Mediation analysis with multiple mediators under unmeasured mediator-outcome confounding

Deshanee Senevirathne Wickramarachchi | Laura Huey Mien Lim | Baoluo Sun

Department of Statistics and Data Science, National University of Singapore, Singapore

Correspondence
Baoluo Sun, Department of Statistics and Data Science, National University of Singapore, 6 Science Drive 2, Singapore 117546.
Email: stasb@nus.edu.sg

Funding information
National University of Singapore, Grant/Award Number: R-155-000-203-133; Singapore Ministry of Education, Grant/Award Number: A-8000452-00-00

It is often of interest in the health and social sciences to investigate the joint mediation effects of multiple post-exposure mediating variables. Identification of such joint mediation effects generally require no unmeasured confounding of the outcome with respect to the whole set of mediators. As the number of mediators under consideration grows, this key assumption is likely to be violated as it is often infeasible to intervene on any of the mediators. In this article, we develop a simple two-step method of moments estimation procedure to assess mediation with multiple mediators simultaneously in the presence of potential unmeasured mediator-outcome confounding. Our identification result leverages heterogeneity of the population exposure effect on the mediators, which is plausible under a variety of empirical settings. The proposed estimators are illustrated through both simulations and an application to evaluate the mediating effects of post-traumatic stress disorder symptoms in the association between self-efficacy and fatigue among health care workers during the COVID-19 outbreak.

KEYWORDS
causal mediation analysis, multiple mediators, unmeasured confounding

1 | INTRODUCTION

Although evaluation of the total effect of a given exposure on an outcome of interest remains the goal of many empirical studies, there has been considerable interest in recent years to gain deeper understanding of causal mechanisms by teasing apart the direct exposure effect and the indirect effect operating through a given post-exposure mediating variable. While mediation analysis has a longstanding tradition primarily in the context of linear structural modeling, the concepts of natural direct and indirect effects have since been formalized under the potential outcomes framework by the seminal works of Robins and Greenland and Pearl, which disentangles the identification assumptions from parametric constraints and allows semiparametric identification and inference for these causal estimands of interest.

Identification of natural direct and indirect effects generally requires no unmeasured confounding of the exposure-mediator, exposure-outcome and mediator-outcome relationships. While the no unmeasured exposure-mediator and exposure-outcome confounding assumptions hold by design when the exposure is randomly assigned by the investigator, it is often infeasible to also intervene on the post-exposure mediator. As a result, one can rarely rule out unmeasured mediator-outcome confounding in empirical settings and sensitivity analysis methods have been developed to assess the impact of departures from this assumption. A growing literature has likewise emerged...
on the development of identification strategies in the presence of potential unmeasured mediator-outcome confounding, typically by leveraging an instrumental variable for the effect of the mediator on the outcome.\textsuperscript{15-23}

While most of the aforementioned literature has been restricted to the setting of a single mediator, researchers often focus on multiple mediators, either to assess the joint mediation effects through multiple mediators or address concerns about mediator-outcome confounding by exposure-induced intermediates.\textsuperscript{24,25} Identification of such joint mediation effects require no unmeasured confounding of the outcome with respect to the whole set of mediators.\textsuperscript{24} This key assumption is arguably more likely to be violated as the number of mediators under consideration grows, and the development of methodology to adequately address this issue remains a priority in mediation analysis. Nonetheless, extension of the instrumental variable identification approach to the multiple mediator setting presents some practical challenges as it generally requires as many instrumental variables as there are mediators.

Recently, Fulcher et al.\textsuperscript{26} proposed an alternative identification approach by using higher order moment conditions which does not require ancillary variables such as instrumental variables.\textsuperscript{27-30} In this article, we build upon and extend the works of VanderWeele and Vansteelandt\textsuperscript{24} and Fulcher et al.\textsuperscript{26} to the multiple mediator setting in the presence of potential unmeasured mediator-outcome confounding. The rest of the article is organized as follows. We introduce notation and assumptions in Section 2 and delineate the conditions for identification of mediation effects with multiple mediators simultaneously in Section 3. For estimation and inference, we propose a simple two-step method of moments procedure in Section 4. We investigate the empirical performance of the proposed estimator through extensive simulation studies in Section 5 and an application to evaluate the effect of self-efficacy on fatigue among health care workers during the COVID-19 outbreak in Section 6, before concluding with a brief discussion about possible extensions in Section 7.

## 2 Notation and Assumptions

Suppose that \((O_1, \ldots, O_n)\) are independent and identically distributed observations of \(O = (Y, A, M, X)\) from a population of interest, where \(Y\) is the outcome, \(A\) is the exposure, \(M = (M_1, \ldots, M_K)^T\) is a vector of \(K\) mediators of interest for some \(K \in \mathbb{Z}^+\) and \(X\) is a vector of measured baseline covariates not affected by the exposure. Robins and Greenland\textsuperscript{5} and Pearl\textsuperscript{6} defined what are now often called natural direct and indirect effects. To formalize these causal effects under the potential outcomes framework,\textsuperscript{31,32} let \(Y(a, m)\) denote the potential outcome that would be observed if \(A = a\) and \(M = m\). Similarly, let \(M(a)\) denote the potential mediator values that would be observed when exposure takes value \(A = a\). Comparing any two values \(a, a'\) of the exposure, the population average natural direct and indirect effects through the \(K\) mediators jointly are given by

\[
NDE(a, a') = E[Y(a, M(a'))] - Y(a', M(a'))], \quad NIE(a, a') = E[Y(a, M(a)) - Y(a, M(a'))],
\]

respectively.\textsuperscript{24}

The average natural direct and indirect effects are particularly relevant for describing the underlying mechanism by which the exposure operates, as their sum \(NDE(a, a') + NIE(a, a') = E[Y(a, M(a)) - Y(a', M(a'))]\) recovers the average total effect comparing any two values \(a, a'\) of the exposure. VanderWeele and Vansteelandt\textsuperscript{24} proposed a parametric regression-based identification approach under the following assumptions that for any values \(a, a'\) of the exposure and \(m\) of the mediators,

\begin{align*}
A1^* &. \quad M(a) \perp A | X; \\
A2^* &. \quad Y(a, m) \perp A | X; \\
A3^* &. \quad Y(a, m) \perp M | A, X; \\
A4^* &. \quad Y(a, m) \perp M(a') | X; \\
A5^* &. \quad Y = Y(a, m) \text{ if } A = a \text{ and } M = m \text{ and } M = M(a) \text{ if } A = a; \\
A6^* &. \quad E(Y | A, M, X) = \vec{\theta}_0 + \vec{\theta}_A A + \vec{\theta}_T M + \vec{\theta}_X X \text{ and } E(M_j | A, X) = \vec{\theta}_{0j} + \vec{\theta}_{1j} A + \vec{\theta}_{2j} X \text{ for } j = 1, \ldots, K.
\end{align*}

Assumptions (\(A1^*\))-(\(A3^*\)) state that the vector of measured covariates \(X\) is sufficiently rich to capture all confounding effects of the exposure-mediator, exposure-outcome and mediator-outcome relationships. Assumption (\(A4^*\)) is a cross-world independence assumption which, in conjunction with Assumption (\(A3^*\)), is equivalent to the absence of exposure-induced mediator-outcome confounders \(L\) on a causal diagram interpreted as a nonparametric structural equation model with independent errors.\textsuperscript{33} If the variables in \(L\) are measured then we can treat them as additional
mediators in $M$ and assess the mediation effects jointly. The potential outcomes are related to the observed data via the consistency Assumption (A5*). Lastly, VanderWeele and Vansteelandt considered the parametric outcome and mediator regressions in Assumption (A6*) for simplicity and ease of estimation. Under Assumptions (A1*)-(A6*), VanderWeele and Vansteelandt showed that the natural direct and indirect effects are encoded by the parameters

$$NDE(a, a') = \theta_1(a - a'), \quad NIE(a, a') = \beta_1 \theta_2(a - a'),$$

respectively, where $\beta_1 = (\beta_{1,1}, \ldots, \beta_{1,K})^T$.

Although Assumptions (A1*) and (A2*) generally hold when the exposure is randomly assigned, in practical settings Assumption (A3*) may be violated as it is often infeasible to also randomize or intervene on any of the $K$ mediators. To develop the identification framework which allows for unmeasured mediator-outcome confounding, we follow Ding and VanderWeele and incorporate an additional set of unmeasured baseline covariates $U$ not affected by the exposure in the assumptions:

A1. $M(a) \perp A|X, U$;
A2. $Y(a, m) \perp A|X, U$;
A3. $Y(a, m) \perp M|A, X, U$;
A4. $Y(a, m) \perp M(a')|X, U$;
A5. $Y = Y(a, m)$ if $A = a$ and $M = m$, and $M = M(a)$ if $A = a$;
A6. $E(Y|A, M, X, U) = \bar{\theta}_1 A + \bar{\theta}_2^T M + H(X, U)$ and $E(M_j|A, X, U) = \bar{\beta}_{1,j} A + G_j(X, U)$ for $j = 1, \ldots, K$;
A7. $U \perp A|X$.

Assumptions (A1)-(A4) are analogues of Assumptions (A1*)-(A4*) conditional on $(X, U)$. The partially linear regression models in Assumption (A6) extend the models in Assumption (A6*) to incorporate $U$, where $H(X, U)$ and $G_j(X, U)$ are arbitrary functions of $(X, U)$ encoding their confounding effects on the outcome and $j$th mediator, respectively. Assumption (A7) states that $U$ is independent of $A$ given $X$. Figure 1 depicts such a scenario, which may arise if the exposure is randomly assigned conditional on values of $X$ under completely randomized or randomized block designs. Based on the derivation in VanderWeele and Vansteelandt, it is straightforward to show that the natural direct and indirect effects are again given by (2) under Assumptions (A1)-(A6), which do not restrict the causal structure among the mediators under consideration but preclude identification and estimation of path-specific effects. However, it is now more challenging to identify the parameters indexing the regressions in Assumption (A6) as they depend on unmeasured variables $U$. In the next section, we discuss how identification may be achieved in conjunction with Assumption (A7). Extensions of the proposed framework to allow for binary outcome or mediators as well as exposure-mediator interactions are described in the Supplementary material of this article.

### 3 IDENTIFICATION

Under Assumptions (A6) and (A7), the observed data mediator regression models are

$$E(M_j|A, X) = \bar{\beta}_{1,j} A + G_j(X), \quad j = 1, \ldots, K,$$

**Figure 1** Causal diagram with unmeasured mediator-outcome confounding for two mediators within strata defined by $X$.
where \( \overline{G}_j(X) = E\{G_j(X, U) | A, X\} = E\{G_j(X, U) | X\} \) is an arbitrary function of the measured baseline covariates for \( j = 1, \ldots, K \). The partially linear regression curves and hence the parameters \( \overline{\beta}_1 \) and \( \overline{G} = (\overline{G}_1, \ldots, \overline{G}_K) \) in (3) are identified from the observed data. It remains to identify the parameters \( \overline{\theta} = (\overline{\theta}_2, \overline{\theta}_3)^T \) indexing the outcome regression. Fulcher et al\(^{26}\) proposed a higher moment condition to identify the natural indirect effect, which can be extended under Assumptions (A6) and (A7) to yield the following \( K \) conditional mean independence conditions,

\[
E\{(M_j - \overline{\beta}_1 A - \overline{G}_j(X))(Y - \overline{\theta}_2^T M) | A, X\} = \text{Cov}\{G_j(X, U), H(X, U) | A, X\} = \text{Cov}\{G_j(X, U), H(X, U) | X\},
\]

for \( j = 1, \ldots, K \). While (4) suffices for identification of \( \overline{\theta}_2 \), it is generally insufficient for identification of \( \overline{\theta}_3 \). For example, when there are no measured baseline covariates and \( A \) is binary, (4) places \( K \) restrictions on the observed data law, but there are \( K + 1 \) parameters in \( \overline{\theta} \). As a remedy, we propose the following additional conditional mean independence condition under Assumptions (A6) and (A7),

\[
E(Y - \overline{\theta}_1 A - \overline{\theta}_2^T M | A, X) = E\{H(X, U) | A, X\} = E\{H(X, U) | X\}.
\]

The moment conditions in (4) and (5) can be summarized as

\[
E\{\psi(O; \overline{\beta}_1, \overline{\theta}, \overline{G}) | A, X\} = E\{\psi(O; \overline{\beta}_1, \overline{\theta}, \overline{G}) | X\},
\]

where

\[
\psi(O; \overline{\beta}_1, \overline{\theta}, G) = \begin{bmatrix} (M_1 - \beta_{1,1} A - G_1(X)) (Y - \theta_2^T M) \\ \vdots \\ (M_K - \beta_{1,K} A - G_K(X)) (Y - \theta_2^T M) \\ Y - \overline{\theta}_1 A - \theta_2^T M \end{bmatrix} = \begin{bmatrix} (M - E(M | A, X; \overline{\beta}_1, G)) (Y - \theta_2^T M) \\ Y - \overline{\theta}_1 A - \theta_2^T M \end{bmatrix}.
\]

Following the G-estimation approach of Robins,\(^{34}\) we consider the unconditional form of (6) for identification and estimation. Let \( \overline{\pi}(X) = E\{A | X\} \) denote the propensity score\(^{35}\) which is either known by design or identified from observed data, and let \( \Psi(O; \overline{\pi}, \overline{\beta}_1, \overline{\theta}, G) = \{A - \pi(X)\} \psi(O; \overline{\beta}_1, \overline{\theta}, G) \).

**Lemma 1.** Suppose Assumptions (A6) and (A7) hold and the matrix \( G = \partial E\{\Psi(O; \overline{\pi}, \overline{\beta}_1, \overline{\theta}, G)\} / \partial \theta \mid_{\theta = \overline{\theta}} \) has full rank or equivalently \( \text{det}(G) \neq 0 \). Then there exists a neighborhood of \( \overline{\theta} \) such that \( \overline{\theta} \) is the unique solution to the moment condition

\[
E\{\Psi(O; \overline{\pi}, \overline{\beta}_1, \overline{\theta}, G)\} = 0,
\]

for all \( \theta \) in that neighborhood.

**Proof.** Suppose moment condition (6) holds. By the law of iterated expectations,

\[
E\{\Psi(O; \overline{\pi}, \overline{\beta}_1, \overline{\theta}, G) | A, X\} = E\{A - \overline{\pi}(X)\} E\{\psi(O; \overline{\beta}_1, \overline{\theta}, G) | A, X\} = E\{A - \overline{\pi}(X)\} E\{\psi(O; \overline{\beta}_1, \overline{\theta}, G) | X\} = 0.
\]

The rank condition is sufficient for local identification.\(^{36}\)
following mediator structural equation models:  

\[ M_j = \beta_{1,j}(\epsilon_j)A + G_j(X, U), \quad j = 1, \ldots, K, \]

where \( \epsilon_j \perp A, X, U \) represents heterogeneity in the effect of \( A \) on \( M_j \). Tchetgen Tchetgen et al.\(^\text{30} \) showed that \( \text{Var}(M_j|A, X) \) remains a function of \( A \) provided that the distribution of \( \epsilon_j \) is nondegenerate, for \( j = 1, \ldots, K \).

4 | A TWO-STEP METHOD OF MOMENTS ESTIMATION APPROACH

The semiparametric identification result in Lemma 1 involves the nonparametric nuisance components \( \pi \) and \( G \). In principle it is possible to estimate these components nonparametrically under sufficient smoothness conditions. However, if \( X \) contains numerous continuous variables, the resulting estimators may exhibit poor finite sample behavior in moderately sized samples as the data are too sparse to conduct stratified estimation.\(^\text{37} \) For this reason as well as simplicity and ease of implementation, in this article, we propose a two-step method of moments approach in which \( \pi \) and \( G \) are estimated under the following parametric assumptions in the first step, \( \text{A8.} \quad \pi(X) = \pi(C; \gamma) \) and \( G_j(X; \alpha_j) \) for \( j = 1, \ldots, K \), where \( \pi(C; \gamma) \) and \( G_j(X; \alpha_j) \) are known functions smooth in the finite-dimensional parameters \( \gamma \) and \( \alpha_j \) respectively.

Throughout let \( \hat{E}_n\{g(O)\} = n^{-1} \sum_{i=1}^n g(O_i) \) denote the empirical mean operator and \( a = (a_1^T, \ldots, a_K^T)^T \).

4.1 | First step

Let \( S(A, C; \gamma) \) and \( S(M, A, X; a, \beta_1) \) denote some user-specified unbiased estimating functions for the parameters \( \gamma, \alpha \) and \( \beta_1 \) respectively based on Assumption (A8). For example, typically a logistic regression model is used for binary \( A \), say \( \pi(X; \gamma) = \{1 + \exp(-\gamma^T X)\}^{-1} \), in which case the score function of \( \gamma \) is

\[ S(A, X; \gamma) = X[A - \pi(X; \gamma)]. \]  

(8)

In addition, if we specify the linear model \( G_j(X; \alpha_j) = a_j^T X \) for the \( j \)th mediator regression, \( j = 1, \ldots, K \), then

\[ S(M, A, X; a, \beta_1) = \begin{bmatrix} \langle A, X^\top \rangle^T(M_1 - \beta_{1,1}A - a_1^T X) \\ \vdots \\ \langle A, X^\top \rangle^T(M_K - \beta_{1,K}A - a_K^T X) \end{bmatrix}. \]  

(9)

Let \( \hat{\gamma}, \hat{\alpha}, \) and \( \hat{\beta}_1 \) denote the estimators which jointly solve the empirical moment condition

\[ \hat{E}_n\{S^T(A, C; \gamma), S^T(M, A, X; a, \beta_1)\}^T = 0, \]

and let \( \hat{\pi}(X) = \pi(X; \hat{\gamma}) \) and \( \hat{G}(X) = \{G_1(X; \hat{\alpha}_1), \ldots, G_K(X; \hat{\alpha}_K)\} \).

4.2 | Second step

In the second stage, we obtain \( \hat{\theta} \) which solves the following empirical version of moment condition (7),

\[ \hat{E}_n[\Psi(O; \hat{\pi}, \hat{\beta}_1, \theta, \hat{G})] = 0. \]
An appealing feature of the two-step procedure is that the nuisance components $\pi$ and $\overline{C}$ can be estimated before the second stage involving the outcome data, and therefore mitigates potential for “data-dredging” exercises that comes with a fully specified outcome model.\textsuperscript{38}

4.3 Inference

For inference, the empirical moments involved in the two-step procedure can be “stacked” to account for the variability associated with the first-step estimation of nuisance parameters.\textsuperscript{39} Specifically, let $\varphi = (\gamma^T, \alpha^T, \beta^T_1, \theta^T)^T$. Then the estimator $\hat{\varphi}$ may be viewed as solving the joint empirical moment condition $\hat{E}_n(\Phi(O; \varphi)) = 0$, where

$$\Phi(O; \varphi) = \{S^T(A, C; \gamma), S^T(M, A, X; \alpha, \beta_1), \Psi^T(O; \pi(\gamma), \beta_1, \theta, G(\alpha))\}^T.$$  

The following result on asymptotic normality of $\hat{\varphi}$ follows from standard M-estimation theory.\textsuperscript{39}

**Lemma 2.** Suppose Assumptions (A6)-(A8) and the rank condition in Lemma 1 hold. Then under standard regularity conditions and as $n \to \infty$,

$$\sqrt{n}(\hat{\varphi} - \varphi) \overset{D}{\rightarrow} N\left(0, \left[\frac{\partial E(\Phi(O; \varphi))}{\partial \varphi} \bigg|_{\varphi=\varphi}\right]^{-1} E\left[\Phi(O; \varphi)\Phi^T(O; \varphi)\right] \left[\frac{\partial E(\Phi(O; \varphi))}{\partial \varphi} \bigg|_{\varphi=\varphi}\right]^{-1T}\right).$$

It follows that, in conjunction with Assumptions (A1)-(A5), the estimators of the natural direct and indirect effects per unit change in the exposure are given by

$$\hat{NDE} = \hat{\theta}_1, \quad \hat{NIE} = \hat{\beta}_1^T \hat{\theta}_2,$$

respectively, and their limiting variances can be derived using the multivariate delta method. Alternatively, nonparametric bootstrap may also be used to perform inference.

5 SIMULATION STUDY

In order to investigate the numerical performance of the proposed estimators, we perform Monte Carlo experiments involving independent and identically distributed data $(Y, A, M, X, U)$ with three mediators $M = (M_1, M_2, M_3)^T$ generated as follows:

$$X \sim N(0, 1); \quad U|X \sim N(1 + 0.5X, 1),$$

$$A|X, U \sim \text{Bernoulli}\left(p = \frac{1}{1 + \exp(-0.8 - 1.2X)}\right),$$

$$M|A, X, U \sim N\left(\begin{array}{c} 1.2 + 1.5A + 1.1X + \eta U \\ 0.5 + 1.2A + 1.8X + \eta U \\ 1.3 + 1.0A + 0.5X + \eta U \end{array}\right), \Sigma,$$

$$Y|A, M, X, U \sim N(1.3 + 2.5A + 1.2M_1 + 0.8M_2 + M_3 + 1.5X + \eta U, 1),$$

where $\Sigma$ has the Toeplitz covariance structure $\text{Cov}(M_i, M_j) = 2^{-|i-j|}(1 + \delta A)$ for $1 \leq i, j \leq 3$. Under this data generating mechanism, we vary (i) the sample size $n = 400$ or 800, (ii) the degree of unmeasured mediator-outcome confounding encoded by $\eta = 0$ or 0.5 as well as (iii) the magnitude of heteroscedasticity encoded by $\delta = 2.0$ or 5.0. We compare the proposed estimators $\hat{NDE}$ and $\hat{NIE}$ implemented using the correctly specified models (8) and (9) with the regression-based estimators $\hat{NDE}_{\text{reg}}$ and $\hat{NIE}_{\text{reg}}$ of VanderWeele and Vansteelandt.\textsuperscript{24} The estimating equations are solved using the R package “BB”\textsuperscript{40} and standard errors are obtained using the empirical sandwich estimator.\textsuperscript{39}
TABLE 1 Comparison of methods for estimation of the natural direct and indirect effects in the absence of unmeasured outcome-mediator confounding ($\eta = 0$)

| $\hat{\text{NDE}}$ | $\hat{\text{NIE}}$ | $\hat{\text{NDE}}_{\text{reg}}$ | $\hat{\text{NIE}}_{\text{reg}}$ |
|---------------------|---------------------|-------------------------------|-------------------------------|
| $\delta = 2.0$      |                     |                               |                               |
| [Bias]              | 0.001               | 0.006                         | 0.007                         | 0.000                         |
|                     | 0.003               | 0.003                         | 0.001                         | 0.000                         |
| $\sqrt{\text{Var}}$| 0.224               | 0.405                         | 0.134                         | 0.378                         |
|                     | 0.148               | 0.260                         | 0.095                         | 0.253                         |
| $\sqrt{\text{EVar}}$| 0.225               | 0.405                         | 0.129                         | 0.375                         |
|                     | 0.149               | 0.280                         | 0.091                         | 0.265                         |
| Cov95               | 0.956               | 0.957                         | 0.937                         | 0.945                         |
|                     | 0.958               | 0.961                         | 0.941                         | 0.956                         |
| $\delta = 5.0$      |                     |                               |                               |
| [Bias]              | 0.003               | 0.003                         | 0.007                         | 0.001                         |
|                     | 0.002               | 0.002                         | 0.001                         | 0.000                         |
| $\sqrt{\text{Var}}$| 0.152               | 0.479                         | 0.129                         | 0.483                         |
|                     | 0.105               | 0.321                         | 0.091                         | 0.328                         |
| $\sqrt{\text{EVar}}$| 0.148               | 0.478                         | 0.124                         | 0.482                         |
|                     | 0.103               | 0.337                         | 0.088                         | 0.340                         |
| Cov95               | 0.940               | 0.946                         | 0.938                         | 0.945                         |
|                     | 0.947               | 0.955                         | 0.941                         | 0.954                         |

Note: The two rows of results for each estimator correspond to sample sizes of $n = 400$ and $n = 800$ respectively. [Bias] and $\sqrt{\text{Var}}$ are the Monte Carlo absolute bias and standard deviation of the point estimates, $\sqrt{\text{EVar}}$ is the square root of the mean of the variance estimates and Cov95 is the coverage proportion of the 95% confidence intervals, based on 1000 repeated simulations. Zeros denote values smaller than 0.0005.

The following remarks can be made based on the results of 1000 simulation replicates summarized in Tables 1 and 2. The proposed estimators $\hat{\text{NDE}}$ and $\hat{\text{NIE}}$ show negligible bias and coverage proportion close to the nominal level of 0.95 once sample size reaches $n = 800$ across all simulation settings, but with substantially higher variance than $\hat{\text{NDE}}_{\text{reg}}$ and $\hat{\text{NIE}}_{\text{reg}}$. In agreement with theory, $\hat{\text{NDE}}_{\text{reg}}$ and $\hat{\text{NIE}}_{\text{reg}}$ perform well in the absence of unmeasured outcome-mediator confounding but exhibit noticeable bias and undercoverage otherwise. The variance of the proposed estimators generally decreases when sample size or the magnitude of mediator heteroscedasticity increases. The Supplementary material contains results of additional Monte Carlo experiments in the presence of an exposure-induced unmeasured mediator-outcome confounder $U$. When $U$ is affected by $A$ and $\eta = 0.5$, both $\hat{\text{NDE}}$ and $\hat{\text{NIE}}$ show noticeable bias and undercoverage of confidence intervals. Furthermore, the target parameters $\tilde{\theta}_1$ and $\tilde{\theta}_1^T \tilde{\theta}_2$ can no longer be interpreted as the natural direct and indirect effects, respectively, due to violation of Assumption (A4).

6 APPLICATION

We apply the proposed methods in reanalysing an observational dataset from Hou and Deng41 on $n = 527$ healthcare workers from Anhui Province, China during the COVID-19 outbreak in March 2020 to evaluate the mediating effects of post-traumatic stress disorder symptoms in the association between self-efficacy and fatigue. We refer interested readers to Hou et al42 for further details on the study design. The binary exposure $A$ is the total score on the General Self-Efficacy Scale dichotomized at the sample median while the outcome of interest $Y$ is the standardized total score on the 14-item Fatigue Scale. The three mediators under consideration ($M_1, M_2, M_3$) are standardized scores in post-traumatic stress disorder symptoms reported in three categories; re-experiencing, avoidance and hyperarousal. The vector of measured
TABLE 2  Comparison of methods for estimation of the natural direct and indirect effects in the presence of unmeasured outcome-mediator confounding ($\eta = 0.5$)

|       | $\hat{NDE}$ | $\hat{NIE}$ | $\hat{NDE}_{\text{reg}}$ | $\hat{NIE}_{\text{reg}}$ |
|-------|-------------|-------------|--------------------------|--------------------------|
| $\delta = 2.0$ |             |             |                          |                          |
| | $\frac{|\text{Bias}|}{\sqrt{\text{Var}}}$ | $\sqrt{\text{EVar}}$ | Cov95 | $\frac{|\text{Bias}|}{\sqrt{\text{Var}}}$ | $\sqrt{\text{EVar}}$ | Cov95 |
| 0.027 | 0.009 | 0.295 | 0.191 | 0.317 | 0.192 | 0.957 | 0.957 |
| 0.040 | 0.027 | 0.467 | 0.301 | 0.486 | 0.324 | 0.953 | 0.962 |
| 0.198 | 0.194 | 0.146 | 0.101 | 0.140 | 0.099 | 0.697 | 0.496 |
| 0.186 | 0.176 | 0.434 | 0.296 | 0.434 | 0.307 | 0.926 | 0.923 |
|       |       |       |       |       |       |       |       |
| $\delta = 5.0$ |             |             |                          |                          |
| | $\frac{|\text{Bias}|}{\sqrt{\text{Var}}}$ | $\sqrt{\text{EVar}}$ | Cov95 | $\frac{|\text{Bias}|}{\sqrt{\text{Var}}}$ | $\sqrt{\text{EVar}}$ | Cov95 |
| 0.004 | 0.000 | 0.176 | 0.123 | 0.174 | 0.121 | 0.940 | 0.948 |
| 0.014 | 0.018 | 0.506 | 0.342 | 0.510 | 0.359 | 0.954 | 0.951 |
| 0.119 | 0.115 | 0.142 | 0.099 | 0.136 | 0.096 | 0.839 | 0.763 |
| 0.108 | 0.097 | 0.526 | 0.361 | 0.528 | 0.373 | 0.943 | 0.947 |

Note: The two rows of results for each estimator correspond to sample sizes of $n = 400$ and $n = 800$ respectively. See the footnote of Table 1.

TABLE 3  Estimates of direct and indirect effects of self-efficacy on fatigue mediated though three post-traumatic stress disorder symptoms jointly.

|       | $\hat{NDE}$ | $\hat{NIE}$ | $\hat{NDE}_{\text{reg}}$ | $\hat{NIE}_{\text{reg}}$ |
|-------|-------------|-------------|--------------------------|--------------------------|
|       | $\pm 1.96 \times \text{standard error}$ | $\pm 1.96 \times \text{standard error}$ | $\pm 1.96 \times \text{standard error}$ | $\pm 1.96 \times \text{standard error}$ |
| $-0.745 \pm 0.557$ | 0.050 $\pm 0.548$ | $-0.405 \pm 0.145$ | $-0.289 \pm 0.088$ |

Note: Estimate $\pm 1.96 \times$ standard error.

baseline variables $X$ include age, level of negative coping, gender, marital status, education level, years of working experience and technical title.

We implement the proposed estimators $\hat{NDE}$ and $\hat{NIE}$ using generalized linear models (8) and (9) together with the regression-based estimators $\hat{NDE}_{\text{reg}}$ and $\hat{NIE}_{\text{reg}}$ of VanderWeele and Vansteelandt$^{24}$ to estimate the natural direct and indirect effects of self-efficacy on fatigue mediated though the three categories of post-traumatic stress disorder symptoms jointly. Table 3 shows the analysis results. The regression-based estimators of VanderWeele and Vansteelandt$^{24}$ have noticeably smaller standard errors than the proposed estimators, in agreement with the Monte Carlo results, and yield statistically significant natural direct and indirect effect estimates of $-0.405$ and $-0.298$ respectively at $\alpha$-level of 0.05, which suggests that high self-efficacy alleviates fatigue not only directly, but also indirectly through mitigation of post-traumatic stress disorder symptoms. These results are consistent with those reported in the original study by Hou et al.$^{42}$ On the other hand, the point estimate of $\hat{NIE}$ is close to zero and statistically insignificant at the same $\alpha$-level. One reason for this discrepancy may lie in the presence of unmeasured mediator-outcome confounding even after accounting for the measured baseline variables in $X$, a scenario which is posited in Figure 2. For example, sleep quality is known to be associated with both post-traumatic stress disorder symptoms and fatigue.$^{43,44}$ Other potential unmeasured mediator-outcome confounders include occupation and hospital working unit, which were not recorded in the study due to confidentiality concerns.$^{42}$ Failure to account for these unmeasured mediator-outcome confounders may generate a biased estimate.
under a null natural indirect effect mediated through post-traumatic stress disorder symptoms. In closing, we acknowledge several limitations of the present application. We use main effects generalized linear models in the analysis due to their simplicity for illustration, although in principle they can be checked against the observed data using existing goodness-of-fit tests. In addition, the baseline variables measured in the observational study may be insufficient to ensure independence of $U$ and $A$ within strata of $X$. Further observational studies with richer baseline information are warranted to elucidate the mediating role of post-traumatic stress disorder symptoms in the association between self-efficacy and fatigue.

## DISCUSSIONS

One of the main concerns with causal mediation analysis with multiple mediators is the inability to categorically rule out the existence of unmeasured mediator-outcome confounders, as it is generally infeasible to randomize any of the mediators under consideration. In this article, we build upon the works of VanderWeele and Vansteelandt\cite{24} and Fulcher et al\cite{26} to develop a simple two-step method of moments estimation procedure to assess the joint mediation effects of multiple mediators in the presence of potential unmeasured mediator-outcome confounding. Moment condition (7) can in principle also be “stacked” with the moment conditions of existing instrumental variable methods to improve efficiency under a generalized methods of moments framework.\cite{45} Our work can be further extended in several important directions, including identification and estimation of path-specific effects\cite{46-48} or their interventional analogues\cite{49-51} as well as causal mediation analysis of survival outcome with multiple mediators,\cite{52,53} which we plan to pursue in future research.

## ACKNOWLEDGMENT

Baoluo Sun is supported by National University of Singapore Start-Up Grant (R-155-000-203-133) and Singapore Ministry of Education Academic Research Fund Tier 1 Grant (A-8000452-00-00).

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in openICPSR at https://doi.org/10.3886/E119156V1, reference number 119156.

## ORCID

Baoluo Sun https://orcid.org/0000-0003-0122-9993

## REFERENCES

1. VanderWeele T. Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford: Oxford University Press; 2015.
34. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat Theory Methods*. 1994;23(8):2379-2412. doi:10.1080/03610929408831393

35. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. doi:10.1093/biomet/70.1.41

36. Rothenberg TJ. Identification in parametric models. *Econometrica*. 1971;39(3):577-591. doi:10.2307/1913267

37. Robins JM, Ritov Y. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Stat Med*. 1997;16(3):285-319. doi:10.1002/(SICI)1097-0258(19970215)16:3<285::AID-SIM355>3.0.CO;2-%23

38. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med*. 2007;26(1):20-36. doi:10.1002/sim.2739

39. Newey WK, McFadden D. Large sample estimation and hypothesis testing. In: Engle R, McFadden D, eds. *Handbook of Econometrics*. Vol 4. Amsterdam: Elsevier; 1994:2111-2245.

40. Varadhan R, Gilbert P, BB: an R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *J Stat Softw*. 2009;32(4):1-26. doi:10.18637/jss.v032.i04

41. Hou T, Deng G. Data from: Self-efficacy and fatigue among health care workers during COVID-19 outbreak a moderated mediation model of posttraumatic stress disorder symptoms and negative coping. Inter-Univ Consortium Political Social Res Deposit. 2020;28. doi:10.3886/E119156V1

42. Hou T, Zhang R, Song X, et al. Self-efficacy and fatigue among non-frontline health care workers during COVID-19 outbreak: a moderated mediation model of posttraumatic stress disorder symptoms and negative coping. *PloS one*. 2020;15(12):e0243884. doi:10.1371/journal.pone.0243884

43. Wagner L, Cella D. Fatigue and cancer: causes, prevalence and treatment approaches. *Br J Cancer*. 2004;91(5):822-828. doi:10.1038/sj.bjc.6602012

44. Noel M, Vinall J, Tomfohr-Madsen L, Holley AL, Wilson AC, Palermo TM. Sleep mediates the association between PTSD symptoms and chronic pain in youth. *J Pain*. 2018;19(1):67-75. doi:10.1016/j.jpain.2017.09.002

45. Hansen LP. Large sample properties of generalized method of moments estimators. *Econometrica*. 1982;50(4):1029-1054. doi:10.2307/1912775

46. Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI’05*. San Francisco, CA: Morgan Kaufmann Publishers Inc; 2005:357-363.

47. Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics*. 2011;67(3):1028-1038. doi:10.1111/j.1541-0420.2010.01547.x

48. Daniel RM, De Stavola BL, Cousens S, Vansteelandt S. Causal mediation analysis with multiple mediators. *Biometrics*. 2015;71(1):1-14. doi:10.1111/biom.12248

49. Vansteelandt S, Daniel RM. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*. 2017;28(2):258-265. doi:10.1097/EDE.0000000000000596

50. Lin SH, VanderWeele T. Interventional approach for path-specific effects. *J Causal Inference*. 2017;5(1):20150027. doi:10.1515/jci-2015-0027

51. Loh WW, Moerkerke B, Loeys T, Vansteelandt S. Heterogeneous indirect effects for multiple mediators using interventional effect models. *Epidemiol Methods*. 2020;9(1):2020023. doi:10.1515/em-2020-0023

52. Huang YT, Yang HI. Causal mediation analysis of survival outcome with multiple mediators. *Epidemiology*. 2017;28(3):370-378. doi:10.1097/EDE.0000000000000651

53. Lin SH, Young JG, Logan R, VanderWeele TJ. Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Stat Med*. 2017;36(26):4153-4166. doi:10.1002/sim.7426

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wickramarachchi DS, Lim LHM, Sun B. Mediation analysis with multiple mediators under unmeasured mediator-outcome confounding. *Statistics in Medicine*. 2023;42(4):422-432. doi: 10.1002/sim.9624