Invited Commentary

Invited Commentary: Frontiers of Power Assessment in Mediation Analysis

Tyler J. VanderWeele*

* Correspondence to Dr. Tyler J. VanderWeele, Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston MA 02115 (e-mail: tvanderw@hsph.harvard.edu).

Initially submitted April 2, 2020; accepted for publication April 15, 2020.

The development of tools for power and sample-size calculations for mediation analysis has lagged far behind the development of methods. The accompanying paper by Rudolph et al. (*Am J Epidemiol.* 2020;189(12):1559–1567) is a helpful contribution in using simulations as a tool for power calculations for more complex methods and settings. Much work remains to be done in the development of easy-to-use packages and simple online websites for carrying out power and sample-size calculations for mediation analysis. Much remains to be learned with respect to the relative power of different methods in different settings. There will likely be feedback between these 2 important frontiers of the tools we have available and of our understanding of power when conducting mediation analysis.

direct and indirect effects; mediation; power; sample size; software; statistics

Abbreviation: TMLE, targeted maximum likelihood estimation.

Editor's note: *The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the American Journal of Epidemiology.*

The accompanying paper by Rudolph et al. (1) is a helpful contribution in advancing the work on statistical power in mediation analysis. As the authors note, and as also discussed in my book (2), the state of power and sample-size calculation tools for mediation analysis has lagged considerably behind the development of methods. Many researchers have relied on the tables provided by Fritz and MacKinnon (3), obtained by simulations with linear models, though they have sometimes been employed even when other models are in fact used. As Rudolph et al. point out (1), investigators have also sometimes used various power and sample-size formulas for detection of mediation by employing the “difference method” of adding the mediator to the outcome regression model and seeing whether this changes the coefficient of the exposure (4). However, methods for mediation have expanded dramatically over the past couple of decades, both in psychology (5) and in the causal inference literature (1, 2). The literature on power and sample-size calculations has not been keeping pace.

Rudolph et al. help begin to correct this by providing software code for, and an example of the use of, simulations to assess the power of different estimation methods under different data-generating models (1). Attempting to derive analytical formulas for power and sample-size calculations for many of the more recent methods will prove challenging or impossible. Many targeted maximum likelihood estimation (TMLE) methods now employ ensemble approaches and choose among or weight different statistical models based on fit using cross-validation. Moreover, the possibility of mediator-outcome confounders affected by the exposure, a situation considered by Rudolph et al. (1), will likewise render analytical formulae for power and sample-size calculations more challenging. In situations like these, simulations will probably be needed. Simulations might also be preferable because often the power and sample-size formulae that are derived analytically apply only asymptotically, and when sample sizes are more modest, there can be considerable deviation from the analytical formulae. The paper by Rudolph et al. (1), using simulations, thus arguably points the way forward for having power and sample-size calculation tools “catch up” with the vast array of methods now available for mediation analysis. I would like to point towards 2 frontiers that I think will be important to make

progress on before we are at the point where the tools for power and sample-size calculations for mediation analysis can be considered satisfactory.

First, while Rudolph et al. helpfully make the code for their simulations publicly available and also even annotate it (1), more will be needed if such tools are to be employed. Experience suggests that unless very easy-to-use tools are available, uptake will likely be very low. For the computationally adept epidemiologist or methodologist, the code Rudolph et al. provide will be a helpful starting point for adaptation. However, for most clinical or public health researchers interested in carrying out mediation analyses to assess mechanisms, more will be needed. They will probably not proceed with power and sample-size calculations for mediation analyses unless more accessible tools are available. We need either a simple online website or, alternatively, a very simple 1-line R (R Foundation for Statistical Computing, Vienna, Austria) or Stata (StataCorp LLC, College Station, Texas) command or SAS (SAS Institute, Inc., Cary, North Carolina) or SPSS (SPSS, Inc., Chicago, Illinois) macro with which to implement these approaches more immediately. The sort of code, and the overall approach, that Rudolph et al. provide (1) is a good starting point, but development of simple packages and tools is needed. Ideally these tools would provide not only power calculations but also sample-size calculations, and sample-size calculations not only for detecting effects of a specified magnitude with a given power but also for attaining a given level of precision (6). Of course, one of the challenges, already present in the paper by Rudolph et al. (1), is that the array of potential settings and methods that might be employed is vast. For a practical tool or package, and given the work required to create these, some choice and narrowing of options will be needed.

Thus, a second important frontier for power and sample-size calculations for mediation analysis will be to try to discern which methods work best in which settings, and to thereby consequently choose which power and sample-size calculations to implement. In terms of power and sample-size calculations, we are arguably at the very beginning of the process of trying to advance this second frontier. The simulations and figures presented by Rudolph et al. in their main paper (1) and in their Web Appendix (available at <https://academic.oup.com/aje>) do provide some helpful information on the relative performance—in terms of both power and coverage—of different methods under different scenarios. Much more remains to be done.

For better or for worse, Rudolph et al. did not choose the simplest setting for their work. They considered the setting of a binary mediator and a binary outcome in which to carry out their simulations (1). Methodological and interpretative issues would likely have been simpler in the setting of a continuous mediator and a continuous outcome. This is the classic setting for the popular “Baron and Kenny” approach (7). The continuous mediator and continuous outcome setting constitutes the vast majority of the applications of that approach. With a binary outcome, additional complexities are introduced by the consideration of a nonrare outcome, including issues of scale, interpretation, approximation, and noncollapsibility of certain effect scales (2, 8–10). Rudolph

et al. get around these to some extent by reporting effect estimates for the binary outcome on the risk difference scale (1). However, this is not the scale on which direct and indirect effects for a binary outcome have been most frequently reported within epidemiology. Moreover, the linear risk models that they consider in their Web Appendix will be more challenging to implement in practice with continuous covariates, so the relevance of those simulations to practical situations is less clear. While the formulas for the exposure-mediator interaction extension to the Baron and Kenny approach come from the paper by Valeri and VanderWeele (11), this formula was originally intended for the case of a continuous mediator and a continuous outcome; for a binary mediator and a binary outcome, an alternative formula was given for the odds ratio scale. If risk differences were desired for the extension of the Baron and Kenny approach to allow for interaction, the estimator and package of Imai et al. (12), which estimates direct and indirect effects on the risk difference scale, might have been preferable. In any case, all of this is to say that while the present simulations of Rudolph et al. provide some insight as to the relative performance of different methods in different contexts, much more still needs to be done to evaluate the performance of the methods in the settings in which they are used most commonly.

And indeed, that may be a good place to begin the further exploration of relative power across settings: with the settings and methods that are currently most commonly encountered and employed. I would thus suggest, for the time being, focusing on methods and settings without mediator-outcome confounders affected by the exposure rather than with such confounders—that is, settings in the Rudolph et al. paper (1) corresponding to their data-generating mechanisms 1–5, instead of mechanisms 6–13. Once the case without mediator-outcome confounders affected by the exposure is better understood, we can move on to those with such more complex confounders. However, we have a lot to learn even in the simpler setting first. Likewise, I would propose trying to understand the relative power and sample-size properties of the methods and packages most commonly being employed at present, such as those (based on citation counts) by Valeri and VanderWeele (11) and Imai et al. (12) and the difference methods (2, 4). Work should not stop here, and it is important to consider in what settings other methods such as TMLE or inverse odds ratio weighting may be preferable. Rudolph et al. are to be commended for considering these cases (1)—and the TMLE approaches, perhaps unsurprisingly, do reasonably well in at least a number of the settings—but again, more will be needed to discern what power and sample-size calculation tools to implement in which settings. Once we move to the setting with mediator-outcome confounders affected by the exposure, it will be important to examine a variety of methods that offer unbiased estimates of the direct and indirect effects. At present, Rudolph et al. consider only TMLE for this (1), but other estimators and approaches are also available (13–15).

There will inevitably be tradeoffs across methods. When models are correctly specified, parametric approaches such as those of Valeri and VanderWeele (11) will tend to have high power compared with other methods. However, they risk model misspecification, and TMLE methods can be

very helpful addressing this, even if sometimes at the cost of some power. Understanding such tradeoffs through a more extensive and systematic set of simulations will be important both in choosing the power and sample-size tools to implement in easy-to-use packages and in providing practical guidance. There will, however, inevitably be feedback between the tools and our understanding—that is, between these 2 important frontiers. As our understanding advances we will have a better sense as to what tools to build. However, as tools become available, we will more easily be able to carry out extensive ranges of simulations to enhance our understanding.

The field of mediation analysis is long overdue for better tools for power and sample-size calculations. Rudolph et al. (1) have helped move us somewhat closer towards this. Much important work remains to be done. The practice of mediation analysis, and thus our ultimate insight into mechanisms, would benefit from this work's being carried out.

ACKNOWLEDGMENTS

Author affiliations: Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Tyler J. VanderWeele).

This work was funded by National Institutes of Health grant R01 CA222147.

Conflict of interest: none declared.

REFERENCES

1. Rudolph KE, Goin DE, Stuart EA. The peril of power: a tutorial on using simulation to better understand when and how we can estimate mediating effects. *Am J Epidemiol*. 2020;189(12):1559–1567.
2. VanderWeele T. *Explanation in Causal Inference: Methods for Mediation and Interaction*. 1st ed. New York, NY: Oxford University Press; 2015.
3. Fritz MS, MacKinnon DP. Required sample size to detect the mediated effect. *Psychol Sci*. 2007;18(3):233–239.
4. MacKinnon D. *Introduction to Statistical Mediation Analysis*. Abingdon, United Kingdom: Routledge Press; 2012.
5. Vittinghoff E, Sen S, McCulloch C. Sample size calculations for evaluating mediation. *Stat Med*. 2009;28(4):541–557.
6. Rothman KJ, Greenland S. Planning study size based on precision rather than power. *Epidemiology*. 2018;29(5):599–603.
7. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986;51(6):1173–1182.
8. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis with a dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339–1348.
9. Jiang Z, VanderWeele TJ. When is the difference method conservative for mediation? *Am J Epidemiol*. 2015;182:105–108.
10. VanderWeele TJ, Valeri L, Ananth CV. Mediation formulas with binary mediators and outcomes and the “rare outcome assumption”. *Am J Epidemiol*. 2019;188:1204–1205.
11. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18(2):137–150.
12. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010;15(4):309–334.
13. VanderWeele TJ, Vansteelandt S, Robins JM. Methods for effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*. 2014;25(2):300–306.
14. De Stavola BL, Daniel RM, Ploubidis GB, et al. Mediation analysis with intermediate confounding: structural equation modeling viewed through the causal inference lens. *Am J Epidemiol*. 2015;181(1):64–80.
15. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*. 2017;28(2):258–265.