Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Simulation-Based Sensitivity Analysis for Causal Mediation Studies

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Qin, Xu	University of Pittsburgh	
Yang, Fan	University of Colorado, Denver	

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2022

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Psychological Methods, Vol. 27, No. 6

DOI or URL to published work (if available) https://doi.org/10.1037/met0000340

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Statistical and Research M through **[Grant number]** R305D2000 to **Institution]** University of P. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Statistical and Research M or the U.S. Department of Education.

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

2022, Vol. 27, No. 6, 1000–1013
https://doi.org/10.1037/met0000340

Simulation-Based Sensitivity Analysis for Causal Mediation Studies

Xu Qin[1] and Fan Yang[2]

[1] Department of Health and Human Development, School of Education, University of Pittsburgh

[2] Department of Biostatistics and Informatics, University of Colorado, Denver

Abstract

Causal inference regarding a hypothesized mediation mechanism relies on the assumptions that there are no omitted pretreatment confounders (i.e., confounders preceding the treatment) of the treatment–mediator, treatment–outcome, and mediator–outcome relationships, and there are no posttreatment confounders (i.e., confounders affected by the treatment) of the mediator–outcome relationship. It is crucial to conduct a sensitivity analysis to determine if a potential violation of the assumptions would easily change analytic conclusions. This article proposes a simulation-based method to assess the sensitivity to unmeasured pretreatment confounding, assuming no posttreatment confounding. It allows one to (a) quantify the strength of an unmeasured pretreatment confounder through its conditional associations with the treatment, mediator, and outcome; (b) simulate the confounder from its conditional distribution; and (c) finally assess its influence on both the point estimation and estimation efficiency by comparing the results before and after adjusting for the simulated confounder in the analysis. The proposed sensitivity analysis strategy can be implemented for any causal mediation analysis method. It is applicable to both randomized experiments and observational studies and to mediators and outcomes of different scales. A visualization tool is provided for vivid representations of the sensitivity analysis results. An R package mediationsens has been developed for researchers to implement the proposed method easily (https://cran.r-project.org/web/packages/mediationsens/index.html).

Translational Abstract

Causal mediation analysis is essential for investigating the mechanisms through which an intervention operates. Causal inference regarding a hypothesized mediation mechanism might be invalidated if there are omitted pretreatment confounders (i.e., confounders preceding the treatment) of the treatment-mediator, treatment-outcome, and mediator–outcome relationships, or in the presence of posttreatment confounders (i.e., confounders affected by the treatment) of the mediator-outcome relationship. However, this has not received enough attention in psychological studies. After a review of the existing causal mediation analysis methods and the approaches that assess the sensitivity of mediation analysis results to unmeasured pretreatment confounding, we propose a simulation-based sensitivity analysis strategy, assuming no posttreatment confounding. The method has five primary advantages. First, it enables applied researchers to intuitively quantify the strength of an unmeasured pretreatment confounder. Second, by simulating the unmeasured confounder from its conditional distribution and adjusting for it in the analysis, the method accurately reflects the influence of unmeasured pretreatment confounding on both the causal effect estimates and their sampling variability, while most existing sensitivity analysis methods ignore the latter. Third, a convenient tool is provided for visualization of sensitivity analysis results. Fourth, it is applicable to both randomized experiments and observational studies and to mediators and outcomes of different scales. Fifth, it can assess the sensitivity of results obtained from different causal mediation analysis approaches. The broad utility of the proposed method is illustrated through a re-analysis of the Job Search Intervention Study. We have also developed an R package that implements the proposed method (https://cran.r-project.org/web/packages/mediationsens/index.html).

Keywords: causal mediation analysis, confounders, propensity score, sensitivity analysis, simulation

Supplemental materials: https://doi.org/10.1037/met0000340.supp

This article was published Online First December 16, 2021.

Xu Qin https://orcid.org/0000-0001-6160-1545

An earlier version of this work was presented at the Society for Research on Educational Effectiveness 2020 conference and the 2019 Joint Statistical Meetings. This study received support from a small research grant funded by the Spencer Foundation. The authors thank Matteo Bonvini, Guanglei Hong, Nicholas Jewell, Brian Junker, Trang Nguyen, and Jiebiao Wang for their insightful comments on previous versions of the article.

Correspondence concerning this article should be addressed to Xu Qin, Department of Health and Human Development, School of Education, University of Pittsburgh, Office 5100 WWPH, 230 South Bouquet Street, Pittsburgh, PA 15260, United States. Email: xuqin@pitt.edu

Questions of mediation are essential for understanding causal pathways by which an intervention affects outcomes. A hypothesized mediation mechanism often involves a change in the mediator induced by the treatment, subsequently leading to a change in the outcome. The total treatment effect can be decomposed into an indirect effect operating through the mediator and a direct effect transmitted through all the other possible mechanisms. Identification of the indirect and direct effects relies on assumptions that (a) there are no omitted pretreatment covariates (i.e., covariates preceding the treatment) that confound the treatment–mediator and the treatment–outcome relationships; and (b) there are no posttreatment covariates (i.e., covariates that are affected by the treatment) or omitted pretreatment covariates that confound the mediator–outcome relationship. These assumptions are referred to as *sequential ignorability* (e.g., Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010; Ten Have et al., 2004). The validity of Assumption (b) is a major concern. This is because even if Assumption (a) can be satisfied in a randomized experiment, Assumption (b) typically does not hold, given that mediator values are usually generated through a natural process. It is crucial to conduct a sensitivity analysis to determine if potential violations of an identification assumption would easily change causal inference regarding the hypothesized mediation mechanism. However, its importance has not received enough attention in psychological studies.

There are two types of violations of the identification assumptions. First, one may arbitrarily omit some observed confounders from the analysis to avoid model overfitting. To assess the influence of such omissions, one may simply compare the results before and after including the omitted variables in the analysis. Second, some confounders are unmeasured. We focus on the latter in this study. Various strategies have been developed to evaluate potential bias in the indirect and direct effect estimates as functions of sensitivity parameters, which imply departures from the identification assumptions due to unmeasured confounding. A review can be found in the section of existing sensitivity analysis methods for mediation analysis.

In addition to the point estimation, unmeasured confounding would also affect the sampling variability of the indirect and direct effect estimates. When assessing the sensitivity of a total treatment effect to an unmeasured confounder at a given strength, Cinelli and Hazlett (2020) proved that, accounting for the unmeasured confounder in the analysis would reduce the standard error of the treatment effect estimate by reducing the variance of residuals, while increasing the standard error via the decrease in the degrees of freedom and the partial correlation of the unmeasured confounder with the treatment. We argue that it would affect the sampling variability of the indirect and direct effect estimates in the same way, while the uncertainty of the unmeasured confounder and the partial correlation between the unmeasured confounder and the mediator would also play a role.

Ignoring such changes in the sampling variability would lead to an inaccurate assessment of the influence of unmeasured confounding on statistical inference. However, in the literature of mediation analysis, only the L.O.V.E.-based methods (Cox et al., 2013; Liu and Wang, 2020) and the method that Imai and colleagues developed (Imai, Keele, and Tingley, 2010; Imai, Keele, and Yamamoto, 2010) accounted for the influence of unmeasured pretreatment confounding on the sampling variability of the indirect and direct effect estimates. Nevertheless, the former is not applicable when the treatment interacts with the mediator in affecting the outcome, while the latter relies on a sensitivity parameter that lacks intuitive interpretations. Not only in causal mediation analysis, but in causal inference in general, there have been few discussions on how potential violations of identification assumptions would affect sampling variability. With a focus on total treatment effect, Carnegie et al. (2016) assessed the influence of unmeasured confounding on both the point estimate and its standard error by generating an unmeasured confounder from its conditional distribution and adjusting for it in the estimation. Dorie et al. (2016) further incorporated Bayesian additive regression trees into this strategy.

This article, motivated by Cinelli and Hazlett (2020) and Carnegie et al. (2016), proposes a simulation-based sensitivity analysis method for causal mediation analysis, assuming no posttreatment confounders of the mediator–outcome relationship. It allows one to specify a departure from the sequential ignorability assumption through the conditional associations of an unmeasured pretreatment confounder with the treatment, mediator, and outcome, simulate the confounder from its conditional distribution, and finally assess its influence by comparing the indirect and direct effect estimates before and after adjusting for it in the analysis. The proposed approach (a) accurately reflects the influence of unmeasured pretreatment confounding at a given strength on estimation efficiency, (b) enables psychological researchers to intuitively quantify sensitivity parameters, (c) provides a convenient tool for visualization of sensitivity analysis results, (d) is applicable to both randomized experiments and observational studies and to mediators and outcomes of different scales, and (e) can be implemented for different causal mediation analysis methods, as reviewed in the Mediation Analysis Methods section.

We organize this article as follows. We first introduce an application example. After defining the causal mediation effects under the potential outcomes framework and clarifying the identification assumptions, we review different causal mediation analysis methods and the existing sensitivity analysis methods for assessing the sensitivity of causal mediation analysis results to unmeasured pretreatment confounders. Based on the derived conditional distribution of the unmeasured pretreatment confounder, we delineate the sensitivity analysis algorithm. We assess the performance of the proposed sensitivity analysis method through simulations. We also illustrate the method with a real-data application and visualize the sensitivity analysis results. Finally, we discuss the strengths and limitations of the method.

Motivating Example

This work is motivated by the Job Search Intervention study (JOBS II; Vinokur et al., 1995), which examined the impact of a job training intervention on unemployed job seekers through a randomized field experiment. The intervention program consisted of five 4-hour training sessions, aiming at enhancing participants' capability and motivation for obtaining new jobs and improving their mental health. The study randomly assigned a sample of 1,801 unemployed workers who lost their jobs no longer than 13 weeks ago. Six-hundred and 71 individuals were assigned to the experimental group, which participated in the JOBS intervention program, and 1,130 were assigned to the control group, which received a booklet very briefly introducing job search methods.

The goal of this application is to reevaluate the hypothesized mediation mechanism that Imai, Keele, and Tingley (2010) tested. Under the hypothesized mechanism, the JOBS intervention program enhances one's job search self-efficacy, which further decreases his or her depression level. In other words, participants' confidence in their job searching ability mediates the effect of the job training intervention on their depression level. The mediator and the outcome were respectively measured by one's confidence in six job search skills and symptoms of depression in follow-up interviews. The baseline covariates, collected 2 weeks before the intervention, contain participants' level of depression and demographics, such as age, gender, education, ethnic/racial identification, marital status, occupation, family income, and economic hardship.

Definition of the Causal Mediation Effects

Let T denote the treatment assignment. In the JOBS II example, $T = 1$ (or $T = 0$) implies that an individual was randomly assigned to the intervention (or the control group). We use M to indicate the focal mediator, which was measured after the treatment assignment and before the assessment of the outcome. It takes the values of 1 if an individual's job search self-efficacy (i.e., confidence level in job search skills) is high and 0 if not. We use Y to represent the outcome, that is, one's depression level. The goal of a mediation analysis is to decompose the total effect of T on Y into an indirect effect that operates through M and a direct effect transmitted through other pathways.

We define the causal indirect and direct effects under the potential outcomes framework (Neyman & Iwaszkiewicz, 1935; Rubin, 1978). We use $M_i(t)$ to denote individual i's potential job search self-efficacy under the treatment *condition* t, where $t = 0, 1$. For each individual, there are two potential mediator values, while only the one under the individual's actual treatment condition is observable. Similarly, the same individual's potential depression level is defined as $Y_i(t)$ when the treatment condition is set to t. Given that the potential outcome depends on both the treatment and the potential mediator, the potential outcome under treatment *condition* t can be alternatively written as $Y_i(t, M_i(t))$. These are defined under the stable unit treatment value assumption (SUTVA; Rubin, 1980, 1986, Rubin, 1990), which implies that there is only one version of each treatment condition, and there is no interference between individuals.

Additionally, we use $Y_i(1, M_i(0))$ to denote individual i's potential depression level if assigned to the intervention group while his or her job search self-efficacy took the value as if under the control condition. This enables us to define individual i's "natural indirect effect" (Pearl, 2001) as

$$\mathrm{NIE}_i = Y_i\big(1, M_i(1)\big) - Y_i\big(1, M_i(0)\big),$$

which represents the impact of the intervention on individual i's depression level transmitted solely through the intervention-induced change in job search self-efficacy from $M_i(0)$ to $M_i(1)$, while the treatment status stays at the intervention condition. Similarly, individual i's "natural direct effect" (Pearl, 2001) is defined as

$$\mathrm{NDE}_i = Y_i\big(1, M_i(0)\big) - Y_i\big(0, M_i(0)\big),$$

which represents the impact of the intervention on individual i's depression level while his or her potential mediator is held constant at the value that would be realized under the control condition, $M_i(0)$. By averaging each effect over all the individuals in the population, we define the population average natural indirect and direct effects as

$$\mathrm{NIE} = \mathrm{E}\big[Y_i\big(1, M_i(1)\big)\big] - \mathrm{E}\big[Y_i\big(1, M_i(0)\big)\big],$$

$$\mathrm{NDE} = \mathrm{E}\big[Y_i\big(1, M_i(0)\big)\big] - \mathrm{E}\big[Y_i\big(0, M_i(0)\big)\big].$$

The sum of NIE and NDE is equal to the total effect of treatment assignment on the outcome. Alternatively, the total effect can be decomposed into the sum of $\mathrm{E}[Y_i(0, M_i(1))] - \mathrm{E}[Y_i(0, M_i(0))]$ and $\mathrm{E}[Y_i(1, M_i(1))] - \mathrm{E}[Y_i(0, M_i(1))]$. The two decompositions would be different in the presence of an interaction between the treatment and the mediator. We focus on the NIE and NDE defined as above in this article, while the extensions to the alternative decomposition are straightforward.

Identification Assumptions

While $Y_i(t, M_i(t))$ is observed only if individual i was assigned to treatment group t, $Y_i(t, M_i(t'))$, where $t \neq t'$, is never observable. To identify the population average natural indirect and direct effects, we need to relate the unobservable quantities to observed data. The identification[1] of the natural indirect and direct effects relies on the sequential ignorability assumption (e.g., Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010; Ten Have et al., 2004). The assumption includes two ignorability assumptions.

First, the treatment assignment is ignorable given pretreatment covariates $\boldsymbol{X}$. This assumption can be formally written as

$$\big\{Y_i(t', m),\ \ M_i(t)\big\} \perp\!\!\!\perp T_i | \boldsymbol{X}_i = \mathbf{x},$$

where $0 < \Pr(T_i = t | \boldsymbol{X_i} = \mathbf{x}) < 1$ for $t, t' = 0, 1$. That is, given pretreatment covariates $\boldsymbol{X}$, the treatment assignment is independent of potential outcomes and potential mediators. In other words, there is no omitted confounding of the treatment-mediator or treatment–outcome relationship. The assumption holds in randomized experiments by design.

Second, the mediator is ignorable within and across treatment conditions given pretreatment covariates $\boldsymbol{X}$. It can be formalized as

$$\big\{Y_i(t', m)\big\} \perp\!\!\!\perp M_i(t) | T_i = t, \boldsymbol{X}_i = \mathbf{x},$$

where $0 < \Pr(M_i(t) = m | T_i = t, \boldsymbol{X_i} = \mathbf{x}) < 1$ for $t, t' = 0,1$. That is, given pretreatment covariates $\boldsymbol{X}$, the potential mediator is independent of potential outcomes within and across treatment conditions. In other words, there are no omitted pretreatment covariates that confound the mediator–outcome relationship, and there are no posttreatment confounders of the mediator–outcome relationship. The latter is a strong assumption and can be relaxed, which we will discuss in the last section. The ignorability assumption of the

[1] "Identification" refers to the identification of causal effects, rather than model identification as in locating a unique parameter solution.

mediator implies that, among the individuals who were assigned to the same treatment group and share the same pretreatment covariates, the mediator is as if randomized. Unlike the ignorability of the treatment, the ignorability of the mediator may not hold even in a randomized experiment because the mediator is generated in a natural process. For example, among the unemployed job seekers with the same observed pretreatment covariates, those who were more motivated to find new jobs before the intervention might be more confident in their job search skills after the intervention and are also expected to have lower depression levels, no matter which treatment group they were assigned to. Therefore, the observed association between the job search self-efficacy and the final depression level might be partly due to the confounding of one's motivation to find new jobs at baseline. Ignoring it would bias the indirect and direct effects estimates.

Mediation Analysis Methods

Methods Under the Traditional Linear Additive Framework

In the traditional mediation analysis (MacKinnon, 2008; MacKinnon & Dwyer, 1993), one usually regresses a continuous mediator on the treatment and pretreatment covariates and regresses a continuous outcome on the treatment, mediator, and pretreatment covariates, ignoring a possible interaction between the treatment and the mediator in the outcome model. The direct effect is evaluated via the coefficient of the treatment in the outcome model, and the indirect effect is often evaluated via the product of the coefficient of the treatment in the mediator model and the coefficient of the mediator in the outcome model. In addition to the identification assumptions, causal interpretations of the effects also rely on a strong assumption that the treatment does not interact with the mediator when affecting the outcome, which is usually violated in real applications. For example, in JOBS II, the impact of job search self-efficacy on depression level may be stronger for participants assigned to the job training intervention than for those assigned to the control group, because of more support in the training program.

In the recent years, various causal mediation analysis methods have been developed to accommodate the treatment-by-mediator interaction and carefully adjust for confounders of the treatment–mediator, treatment–outcome, and mediator–outcome relationships, while they vary in the degree of reliance on correct model specifications.

Regression-Based Methods

Some researchers modified the traditional mediation analysis by incorporating a treatment-by-mediator interaction in the outcome model.

$$M = \beta_0^m + \beta_t^m T + \mathbf{X}\boldsymbol{\beta}_x^m + \varepsilon_m, \varepsilon_m \sim N\left(0, \sigma_m^2\right), \tag{1}$$

$$Y = \beta_0^y + \beta_t^y T + \beta_m^y M + \beta_{\text{tm}}^y TM + \mathbf{X}\boldsymbol{\beta}_x^y + \varepsilon_y, \ \ \varepsilon_y \sim N\left(0, \sigma_y^2\right). \tag{2}$$

Under the sequential ignorability assumption, VanderWeele and Vansteelandt (2009) identified the causal mediation effects for a binary treatment, a continuous mediator, and a continuous outcome as:

$$\text{NIE} = \left(\beta_m^y + \beta_{\text{tm}}^y\right)\beta_t^m,$$

$$\text{NDE} = \beta_t^y + \beta_{\text{tm}}^y\left(\beta_0^m + E[\mathbf{X}]\boldsymbol{\beta}_x^m\right),$$

where $E[\mathbf{X}] = 0$ if $\mathbf{X}$ is standardized. There are more than two parameters involved in the estimands, rendering the estimation and inference more complex than the traditional mediation analysis that ignores the treatment-by-mediator interaction. Valeri and Vanderweele (2013) extended the above identification results to the scenarios where one or both of the mediator and outcome are binary, by replacing the linear models in Equation 1 or/and Equation 2 with logistic regression(s).

Imai, Keele, and Tingley (2010) and Imai, Keele, and Yamamoto (2010) developed a simulation-based strategy that does not require closed forms of the NIE and NDE estimands and is thus more flexible. By simulating model parameters from their sampling distributions, one could simply simulate the potential outcomes and estimate the NIE and NDE through mean contrasts of the potential outcomes. If the mediator and outcome models are the same as Equations 1 and 2, the NIE and NDE estimates are expected to be identical with those obtained based on VanderWeele and Vansteelandt (2009). By applying the algorithm to bootstrapped samples, Imai, Keele, and Tingley (2010) and Imai, Keele, and Yamamoto (2010) further enabled its application to semiparametric or nonparametric mediator and outcome models and thus relaxed functional form assumptions. The algorithm is applicable to continuous and discrete mediators and outcomes.

Weighting-Based Method

While the regression-based methods are vulnerable to model misspecifications, the weighting-based method does not rely on a parametric outcome model and thus relaxes functional and distributional assumptions. Under the ignorability assumption of the treatment assignment,

$$E[Y_i(t, M_i(t))] = E[W_{Ti} Y_i \mid T_i = t] \tag{3}$$

for $t = 0,1$, where $W_{Ti} = \frac{Pr(T_i = t)}{Pr(T_i = t|\mathbf{X}_i = \mathbf{x})}$ is well known as an inverse probability of treatment weighting (IPTW) scheme (Horvitz & Thompson, 1952; Robins, 2000; Rosenbaum, 1987; Schafer & Kang, 2008), which removes treatment selection by equalizing the treatment assignment probability of all the individuals, as in a randomized experiment. The denominator can be predicted based on a treatment model of T on $\mathbf{X}$. $W_{ti} = 1$ if the treatment is randomized.

Under the ignorability assumption of both the treatment and the mediator,

$$E\left[Y_i\left(1, \ M_i(0)\right)\right] = E[W_{Ti} W_{Mi} Y_i \mid T_i = 1], \tag{4}$$

where $W_{Mi} = \frac{Pr(M_i = m|T_i = 0, \mathbf{X}_i = \mathbf{x})}{Pr(M_i = m|T_i = 1, \mathbf{X}_i = \mathbf{x})}$, which is named as ratio-of-mediator-probability weighting (RMPW) by Hong (2010), is equivalent to the weights proposed by others (e.g., Huber, 2014; Lange et al., 2012; Tchetgen & Shpitser, 2012). W_{Mi} transforms a treated individual's probability of having high job search self-efficacy to resemble that under the control condition within levels of

pretreatment covariates. For each treated individual, the denominator and the numerator of the weight can be predicted based on the mediator model fitted to the treated group and that fitted to the control group, respectively. The counterfactual quantity, $E[Y_i\ (1, M_i\ (0))]$, can therefore be related to the observed values of the outcome in the treated group.

The NIE and NDE can be estimated via weighted mean contrasts of the outcome, which does not require an outcome model. Therefore, the weighting-based method reduces model-based assumptions. It is applicable to continuous and discrete mediators and outcomes. If the mediator is continuous, one may estimate RMPW based on the ratio of conditional densities of M or use a mathematical equivalent of RMPW that is constructed based on conditional probabilities of T given M and $\mathbf{X}$ (Huber, 2014).

Imputation-Based Method

Vansteelandt et al. (2012) developed an imputation-based method that does not require a mediator model. They imputed the potential outcome $Y(t, M(t))$ with the observed outcome in treatment *group* t. To impute the potential outcome $Y(t, M(t'))$, where $t \neq t'$, they fitted an outcome model as shown in Equation 2 and predicted the outcome for everyone in the treatment group t' while forcing T to be equal to t. Subsequently, they estimated the NIE and NDE through a so-called natural effect model, which regresses the imputed potential outcomes on t, t', and $\mathbf{X}$. As Vansteelandt et al. (2012) pointed out, the method ignores extrapolation uncertainty and thus may underestimate the sampling variability of the estimates.

Multiply Robust Methods

By combining the weighting-based and the imputation- or regression-based methods, multiply robust estimation strategies can provide consistent indirect and direct effect estimates when at most one of the treatment, mediator, and outcome models is misspecified (e.g., Tchetgen & Shpitser, 2012; Vansteelandt et al., 2012; Zheng & van der Laan, 2012).

All these causal mediation analysis methods require that there be no unmeasured pretreatment confounding of the treatment–mediator, treatment–outcome, or mediator–outcome relationship, and that there be no posttreatment confounding of the mediator–outcome relationship. Even though the treatment assignment is random in the JOBS II study, it does not guarantee the ignorability of the mediator. The mediator–outcome relationship may still be confounded after conditioning on observed pretreatment covariates. It becomes necessary to conduct a sensitivity analysis to assess the extent to which causal inference about the natural indirect and direct effects would be invalidated by potential violations of the identification assumptions.

Existing Sensitivity Analysis Methods for Mediation Analysis

Various strategies have been developed to evaluate the sensitivity of causal mediation analysis results to unmeasured pretreatment confounding under the assumption that there is no posttreatment confounder of the mediator–outcome relationship. By reviewing the existing methods, this section aims to highlight the limitations that we will address in this study.

Methods Under the Traditional Linear Additive Framework

Researchers have developed sensitivity analysis methods under the traditional mediation analysis framework that focuses on continuous mediators and outcomes and ignores the treatment-by-mediator interaction. Harring et al. (2017) utilized a phantom variable, which is a latent variable with predetermined mean and variance, to assess a model's sensitivity to an unmeasured pretreatment confounder given its conditional associations with the mediator and the outcome. However, they did not offer a solution to evaluating its impact on statistical inference of the indirect effect. Neither did they assess the influence of a violation of the treatment ignorability assumption. Other researchers (Cox et al., 2013; Liu & Wang, 2020) extended the L.O.V.E. (left out variables error) method (Mauro, 1990). Cox et al. (2013) studied how unmeasured pretreatment confounding affects both the estimation and inference of the indirect and direct effects given its correlations with the treatment, mediator, and outcome. The method relies on sample correlations among all the observed variables in the mediator and outcome models. As Mauro (1990) acknowledged, it may become unwieldy when the number of observed pretreatment covariates is very large. Even though the large pool of observed pretreatment covariates can be replaced with their linear combination, it may result in bias in the sensitivity analysis results.

Regression-Based Methods

Multiple sensitivity analysis methods have been developed for the regression-based causal mediation analysis that accounts for the treatment-by-mediator interaction. Assuming that the ignorability assumption of the treatment assignment holds, and that there is an unmeasured pretreatment confounder of the mediator–outcome relationship U, which is binary and independent of $\mathbf{X}$, VanderWeele (2010) derived the bias in each of the NIE and NDE estimates as the product of two sensitivity parameters, (a) the conditional association between U and Y given T, M, and $\mathbf{X}$, and (b) the conditional association between U and T given M and $\mathbf{X}$. M would become a collider of the U–T relationship if both U and T affect M. Correspondingly, conditioning on M would create an association between U and T (Pearl, 1988). Therefore, sensitivity parameter (b) reflects the conditional association between U and M, but in a nonintuitive way. In addition, the derivation of the bias relies on strong model-based assumptions that the sensitivity parameter (a) is constant across levels of T, M, and $\mathbf{X}$, and the sensitivity parameter (b) is constant across levels of M and $\mathbf{X}$.

By relaxing the strong assumptions, Ding and VanderWeele (2016) derived bounds of the bias as functions of the sensitivity parameters similar to those proposed by VanderWeele (2010). An alternative sensitivity parameter that can directly reflect the U–M association results in weaker bounds than the true bounds (Smith & VanderWeele, 2019).

An important limitation of the above methods is that they ignore the influence of unmeasured confounding on the sampling variability of the effect estimates. In contrast, Imai, Keele, and Tingley (2010) and Imai, Keele, and Yamamoto (2010) took the change into account. Assuming that the ignorability assumption of the

treatment assignment holds, Imai, Keele, and Tingley (2010) and Imai, Keele, and Yamamoto (2010) proposed the correlation between the error terms of the two models, ρ, as the only sensitivity parameter. Hence, we refer to the method as the "ρ-based method" in the rest of the article. The magnitude of ρ increases as unmeasured pretreatment confounding of the mediator–outcome relationship becomes stronger. They derived both the point estimators of NIE and NDE and the corresponding standard error estimators as functions of ρ. The derivation relies on the mediator and outcome models in Equations 1 and 2. If the mediator or the outcome is binary, the corresponding model is replaced with a probit model. Unlike the mediation analysis method that Imai, Keele, and Tingley (2010) and Imai, Keele, and Yamamoto (2010) developed, the sensitivity analysis method is not applicable to semiparametric or nonparametric mediator and outcome models. In addition, because it is hard to determine whether the value of ρ for removing the effects or changing their significance is likely to exist or not, it is implausible to conclude whether the analytic results are sensitive based on ρ without comparisons to other studies. If researchers want to assess the influence of the omission of a particular confounder, they may find it difficult to quantify the corresponding error correlation based on prior knowledge about the confounder.

To ease interpretations, Imai, Keele, and Tingley (2010) and Imai, Keele, and Yamamoto (2010) proposed another R^2-based method with two optional sets of sensitivity parameters, the proportions of the unexplained variances in the initial mediator and outcome models that are explained by unmeasured pretreatment confounders (partial R^2) or the proportions of the variances that unmeasured pretreatment confounders can explain after being included in the mediator and outcome models (R^2). Because their magnitudes can be assessed independently, it becomes more intuitive to interpret the sensitivity analysis results. However, the method does not consider the change in sampling variability.

Weighting-Based Methods

All the above regression-based methods rely heavily on correct specifications of the mediator and outcome models and do not assess the sensitivity to a potential violation of the ignorability assumption of the treatment in observational studies. Hong et al. (2018) overcame these limitations by utilizing the weighting-based identification results in Equations 3 and 4. Sensitivity parameters, constructed based on weights, indirectly reflect the $U - T$, $U - M$, and $U - Y$ associations. The sensitivity parameters can be used to assess the degree of violation of the identification assumptions only through comparisons across variables or studies. Also based on the weighting-based mediation analysis method, Tchetgen and Shpitser (2012) proposed sensitivity parameters that involve counterfactual terms and are thus hard to quantify. In addition, these strategies do not assess the extent to which unmeasured confounding would affect the sampling variability of the causal effect estimates, and they are only applicable to the weighting-based causal mediation analysis.

Limitations

While each of the above sensitivity analysis methods has its unique strengths, it has at least three of the following six limitations:

1. Ignoring the treatment-by-mediator interaction.
2. Lacking intuitive interpretations of the sensitivity parameters.
3. Failing to consider the influence of unmeasured confounding on the sampling variability of the causal effect estimates.
4. Failing to assess the sensitivity to a potential violation of the ignorability assumption of the treatment in observational studies.
5. Can be applied to only one mediation analysis method.
6. Relying on correct specifications of both the mediator and outcome models.

Table 1 lists the limitations of each method.

To overcome limitations 1 - 5, we develop a simulation-based sensitivity analysis method, which quantifies the strength of an unmeasured pretreatment confounder through its conditional associations with the treatment, mediator, and outcome. The idea is to repeatedly generate an unmeasured pretreatment confounder at a given strength from its conditional distribution and assess its influence by comparing the estimation results before and after adjusting for the simulated unmeasured confounder. Same as the methods reviewed above, we assume that there are no posttreatment confounders of the mediator–outcome relationship. A discussion of sensitivity analysis for assessing the influence of posttreatment confounding can be found in the Discussion section.

Conditional Distribution of an Unmeasured Pretreatment Confounder

For illustration purposes, we consider a randomized experiment and extend our method to observational studies in Supplemental Appendix A. When the treatment is randomized, the ignorability of treatment naturally holds, while the plausibility of the ignorability assumption of the mediator relies on the richness of the observed pretreatment covariates. Same as the existing sensitivity analysis methods, we now consider the case in which there is an additional unmeasured pretreatment covariate U that is independent of the observed covariates $\mathbf{X}$ and may confound the mediator–outcome relationship. In other words, U represents the part of the unmeasured pretreatment confounder that remains unexplained by $\mathbf{X}$, and the ignorability of the mediator will be satisfied given both $\mathbf{X}$ and U.

Table 1
Limitations of the Existing Sensitivity Analysis Methods

Methods	Limitations
Traditional—Harring et al. (2017)	(1), (3), (4), (5), (6)
Traditional—Cox et al. (2013)	(1), (5), (6)
Regression-based—VanderWeele (2010)	(2), (3), (4), (5), (6)
Regression-based—Ding and VanderWeele (2016)	(2), (3), (4), (5), (6)
Regression-based—Imai, Keele, and Tingley (2010; ρ-based)	(2), (4), (5), (6)
Regression-based—Imai, Keele, and Yamamoto (2010; R^2-based)	(3), (4), (5), (6)
Weighting-based	(2), (3), (5)

$$\{Y_i(t', m)\} \perp\!\!\!\perp M_i(t) | T = t,\ \mathbf{X} = \mathbf{x},\ U = u.$$

The goal of this study is to understand what the estimates of NIE and NDE and their sampling variability would have been had we accounted for unmeasured pretreatment confounding at various strengths, so that we could assess the degree to which the ignorability assumption must be violated for the original conclusion to be changed. To reach this goal, we derive a conditional distribution of U based on our assumptions about its relationship with the outcome and the mediator. Given random draws of U from its conditional distribution, we can estimate the NIE and NDE after adjusting for U in the analysis.

When T is randomized, we obtain the following complete data likelihood,

$$\Pr(Y, M, U, T \mid \mathbf{X}) = \Pr(Y \mid M, U, T, \mathbf{X}) \times \Pr(M \mid U, T, \mathbf{X}) \times \Pr(U) \times \Pr(T). \tag{5}$$

For mathematical convenience, it is usually assumed that U is binary (e.g., Imbens, 2003; VanderWeele, 2010). Similarly, we focus on a binary U, so that we can factorize the distribution of U conditional on the observed data as

$$\Pr(U = 1 \mid Y, M, T, \mathbf{X}) = \frac{f(Y \mid M, T, \mathbf{X}, U = 1) \times \Pr(M \mid T, \mathbf{X}, U = 1) \times \Pr(U = 1)}{\sum_{u=0}^{1} f(Y \mid M, T, \mathbf{X}, U = u) \times \Pr(M | T, \mathbf{X}, U = u) \times \Pr(U = u)}. \tag{6}$$

An extension for applications to a continuous U can be found in Supplemental Appendix A.

By fitting a regression of Y on M, T, $\mathbf{X}$, and U and a regression of M on T, $\boldsymbol{X}$, and U, we could use the coefficients of U in the two models to intuitively represent the confounding role of U, or in other words, how severe the ignorability assumption of the mediator is violated. Different from the analytic models as reviewed in the section of Mediation Analysis Methods, the models here serve as data generating models, which rely on parametric assumptions of the U-M and U-Y relationships. The analytic models are consistent with the observed part of the data generating models (1) for both the mediator and the outcome in application of the regression-based mediation analysis methods; (2) for the mediator in application of the weighting-based mediation analysis method; and (3) for the outcome in application of the imputation-based mediation analysis method.

We focus on a continuous outcome and a binary mediator, as in the JOBS II example, while making extensions for mediators and outcomes of different scales in Supplemental Appendix A. Specifically, we assume that

$$Y \mid M, T, \mathbf{X}, U \sim \mathrm{N}\left(\beta_0^y + \beta_t^y T + \beta_m^y M + \beta_{\mathrm{tm}}^y TM + \mathbf{X}\boldsymbol{\beta}_x^y + \beta_u^y U,\ \sigma^2_{Y \mid M,T,\mathbf{X},U}\right), \tag{7}$$

$$M \mid T, \mathbf{X}, U \sim \mathrm{Bernoulli}\left(\frac{1}{1 + \exp\{-(\beta_0^m + \beta_t^m T + \mathbf{X}\boldsymbol{\beta}_x^m + \beta_u^m U)\}}\right), \tag{8}$$

$$U \sim \mathrm{Bernoulli}(\pi), \tag{9}$$

where sensitivity parameters are denoted by β_u^y, which represents the association between U and Y conditional on M, T, and $\mathbf{X}$; β_u^m, which represents the association between U and M conditional on T, and $\mathbf{X}$; and π, which is the probability of $U = 1$ and determines the marginal distribution of the binary variable U. The interpretations of the sensitivity parameters are straightforward and intuitive, especially for applied researchers. If the regressions are standardized, the magnitudes of β_u^y and β_u^m can be directly used to assess the extent to which the ignorability assumption of the mediator is violated. In observational studies, the ignorability assumption of the treatment may be violated, and thus the conditional association between the unmeasured confounder and the treatment needs to be introduced as an additional sensitivity parameter. Details can be found in Supplemental Appendix A.

Based on Equation 6 and the distribution assumptions in Equations 7–9,[2] we could easily obtain the distribution of U conditional on Y, M, T, $\mathbf{X}$. The challenge is that, except for the given sensitivity parameters β_u^y and β_u^m, the parameters in Equations 7 and 8 are unknown. Let the vector of these parameters be $\boldsymbol{\theta} = (\boldsymbol{\beta}^y, \boldsymbol{\beta}^m, \sigma^2_{Y \mid M,T,\mathbf{X},U})'$, where $\boldsymbol{\beta}^y = \left(\beta_0^y, \beta_t^y, \beta_m^y, \beta_{tm}^y, \boldsymbol{\beta}_x^{y\prime}\right)'$ is a vector of the outcome model coefficients and $\boldsymbol{\beta}^m = \left(\beta_0^m, \beta_t^m, \boldsymbol{\beta}_x^{m\prime}\right)'$ is a vector of the mediator model coefficients. The estimation of these parameters relies on U, while U needs to be drawn from the conditional distribution determined by these parameters. To solve the problem, we adopt a stochastic EM algorithm (Nielsen, 2000) that iterates between the following steps:

Stochastic E-Step

Given values of $\boldsymbol{\theta}^{(k)}$ and the specified sensitivity parameter values, we simulate U for each individual from its conditional distribution.

M-Step

Given the simulated values of U, we find the parameters that maximize the complete data log-likelihood and let them be $\boldsymbol{\theta}^{(k+1)}$.

To obtain initial values $\boldsymbol{\theta}^{(1)}$, we simulate U for each individual from its marginal distribution, as represented in (9). We then iterate between Stochastic E-step and M-step until convergence.

Sensitivity Analysis Algorithm

Sensitivity assessment becomes possible through a comparison of the analysis results obtained from any causal mediation analysis method before and after adjusting for U. To evaluate the influence of U with different strengths and marginal distributions, we specify a plausible range of sensitivity parameter values. Given each combination of sensitivity parameters, we repeatedly generate U from its conditional distribution, obtain the indirect and direct effect estimates and their standard errors by adjusting for each random draw of U, and finally use Rubin's (1987) rules to combine the estimates

[2] The outcome and mediator models are not restricted to Equations (7) and (8). The model specifications can be modified based on one's assumptions about the data generating process. The components of the models that are unrelated to U can be semiparametric or nonparametric. As the simulation section concludes, if a potential interaction between U and T or M exists theoretically, we need to control for these interactions in Equations (7) and (8). Correspondingly, more sensitivity parameters are involved.

across replications. This allows us to capture the influence of U on not only estimation bias but also estimation efficiency. To be specific, the sensitivity analysis takes the following steps:

Step 1. At each given value of π, we specify a range of possible values for β_u^y and β_u^m and divide them into a grid.

Step 2. Given the sensitivity parameter values in each cell of the grid, we generate U based on its conditional distribution, as described in the previous section.

Step 3. Given the sensitivity parameter values and the simulated U, we estimate the adjusted NIE (or NDE), δ, and its standard error, $\sigma_{\hat{\delta}}$, using a method as introduced in the section of Mediation Analysis Methods, depending on which method is used in the original analysis.

Step 4. The estimates obtained from Step 3 are based on one random draw from the conditional distribution of U. To account for the uncertainty of U and to obtain more accurate estimates of the NIE and NDE, we repeat Step 2 and Step 3 K times for each pair of sensitivity parameters.

Step 5. Let the adjusted estimate of NIE (or NDE) in each replication be $\hat{\delta}_k$, where $k = 1, \ldots, K$. By taking the average over the K estimates, we obtain the final estimate $\hat{\delta} = \frac{1}{K}\sum_{k=1}^{K}\hat{\delta}_k$. To reflect both the average uncertainty of the estimate *within* each random draw of U, $V_W = \frac{1}{K}\sum_{k=1}^{K}\sigma^2_{\hat{\delta}_k}$, and the variation in the estimate *between* multiple draws of U, $V_B = \frac{1}{K-1}\sum_{k=1}^{K}(\hat{\delta}_k - \hat{\delta})^2$, we estimate the standard error estimate of $\hat{\delta}$ as $\hat{\sigma}_{\hat{\delta}} = \sqrt{V_W + \left(1 + \frac{1}{K}\right)V_B}$, by following Rubin's rules as widely used in multiple imputation. The higher K is, the more precise $\hat{\delta}$ will be.

Step 6. The above procedure assesses the sensitivity to U at different strengths, given its marginal distribution. To further evaluate the influence of the marginal distribution of U, we repeat the procedure at different values of π.

When the sampling distributions of the NIE and NDE estimates are nonsymmetric, as is usually encountered in small samples that are common in psychological research, it is inaccurate to make statistical inference simply based on standard errors of the estimates. Unlike the ρ-based method, which bounds the 95% confidence intervals by two times standard errors away from the adjusted effect estimates, we recommend a combination of bootstrap estimation with the above procedure (Schomaker & Heumann, 2018). Specifically, we generate B bootstrap samples and apply Steps 2–4 to each bootstrap sample. Therefore, there are K adjusted NIE and NDE estimates associated with each bootstrap sample. Taking the average over the K estimates in each bootstrap sample yields B adjusted estimates, based on which we construct the confidence interval for each effect.

Simulations

The basic idea of the proposed sensitivity analysis method is to generate the unmeasured pretreatment confounder from its conditional distribution and adjust for it in the estimation of NIE and NDE based on the causal mediation analysis method adopted in the initial analysis. Whether the method can accurately evaluate the bias due to unmeasured pretreatment confounding is determined by whether it can recover the true NIE and NDE. This depends on (a) the specification of generating models of the outcome and mediator, which determines the conditional distribution of the unmeasured confounder, and (b) the robustness of the chosen causal mediation analysis method to possible misspecifications of the models.

Therefore, with a focus on a randomized binary treatment, a binary mediator, a continuous outcome, and a binary unmeasured pretreatment confounder, we run simulations to assess, if all the functional and distributional assumptions of the mediator and outcome are satisfied or if part of them is violated, to what extent the proposed sensitivity analysis strategy would approximate the true NIE and NDE when incorporated with the regression-based and weighting-based causal mediation analysis methods. As introduced in the section of Mediation Analysis Methods, there are two estimation strategies for the regression-based analysis, one based on closed forms of the estimands and the other based on simulations. We choose the former because the latter is more time consuming. We expect that the simulation results have implications for the multiply robust methods, as they are combinations of the methods.

In addition, an important goal of the proposed strategy is to keep the advantage of the ρ-based method that it can account for the influence of unmeasured confounding on the estimation efficiency, while overcoming its limitations as listed in Table 1. Therefore, we also assess the influence of unmeasured pretreatment confounding on the sampling variability of the NIE and NDE estimates and compare the proposed method with the ρ-based method under various scenarios.

Simulation Setup

We begin the simulations by generating the treatment indicator T from a Bernoulli distribution with $\Pr(T) = .5$, generating three independent observed confounders, X_1, X_2, and X_3, each from a standard normal distribution, and generating one confounder that will be omitted from the original analysis, U, from a Bernoulli distribution with $\Pr(U) = .5$. We then generate the mediator and the outcome based on the following models:

$$\begin{aligned} Y = {} & \beta_0^y + \beta_t^y T + \beta_m^y M + \beta_{tm}^y TM + \beta_{x1}^y X_1 + \beta_{x2}^y X_2 + \beta_{x3}^y X_3 \\ & + \beta_{tx1}^y TX_1 + \beta_{tx2}^y TX_2 + \beta_{tx3}^y TX_3 + \beta_u^y U + \beta_{ut}^y UT + \varepsilon, \\ & \varepsilon \sim N\left(0, \sigma^2\right), \end{aligned}$$

$$\begin{aligned} \log\left(\frac{p}{1-p}\right) = {} & \beta_0^m + \beta_t^m T + \beta_{x1}^m X_1 + \beta_{x2}^m X_2 + \beta_{x3}^m X_3 + \beta_{tx1}^m TX_1 \\ & + \beta_{tx2}^m TX_2 + \beta_{tx3}^m TX_3 + \beta_u^m U + \beta_{ut}^m UT, \end{aligned}$$

where

$$\begin{aligned} & \beta_0^y = 1,\ \beta_t^y = 1,\ \beta_m^y = 1,\ \beta_{tm}^y = 0.5,\ \beta_{x1}^y = 0.3, \\ & \beta_{x2}^y = 0.2,\ \beta_{x3}^y = 0.1,\ \beta_0^m = 0.1, \\ & \beta_t^m = 0.2,\ \beta_{x1}^m = 0.3,\ \beta_{x2}^m = 0.2,\ \beta_{x3}^m = -0.1,\ \sigma = 0.6. \end{aligned}$$

In Scenario 1, to evaluate the performance of the proposed method when both the outcome model and the mediator model are correctly specified, we specify $\beta^y_{tx1} = \beta^y_{tx2} = \beta^y_{tx3} = \beta^m_{tx1} = \beta^m_{tx2} = \beta^m_{tx3} = \beta^y_{ut} = \beta^m_{ut} = 0$ in both the data generation and sensitivity analysis.

In Scenario 2, to assess the influence of a violation of the distributional assumption of the continuous outcome, we modify Scenario 1 by changing the distribution of ε to $\Gamma(.01, .01)$ and rescaling it to keep its mean at 0 and standard deviation at $\sigma = .6$ in the data generation but keep assuming ε to be normal in the sensitivity analysis. For the evaluation of a violation of the distributional assumption of the binary mediator, we replace the logit model with the following probit model when generating the mediator,

$$p = \Phi\big(\beta^m_0 + \beta^m_t T + \beta^m_{x1} X_1 + \beta^m_{x2} X_2 + \beta^m_{x3} X_3 + \beta^m_{tx1} TX_1 + \beta^m_{tx2} TX_2 + \beta^m_{tx3} TX_3 + \big(\beta^m_u/1.6\big)U + \beta^m_{ut} UT\big),$$

where Φ denotes the cumulative distribution function of the standard normal distribution. In the implementation of the proposed method, we keep using a logit model as shown in Equation 8. Because a logit coefficient equals 1.6 times a probit coefficient (e.g., Amemiya, 1981), we specify the sensitivity parameter reflecting the conditional association between U and M, which is a logit coefficient, as $1.6 \times (\beta^m_u/1.6) = \beta^m_u$.

Because the ρ-based method only allows a probit model to be fitted to a binary mediator, Scenario 1, which generates the mediator from a logit model, assesses how the ρ-based method is affected by a violation of the distributional assumption of the mediator. Scenario 2, which generates the mediator from a probit model, assesses the performance of the ρ-based method when all the functional and distributional assumptions are met.

In Scenario 3, to assess the influence of violations of the functional assumptions of the mediator or the outcome, we vary the coefficients of X-by-T and U-by-T interactions in the data generation while always setting them to 0 in the sensitivity analysis. Specifically, we separately evaluate the influence of misspecifications in (a) the observed part of the outcome model, (b) the observed part of the mediator model, (c) the unobserved part of the outcome model, and (d) the unobserved part of the mediator model, by respectively specifying in the data generation (a) $\beta^y_{tx1} = \beta^y_{tx2} = \beta^y_{tx3} = 1, \beta^m_{tx1} = \beta^m_{tx2} = \beta^m_{tx3} = \beta^y_{ut} = \beta^m_{ut} = 0$, (b) $\beta^m_{tx1} = \beta^m_{tx2} = \beta^m_{tx3} = 1, \beta^y_{tx1} = \beta^y_{tx2} = \beta^y_{tx3} = \beta^y_{ut} = \beta^m_{ut} = 0$, (c) $\beta^y_{ut} = 1, \beta^y_{tx1} = \beta^y_{tx2} = \beta^y_{tx3} = \beta^m_{tx1} = \beta^m_{tx2} = \beta^m_{tx3} = \beta^m_{ut} = 0$, and (d) $\beta^m_{ut} = 1, \beta^y_{tx1} = \beta^y_{tx2} = \beta^y_{tx3} = \beta^m_{tx1} = \beta^m_{tx2} = \beta^m_{tx3} = \beta^y_{ut} = 0$.

To evaluate the performance of the sensitivity analysis strategies with the change of the strength of unmeasured confounding, we set each of the sensitivity parameters, β^m_u and β^y_u, to -2, -1, 0, 1, and 2. This yields a 5×5 grid, including 25 conditions within each of the above scenarios. For the application of the ρ-based method, we calculate ρ as the correlation between the error terms of the outcome model and the latent mediator model that do not adjust for U, given each pair of β^m_u and β^y_u.

We make 1,000 replications for each combination of the sensitivity parameters in each scenario. For each replication, we obtain the adjusted NIE and NDE estimates that account for unmeasured confounding at a given strength by applying the proposed and ρ-based methods. For the proposed approach, we repeatedly draw U 100 times ($K = 100$) from its conditional distribution and adopt various mediation analysis methods to estimate the NIE and NDE by adjusting for U. By comparing the adjusted estimates to the true effects, which can be calculated based on the true model parameter values, we assess the ability of the sensitivity analysis methods to recover the true effects.

To better illustrate the influence of unmeasured confounding on both the point estimation and the sampling variability, we also estimate the NIE and NDE without U and with true values of U based on each mediation analysis method in Scenario 1. Such estimations are not considered in the other two scenarios, which are targeted at assessing the influence of violations of the distributional or functional assumptions on the ability of the sensitivity analysis methods to recover the true effects.

Simulation Results

We present the simulation results in Figures B1–B14 in Supplemental Appendix B, each of which is a 5×5 grid. The columns and rows respectively represent the sensitivity parameters β^y_u and β^m_u. Each cell of the grid is composed of multiple boxplots, each displaying the sampling distribution of the NIE/NDE estimate obtained from the method as labeled. A detailed introduction to the labels can be found below each figure. Each red line represents the true NIE/NDE in the corresponding cell.

Both the Mediator and Outcome Models Are Correctly Specified

As represented in Figures B1 and B2, when either sensitivity parameter is 0, the red lines in the first two boxplots in each grid align with the medians of the boxplots. This indicates that, when there is no unmeasured confounding and when the models are correctly specified, both the regression- and weighting-based analysis methods provide unbiased NIE and NDE estimates.

When both sensitivity parameters are nonzero, the red lines deviate from the medians of the first two boxplots in each grid, and the deviations increase as the magnitudes of the sensitivity parameters become larger. This reveals that, when the confounding role of U is not negligible, ignoring U in the analysis would bias both the NIE and NDE estimates, no matter whether the regression- or weighting-based analysis method is adopted. The bias increases with the strength of the unmeasured confounding. Nevertheless, both the proposed and ρ-based sensitivity analysis strategies can recover the true effects (as represented by boxplots 5–6 in each grid in Figures B1 and B2 and boxplot 3 in each grid in Figures B3 and B4), just as if the true U is adjusted for in the analysis (as represented by boxplots 3 and 4 in each grid in Figures B1 and B2).

A comparison of the spreads of the sampling distributions as represented in the first two boxplots and those in boxplots 3 and 4 in each grid in Figures B1 and B2 shows that, should U be observed, controlling for U in the analysis would either increase or decrease the sampling variability of the NIE and NDE estimates, depending on the factors explicated in the introduction section. This verifies the importance of considering the change in the sampling variability of the focal effect estimates in a sensitivity analysis. As represented by boxplots 5–6 in each grid in Figures B1 and B2 and boxplot 3 in each grid in Figures B3 and B4, due to the uncertainty of U, sensitivity analysis strategies sometimes provide

slightly less precise effect estimates than the analysis adjusting for true U.

The Distributional Assumption of the Outcome or the Mediator Is Violated

As represented in the first two boxplots in each grid in Figures B3 and B4 and boxplot 7 in each grid in Figures B1 and B2, using a link function different from the one used in the generation of a binary mediator does not affect the proposed or ρ-based strategies' ability to recover the true NIE and NDE. Similarly, Figures B5 and B6 reflect that, when the normal assumption of a continuous outcome is violated, both the ρ-based and proposed methods can still recover the true effects. We expect that the same conclusion applies to a continuous mediator or a binary outcome.

The Functional Assumptions of the Outcome or the Mediator Are Violated

As shown in Figures B7–B14, the omission of the X-by-T interaction or the U-by-T interaction from the outcome or mediator model would bias the adjusted estimates of both NIE and NDE. When the outcome model is misspecified, the proposed method generates a slightly smaller bias in the adjusted estimates when implemented in the weighting-based causal mediation analysis than in the regression-based analysis. Nevertheless, the difference is small. This is because, although the weighting-based approach is robust to misspecifications of the outcome model, such misspecifications would pull the simulated U away from its true conditional distribution.

Summary

The proposed sensitivity analysis strategy performs similarly to the ρ-based method in most scenarios. The two methods capture the influence of unmeasured pretreatment confounding on the estimation efficiency to the similar extent. Both methods can recover the true effects when the mediator and outcome models are correctly specified or under violations of distributional assumptions of the outcome or the mediator. Misspecification of the mediator model or the outcome model would bias the adjusted effect estimates. If a potential interaction between the unmeasured confounder and the treatment or the mediator exists, we need to modify the conditional distribution of the unmeasured confounder accordingly and control for these interactions in the adjusted causal mediation analysis. Correspondingly, more sensitivity parameters are involved.

Application

In this application, we use the same data as Imai, Keele, and Tingley (2010). As described in the section about the JOBS II intervention study, we investigate whether a job training intervention (T) reduces participants' level of depression (Y) by enhancing their job search self-efficacy (M). There are 899 individuals remaining in the sample after deletion of all the observations that contain missing values. Hence, as Imai, Keele, and Tingley (2010) acknowledged, the analysis is for illustrative purposes and not for inference about the program efficacy. The data include all the pretreatment covariates to increase the credibility of the sequential ignorability assumption. A binary mediator was constructed by splitting one's original score of job search self-efficacy at the sample median. Different from Imai, Keele, and Tingley (2010), we standardize the outcome and continuous covariates, to facilitate the interpretations of the sensitivity parameters. Hence, the following analytic results are in different scales from those reported in Imai, Keele, and Tingley (2010).

We employ in the original analysis the weighting-based causal mediation analysis method. The intervention program increased the rate of high confidence level in job search skills by 8% (SE = .03, t = 2.27, p = .02).[3] The natural indirect effect is estimated to be −.033 (SE = .015, t = −2.20, p = .03), and the natural direct effect is estimated to be −.064 (SE = .071, t = −.90, p = .37). The results indicate that the increase in job search self-efficacy induced by the intervention significantly decreased participants' depression level.

The above results are obtained based on the assumption that the relationship between job search self-efficacy and depression level is unconfounded given the observed pretreatment covariates. However, the assumption may not hold. As illustrated in the section of identification assumptions, one's motivation to find new jobs at baseline is a potential unmeasured pretreatment confounder. Ignoring it in the analysis may bias the NIE and NDE estimates. With such a specific unmeasured confounder in mind, an analyst may determine the sensitivity parameter values based on existing data or previous empirical findings. Alternatively, one may compare its unique confounding role to those of the observed confounders based on existing data, previous empirical findings, or theoretical reasoning, so that the coefficients of the observed covariates in the mediator and outcome models could be used as referent values for determining specific values or a plausible range of the sensitivity parameters (e.g., Carnegie et al., 2016; Hong et al., 2021; Imbens, 2003). For example, one may argue that the conditional associations of motivation with the mediator and the outcome do not exceed those of the baseline depression level. By respectively setting β_u^y and β_u^m to be equal to the coefficient estimates of the baseline depression level in the outcome and mediator models, we estimate the NIE to be −.031 (SE = .014, t = −2.21, p = .03) and the NDE to be −.066 (SE = .072, t = −.92, p = .36). The results indicate that an additional adjustment of motivation would have little influence on the estimation and inference results, and thus the original analytic results are expected to be robust to the omission of motivation to find new jobs at baseline from the analysis. Similarly, should there be a set of potential unmeasured pretreatment confounders, one could evaluate their collective influence by comparing their joint confounding role to the observed covariates.

In addition, we provide a convenient tool for researchers to visually assess how strong the unmeasured confounding needs to be for the sign or statistical significance of the original conclusions to be changed. We first set the unconditional probability of U at π = .5 and specify for each sensitivity parameter a range three times wider than that of the coefficients of the observed covariates in the corresponding model. In this example, we set β_u^m to range between −3 and 3 and β_u^y to range between −1.5 and 1.5. We then divide all the possible values of β_u^m and β_u^y within the range into a 20 × 20 grid and generate 5,000 draws of the unmeasured pretreatment

[3] Since the estimator of the weight is consistent, the weighting-based NIE and NDE estimates, as weighted mean contrasts of the outcome, are approximately normal as the sample size increases. Given the sample size of this example, we assess the significance of the effects based on t tests rather than bootstrapping.

confounder U for each cell.[4] Figures 1 and 2 respectively illustrate the sensitivity analysis result for the natural indirect effect and that for the natural direct effect.

Each black contour represents the combinations of sensitivity parameters that lead to the same effect estimate as indicated by the number on the contour. For example, as shown in Figure 1, if the coefficient of U in the standardized outcome regression on the treatment, mediator, and observed covariates (β_u^y) is $-.5$, and if the coefficient of U in the standardized logistic regression of the mediator on the treatment and observed covariates (β_u^m) is 1, the NIE estimate is increased from $-.033$ to $-.024$ after adjustment for U. Because the treatment is randomized in this application, the total effect of treatment assignment on the outcome, that is, the sum of NIE and NDE, is unbiased. Hence, the number on the contour in Figure 1 and the number on the corresponding contour in Figure 2 always add up to the total effect estimate, .097. At the same parameter values in the above example ($\beta_u^y = -0.5,\ \beta_u^m = 1$), the NDE estimate is $-.073$.

The sensitivity parameters along the red dashed curves reduce the estimate to zero. As shown in Figure 1, if β_u^y is equal to 1, for the NIE to be removed, β_u^m needs to be around -2. Each blue dotted curve corresponds to the boundary at which the significance of the effect is changed at the significance level of .05. The effect is insignificant on the side that contains the zero line. For example, the NIE is significant in the original analysis. If β_u^y is equal to .5, for the NIE to become insignificant, β_u^m must be smaller than -1. The larger the magnitudes of the sensitivity parameters are for removing the effects or changing their significance, the less sensitive the results are. Each black curve tends to be parallel to an adjacent black curve but not to an adjacent blue dotted curve, especially in Figure 1 for the NIE in this application. This indicates the change in sampling variability after adjustment for U. Ignoring it would lead to misleading conclusions about the influence of unmeasured pretreatment confounding on the significance of the causal effects. Therefore, it is crucial to account for the change in sampling variability in the sensitivity analysis.

Each dot corresponds to the conditional associations of each observed covariate with Y and M, which are used to calibrate the strength of the sensitivity parameters. Both Figure 1 and Figure 2 indicate that, for the original causal conclusions about the indirect and direct effects to be changed, an unmeasured pretreatment confounder must be much stronger than the most important observed pretreatment confounder. Given that this is highly unlikely, the results are insensitive to an unmeasured pretreatment confounder whose marginal distribution is Bernoulli (.5). To further evaluate the influence of the marginal distribution of U, we draw two additional sets of sensitivity plots when $\pi = .1$ and $\pi = .9$, respectively, as shown in Supplemental Appendix C. Clearly, it becomes even harder to change the original conclusion when the marginal probability of U is further away from .5.

Alternative Sensitivity Parameters Based on Partial R^2

In addition to the conditional association between U and Y and that between U and M, β_u^y and β_u^m, we could also assess the confounding role of U through partial R^2 values, that is, the proportions of the unexplained variances in the initial mediator and outcome models that are explained by U (e.g., Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010). Some researchers prefer partial R^2 values (e.g., Imbens, 2003) because partial R^2 values always range between 0 and 1. When specific information about the confounding role of U is unavailable, one can choose common partial R^2 values within the range and thus avoid specifying implausible values for the sensitivity parameters. It also eases the specification of sensitivity parameter values if there is more than one potential unmeasured pretreatment confounders. When Y or M is binary, one has to construct a pseudo partial R^2 (e.g., McKelvey & Zavoina, 1975) based on the variation in the latent index of the binary variable, which lacks intuitive interpretations. Hence, we propose partial R^2 as alternative sensitivity parameters only for continuous Y and M, while β_u^y and β_u^m are more intuitive sensitivity parameters when Y or M is binary.

Following Cinelli and Hazlett (2020), we express β_u^y and β_u^m as functions of partial R^2 values, R_Y^{*2} and R_M^{*2}. Specifically,

$$\beta_u^y = \frac{\mathrm{cov}(Y^{\perp T,M,X},\ U^{\perp T,M,X})}{\mathrm{var}(U^{\perp T,M,X})} = \frac{\mathrm{sd}(Y^{\perp T,M,X})R_Y^*}{\mathrm{sd}(U^{\perp T,M,X})} = \frac{\mathrm{sd}(Y^{\perp T,M,X})R_Y^*}{\mathrm{sd}(U^{\perp T,X})\sqrt{1-R_M^{*2}}} = \frac{\mathrm{sd}(Y^{\perp T,M,X})R_Y^*}{\mathrm{sd}(U)\sqrt{1-R_M^{*2}}}, \tag{10}$$

where $\mathrm{sd}(Y^{\perp T,M,X})$ denotes the standard deviation of Y after removing the components linearly explained by T, M, and X; $R_Y^* = \mathrm{cor}(Y^{\perp T,M,X},\ U^{\perp T,M,X})$ indicates the partial correlation between Y and U, with the effects of T, M, X removed from both Y and U, and correspondingly R_Y^{*2} represents the proportion of the unexplained variance in the initial outcome model that is explained by U. Similarly, $R_M^* = \mathrm{cor}(M^{\perp T,X},\ U^{\perp T,X})$ indicates the partial correlation between M and U, with the effects of T and X removed from both M and U, and correspondingly R_M^{*2} represents the proportion of the unexplained variance in the initial mediator model that is explained by U, that is, $R_M^{*2} = 1 - \frac{\mathrm{var}(M^{\perp T,X,U})}{\mathrm{var}(M^{\perp T,X})}$. Due to the symmetry of partial R^2, $R_M^{*2} = 1 - \frac{\mathrm{var}(U^{\perp T,X,M})}{\mathrm{var}(U^{\perp T,X})}$, and thus $\mathrm{sd}(U^{\perp T,M,X}) = \mathrm{sd}(U^{\perp T,X}) \times \sqrt{1-R_M^{*2}}$, in which $\mathrm{sd}(U^{\perp T,X}) = \mathrm{sd}(U)$ because $U \perp \{T, X\}$. Similarly,

$$\beta_u^m = \frac{\mathrm{cov}(M^{\perp T,X},\ U^{\perp T,X})}{\mathrm{var}(U^{\perp T,X})} = \frac{R_M^*\mathrm{sd}(M^{\perp T,X})}{\mathrm{sd}(U^{\perp T,X})} = \frac{R_M^*\mathrm{sd}(M^{\perp T,X})}{\mathrm{sd}(U)}, \tag{11}$$

where $\mathrm{sd}(M^{\perp T,X})$ denotes the standard deviation of M unexplained by T and X.

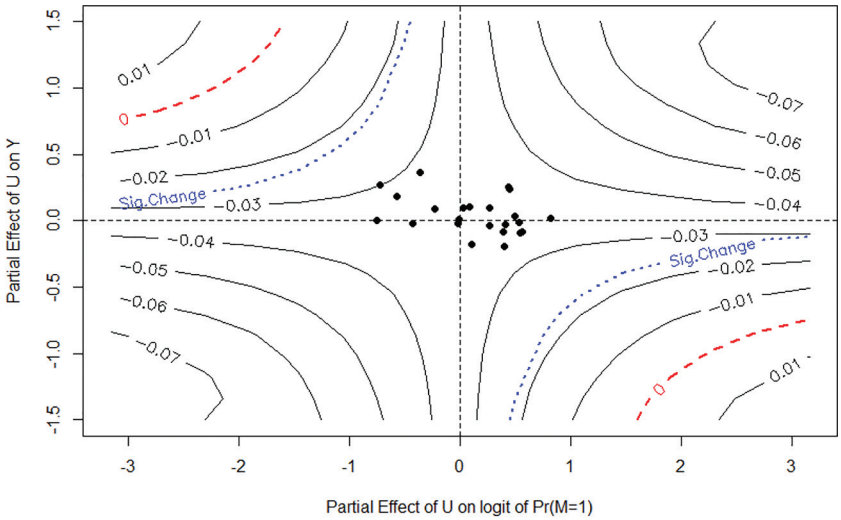
With R_Y^{*2} and R_M^{*2} as sensitivity parameters, one can divide the whole range of their plausible values between 0 and 1 into a grid. Because $\mathrm{sd}(Y^{\perp T,M,X})$ and $\mathrm{sd}\ (M^{\perp T,X})$ can be calculated based on observed data, and $\mathrm{sd}(U)$ is known given the marginal distribution of U, one can calculate the values of β_u^y and β_u^m corresponding to each pair of R_Y^{*2} and R_M^{*2} and apply the proposed sensitivity analysis algorithm.

In observational studies, one may further express the conditional association between the unmeasured confounder and the treatment as a function of the proportion of the unexplained variance in the initial treatment model that is explained by U.

[4] In this application, it is sufficient to divide all the possible values of β_u^m and β_u^y within the range into a 10 by 10 grid and generate 20 draws of the unmeasured pretreatment confounder U for each cell. We refined the grid and increased the number of repetitions for improving the smoothness of the curves in the following visualization of the sensitivity analysis results.

Figure 1

Sensitivity Analysis Plot for the Natural Indirect Effect When π = 0.5



Note. π denotes the unconditional probability of *U*. Each black contour represents the combinations of sensitivity parameters that lead to the same effect estimate as indicated by the number on the contour. The sensitivity parameters along the red dashed curves reduce the estimate to zero. Each blue dotted curve corresponds to the boundary at which the significance of the effect is changed at the significance level of 0.05. The effect is insignificant on the side that contains the zero line. Each dot corresponds to the conditional associations of each observed covariate with *Y* and *M*, which are used to calibrate the strength of the sensitivity parameters. See the online article for the color version of this figure.

Discussion

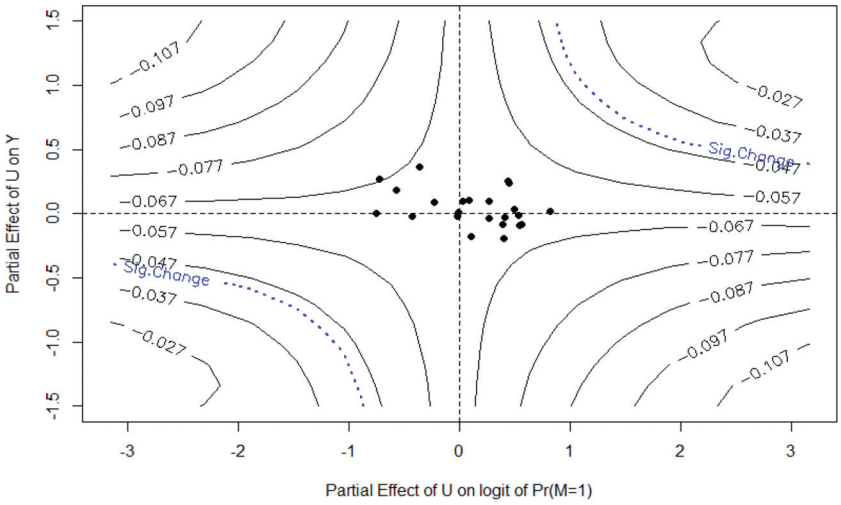
In this article, we review the existing causal mediation analysis methods under the sequential ignorability assumption and introduce sensitivity analysis methods for assessing the sensitivity of mediation analysis results to unmeasured pretreatment confounding. To overcome the limitations of the existing sensitivity analysis methods, we propose a simulation-based strategy and provide a convenient tool for psychological researchers to visually represent the sensitivity analysis results. The method allows users to intuitively quantify the strength of unmeasured pretreatment confounding through unmeasured pretreatment confounders' conditional associations with the treatment, mediator, and outcome or the corresponding partial R^2 values. Given values of the sensitivity parameters, we simulate the unmeasured confounder from its conditional distribution. By comparing the causal mediation analysis results before and after adjusting for the simulated confounder, we capture the influence of unmeasured pretreatment confounding on both estimation bias and estimation efficiency, while the latter is nontrivial but usually ignored in most existing sensitivity analysis strategies. We have verified this through simulations. Although the ρ-based method shares the same strength, our method has its own advantages. First, while the magnitude of ρ is meaningful only through comparisons across studies, our sensitivity parameters can be used independently to intuitively assess the confounding role of unmeasured pretreatment confounders. Second, while the ρ-based method assumes that the ignorability assumption of the treatment holds, our method can assess sensitivity to a potential violation of the assumption. Third, unlike the ρ-based method and the other existing sensitivity analysis strategies, our method is compatible with various causal mediation analysis methods and therefore enjoys broad applicability. Fourth, while the ρ-based method assumes the adjusted NIE and NDE estimates to be normally distributed, which may be violated in small samples and thus lead to inaccurate statistical inference, we offer a solution based on a bootstrap procedure.

For illustration purposes, we have presented our method with a binary treatment, a binary mediator, a continuous outcome, and a binary unmeasured pretreatment confounder, in a randomized experiment. The strategy can be easily extended for applications to (a) a continuous outcome and a continuous mediator, (b) a binary outcome and a binary mediator, or (c) a binary outcome and a continuous mediator, by using appropriate regression models in the derivation of the conditional distribution of the unmeasured confounder. The unmeasured pretreatment confounder can be either binary or continuous in each scenario. We have also extended the approach to observational studies in which the relationship between the treatment and the outcome and that between the treatment and the mediator are also confounded. Details can be found in Supplemental Appendix A. All the extensions have been incorporated into the R package mediationsens.

Important topics remain. First, we assume no posttreatment confounding of the mediator–outcome relationship and focus on evaluating sensitivity to unmeasured pretreatment confounding. However, in practice, covariates that confound the relationship between the mediator and the outcome might be affected by the treatment. For example, the job training program may stimulate one's preference for working, which might lead to higher job search self-efficacy and lower depression levels. Nevertheless, sensitivity analysis strategies for posttreatment confounders are still minimal. As Avin et al. (2005) showed, the indirect and direct effects are not identifiable in the presence of posttreatment

Figure 2

Sensitivity Analysis Plot for the Natural Direct Effect When π = 0.5



Note. π denotes the unconditional probability of *U*. Each black contour represents the combinations of sensitivity parameters that lead to the same effect estimate as indicated by the number on the contour. Each blue dotted curve corresponds to the boundary at which the significance of the effect is changed at the significance level of 0.05. The effect is insignificant on the side that contains the zero line. Each dot corresponds to the conditional associations of each observed covariate with *Y* and *M*, which are used to calibrate the strength of the sensitivity parameters. See the online article for the color version of this figure.

confounders if an interaction exists between the treatment and the mediator. Nevertheless, one may still hope to examine plausible estimates of the indirect and direct effects in such settings and evaluate their sensitivity to the omission of observed or unmeasured posttreatment confounding. Imai and Yamamoto (2013) and Vansteelandt and Vanderweele (2012) proposed sensitivity analysis methods for evaluating the influence of an observed posttreatment confounder. VanderWeele and Chiba (2014) is also applicable to unmeasured posttreatment confounding. However, all these strategies ignore the influence of posttreatment confounders on the estimation efficiency. We leave it to our future research for investigating to what extent a posttreatment confounder must be associated with the mediator and the outcome for the original causal inference about the indirect and direct effects to be altered. Second, we assume that the treatment, mediator, and outcome are measured without errors. However, this assumption is often violated, especially in psychological and behavioral sciences. Liu and Wang (2020) assessed the joint consequences of measurement errors and unmeasured pretreatment confounding, by using the reliability levels of the treatment, mediator, and outcome to quantify the degree of violation of the no-measurement-error assumption. However, this method does not apply when the treatment is not randomized, the mediator or the outcome is discrete, or there exists an treatment-by-mediator interaction. It would be of future research interest to extend the proposed method for examining the influence of co-occurrence of measurement errors and unmeasured pretreatment confounding.

References

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, *19*(4), 1483–1536.

Avin, C., Shpitser, I., & Pearl, J. (2005). *Identifiability of path-specific effects*. Department of Statistics, UCLA.

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. https://doi.org/10.1080/19345747.2015.1078862

Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *82*(1), 39–67. https://doi.org/10.1111/rssb.12348

Cox, M. G., Kisbu-Sakarya, Y., Miočević, M., & MacKinnon, D. P. (2013). Sensitivity plots for confounder bias in the single mediator model. *Evaluation Review*, *37*(5), 405–431. https://doi.org/10.1177/0193841X14524576

Ding, P., & Vanderweele, T. J. (2016). Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. *Biometrika*, *103*(2), 483–490. https://doi.org/10.1093/biomet/asw012

Dorie, V., Harada, M., Carnegie, N. B., & Hill, J. (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in Medicine*, *35*(20), 3453–3470. https://doi.org/10.1002/sim.6973

Harring, J. R., McNeish, D. M., & Hancock, G. R. (2017). Using phantom variables in structural equation modeling to assess model sensitivity to external misspecification. *Psychological Methods*, *22*(4), 616–631. https://doi.org/10.1037/met0000103

Hong, G. (2010). Ratio of mediator probability weighting for estimating natural direct and indirect effects. *Proceedings of the American Statistical Association, Biometrics Section* (pp. 2401–2415). 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. https://doi.org/10.3102/1076998617749561

Hong, G., Yang, F., & Qin, X. (2021). Did you conduct a sensitivity analysis? A new weighting-based approach for evaluations of the average treatment effect for the treated. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, *184*(1), 227–254. https://doi.org/10.1111/rssa.12621

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685. https://doi.org/10.1080/01621459.1952.10483446

Huber, M. (2014). Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics*, *29*(6), 920–943. https://doi.org/10.1002/jae.2341

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

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71. https://doi.org/10.1214/10-STS321

Imai, K., & Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*, *21*(2), 141–171. https://doi.org/10.1093/pan/mps040

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *The American Economic Review*, *93*(2), 126–132. https://doi.org/10.1257/000282803321946921

Lange, T., Vansteelandt, S., & Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *American Journal of Epidemiology*, *176*(3), 190–195. https://doi.org/10.1093/aje/kwr525

Liu, X., & Wang, L. (2020). The impact of measurement error and omitting confounders on statistical inference of mediation effects and tools for sensitivity analysis. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000345

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Erlbaum.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*(2), 144–158. https://doi.org/10.1177/0193841X9301700202

Mauro, R. (1990). Understanding LOVE (left out variables error): A method for estimating the effects of omitted variables. *Psychological Bulletin*, *108*(2), 314–329. https://doi.org/10.1037/0033-2909.108.2.314

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, *4*(1), 103–120. https://doi.org/10.1080/0022250X.1975.9989847

Neyman, J., & Iwaszkiewicz, K. (1935). Statistical problems in agricultural experimentation. *Supplement to the Journal of the Royal Statistical Society*, *2*(2), 107–180.

Nielsen, S. F. (2000). The stochastic EM algorithm: Estimation and asymptotic results. *Bernoulli*, *6*(3), 457–489. https://doi.org/10.2307/3318671

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann.

Pearl, J. (2001). Direct and indirect effects. In J. Breese & D. Koller (Eds.), *Proceedings of the seventeenth conference on uncertainty in artificial intelligence* (pp. 411–420). Morgan Kaufmann.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In M. E. Halloran & D. Berry (Eds.), *Statistical models in epidemiology, the environment, and clinical trials* (pp. 95–133). Springer. https://doi.org/10.1007/978-1-4612-1284-3_2

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, *82*(398), 387–394. https://doi.org/10.1080/01621459.1987.10478441

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, *6*(1), 34–58.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593. https://doi.org/10.2307/2287653

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*(396), 961–962. https://doi.org/10.2307/2289065

Rubin, D. B. (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, *25*(3), 279–292. https://doi.org/10.1016/0378-3758(90)90077-8

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*(4), 279–313. https://doi.org/10.1037/a0014268

Schomaker, M., & Heumann, C. (2018). Bootstrap inference when using multiple imputation. *Statistics in Medicine*, *37*(14), 2252–2266. https://doi.org/10.1002/sim.7654

Smith, L. H., & VanderWeele, T. J. (2019). Mediational E-values: Approximate sensitivity analysis for unmeasured mediator–outcome confounding. *Epidemiology*, *30*(6), 835–837. https://doi.org/10.1097/EDE.0000000000001064

Tchetgen, E. J., & Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of Statistics*, *40*(3), 1816–1845. https://doi.org/10.1214/12-AOS990

Ten Have, T. R., Elliott, M. R., Joffe, M., Zanutto, E., & Datto, C. (2004). Causal models for randomized physician encouragement trials in treating primary care depression. *Journal of the American Statistical Association*, *99*(465), 16–25. https://doi.org/10.1198/016214504000000034

Valeri, L., & Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*(2), 137–150. https://doi.org/10.1037/a0031034

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, *21*(4), 540–551. https://doi.org/10.1097/EDE.0b013e3181df191c

VanderWeele, T. J., & Chiba, Y. (2014). Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. Epidemiology. *Biostatistics, and Public Health*, *11*(2), 1–16.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*(4), 457–468. https://doi.org/10.4310/SII.2009.v2.n4.a7

Vansteelandt, S., Bekaert, M., & Lange, T. (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods*, *1*(1), 131–158. https://doi.org/10.1515/2161-962X.1014

Vansteelandt, S., & Vanderweele, T. J. (2012). Natural direct and indirect effects on the exposed: Effect decomposition under weaker assumptions. *Biometrics*, *68*(4), 1019–1027. https://doi.org/10.1111/j.1541-0420.2012.01777.x

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. https://doi.org/10.1007/BF02506922

Zheng, W., & van der Laan, M. J. (2012). Targeted maximum likelihood estimation of natural direct effects. *The International Journal of Biostatistics*, *8*(1), 1–40. https://doi.org/10.2202/1557-4679.1361

Received May 23, 2020
Revision received March 11, 2021
Accepted June 21, 2021 ■