

is assumed.

In linear structural equation models, β_1 captures the direct effect of exposure on the outcome, while the indirect effect is $\gamma_1\beta_2$. However, Sobel's (Sobel 1982) direct test of $\gamma_1\beta_2 = 0$ is complicated by the distribution of the test statistic, which has no clear interpretation except in the linear model. In contrast, joint testing of $\gamma_1 = 0$ and $\beta_2 = 0$ may be more powerful (Wang and Xue 2012; MacKinnon et al. 2002; Mallinckrodt et al. 2006) and is consistent with a counterfactual definition of the indirect effect for continuous as well as binary and count outcomes (Pearl 2011). Wang and Xue (2012) also show that the type I error rate of the joint test is asymptotically bounded by a common nominal rate used for each test separately. Under the assumption that $\hat{\gamma_1}$ and $\hat{\beta_2}$ are asymptotically independent, the power of the joint test P_J is approximately $P_{\gamma 1}\, P_{\beta 2}$, the product of the powers of each test considered separately.

For any given sample size, estimation of $P_{\gamma 1}$ is straight-forward for continuous mediators. Moreover, $P_{\beta 2}$ could be estimated using existing methods for linear (Hsieh et al. 1998), logistic (Hsieh et al. 1998), Poisson (Vittinghoff et al. 2008), and Cox (Hsieh and Lavori 2000; Schmoor et al. 2000; Bernardo et al. 2000) models using a correction based on the variance inflation factor from linear regression (Hsieh et al. 1998) to account for the correlation of X_2 with X_1 when $\gamma_1 \neq 0$, which limits the effect of increasing $|\gamma_1|$ on power, as investigated by Fritz et al. (2012). While the adjustment is exact for the linear model, it can break down for logistic, Poisson, and Cox models (Vittinghoff et al. 2008). This motivates using Monte Carlo integration to approximate the standard error of $\hat{\beta_2}$, accounting for adjustment for X_1, which can then be used to calculate $P_{\beta 2}$.

However, confounding of the mediator-outcome relationship by other factors must also be controlled, even when exposure is randomized (Judd and Kenny 1981; Pearl 1998; Cole and Hernán 2002). More generally, when exposure is not randomized, confounding of the exposure-outcome and exposure-mediator relationships must also be controlled (VanderWeele 2009; Pearl 2011). Accordingly, we assume that analysts will carefully control confounding in both the mediator and outcome models. However, full specification of the joint distribution of exposure, mediator, confounders, and outcome is usually not possible. Thus, we use the approximate correction based on the variance inflation factor to account for the resulting loss of precision. Similarly, design effects due to clustering are accommodated as a second, optional variance inflation factor, with independence assumed by default.

Although no explicit solution is available for the sample size ensuring that $P_{\gamma 1}\, P_{\beta 2}$ equals the targeted power, a line search can be used to find the solution. This requires specification of γ_1 and β_2 as well as case-specific nuisance parameters including σ_1 or f_1 and σ_2 or f_2; σ_e for continuous outcomes; E(Y) for binary and count outcomes; φ, the over-dispersion of count outcomes; ψ, the proportion of uncensored observations, for failure time outcomes; β_1 for all except the linear model; ρ_1 and ρ_2, respectively, the multiple correlation of X_1 and X_2 with confounders of the exposure-mediator and mediator-outcome relationships; and finally, the 2-sided type I error rate α and the target power.

Estimating $P_{\gamma 1}$ and $P_{\beta 2}$

Continuous Outcomes

$P_{\gamma 1}$ can be estimated for a continuous mediator using

$$P_{\gamma_1}=\Phi\left[|\gamma_1|\sigma_1\sqrt{\frac{n(1-\rho_1^2)}{\sigma_2^2(1-\rho^2)\delta}}+z_{\alpha/2}\right] \quad (4)$$

In (4), Φ is the standard normal cumulative distribution function, γ_1, σ_1, and σ_2 are defined as before, n is the sample size, $(1-\rho_1^2)$ is the approximate correction for confounding of the exposure-mediator relationship, ρ is the correlation of X_1 and X_2, calculated using γ_1, σ_1, and σ_2, δ is the design effect, defined as the ratio of the actual variance of the outcome in the presence of clustering to its variance under independence, and finally, $z_{\alpha/2}$ is the quantile of the standard normal distribution corresponding $\alpha/2$.

In our previous work (Vittinghoff et al. 2008), Monte Carlo integration slightly outperformed approximate adjustment for confounding, provided the joint distribution of exposure, mediator, and outcome can be fully specified, for example using (1) and (2) or (3). Thus, we use this technique to estimate the standard error of $\hat{\beta_2}$, accounting only for confounding of X_2 by X_1. Then, if there are additional confounders of the mediator-outcome relationship, we use the approximate method to inflate the result. Specifically, for the linear model, $\mathrm{Cov}(\hat{\boldsymbol{\beta}})=(\mathbf{X}'\mathbf{X})^{-1}\sigma_e^2$. To estimate P_{β_2}, our implementation simulates n_s = 10, 000 observations under the assumed joint distribution of X_1 and X_2 and then computes $n_s(\mathbf{X}'\mathbf{X})^{-1}\sigma_e^2$. Then, we calculate

$$P_{\beta_2}=\Phi\left[|\beta_2|\sqrt{\frac{n(1-\rho_2^2)}{\tilde{\sigma}^2_{\beta_2}\delta}}+z_{\alpha/2}\right] \quad (5)$$

where $(1-\rho_2^2)$ is the approximate correction for confounding of the mediator-outcome relationship, and $\tilde{\sigma}^2_{\beta_2}$ is the diagonal element of $n_s(\mathbf{X}'\mathbf{X})^{-1}\sigma_e^2$ corresponding to β_2.

As an example, suppose that both X_1 and X_2 are continuous with $\sigma_1 = \sigma_2 = \sigma_e = 1.0$, $\gamma_1 = 0.25$, $\beta_2 = 0.20$, $\rho_1 = 0$, as in a randomized trial, but $\rho_2 = 0.3$. The R function call and result are

```
> sampsi(1, 1, 1, g1=.25, b2=.20, rho2=.3)
N = 240 Power g1=0: 97.9 b2=0: 81.9
joint: 80.2
```

The first three arguments specify continuous exposure, mediator, and outcome, respectively, and g1, b2, and rho2 are the arguments specifying γ_1, β_2, and ρ_2; because the default value of rho1, used to specify ρ_1, is zero, it can be omitted. Thus, in two-sided tests with $\alpha = 0.05$,

a sample size of 240 is estimated to provide 97.9 % power to reject $\gamma_1 = 0$, 81.9 % power to reject $\beta_2 = 0$, and 80.2 % power to reject both null hypotheses.

Binary Outcomes

For analysis of binary outcomes with marginal prevalence E(Y) using the logistic model, we also use Monte Carlo integration to estimate power. For this GLM, $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$, where $\mathbf{V}$ is the diagonal covariance matrix of the outcome, with $\mathbf{V}_{i,i} = \mathrm{Var}(Y_i \mid \mathbf{X}_i) = \mathrm{E}(Y_i \mid \mathbf{X}_i)(1 - \mathrm{E}(Y_i \mid \mathbf{X}_i))$ and $\mathbf{X}_i = (X_{1i}, X_{2i})$. To estimate P_{β_2}, the implementation simulates n_s = 10, 000 observations under the assumed joint distribution of X_1 and X_2, calculates $\mathrm{Var}(Y_i \mid \mathbf{X}_i)$ under the assumed logistic model, and finally computes $n_s(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$. This procedure requires calculating the value of the intercept parameter β_0 consistent with β_1, β_2, and the marginal outcome prevalence, E(Y). Then

$$P_{\beta_2} = \Phi\left[|\beta_2|\sqrt{\frac{n(1-\rho_2^2)}{\tilde{\sigma}^2_{\beta_2}\delta}} + z_{\alpha/2}\right] \quad (6)$$

where $\tilde{\sigma}^2_{\beta_2}$ is now the diagonal element of $n_s(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$corresponding to β_2.

This approach can also be used to estimate P_{γ_1} when the mediator is binary. In this case, only X_1 is included in $\mathbf{X}$, $\mathrm{E}(X_{2i} \mid X_{1i})$ plays the role of $\mathrm{E}(Y_i \mid \mathbf{X}_i)$, and the subroutine used to calculate the intercept γ_0 is also simplified. Then

$$P_{\gamma_1} = \Phi\left[|\gamma_1|\sqrt{\frac{n(1-\rho_1^2)}{\tilde{\sigma}^2_{\gamma_1}\delta}} + z_{\alpha/2}\right] \quad (7)$$

with $\tilde{\sigma}^2_{\gamma_1}$ defined analogously to $\tilde{\sigma}^2_{\beta_2}$ in (6).

As an example, suppose that both X_1 and X_2 are binary with prevalence 50 and 35 %, respectively, $\gamma_1 = \log(2.1)$, $\beta_1 = \log(1.5)$, $\beta_2 = \log(1.9)$, EY = 0.4, $\rho_1 = 0.25$, $\rho_2 = 0.35$, and the design effect $\delta = 1.5$. In this example, γ_1, β_1, and β_2 are log odds ratios. The R function call and result are

```
> sampsi(2, 2, 2, f1=.5, f2=.35, g1=log(2.1),
b1=log(1.5), b2=log(1.9), EY=.4, rho1=.25,
rho2=.35, de=1.5)
N = 690 Power g1=0: 94.9 b2=0: 84.3
joint: 80
```

Here, the first three arguments specify binary exposure, mediator, and outcome, f1, f2, b1, and de are the arguments specifying the prevalence of X_1 and X_2, β_1, and the design effect, and g1and b2 are defined as before. A sample size of 690 is estimated to provide 94.9 %

power to reject $\gamma_1 = 0$, 84.3 % power to reject $\beta_2 = 0$, and 80.0 % power to reject both null hypotheses.

Count Outcomes

For analysis of count outcomes with marginal mean E(Y) using the Poisson model, the implementation again uses Monte Carlo integration to estimate power. In brief, it simulates $n_s = 10,000$ observations under the assumed joint distribution of X_1 and X_2, calculates $\text{Var}(Y_i|\mathbf{X}_i) = \text{E}(Y_i|\mathbf{X}_i)$ for each observation under the assumed Poisson model, and finally computes $n_s(\mathbf{X'VX})^{-1}$. Then P_{β_2} is calculated using (6), with $\tilde{\sigma}^2_{\beta_2}$ again defined as the diagonal element $n_s(\mathbf{X'VX})^{-1}$ corresponding to β_2. To account for over-dispersion with respect to the Poisson distribution, we inflate $\tilde{\sigma}^2_{\beta_2}$ by the scale factor φ.

As an example, suppose that X_1 is continuous with $\sigma_1 = 1.25$, X_2 is binary with prevalence 35 %, $\gamma_1 = \log(1.4)$, $\beta_1 = \log(1.5)$, $\beta_2 = \log(1.35)$, EY = 2.0, $\varphi = 1.5$, $\rho_1 = 0.35$, and $\rho_2 = 0.25$. Here, γ_1 is a log odds-ratio, and β_1 and β_2 are log rate ratios. The R function call and result are

```
> sampsi(1, 2, 3, sdx1=1.25, f2=.35,
g1=log(1.4), b1=log(1.5), b2=log(1.35),
EY=2, scale=1.5, rho1=.35, rho2=.25)
N = 351 Power g1=0: 91.6 b2=0: 87.3
joint: 80.2
```

Here, the first three arguments specify continuous exposure, binary mediator, and count outcome, respectively, and scale is the argument corresponding to φ. A sample size of 351 is estimated to provide 91.6 % power to reject $\gamma_1 = 0$, 87.3 % power to reject $\beta_2 = 0$, and 80.2 % power to reject both null hypotheses.

Survival Outcomes

For analysis of survival outcomes using the Cox model, Monte Carlo integration is also used. The implementation simulates $n_s = 10,000$ exponential failure times under the assumed joint distribution of the X_1 and X_2 and the specified proportional hazards model, censors all but the shortest fraction ψ of the failure times, a form of type II independent censoring (Kalbfleisch and Prentice 1980), and fits a Cox model to the simulated data. Then, P_{β_2} is again calculated using (6), with $\tilde{\sigma}^2_{\beta_2} = n_s \hat{\text{Var}}(\hat{\beta}_2)$, where $\hat{\text{Var}}(\hat{\beta}_2)$ is the variance of $\hat{\beta_2}$ obtained from the Cox model fit to the simulated data. The implementation also allows for additional independent censoring before the last failure, but this has minor effects (Vittinghoff et al. 2008).

As an example, suppose that X_1 is binary with prevalence 20 %, X_2 is continuous with $\sigma_2 = 1.2$, $\gamma_1 = 0.35$, $\beta_1 = \log(1.5)$, $\beta_2 = \log(1.4)$, $\psi = 0.3$, $\rho_1 = 0.25$, and $\rho_2 = 0.45$. In this example, β_1 and β_2 are log hazard ratios. The R function call and result are

```
> sampsi(2, 1, 4, f1=.2, sdx2=1.2, g1=.35,
b1=log(1.5), b2=log(1.4), psi=.3, rho1=.25,
rho2=.45)
N = 610 Power g1=0: 80.2 b2=0: 99.8
joint: 80
```

Here, the first three arguments specify binary exposure, continuous mediator, and survival outcome, respectively, and psi gives the value ψ, the proportion of uncensored failure times. A sample size of 610 is estimated to provide 80.2 % power to reject $\gamma_1 = 0$, 99.8 % power to reject $\beta_2 = 0$, and 80.0 % power to reject both null hypotheses. In this case, where the second link in the indirect pathway is much stronger than the first, virtually all type II error is expected to come from testing $\gamma_1 = 0$.

Illustrative Example

In a study of 87 methamphetamine-using men who have sex with men (MSM) seeking drug treatment, Carrico et al. (2013) showed that positive emotions were associated with a lower frequency of self-reported methamphetamine use in the past 30 days and lower odds of testing positive for stimulant drugs in urine samples, adjusting for age, ethnicity, HIV status, and negative affect.

A possible next step is to propose an intervention to reduce methamphetamine use by increasing positive emotions (R01 DA033854; ClinicalTrials.Gov registration #NCT01926184). We used the proposed methods to estimate the number of participants needed to detect mediation of the intervention effect by a standardized continuous post-randomization measure of positive emotions. Here, the intervention exposure is binary, with 50 % assigned to the active arm. Focusing first on frequency of methamphetamine use, a standardized continuous outcome, we set b2 = 0.29, as estimated by Carrico et al. (2013). Because the intervention will be randomized, rho1was assumed to be zero, the default value, while rho2 was set to .30, following the convention established by Hsieh et al. (1998) to account for mediator-outcome confounding. The key unknown parameter is g1, the effect of the intervention on the mediator. Following Cohen's conventions for effect sizes, we specified g1 as the square root of R^2 = 13 %, a medium standardized effect size. The R function call and result are

```
> sampsi(2, 1, 1, f1=.5, g1=sqrt(.13),
b2=.29, rho2=.30)
N = 241 Power g1=0: 81.1 b2=0: 98.8
joint: 80.1
```

Under these assumptions, 241 participants would be needed to detect mediation using the joint test. Additional calculations show that sample sizes of 621 and 149 would be needed to detect mediation with small (R^2 = 5 %) and large (R^2 = 25 %) effects of the intervention on positive emotions, respectively.

Carrico et al. (2013) also reported positive urine tests among 31 % of participants and an odds ratio of 1.29 per SD for the effect of the positive emotions on this binary outcome. We also hypothesized an odds ratio of 1.1 for the direct effect of the intervention on the outcome, a nuisance parameter in this calculation. Again setting $g1 = \sqrt{.13}$, the R function call and result are

```
> sampsi (2, 1, 2, f1=.5, g1=sqrt(.13),
b1=log(1.1), b2=log(1.29), EY=.31,
+ rho2=.30)
N = 666 Power g1=0: 99.7 b2=0: 80.3
joint: 81.2
```

Thus, we find that 666 participants would be needed to detect mediation of the intervention effect on urine positivity by positive emotions. An additional calculation shows that with a direct treatment effect $\beta_1 = \log(1.5)$ on urine positivity, the required sample size would increase to 691.

Simulation Studies

Simulations were used to assess the performance of the sample size estimators. In the first step, we used the R implementation to compute sample sizes supposed to provide 80 % power in tests with a two-sided type I error rate of 5 %, using (4), (5), (6), and (7), for a range of the relevant parameters. For binary predictors, f_1 and $f_2 = 0.25$ or 0.5, while for continuous X_1 and X_2, $\sigma_1 = \sigma_2 = 1$. For the linear model $\sigma_e = 1$, for the logistic and Poisson models, $E(Y) = 0.2$ or 0.5, and for the Cox model, $\Phi = 0.2$ or 0.5. A common range of values for γ_1 was used for each configuration of exposure and mediator. Values of β_2 were chosen so that the resulting sample size or number of events was in the small-to-moderate range. We also modeled moderate confounding of both the exposure-mediator and mediator-outcome relationships ($\rho_1 = \rho_2 = 0.3$). To simplify data generation, we assumed Poisson-distributed count outcomes without over-dispersion ($\varphi = 1$) and independence ($\delta = 1$). For each configuration, 16 or 18 scenarios were examined.

For each model and set of parameters, 1,000 datasets were generated of the size specified by the sample size estimation procedures. In each dataset, X_1 was generated as standard normal or binary with exact prevalence f_1, then X_2 given X_1 was generated using (1). Continuous, binary, and count outcomes were generated using (2). Finally, survival times were generated using (3), assuming a constant baseline hazard, with the shortest $d = n\psi$ treated as events, and the rest censored. In a final step, Wald tests of $\gamma_1 = 0$ and $\beta_2 = 0$ were conducted, and P_J, the power of the joint test, estimated by the proportion of datasets in which both null hypotheses were rejected.

Results

Figure 1 plots $\hat{P_J}$ against the calculated sample size or number of events for each type of outcome. The two dashed horizontal lines give approximate 95 % point-wise margins of simulation error for a true power of 80 %. The figure shows that in all of the scenarios, a

calculated power of 80 % corresponded to simulated power of 75–85 %, indicating that the formulas may be very useful in practice. Excursions from the expected point-wise simulation error with a range of 77.5–82.5 % were observed in 8 % of scenarios, respectively, only slightly more than the expected 5 %, and fairly mild. Tabular results corresponding to the figure are available online. We also used regression to assess patterns in the simulation results. Table 1 shows linear regression effects on the absolute deviations from 80 % power in the simulations presented in Fig. 1. We found evidence for small increases in average absolute error with stronger exposure-mediator and weaker mediator-outcome relationships, count outcomes, and scenarios with continuous exposure and binary mediator.

Discussion

In this paper, simulation methods have been proposed for estimating sample size and power for joint testing of both links in an indirect pathway from exposure to mediator to outcome. The new methods, which build on the work of Wang and Xue (2012), ensure adequate power when both links in the indirect pathway must be established.

Many approaches have been proposed for computing power and sample size for generalized linear models with multiple predictors. In contrast to methods based on score (Self and Mauritsen 1988; Lyles et al. 2007) and likelihood-ratio tests (Lyles et al. 2007; Self et al. 1992; Shieh 2000), our proposed methods, like those of several other authors (Bernardo et al. 2000; Whittemore 1981; Wilson and Gordon 1986; Signorini 1991; Shieh 2005; Schoenfeld and Borenstein 2005; Demidenko 2007), are based on Wald tests, which are the default in regression routines in most statistical packages and commonly used in practice, although they may be less reliable in small samples and at alternatives far from the null (Hauck and Donner 1977). The proposed methods rely on approximation of the expected information matrix. Analytic results for the information matrix have previously been obtained for a limited number of cases (Demidenko 2007; Signorini 1991; Whittemore 1981; Bernardo et al. 2000; Vittinghoff et al. 2008). We also follow others (Schoenfeld and Borenstein 2005; Demidenko 2007) in estimating the information matrix only under the alternative rather than the null hypothesis (Wilson and Gordon 1986; Shieh 2005). While numeric integration has been implemented (Schoenfeld and Borenstein 2005), the simulation approach we propose is simpler and flexible, but not without limitations (Lyles et al. 2007; Glueck and Muller 2003).

Likewise, several approaches have been used to test for indirect effects. These include tests of $\gamma_1\beta_2 = 0$, sometimes implemented within structural equation models, and tests of the difference in coefficients $\beta_1^* - \beta_1 = 0$, where β_1^* is the coefficient for exposure in a reduced model excluding the mediator. In addition, Judd and Kenny (1981) and Baron and Kenny (1986) proposed more elaborate, but less powerful, joint testing setups, requiring an additional test of $\beta_1^* = 0$ with Judd and Kenny (1981) also requiring that β_1 should not be statistically different from zero in the full outcome model adjusting for X_2. An anonymous reviewer helpfully noted that a test for the total effect is not needed to establish mediation, nor is demonstration that the direct effect is zero, because partial mediation is common in practice. In analyses using logistic and Cox models for the outcome, tests of the difference

in coefficients are complicated by so-called *collapsibility* (Robins and Greenland 1992): that is, when $\beta_2 \neq 0$, β_1 and β_1^* may differ systematically even if $\gamma_1 = 0$, especially when X_2 is common but not ubiquitous. Moreover, for each testing approach, power and type I error rates also differ across many proposed standard error estimators, including bootstrap resampling; see MacKinnon et al. (2002) for a review. Joint testing of $\gamma_1 = 0$ and $\beta_2 = 0$ avoids problems with collapsibility in logistic and Cox models, as well as the non-normal distribution of the test statistic $\hat{\gamma_1}\hat{\beta_2}$, and provides a good tradeoff of power and type I error (MacKinnon et al. 2002).

The proposed methods have several limitations. First, the sample size estimates may be less accurate for other methods of testing for an indirect effect. MacKinnon et al. (2002) show that power for the alternative tests can differ from the power of the joint test; similarly, Fritz and MacKinnon (2007) document substantial variation in sample size estimates across testing approaches. Second, the proposed methods do not accommodate cases where exposure and mediator interact in the model for the outcome; this will be the focus of future work. Third, the approximate method used to account for confounding of the exposure-mediator and mediator-outcome relationships may sometimes be inaccurate. Our simulation results suggest that implementations involving strong exposure-mediator and weak mediator-outcome relationships, count outcomes, and scenarios with continuous exposure and binary mediator may warrant extra caution.

In addition, while Hsieh et al. (1998) propose 0.3 as a default value for the multiple correlation with confounders, this input, as well as the over-dispersion of count outcomes and the design effect in clustered studies, may be unknown or hard to estimate. A general problem in all sample size planning is identifying reasonable values for the required inputs, which are more numerous in the mediation context. As in other contexts, some inputs may be chosen by convention (e.g., $\alpha = 0.05$ and power of 80 %). Others, however, may need to be specified based on relevant findings in the literature, pilot study data, or clinically meaningful differences. When specific values for input parameters are difficult to identify, a range of plausible values may be considered, which our easy-to-use methods facilitate.

Finally, the asymptotic independence of the two Wald statistics for testing $\gamma_1 = 0$ and $\beta_2 = 0$ may not hold exactly with small or moderate sample sizes. In particular, when $\hat{\gamma_1}$ is large by chance, reflecting greater than expected correlation between X_1 and X_2, Var($\hat{\beta_2}$) should increase, making it more difficult to reject $\beta_2 = 0$ and thus both null hypotheses (Fritz et al. 2012). However, the simulations suggest reasonable robustness to such violations for the range of inputs examined. While they show slight inaccuracies in some cases, these are likely small compared to the effects of uncertainty in prior knowledge about γ_1 and β_2 as well as other required parameters.

In summary, we have proposed easy-to-use methods for calculating sample sizes for joint testing of both links in an indirect mediating pathway analyzed using linear, logistic, Poisson, and Cox models. These methods can also be used to compute power when investigators propose examining mediation using already collected data. The implementation requires straightforward inputs, accounts for confounding of the exposure-mediator and mediator-outcome relationships, over-dispersion of count outcomes, and

design effects due to clustering, and has been validated using simulation. R code and documentation are available online.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work of Torsten B. Neilands on this research was supported by National Institutes of Health Grant P30 MH062246. The authors thank Charles E. McCulloch and Steven Gregorich for helpful discussions.

References

Baron R, Kenny D. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986; 51:1173–1182. [PubMed: 3806354]

Bernardo M, Lipsitz S, Harrington D, Catalano P. Sample size calculations for failure time random variables in non-randomized studies. Journal of the Royal Statistical Society (Series D): The Statistician. 2000; 49:31–40.

Breen R, Karlson K, Holm A. Total, direct, and indirect effects in logit and probit models. Sociological Methods & Research. 2013; 42:164–191.

Carrico A, Woods W, Siever M, Discepola M, Dilworth S, Neilands T, Miller N, Moskowitz J. Positive affect and processes of recovery among treatment-seeking methamphetamine users. Drug and Alcohol Dependence. 2013; 132:624–629. [PubMed: 23684632]

Cole S, Hernán M. Fallibility in estimating direct effects. International Journal of Epidemiology. 2002; 31:163–165. [PubMed: 11914314]

Demidenko E. Sample size determination for logistic regression revisited. Statistics in Medicine. 2007; 26:3385–3397. [PubMed: 17149799]

Freedman L, Schatzkin A. Sample size for studying intermediate endpoints within intervention trials or observational studies. American Journal of Epidemiology. 1992; 136:1148–1159. [PubMed: 1462974]

Fritz M, MacKinnon D. Required sample size to detect the mediated effect. Psychological Science. 2007; 18:233–239. [PubMed: 17444920]

Fritz M, Taylor A, MacKinnon D. Explanation of two anomalous results in statistical mediation research. Multivariate Behavioral Research. 2012; 47:61–87. [PubMed: 24049213]

Glueck D, Muller K. Adjusting power for a baseline covariate in linear models. Statistics in Medicine. 2003; 22:2535–2551. [PubMed: 12898543]

Hauck W, Donner A. Wald's test as applied to hypotheses in logit analyses. Journal of the American Statistical Association. 1977; 72:851–853.

Hicks R, Tingley D. Causal mediation analysis. The Stata Journal. 2011; 11:605–619.

Hsieh F, Bloch D, Larsen M. A simple method of sample size calculation for linear and logistic regression. Statistics in Medicine. 1998; 17:1623–1634. [PubMed: 9699234]

Hsieh F, Lavori P. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. Controlled Clinical Trials. 2000; 21:552–560. [PubMed: 11146149]

Imai K, Keele L, Yamamoto T. Identification inference, and sensitivity analysis for causal mediation effects. Statistical Science. 2010; 25:51–71.

Judd C, Kenny D. Process analysis: Estimating mediation in treatment evaluations. Evaluation Review. 1981; 5:602–619.

Kalbfleisch, J.; Prentice, R. The Statistical Analysis of Failure Time Data. New York: Wiley; 1980.

Kenny, D. PowMedR. R program to compute power of joint test for continuous exposure, mediator, and outcome. 2013. Available at http://davidakenny.net/progs/PowMedR.txt

Kohler U, Karlson K, Holm A. Comparing coefficients of nested nonlinear probability models. The Stata Journal. 2011; 11:420–438.

Lyles R, Lin HM, Williamson J. A practical approach to computing power for generalized linear models with nominal, count, or ordinal responses. Statistics in Medicine. 2007; 26:1632–1648. [PubMed: 16817148]

MacKinnon D, Lockwood C, Brown C, Wang W, Hoffman J. The intermediate endpoint effect in logistic and probit regression. Clinical Trials. 2007; 4:499–513. [PubMed: 17942466]

MacKinnon D, Lockwood C, Hoffman J, West S, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychological Methods. 2002; 7:83–104. [PubMed: 11928892]

Mallinckrodt B, Abraham W, Wei M, Russell D. Advances in testing the statistical significance of mediation effects. Journal of Counseling Psychology. 2006; 53:372–378.

Muthén, B. Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. 2011. Available at http://www.statmodel.com/examples/penn.shtml#extendSEM

Pearl J. Graphs, causality, and structural equation models. Sociological Methods and Research. 1998; 27:226–284.

Pearl, J. Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence; San Francisco: CA. 2001.

Pearl, J. Tech rep. University of California; Los Angeles: Computer Science Department; 2011. The mediation formula: A guide to the assessment of causal pathways in nonlinear models.

Pearl J. The causal mediation formula—a guide to the assessment of pathways and mechanisms. Prevention Science. 2012; 13:426–436. [PubMed: 22419385]

Petersen M, Sinisi S, van der Laan M. Estimation of direct causal effects. Epidemiology. 2006; 17:276–284. [PubMed: 16617276]

R Development Core Team. R Foundation for Statistical Computing. Austria: Vienna: 2014. R: A language and environment for statistical computing. http://www.R-project.org

Robins J, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992; 3:143–155. [PubMed: 1576220]

Schmoor C, Sauerbrei W, Schumacher M. Sample size considerations for the evaluation of prognostic factors in survival analysis. Statistics in Medicine. 2000; 19:441–452. [PubMed: 10694729]

Schoenfeld D, Borenstein M. Calculating the power or sample size for the logistic and proportional hazards models. Journal of Statistical Computation and Simulation. 2005; 75:771–785.

Self S, Mauritsen R. Power/sample size calculations for generalized linear models. Biometrics. 1988; 44:79–86.

Self S, Mauritsen R, Ohara J. Power calculations for likelihood ratio tests in generalized linear models. Biometrics. 1992; 48:31–39.

Shieh G. On power and sample size calculations for likelihood ratio tests in generalized linear models. Biometrics. 2000; 56:1192–1196. [PubMed: 11129478]

Shieh G. On power and sample size calculations for wald tests in generalized linear models. Journal of Statistical Planning and Inference. 2005; 128:43–59.

Signorini D. Sample size for Poisson regression. Biometrika. 1991; 78:446–450.

Sobel, M. Asymptotic confidence intervals for indirect effects in structural equation models. In: Leinhardt, S., editor. Sociological Methodology. American Sociological Association; 1982. p. 290-312.

Valeri L, VanderWeele T. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods. 2013; 18:1–14. [PubMed: 23148474]

VanderWeele T. Marginal structural models for the estimation of direct and indirect effects. Epidemiology. 2009; 20:18–26. [PubMed: 19234398]

Vittinghoff E, Sen S, McCulloch C. Sample size calculations for evaluating mediation. Statistics in Medicine. 2008; 28:541–557. [PubMed: 19065627]

Wang, E.; Xue, X. Power and sample size calculations for evaluating mediation effects in longitudinal studies. Statistical Methods in Medical Research. 2012. Available at http://www.smm.sagepub.com/content/early/2012/12/05/0962280212465163.full.pdf+html

Whittemore A. Sample size for logistic regression with small response probability. Journal of the American Statistical Association. 1981; 76:27–32.

Wilson S, Gordon I. Calculating sample sizes in the presence of confounding variables. Applied Statistics. 1986; 35:207–213.

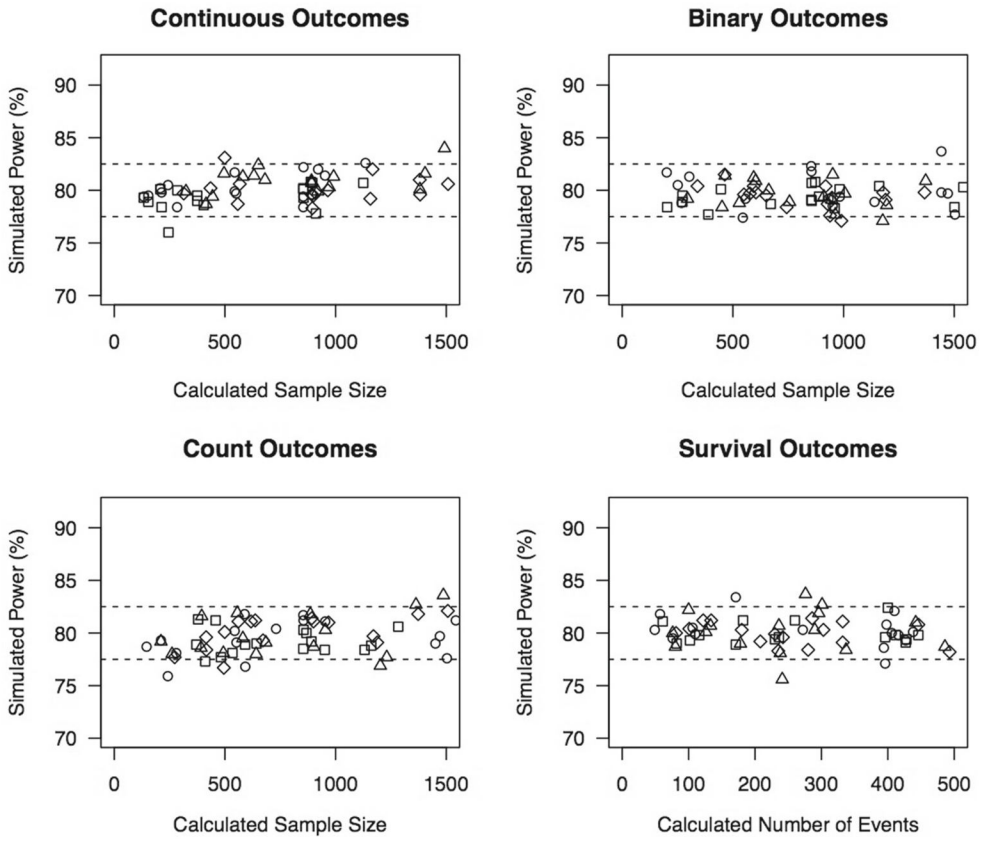

Fig. 1.

Estimated power for continuous, binary, count, and survival outcomes, with moderate confounding of exposure-outcome and mediator-outcome relationships. *Dashed horizontal lines* give approximate 95 % point-wise margins of simulation error for a true power of 80 %; in all scenarios, simulated power is consistent or nearly so with this targeted value. Key to scenarios: □ continuous exposure and mediator; ○ binary exposure, continuous mediator; △ continuous exposure, binary mediator; ◇ binary exposure and mediator

Table 1

Regression effects on absolute percentage point deviations of simulated power from target of 80 %

Factor	Effect	*P* value
Model coefficients		
γ_1	0.88	0.07
β_2	−0.83	0.03
Exposure/mediator		
Continuous/continuous	Ref	–
Binary/continuous	0.02	0.91
Continuous/binary	0.41	0.04
Binary/binary	−0.27	0.35
Outcome		
Continuous	Ref	–
Binary	0.26	0.13
Count	0.60	0.0003
Survival	0.14	0.39

Estimates for γ_1 and β_2 assume standardized continuous exposures and mediators