Responses to Unanswered Questions Asked During the Webinar

Provided by: David P. MacKinnon, Ph.D.
Arizona State University
Methods: Mind the Gap
June 18, 2020

1. **Could you please clarify the distinction between confounder variables and covariates? Is the difference that a covariate is associated with one variable only, whereas confounder is associated with both X and Y?**

Yes, that is how I have used the term covariates. Covariates are variables that when added to an analysis do not have an appreciable effect on the relation of X to Y. Often covariates are included to reduce error variance. Note that with a large enough sample size, a covariate may have a reliable effect on the relation of X to Y. Therefore, in this sense, all the variables—collider, mediator, and confounder—are examples of covariates or any covariate is one of these three variables. Generally, covariates refer to all the other variables that could be added to a model. There is more in Chapter 1 of my book.

2. **Any thoughts on causal mediation analysis using observational data collected for another purpose? One limitation may be data on X, M, and Y were collected at different time points. There may be unmeasured confounders.**

That is an interesting question. In general, alternative explanations of results are addressed with design before a study is conducted. Using data for a different purpose than the original purpose may not have addressed alternative explanations of study results in the design of the study. As you mention, accurate interpretation of the new project may require measurement of important confounders that were not measured in the original study. There are so many decisions made in the planning of a research study that may not apply for a new research question. On the other hand, there are many reasons to use existing data to test additional scientific questions. Most studies require enormous monetary and personnel resources so making the most of each data set is important. Secondary analysis is an efficient use of data for the maximum amount of scientific knowledge. In some ways, an independent research question may be ideal in this context as long as there are not aspects of the original study that may compromise conclusions from the study. For example, some demand characteristics of the original study may not apply for the new research study.

3. **Does the model change if the design is cross-sectional or longitudinal? Observational studies are more common than longitudinal.**

The model may not change, but many additional assumptions come in to play such as the stability of the relations among variables. Mediation analysis is appropriate for all of these designs. It is the quality of the conclusions that vary with different designs. Keep in mind that with cross-sectional data, relations between different values for persons are the information used in analysis. With longitudinal designs, relations between levels of a variable for individuals and change across time are available for analysis. The two cross-sectional data collections measure relations between levels at each time, which may be very different from relations between changes over time. Information on assumptions of longitudinal data is covered in Chapter 8 of my 2008 book. There is some work on longitudinal models from a potential outcomes perspective as well—generally called analyses with time dependent covariates (see work by Robins and colleagues and sections of the VanderWeele text for an explanation on causal mediation and moderation to address longitudinal data).

4. You mentioned that assumptions three and four in causal mediation analysis cannot be checked via the data and require sensitivity analysis. So how would one interpret the causal mediation analysis results with these assumptions?

If the assumptions are met, then the quantities are estimates of causal effects. If the assumptions are not satisfied, a researcher could conduct sensitivity analysis, for example, to find the size of the confounder effect that would make a mediated effect become nonsignificant or to become zero. It is an important question to address how bad the violation of assumptions have to be to compromise conclusions. Often, we refer to assumptions as met or not met when the violation of assumptions may not be a yes or no answer. Keep in mind that all models are abstractions of reality so the model may be off in one way or another. If all confounders are measured and included in the statistical analysis, then assumptions are satisfied, but this may not be possible in practice.

5. Is there a relationship between these causal estimates and predicted marginals? I'm mostly referring to predicted values.

If you are referring to the predicted potential outcomes, the answer is yes. The causal estimates are differences in potential outcomes. The Marginal Structural Model used to estimate causal effects is a general model based on marginal quantities to estimate results in all potential outcomes.

6. The total and pure natural indirect effects sound similar to causal estimates of the ATT (average treatment effect for the treated) and ATU (average treatment effect for the untreated). Are they somehow related?

Yes, except that the ATT usually refers to a non-randomized intervention effect so that effect can be obtained at different levels of treatment uptake. Yes, the ATT and ATU can be based on the potential outcomes/counterfactual framework.

7. How is the causal analysis for linear regression?

Causal analysis focuses on conditions necessary to conclude there is a causal effect. Linear regression estimates are associations and do not necessarily have an interpretation as a causal effect. The best book description of this topic in my opinion is *Regression Analysis: A Constructive Critique* (Berk R. Sage Publications, Inc.; 2003). I think this book was ideal for me because the ideas were described from a traditional regression perspective, from which I am most familiar.

8. Would statistical power of the causal mediation approach be comparable to conventional mediation reported in MacKinnon et al. (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. doi: 10.1037/1082-989X.7.1.83)?

Yes, they are comparable. There are power results in the MacKinnon et al. paper cited at the beginning of the presentation (MacKinnon DP, Valente MJ, Gonzalez OJ. The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. Prevention Science. 2020:21;147–157. doi: 10.1007/s11121-019-01076-4). The supplemental materials for that paper include a program to compute power empirically for causal estimators and traditional single mediator models. There is some loss of power when adding the estimation of the XM interaction in the statistical model.

9. **Can you just confirm that when M is continuous and there is no XM interaction the following is correct?**
   TDIE=indirect effect among treated in simple analysis
   PDIE=indirect effect among controls in simple analysis
   TDE=direct effect among treated in simple analysis
   PDE=direct effect among controls in simple analysis

If the XM interaction is zero and assumptions are met, ab=TNIE=PNIE and c'=TNDE=PNDE. If the XM interaction is not zero then TNIE=simple mediated effect in the treatment group, PNIE=simple mediated effect in the control group, PNDE=simple direct effect in the control group, and TNDE=simple direct effect in the treatment groups. Please see slides 49 and 51 of the presentation.

10. **Can traditional mediation analysis method be used for multilevel models? Would it be helpful to use it in a structural equation modeling (SEM) approach?**

The usual multilevel mediation analysis is a traditional mediation analysis. Multilevel structural equation modeling (MSEM) is a good approach. See MacKinnon and Valente for a description of multilevel mediation approaches that includes additional references (MacKinnon DP, Valente MJ. Mediation from multilevel to structural equation modeling. Annals of Nutrition and Metabolism. 2014;65:198–204. doi: 10.1159/000362505). There are also papers that have examined the integration of causal estimation with the multilevel model (VanderWeele TJ. Direct and indirect effects for neighborhood-based clustered and longitudinal data. Sociological Methods and Research. 2010;38(4):515–544. doi: 10.1177/0049124110366236; Pituch KA, Stapleton LM. Distinguishing between cross- and cluster- level mediation processes in the cluster randomized trial. Sociological Methods and Research. 2012;41(4):630–670. doi: 10.1177/0049124112460380).

11. **Which R package would you recommend for mediation analysis?**

They all have strengths. Imai and colleagues' causal mediation program has many capabilities—including outstanding plots for evaluating sensitivity to confounding (Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. Journal of Statistical Software. 2014; 59(5). doi: 10.18637/jss.v059.i05). Medflex is also outstanding (Steen J, Loeys T, Moerkerke B, Vansteelandt S. medflex: An R package for flexible mediation analysis using natural effect models. Journal of Statistical

Software. 2017;76(11). doi: 10.18637/jss.v076.i11). Mplus—not an R program—can estimate these models and handle missing data and other unique aspects of real data (see examples in Muthén BO, Muthén LK, Asparouhov T. *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén; 2017.). We have a program for plotting mediated effects (Fritz MS, MacKinnon DP. A graphical representation of the mediated effect. Behavior Research Methods. 2008;40(1):55–60. doi: 10.3758/brm.40.1.55.) and another program to build confidence intervals based on the distribution of the product and other methods (Tofighi D., MacKinnon DP. RMediation: An R package for mediation analysis confidence limits. Behavior Research Methods. 2011;43(3):692–700. doi: 10.3758/s13428-011-0076-x).

12. What is the difference between direct effect and total effect between X and Y?

Often these terms are mixed up in different research areas but it is not a big deal in my opinion. In a single mediator model, there are three effects—total effect of X on Y, an indirect effect X to M to Y, and the direct effect from X to Y, which is not through the mediator. The total effect is the entire effect of X on Y not considering the mediator. Often people call the total effect the direct effect, but this name is not consistent with the usual mediation literature where the direct effect is the effect of X on Y that is not through M.

13. These assumptions are very stringent. What are the effects of some of the assumptions (e.g., assumptions that there are not subgroups with different effects between treatment and outcome)?

The violation of assumptions is usually not an either/or decision. It is likely that assumptions are violated to some extent in any analysis. The important point is whether the violation invalidates the conclusion of the study. The assumption of post-treatment confounders can be addressed with multiple mediator models and focusing on accurate estimation of the direct effect rather than the relation of M to Y. Ignoring subgroups that differ in their mediation model may still lead to an overall correct decision about a mediation process but it is possible to come up with situations where two groups have opposite patterns that would be missed when combined (see page 93 of Fairchild AJ, MacKinnon DP. A general model for testing mediation and moderation effects. Prevention Science. 2009;10(2):87–99. doi: 10.1007/s11121-008-0109-6).

14. Can you give an example of a collider situation? Can this be empirically tested?

There are several interesting articles on this topic. Elwert and Winship give an excellent overview of collider effects, which they call endogenous selection effects. Collider effects occur when X and Y cause the collider but the relation between X and Y is adjusted for the collider. One way to adjust for a collider is to select a sample based on scores for the collider, which induces bias when estimating the X to Y relation. Examples are Berkson's paradox, birthweight paradox, and the obesity paradox. We have a couple papers under review that discuss methods for testing colliders. (Elwert F, Winship C. Endogenous selection bias: the problem of conditioning on a collider variable. Annual Review of Sociology. 2014;40: 31–53. doi: 10.1146/annurev-soc-071913-043455).

15. **In your article in 2018, significant relationship X -> Y is not necessary in the complete mediation model. Would you please elaborate this and how to interpret the results X-> M and M -> Y are significant, but not X -> Y?**

The test for mediation can have more power than the test of the overall effect of X on Y. More reasons are described in O'Rourke HP, MacKinnon DP. Reasons for testing mediation in the absence of an intervention effect: A research imperative in prevention and intervention research. Journal of Studies on Alcohol and Drugs. 2018;79(2):171–181. doi: 10.15288/jsad.2018.79.171.

With inconsistent mediation models, the indirect effect and direct effect have opposite signs so that the total effect is less than the mediated effect and the test of mediation may have more power than the test of the overall effect of X on Y. For more on suppression and mediation see MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding, and suppression effect. Prevention Science. 2000;1(4):173–181. doi: 10.1023/a:1026595011371.

16. **Here are few additions to answers that I gave after the presentation.**

Missing Data and Mediation Analysis

Here are a few missing data and mediation papers to help you get started:

Enders CK, Fairchild AJ, MacKinnon DP. A Bayesian approach for estimating mediation effects with missing data. Multivariate Behavioral Research. 2013;48(3):340–369. doi: 10.1080/00273171.2013.784862.

Zhang Z, Wang L. Methods for mediation analysis with missing data. Psychometrika. 2013;78(1):154–184. doi: 10.1007/s11336-012-9301-5.

Causal Mediation Software

When asked if SPSS had a program to estimate causal mediation effects, I was referring to the main SPSS program, which does not have such a program. There is an SPSS macro by Valeri and VanderWeele that will estimate causal mediation effects. I am not sure that it still works with the latest version of SPSS. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. Psychological Methods. 2013;18(2):137–150. doi: 10.1037/a0031034.

Here is the citation for the causal mediation software paper that is in press.
Valente MJ, Rijnhart JJM, Smyth H, Muniz FB, MacKinnon DP. Causal mediation programs in R, Mplus, SAS, SPSS, and STATA. Structural Equation Modeling: A Multidisciplinary Journal. In press.

There are many additional citations in this article but here are a few papers on causal mediation software:

Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychological Methods. 2010;15(4):309–334. doi: 10.1037/a0020761.

Muthén BO, Muthén LK, Asparouhov T. *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén; 2017.

Steen J, Loeys T, Moerkerke B, Vansteelandt S. medflex: An R package for flexible mediation analysis using natural effect models. Journal of Statistical Software. 2017;76(11). doi: 10.18637/jss.v076.i11.

Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. Journal of Statistical Software. 2014;59(5).

Information About Workshops on Causal Mediation Analysis

Last February 2020, Matt Valente and I gave workshops at the University of Miami Department of Public Health Sciences and Coral Gables campuses. Matt Valente, Milica Miocevic, Oscar Gonzalez, and I gave a workshop at the Society for Prevention Research on Bayesian Causal Mediation in 2018. I also gave a workshop at the 2019 Association for Psychological Science meeting in Washington, DC on the Potential Outcomes Approach to Causal Inference. We don't currently have any additional workshops planned.

Thanks to Trang Nguyen at Johns Hopkins for sending the information below about other workshops on causal mediation.

Linda Valeri and Caleb Miles are giving a Causal Mediation Analysis Training August 12–14, 2020.

There is another course by Andrew Li in Budapest on Advanced Causal Mediation Analysis, which seems to closely track the mediation package.