Title: Simulation-Based Sensitivity Analysis for Causal Mediation Analysis

Authors and Affiliations: Xu Qin, University of Pittsburgh; Fan Yang, University of Colorado Denver

Context:

Questions of mediation are important for understanding causal pathways by which an intervention affects outcomes. This study is motivated by the Job Search Intervention Study (JOBS II) (Vinokur et al., 1995), which evaluates the impact of a job training intervention through a randomized experiment. In the hypothesized mediation mechanism (Imai et al., 2010a), as depicted in Figure 1, participants' job search self-efficacy, M, mediates the effect of the intervention, T, on their depression level Y. The total treatment effect can be decomposed into an indirect effect transmitted through the mediator and a direct effect operating through all the other possible mechanisms.

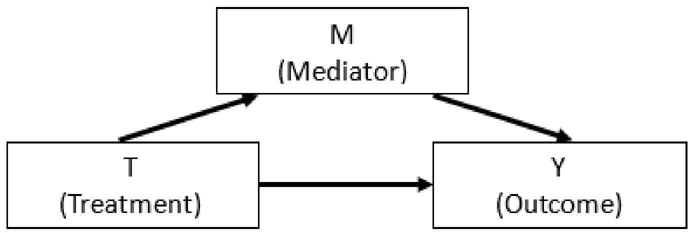

Figure 1. Diagram of a Causal Mediation Process

Identification of direct and indirect effects relies on the assumptions of (1) no unmeasured confounders of the treatment–mediator and treatment–outcome relationships, and (2) no unmeasured confounders of the mediator–outcome relationship. Even if (1) is satisfied in a randomized experiment, (2) typically does not hold, given that mediator values are usually generated naturally. A sensitivity analysis is necessary for determining whether potential violations of identification assumptions would easily alter causal conclusions. However, its importance has not received enough attention in educational studies.

Different sensitivity analysis strategies have been developed for causal mediation analysis. Most are proposed within the regression framework (e.g. VanderWeele, 2010). Mediator and outcome model misspecifications may result in misleading sensitivity assessment. To reduce reliance on parametric assumptions, Hong et al. (2018) developed a weighting-based method that does not require outcome model specification. While these methods evaluate the extent to which unmeasured confounding would bias point estimates, little attention has been paid to its influence on estimation efficiency. This would lead to an inaccurate assessment of the change in the significance of causal effects. Only Imai et al. (2010a, 2010b) took this issue into account. Nevertheless, as they acknowledged, it is difficult to interpret their sensitivity parameter.

Objective/Research Question:

To overcome these limitations, we develop a simulation-based sensitivity analysis strategy. Illustrated through JOBS II, we aim to investigate, for the original causal inference about direct and indirect effects to be altered under treatment randomization, to what extent must an unmeasured confounder be associated with the mediator and outcome. Should a weak unmeasured confounder easily alter the significance of direct and indirect effects, we would conclude that causal conclusions are sensitive to potential violations of the identification assumption.

Intervention/Participants:

The JOBS intervention consisted of five 4-hour training sessions, aiming at enhancing participants' job-search skills and mental health. 1,801 recent job losers were randomly assigned to an experimental group (n = 671), which received the intervention during 22 weeks, and a control group (n = 1,130), which received a booklet briefly introducing job-search methods.

Measures:

The mediator (confidence in six job-search skills) and outcome (depression level) were respectively measured 2 and 4 weeks after randomization. Baseline data contain participants' level of depression and demographics such as age, gender, education, race, marital status, and occupation, etc.

Proposed Method:

For illustration purposes, we consider a simple setup of a randomized binary treatment T, a binary mediator M, and a continuous outcome Y. We assume that a binary pretreatment covariate U is the only unmeasured confounder and is independent of observed pretreatment confounders $\mathbf{X}$. Hence, the conditional distribution of U is factorized as

$$\Pr(U=1|Y,M,T,\mathbf{X})=\frac{f(Y|M,T,\mathbf{X},U=1)\times\Pr(M|T,\mathbf{X},U=1)\times\Pr(U=1)}{\sum_{u=0}^{1}f(Y|M,T,\mathbf{X},U=u)\times\Pr(M|T,\mathbf{X},U=u)\times\Pr(U=u)}.$$

This enables us to specify sensitivity parameters as the conditional associations of U with Y and M, which intuitively reflect the confounding role of U. The challenge is that all the other parameters from $f(Y|M,T,\mathbf{X},U=1)$, $\Pr(M|T,\mathbf{X},U=1)$, and $\Pr(U=1)$ are unknown. The estimation of the parameters relies on U, while U needs to be drawn from the conditional distribution determined by these parameters. We solve the problem with Stochastic EM algorithm (Nielsen, 2000; Carnegie et al., 2016).

To evaluate the influence of U of different strength, we specify a plausible range of sensitivity parameter values. Given each combination of sensitivity parameters, we repeatedly generate U from its conditional distribution, update the original analysis by adjusting for each random draw of U, and use Rubin's (1987) rules to combine the adjusted direct and indirect effect estimates and their standard error estimates. This allows us to capture the influence of U on estimation efficiency.

Sensitivity assessment becomes possible through a comparison of analysis results before and after adjusting for U. The proposed approach is applicable to different causal mediation analysis methods. We have verified through simulations that, the approach is robust to misspecifications of the observed part of the outcome model when applied to the ratio-of-mediator probability weighting (Hong, 2010), which is a propensity score-based weighting method that only requires mediator model specification.

We also provide a convenient tool to visually represent sensitivity analysis results. By applying the method to JOBS II, we obtain Figure 2. Each black contour represents the combinations of sensitivity parameters that lead to the same indirect effect estimate as indicated by the number on the contour. The sensitivity parameters along the red dashed curves reduce the estimate to zero. In the region between the blue dotted curves, the significance is unchanged. The change in the standard error tends to be bigger with the increase in the confounding effect of U. Each dot corresponds to the conditional associations of each observed covariate with Y and M, which are used to calibrate the strength of sensitivity parameters. This plot indicates that, for the original causal conclusions to be reversed, an unmeasured confounder must be stronger than the most important observed confounder. Given that this is highly unlikely, the results are insensitive to unmeasured pretreatment confounding.

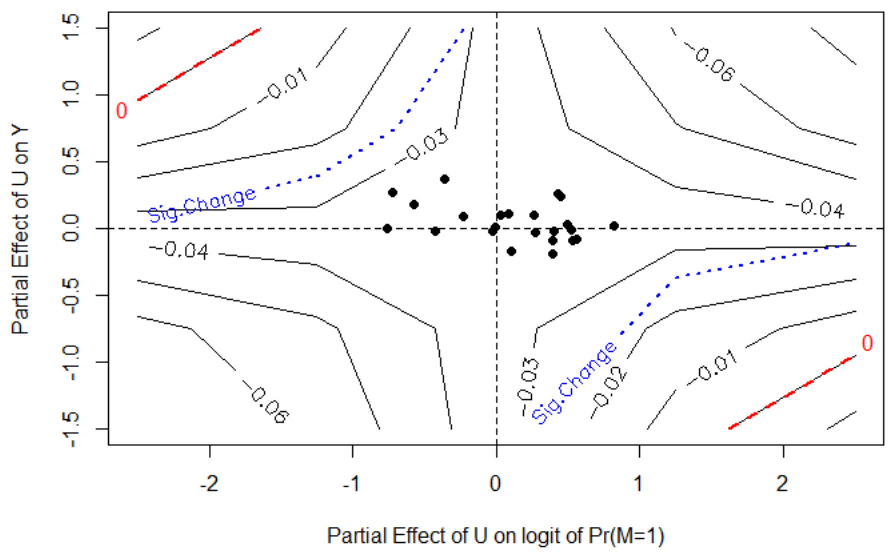

Figure 2. Sensitivity Analysis for the Indirect Effect

Conclusions:

The proposed method allows researchers to intuitively evaluate sensitivity parameters in reference to prior knowledge about the strength of unmeasured confounders, and accurately reflects the influence of unmeasured confounding on estimation efficiency.

We have been focusing on binary M and U. For broader applications, we will extend the method to allow M and U to be discrete or continuous. It can also be extended to observational studies. In addition, we assume that there are no unmeasured posttreatment confounders of the mediator-outcome relationship. We relax this assumption in another ongoing work.

References:

Carnegie, N. B., Harada, M., & Hill, J. L. (2016). Assessing sensitivity to unmeasured confounding using a simulated potential confounder. Journal of Research on Educational Effectiveness, 9(3), 395-420.

Hong, G. (2010). Ratio of mediator probability weighting for estimating natural direct and indirect effects. In Proceedings of the American Statistical Association, Biometrics Section (pp. 2401–2415). Alexandria, VA: American Statistical Association.

Hong, G., Qin, X., & Yang, F. (2018). Weighting-Based Sensitivity Analysis in Causal Mediation Studies. Journal of Educational and Behavioral Statistics, 43(1), 32-56.

Imai, K., Keele, L., & Tingley, D. (2010a). A general approach to causal mediation analysis. Psychological Methods, 15, 309.

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

Nielsen, S. F. (2000). The stochastic EM algorithm: estimation and asymptotic results. Bernoulli, 6(3), 457-489.

Rubin, D. B. (1987). Multiple Imputationfor Nonresponse in Surveys. New York: Wiley.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology, 21, 540–551.

Vinokur, A. D., Price, R. H., & Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. American Journal of Community Psychology, 23(1), 39–74.