Sample size and power calculations for joint testing of indirect effects

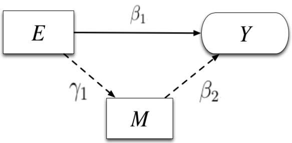
Tor Neilands[1] and Eric Vittinghoff[2]

January 16, 2015

1 Torsten.Neilands@ucsf.edu
2 Eric.Vittinghoff@ucsf.edu

What is mediation?

- Consider the 3-variable sequential mediation model



- *In a mediation model, the independent variable* (E) *causes the mediator* (M)*, which then causes the dependent variable* (Y)[3]
- Other, more complex models available[4]

[3] MacKinnon DP. *Introduction to statistical mediation analysis*. New York, NY, 2008.

[4] Mitchell M, Maxwell SE. A Comparison of the cross-sectional and sequential designs when assessing longitudinal mediation. *Multivariate Behavioral Research*, 2013;48(3):301-339.

Conditions to establish mediation

- Baron and Kenny[5] list 4 steps:
  1. E must be shown to affect Y when M is not included in the analysis
  2. E must be shown to affect M
  3. M must be shown to affect Y, independently of E
  4. The effect of E on Y must be non-significant when M is included in the analysis.
     - or at least differ from the effect when M is omitted (partial mediation)
- (We assume that 2 and 3 suffice)

[5] Baron RM, Kenny DA. The mediator-moderator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 1986;51(6):1173-82.

Quantifying mediation

- For continuous M and Y:

$$
\begin{aligned}
\mathsf{E}[M|E] &= \gamma_0 + \gamma_1 E \\
\mathsf{E}[Y|E,M] &= \beta_0 + \beta_1 E + \beta_2 M \\
\mathsf{E}[Y|E] &= \beta_0^{total} + \beta_1^{total} E
\end{aligned}
$$

- Two ways to quantify the mediated effect:
  1. $\beta_1^{total} - \beta_1$
  2. $\gamma_1 \beta_2$
- Results identical for continuous M and Y but not when either is binary, a count, or a failure time
  - $\beta_1^{total} - \beta_1$ confounded by non-collapsibility of ORs, HRs

Binary, count and failure time M and Y

- $\beta_1^{total} - \beta_1 = \gamma_1\beta_2$ only for continuous M and Y
- What should we do with other types of M and Y?
  - Focus on $\gamma_1\beta_2$?
    - easier to calculate (does not require rescaling)
    - more accurate[6] than $\beta_1^{total} - \beta_1$
    - may lack a clear interpretation
  - KHB method[7] for consistent estimation of $\beta_1^{total} - \beta_1$ using linear probability model for binary and count M
    - implemented in downloadable Stata package `khb`[8]

---

[6] MacKinnon DP. *Introduction to statistical mediation analysis*. New York, NY, 2008, p. 321.

[7] Breen R, Karlson KB, Holm A. Total, direct, and indirect effects in logit and probit models. *Sociological Methods & Research*, 2013;42(2):164-191.

[8] Kohler U, Karlson KB, Holm A. Comparing coefficients of nested nonlinear probability models. *The Stata Journal*, 2011;11(3):420-38. Downloadable from http://www.stata-journal.com/software/sj13-1

Non-normality of $\gamma_1\beta_2$

- $\hat{\gamma}_1$ and $\hat{\beta}_2$ may be normally distributed, but $\widehat{\gamma_1\beta_2}$ is usually not. How to handle this?
  - ignore it
  - product of normal variables method
  - bootstrap-based confidence intervals (CIs)
  - joint testing of $\gamma_1 = 0$ and $\beta_2 = 0$

Sobel's method

- Sobel[9] proposed using the δ-method to compute

$$\mathsf{SE}(\widehat{\gamma_1\beta_2}) = \sqrt{\hat{\gamma}_1^2 \mathsf{SE}^2(\hat{\beta}_2) + \hat{\beta}_2^2 \mathsf{SE}^2(\hat{\gamma}_1)}$$

  then using $\mathsf{SE}(\widehat{\gamma_1\beta_2})$ to calculate Normal-based CI
- Problem: distribution of $\widehat{\gamma_1\beta_2}$ may be badly skewed

[9] Sobel ME. Direct and indirect effects in linear structural equation models. *Sociological Methods and Research*, 1982;16(1):155-176.

Relaxing the normality assumption

- Product-of-normal variables method based on analytic formulas
  - results in asymmetric confidence intervals for $\gamma_1\beta_2$
  - implemented in freeware PRODCLIN[10]
  - only works for 3-variable models

[10] Mackinnon DP, Fritz MS, Williams J, Lockwood C. Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 2007;39(3):384-389. Downloadable from http://amp.gatech.edu/RMediation

Relaxing the normality assumption

- Monte-Carlo method extends to >3 variable models[11]
    - simulate distribution of $\widehat{\gamma_1\beta_2}$ assuming $\hat{\gamma}_1$ and $\hat{\beta}_2$ are joint normal ($\widehat{\gamma_1\beta_2}$ need not be normally distributed)
- Advantages:
    - performance comparable[11] to bootstrap CIs for $\widehat{\gamma_1\beta_2}$
    - only summary information needed
    - runs relatively fast
- Disadvantages:
    - requires point estimates of $\hat{\gamma}_1$ and $\hat{\beta}_2$ plus their asymptotic covariance matrix
    - more conservative than bias-corrected bootstrap[12]

[11] Preacher KJ, Selig JP. Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 2012. Downloadable from http://dx.doi.org/10.1080/19312458.2012.679848

[12] Hayes AF, Scharkow M. The relative trustworthiness of inferential tests of the indirect in statistical mediation analysis: does method really matter? *Psychological Science*, 2013;24(10):1918-27.

Bootstrap CIs for mediated effect

- Allows for multiple mediators
- Debate about best bootstrapping method:
  - bias-corrected CIs may have better power[13]
  - percentile-based CIs better preserve type-I error rate[14]
- Computationally complex
- Slow for sample size calculations[15]

[13] Mackinnon DP, Fritz MS, Williams J, Lockwood C. Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 2007;39(3):384-389.

[14] Fritz MS, Taylor AB, MacKinnon DP. Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 2012;47(1):61-87.

[15] Zhang Z. Monte Carlo based statistical power analysis for mediation models: methods and software. *Behavior Research Methods*, 2014;46:1184-98.

Testing $\beta_2 = 0$ only

- Clogg, Petkova, & Cheng[16], then Vittinghoff, Sen & McCulloch[17] used this shortcut
- Rationale: β_2 reflects the influence of M on Y[18]
- Gregorich,[19] then Wang and Xue[20] showed this underestimates sample size if $\gamma_1 \neq 0$ must be established

---

[16] Clogg CC, Petkova E, and Cheng T. Reply to Allison: More on comparing regression coefficients. *American Journal of Sociology*, 1995;100:1301-12.

[17] Vittinghoff E, Sen Ś, McCulloch CE. Sample size calculations for evaluating mediation. *Statistics in Medicine*, 2008;28(4):541-557.

[18] MacKinnon DP. *Introduction to statistical mediation analysis*. New York, NY, 2008.

[19] personal communication, 2008.

[20] Wang C, Xue X. Power and sample size calculations for evaluating mediation effects in longitudinal studies. *Statistical Methods in Medical Research*, 2012. http://smm.sagepub.com/content/early/2012/12/05/0962280212465163.full.pdf+html.

Joint testing of $\gamma_1 = 0$ and $\beta_2 = 0$

- In many contexts, we can't just assume $\gamma_1 \neq 0$
- Also, large values of γ_1 increase correlation of E and M, reducing power to reject $\beta_2 = 0$[21]
- Joint testing of $\gamma_1 = 0$ and $\beta_2 = 0$
  - establishes both steps in indirect pathway
  - faster and easier than bootstrapping $\gamma_1\beta_2$
  - has good tradeoff of type 1 error rates and power[22,23]
  - achieves performance comparable to bootstrap test[24]

[21] Fritz MS, Taylor AB, MacKinnon DP. Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 2012;47(1):61-87.

[22] 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(1):83-104.

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

[24] Hayes AF, Scharkow M. The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 2013;24:1918-27.

Programs for the 3-variable model

- Kenny R program PowMedR with a graphical user interface[25]
- Vittinghoff R program for testing $\beta_2 = 0$[26]
    - *sample size can be too small if we need to show $\gamma_1 \neq 0$*

[25] http://davidakenny.net/webinars/Mediation/PowMedR/PowMedR.html.

[26] http://www.epibiostat.ucsf.edu/biostat/mediation/

Simulation-based tools using M*plus*

- Monte Carlo programs for 3-variable, multiple-variable, and longitudinal mediation models with observed and latent variables[27]
- Monte Carlo simulations based on causal inference foundation[28]
    - accommodates nominal categorical M

[27] Thoemmes F, MacKinnon DP, Reiser MR. Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, 2010;17(3),510-534.

[28] Muthén BO. Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. http://www.statmodel.com/examples/penn.shtml#extendSEM

Simulations using LSEM, bootstrap CIs

- `R` simulations[29] calling the `R` linear structural equation modeling program `laavan`
  - handles latent variables, multiple mediators, and non-normal distributions
  - currently supports only continuous M and Y
  - time-consuming to set up
  - requires more assumed inputs
  - *can be very slow*

[29] Zhang Z. Monte Carlo based statistical power analysis for mediation models: methods and software. *Behavior Research Methods*, 2014;46:1184-98.

Fast method for joint testing

- Assume GLMs for continuous, binary, and count M and Y

$$
\begin{aligned}
h_1[\mathsf{E}(M|E)] &= \gamma_0 + \gamma_1 E \\
h_2[\mathsf{E}(Y|E,M)] &= \beta_0 + \beta_1 E + \beta_2 M
\end{aligned}
$$

and Cox proportional hazards model for failure time Y

$$
\lambda(t, E, M) = \lambda_0(t)\exp(\beta_1 E + \beta_2 M)
$$

- Joint testing uses Wald tests of $\gamma_1 = 0$ and $\beta_2 = 0$
  - Type 1 error rate asymptotically bounded by common nominal type 1 error rate for both tests[30]

[30] Wang C, Xue X. Power and sample size calculations for evaluating mediation effects in longitudinal studies. *Statistical Methods in Medical Research*, 2012. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3883797/

Crucial assumption

- Define

$$
\begin{aligned}
P_{\gamma_1,\beta_2} &:= \text{probability of rejecting } \gamma_1 = 0 \textit{ and } \beta_2 = 0 \\
P_{\gamma_1} &:= \text{probability of rejecting } \gamma_1 = 0 \\
P_{\beta_2} &:= \text{probability of rejecting } \beta_2 = 0
\end{aligned}
$$

- Easy to estimate P_{γ_1} and P_{β_2} but not P_{γ_1,β_2}
- Following Wang and Xue,[31] *assume* $P_{\gamma_1,\beta_2} \approx P_{\gamma_1} \times P_{\beta_2}$
- Could fail,[32] but simulations suggest that it holds approximately

[31] Wang C, Xue X. Power and sample size calculations for evaluating mediation effects in longitudinal studies. *Statistical Methods in Medical Research*, 2012. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3883797/

[32] Fritz MS, Taylor AB, MacKinnon DP. Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*, 2012;47(1):61-87.

Implementation

- Uses line search in N, stopping when $P_{\gamma_1} \times P_{\beta_2}$ for candidate N equals target power
- Can also be used to estimate power for fixed sample size
- Accommodates continuous and binary E and M
- Assumes linear, logistic, Poisson, and Cox models for continuous, binary, count, and failure time Y
- `R` program `medssp.R`, available with documentation at *Prevention Science* website[33] or from Eric or Tor
- `R` is freeware, can be downloaded from CRAN website[34] for MAC, PC, and Linux machines; easy to install

[33] http://link.springer.com/article/10.1007/s11121-014-0528-5

[34] http://cran.r-project.org

Confounding of $M \to Y$ by E

- Power calculation for Wald test of $\beta_2 = 0$ requires standard error of $\hat{\beta}_2$ accounting for $E - M$ correlation
- SE($\hat{\beta}_2$) estimated by simulating Cov($\hat{\boldsymbol{\beta}}$) $= (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$
  - $\mathbf{X} = (\mathbf{E}, \mathbf{M})$, the design matrix
  - $\mathbf{V} =$ Cov($\mathbf{Y}|\mathbf{X}$), a function of E$[Y|E, M]$
- Three steps:
  1. simulate 10,000 observations from assumed joint distribution of E, M, and E$[Y|E, M]$
  2. calculate $(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$ and rescale to candidate N
  3. extract diagonal element corresponding to $\hat{\beta}_2$
- This method also used to estimate SE($\hat{\gamma}_1$) for binary M

Confounding of $E \to M$ and $M \to Y$

- Calculations must *also* control for confounding of both $E \to M$[35] and $M \to Y$[36,37] by other factors
- Specifying joint distribution of E, M, Y, and additional confounders is too difficult
  - use approximations based on variance inflation factor[38]
- Analogous variance-inflating approximations used for design effects and over-dispersion

[35] VanderWeele TJ, Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 2009;20:18-26

[36] Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. *Evaluation Review*, 1981;5(5):602-19

[37] Cole SR, Hernán MA. Fallibility in estimating direct effects. *International Journal of Epidemiology*, 2002;31:163-5

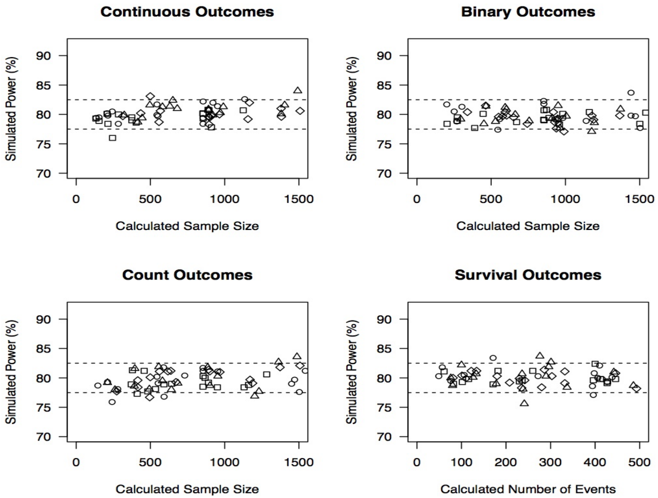
[38] Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic models. *Statistics in Medicine*, 1998;17:1623-34

Nuisance parameters

- Implementation requires you to guesstimate a *lot*, in addition to hypothesized values of γ_1 and β_2
    - SD [prevalence] of continuous [binary] E and M
    - Joint correlation of E and M with additional confounders of $E \rightarrow M$ and $M \rightarrow Y$, respectively
    - SD of continuous Y, marginal mean of binary or count Y, fraction censored for failure time Y
    - Direct effect β_1 of E on Y given M
    - Over-dispersion of count outcome
    - Design effect in clustered data

Simulations to evaluate performance of `medssp.R`

- For each combination of
  - continuous and binary E and M
  - continuous, binary, count, and failure time Y
  - a range of values for γ_1, β_2, and nuisance parameters, including confounding of $E \to M$ and $M \to Y$
- Generate 1,000 datasets with N specified by `medssp.R`
  1. simulate E
  2. simulate M given E and confounder of $E \to M$
  3. simulate Y given E, M, and confounder of $M \to Y$
  4. estimate γ_1 and β_2
- Estimate power by proportion of datasets in which *both* $\gamma_1 = 0$ and $\beta_2 = 0$ are rejected


Continuous Outcomes
Binary Outcomes
Count Outcomes
Survival Outcomes
Simulated Power (%)
90
85
80
75
70
0
500
1000
1500
Calculated Sample Size
100
200
300
400
Calculated Number of Events

Predictors of absolute deviations from 80% power

	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
failure time	0.14	0.39

Variable definitions in medssp.R

- R code written using x1 for E, x2 for M

sdx1, sdx2	SD of continuous E and M
f1, f2	prevalence of binary E and M
g1, b1, b2	γ_1, β_1, and β_2
rho1, rho2	correlation of E and M with confounders
sdy	SD of continuous Y
EY	marginal mean of binary or count Y
psi	fraction of uncensored failure times
scale	over-dispersion scale factor for count Y
de	design effect for clustered data

- Variable type coding for E, M, and Y (in that order):
  - $1 =$ continuous, $2 =$ binary, $3 =$ count, $4 =$ failure time

Default values in `medssp.R`

`sdx1`, `sdx2`	1
`f1`, `f2`	no default
`g1`, `b1`, `b2`	no default
`rho1`, `rho2`	0 (no confounding)
`sdy`	1
`EY`	no default
`psi`	no default
`scale`	1 (no over-dispersion)
`de`	1 (no design effect)

Continuous E, M, and Y

- Variable type codes 1, 1, 1 for E, M, and Y
- SD(E) = SD(M) = SD$(Y) = 1$ (default values)
- $\gamma_1 = \texttt{g1} = .25$
- No need to specify $\beta_1 = \texttt{b1}$ with continuous Y
- $\beta_2 = \texttt{b2} = 0.20$
- No confounding of $E \rightarrow M$, as in a trial (the default)
- Moderate confounding of $M \rightarrow Y$ ($\rho_2 = \texttt{rho2} = 0.3$)[39]

```
> sampsi(1, 1, 1, g1=.25, b2=.20, rho2=.3)

N = 240  Power g1=0: 97.9  b2=0: 81.9  joint: 80.2
```

[39] Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic models. *Statistics in Medicine*, 1998;17:1623-34.

Binary E, M, and Y

- Variable type codes 2, 2, 2 for E, M, and Y
- $\Pr(E = 1) = \texttt{f1} = 0.5$
- $\Pr(M = 1) = \texttt{f2} = 0.35$
- $\gamma_1 = \log(2.1)$ $\beta_1 = \log(1.5)$ $\beta_2 = \log(1.9)$
- Moderate confounding of $E \to M$ $(\rho_1 = 0.25)$
- Moderate confounding of $M \to Y$ $(\rho_2 = 0.35)$
- Design effect $=$ de $= 1.5$
- $\Pr(Y = 1) = \texttt{EY} = 0.4$

```
> sampsi(2, 2, 2, f1=.5, f2=.35, g1=log(2.1), b1=log(1.5),
+       b2=log(1.9), rho1=.25, rho2=.35, de=1.5, EY=.4)

N = 690  Power g1=0: 94.9  b2=0: 84.3  joint: 80
```

Continuous E, binary M, count Y

- Variable type codes 1, 2, 3 for E, M, and Y
- SD(E) $=$ `sdx1` $= 1.25$
- $\Pr(M = 1) =$ `f2` $= 0.35$
- $\gamma_1 = \log(1.4)$ $\beta_1 = \log(1.5)$ $\beta_2 = \log(1.35)$
- Moderate confounding of $E \rightarrow M$ $(\rho_1 = 0.35)$
- Moderate confounding of $M \rightarrow Y$ $(\rho_2 = 0.25)$
- Over-dispersion of Y by `scale` factor of 1.5
- Marginal mean of $Y =$ `EY` $= 2$

```
> sampsi(1, 2, 3, sdx1=1.25, f2=.35, g1=log(1.4), b1=log(1.5),
+        b2=log(1.35), rho1=.35, rho2=.25, scale=1.5, EY=2)

N = 351  Power g1=0: 91.6  b2=0: 87.3  joint: 80.2
```

Binary E, continuous M, failure time Y

- Variable type codes 2, 1, 4 for E, M, and Y
- $\Pr(E = 1) = 0.2$
- SD$(M) = 1.2$
- $\gamma_1 = 0.35 \quad \beta_1 = \log(1.5) \quad \beta_2 = \log(1.4)$
- Moderate confounding of $E \to M$ $(\rho_1 = 0.25)$
- Strong confounding of $M \to Y$ $(\rho_2 = 0.45)$
- 30% of failure times uncensored (ψ = psi = 0.3)

```
> sampsi(2, 1, 4, f1=.2, sdx2=1.2, g1=.35, b1=log(1.5),
+         b2=log(1.4), rho1=.25, rho2=.45, psi=.3)

N = 610  Power g1=0: 80.2  b2=0: 99.8  joint: 80
```

Do positive emotions mediate intervention effect on frequency of methamphetamine use?

- Carrico *et al.*[40] showed that positive emotions were associated a lower frequency of self-reported methamphetamine use in the past 30 days
- Next step: Propose an RCT of intervention E to reduce frequency of methamphetamine use Y by increasing positive emotions M
- Use inputs from Carrico pilot study

[40] 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;132,624-9.

Input assumptions

- Binary E (treatment), continuous M (positive emotions) and continuous Y (frequency of meth use)
- From pilot study:
  - SD(M) $=$ `sdx2` $= 1$ (standardized measure)
  - $\beta_2 =$ `b2` $= 0.29$
- By design or assumption:
  - 50% randomized to active arm ($\Pr[E = 1] =$ `f1` $= 0.5$)
    - so no confounding of $E \rightarrow M$ ($\rho_1 =$ `rho1` $= 0$)
  - $\gamma_1 =$ `g1` $= \sqrt{13\%}$, a medium standardized effect size[41]
  - moderate confounding of $M \rightarrow Y$ ($\rho_2 =$ `rho2` $= 0.3$)

[41] Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey, 1987.

Limitations of `medssp.R`

- Does not handle $E - M$ interactions or multiple mediators
- Assumes normality for continuous outcomes, proportional hazards for failure times
- Requires specification of many nuisance parameters
- May be inaccurate for other ways of testing for mediation
- No GUI

Summary

- We propose a method for sample size and power calculation for joint testing of both steps in indirect mediating pathway $E \rightarrow M \rightarrow Y$
- Implementation in R program medssp.R accommodates
  - continuous and binary E and M
  - continuous, binary, count, and failure time Y
  - confounding of $E \rightarrow M$ and $M \rightarrow Y$
  - design effects
  - over-dispersion of count outcomes
- Accurate, fast, easy to use, freely available

Acknowledgements

- NIMH P30 MH062246 (Lightfoot, PI)
- Steve Gregorich and Chuck McCulloch for many conversations on calculating sample size and power for mediation problems