Biometrika (2023), **00**, 0, *pp.* 1–15
https://doi.org/10.1093/biomet/asad015
Advance Access Publication 3 March 2023

Proximal mediation analysis

BY OLIVER DUKES
Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281 S9, 9000 Ghent, Belgium
oliver.dukes@ugent.be

ILYA SHPITSER
Department of Computer Science, Johns Hopkins University, 160 Malone Hall, 3400 N. Charles Street, Baltimore, Maryland 21218, U.S.A.
ilyas@cs.jhu.edu

AND ERIC J. TCHETGEN TCHETGEN
Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, 265 South 37th Street, Philadelphia, Pennsylvania 19104, U.S.A.
ett@wharton.upenn.edu

SUMMARY

A common concern when trying to draw causal inferences from observational data is that the measured covariates are insufficiently rich to account for all sources of confounding. In practice, many of the covariates may only be proxies of the latent confounding mechanism. Recent work has shown that in certain settings where the standard no-unmeasured-confounding assumption fails, proxy variables can be leveraged to identify causal effects. Results currently exist for the total causal effect of an intervention, but little consideration has been given to learning about the direct or indirect pathways of the effect through a mediator variable. In this work, we describe three separate proximal identification results for natural direct and indirect effects in the presence of unmeasured confounding. We then develop a semiparametric framework for inference on natural direct and indirect effects, which leads us to locally efficient, multiply robust estimators.

Some key words: Causal inference; Mediation; Semiparametric inference; Unmeasured confounding.

1. INTRODUCTION

The last few decades have seen the emergence of a literature on causal mediation analysis (Robins & Greenland, 1992; Pearl, 2001; VanderWeele & Vansteelandt, 2009; Imai et al., 2010; Tchetgen Tchetgen & Shpitser, 2012). This literature provides nonparametric definitions of direct and indirect effects in terms of contrasts of potential outcomes, as well as conditions necessary to identify and estimate these effects from data. Estimands that have received particular focus are natural direct and indirect effects, which are useful for understanding the mechanism underlying the effect of a particular intervention as they combine to produce the total causal effect.

The majority of work on identification of natural direct and indirect effects assumes that the measured covariates are sufficiently rich to account for confounding between the exposure and outcome, the mediator and outcome and the exposure and mediator. In practice, it is likely that many key confounding variables, e.g., disease severity, socio-economic status, cannot be ascertained with certainty from the measured covariates. At best, some of the measured covariates may be confounder proxies, e.g., mismeasured versions of the underlying confounders. This insight has led to work on leveraging proxy variables to help remove confounding bias in observational studies, with focus on the total effect of intervention. Negative control exposures and outcomes are examples of such proxies (Lipsitch et al., 2010; Shi et al., 2020); we refer to Shi et al. (2020) and Tchetgen Tchetgen et al. (2020) for further examples in observational studies.

If we are able to collect data on a sufficient number of proxies, confounding bias can sometimes be successfully removed in settings where standard analyses under a no-unmeasured-confounding assumption would fail. Building on the results of Kuroki & Pearl (2014), Miao et al. (2018) established nonparametric identification of the average treatment effect under a ‘double negative control design’, where both a negative control exposure and outcome are measured. Tchetgen Tchetgen et al. (2020) extended these results to settings with time-varying exposures and potentially unmeasured confounders. For estimation, they proposed proximal g-computation, a generalization of Robins’ parametric g-computation algorithm (Robins, 1986). Under a proximal identification strategy, Cui et al. (2023) developed semiparametric inference for the average treatment effect.

We consider identification and estimation of natural direct and indirect effects in the presence of unmeasured confounding. As an example, we consider the Job Corps study (Schochet et al., 2008). Beyond understanding the total effect of a job training intervention, the investigators were also interested in whether the intervention reduced criminal activity due to increased employment. It was possible that the association between program participation, employment and criminal activity were subject to confounding by a latent factor, such as motivation, that was only partially captured by the pretreatment covariates. In this work, we establish sufficient conditions for nonparametric identification of mediation estimands using a pair of proxy variables, giving three separate identification strategies. These each rely on modelling and estimation of different combinations of ‘confounding bridge’ functions (Miao et al., 2018). To reduce sensitivity to model misspecification, we obtain the efficient influence function under a semiparametric model for the observed data distribution, which leads us to estimators that are multiply robust. Our identification and estimation results allow for continuous or discrete outcomes and mediators. As far as we are aware, this is the first paper to use proxy variables for identification and inference for direct and indirect effects, with the exception of Cheng et al. (2022). However, their identification strategy is distinct from ours, as they relied on deep latent variable models. They also did not consider semiparametric inference and the issues of efficiency and robustness explored here.

2. Nonparametric proximal identification of the mediation functional

2.1. *Preliminaries*

We consider a setting where one is interested in the effect of a binary treatment A on an outcome Y that is mediated via a single intermediate variable M. We use U to refer to an unmeasured, potentially vector-valued confounding variable, which may be discrete, continuous or a combination of both types. Let $Y(a, m)$ refer to the potential outcome that

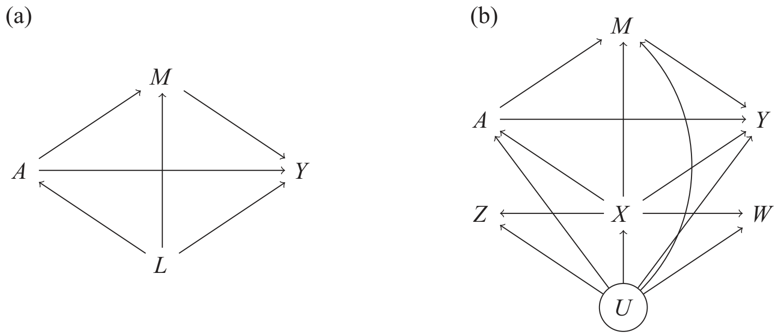

Fig. 1. (a) Directed acyclic graph with measured confounder L. (b) Directed acyclic graph with treatment, proxies and both measured and unmeasured confounders.

would be observed for someone if they were assigned to a given treatment at level a and mediator at m; similarly, $M(a)$ denotes the potential outcome for the mediator if treatment had taken value a. Then the total average treatment effect of A on Y can be decomposed as

$$E\{Y(1) - Y(0)\} = E[Y\{1, M(1)\} - Y\{1, M(0)\}] + E[Y\{1, M(0)\} - Y\{0, M(0)\}].$$

The first term $E[Y\{1, M(1)\} - Y\{1, M(0)\}]$ on the right-hand side of the equality is an example of a natural indirect effect, and captures the expected mean difference in Y if all individuals were assigned treatment $A = 1$, but the mediator was changed to the level it would take with $A = 0$. The second term $E[Y\{1, M(0)\} - Y\{0, M(0)\}]$ is a natural direct effect, and captures the effect of setting $A = 1$ versus $A = 0$ if everyone's mediator were at the level it would take with $A = 0$. Note that $E[Y\{1, M(1)\}] = E\{Y(1)\}$ and $E[Y\{0, M(0)\}] = E\{Y(0)\}$; results on nonparametric identification and inference for these quantities in a proximal learning setting already exist in Tchetgen Tchetgen et al. (2020) and Cui et al. (2023). We therefore focus on the mediation functional $\psi = E[Y\{1, M(0)\}]$ in the remainder of the article.

In order to identify ψ when one has access to a measured, potentially vector-valued covariate L, and supposing that M takes on values in $\mathcal{S}$, then one typically invokes the following conditional exchangeability assumptions: $Y(a, m) \perp\!\!\!\perp A \mid L$ for $a = 0, 1$ and each $m \in \mathcal{S}$; $M(a) \perp\!\!\!\perp A \mid L$ for $a = 0, 1$; and $Y(a, m) \perp\!\!\!\perp M(a) \mid A = a, L$ for $a = 0, 1$ and each $m \in \mathcal{S}$. In addition, the cross-world assumption $Y(a, m) \perp\!\!\!\perp M(a') \mid A = a, L$ for $a, a' = 0, 1$ and each $m \in \mathcal{S}$ is usually invoked (Robins & Richardson, 2010). It is known as such because independence between the counterfactual outcome and mediator values is required to hold across two different worlds of potentially conflicting values of treatment. If these hold, in addition to standard positivity and consistency conditions (Robins, 1986), then ψ can be identified via the mediation formula (Pearl, 2001)

$$\psi = \iint E(Y \mid A = 1, m, l)\, \mathrm{d}F(m \mid A = 0, l)\, \mathrm{d}F(l). \tag{1}$$

If one were to interpret the causal diagram in Fig. 1(a) as a nonparametric structural equation model with independent errors, then the above conditional independencies are consistent with that diagram.

The cross-world assumption has been the subject of much controversy, given that it can never be empirically verified or guaranteed by any study design. This has therefore led to an alternative way of conceptualizing natural direct and indirect effects via 'treatment-splitting' without reference to cross-world counterfactuals (Robins & Richardson, 2010; Robins et al., 2021). In what follows, we adopt the more traditional cross-world framework for mediation analysis, but expect that all of our results for identification of ψ carry over to the split-treatment approach, which we leave to future work.

2.2. *The proximal mediation formula*

Figure 1(b) displays a setting where the previous conditional exchangeability and the cross-world assumptions would not hold due to the presence of the unmeasured variable U, which is a common cause of A, M and Y. We now assume that the observed covariates L can be divided into three buckets (X, Z, W). First, X is a common cause of A, M and Y. Then Z and W are proxy variables that are only associated with A, M and Y via an unmeasured common cause. The failure of conditional exchangeability means that an analysis that would adjust for X, Z and W via X or conditioning on them in the mediation formula would return biased results due to residual confounding from U. In what follows, we therefore adopt a different approach for identification of ψ, based on leveraging the proxy variables Z and W to learn about U. Before giving the first identification result, we discuss the assumptions involved.

Assumption 1 (*Consistency*). Suppose that
(i) $M(a) = M$ almost surely for those with $A = a$;
(ii) $Y(a, m) = Y$ almost surely for those with $A = a$ and $M = m$.

Assumption 2 (*Positivity*). Suppose that
(i) $f_{M|A,U,X}(m \mid A, U, X) > 0$ almost surely for all $m \in \mathcal{S}$;
(ii) $\Pr(A = a \mid U, X) > 0$ almost surely for $a = 0, 1$.

Assumption 3 (*Latent conditional exchangeability*). Suppose that
(i) $Y(a, m) \perp\!\!\!\perp A \mid U, X$ for $a = 0, 1$ and each $m \in \mathcal{S}$;
(ii) $Y(a, m) \perp\!\!\!\perp M(a) \mid A = a, U, X$ for $a = 0, 1$ and each $m \in \mathcal{S}$;
(iii) $M(a) \perp\!\!\!\perp A \mid U, X$ for $a = 0, 1$.

Assumption 4 (*Latent cross-world assumption*). Suppose that $Y(a, m) \perp\!\!\!\perp M(a') \mid U, X$ for $a, a' = 0, 1$ and each $m \in \mathcal{S}$.

Assumptions 1–4 are similar to those made in standard mediation analyses, except that they also allow for the existence of an unmeasured variable U. The following assumption will directly enable us to leverage information from the proxy variables.

Assumption 5 (*Exposure- and outcome-inducing proxies*). Suppose that
(i) $Z \perp\!\!\!\perp Y(a, m) \mid A, M(a), U, X$ for $a = 0, 1$ and each $m \in \mathcal{S}$;
(ii) $Z \perp\!\!\!\perp M(a) \mid A, U, X$ for $a = 0, 1$;
(iii) $W \perp\!\!\!\perp M(a) \mid U, X$ for $a = 0, 1$;
(iv) $W \perp\!\!\!\perp (A, Z) \mid M(a), U, X$ for $a = 0, 1$.

This assumption essentially requires that A and M have no direct causal effect on W, and that Z has no causal effect on M or Y. Also, Z and W are only associated via the unmeasured common cause U. This assumption formally encodes what it means for Z and W to be proxies, in the sense that they become uninformative about confounding conditional on U. It is compatible with the causal diagram in Fig. 1(b). However, there are many other diagrams that may be compatible with Assumption 5, some of which are given in the Supplementary Material. For example, Z can be a direct cause of A, and W can cause Y. In general, this assumption is not testable, given that it involves the unmeasured U.

Identification of the mediation functional requires existence of multiple ‘confounding bridge’ functions (Miao et al., 2020). We also need to ensure that the confounding bridge functions can be identified from the data, given that we have access to Z. Formalizing both of these types of condition in general settings is subtle, since the confounding bridge function will be defined as the solution to an integral equation. We formalize them using the completeness conditions below.

Assumption 6 (*Completeness*).
(i) For any square-integrable function $g(u)$, if $E\{g(U) \mid Z = z, A = 1, M = m, X = x\} = 0$ for any z, m and x almost surely, then $g(U) = 0$ almost surely.
(ii) For any square-integrable function $g(u)$, if $E\{g(U) \mid Z = z, A = 0, X = x\} = 0$ for any z and x almost surely, then $g(U) = 0$ almost surely.

Completeness is a technical condition that arises in conjunction with sufficiency in the theory of statistical inference. In causal inference, it has been invoked for identification in the context of nonparametric regression with an instrumental variable (Newey & Powell, 2003), where it is used as an analogue of the rank and order conditions that arise in the classical instrumental variable set-up. This assumption essentially means that any variation in U is associated with a form of variation in Z, given $A = 1$, M and X, and given $A = 0$ and X. Further discussion of completeness in this context is given in the Supplementary Material.

We are now in a position to give our first identification result.

THEOREM 1. *Suppose that there exist confounding bridge functions* $h_1(w, M, X)$ *and* $h_0(w, X)$ *that satisfy*

$$E(Y \mid Z, A = 1, M, X) = \int h_1(w, M, X)\, \mathrm{d}F(w \mid Z, A = 1, M, X), \tag{2}$$

$$E\{h_1(W, M, X) \mid Z, A = 0, X\} = \int h_0(w, X)\, \mathrm{d}F(w \mid Z, A = 0, X). \tag{3}$$

Then, under Assumptions 1–6, it follows that

$$E(Y \mid U, A = 1, M, X) = \int h_1(w, M, X)\, \mathrm{d}F(w \mid U, A = 1, M, X), \tag{4}$$

$$E\{h_1(W, M, X) \mid U, A = 0, X\} = \int h_0(w, X)\, \mathrm{d}F(w \mid U, A = 0, X), \tag{5}$$

and furthermore that $E[Y\{1, M(0)\}]$ *is identified as*

$$\psi = \iint h_0(w, x)\, \mathrm{d}F(w \mid x)\, \mathrm{d}F(x). \tag{6}$$

The proof of this result, as well as all others in this paper, is given in the Supplementary Material. We name (6) the *proximal mediation formula*, since it generalizes Pearl's fundamental mediation formula (1) to settings where key confounders are unmeasured. Similar to the proximal g-formula in Tchetgen Tchetgen et al. (2020), (6) is expressed in terms of nested bridge functions. Interestingly, although inferring natural direct and indirect effects involves understanding the effects of the exposure on the mediator and the mediator on the outcome, ψ can be identified without a need for additional proxy variables. Nevertheless, compared with proximal identification of the average causal effect, additional restrictions are placed on the exposure- and outcome-inducing proxies, in terms of their relationship to M. For example, neither Z nor W are allowed to cause M. Such assumptions could be relaxed by collecting data on separate proxies for each bridge function; we will investigate this in future work.

Equations (2) and (3) refer to inverse problems that are known as Fredholm integral equations of the first kind. We refer to Miao et al. (2018) and Cui et al. (2023) for mathematical conditions that ensure that these equations admit solutions. The solutions to these equations are not required to be unique, as all solutions yield a unique value of ψ. The identification assumptions may nevertheless be adjusted in order to guarantee a unique solution. Unlike Assumptions 1–6, the condition that (2) and (3) admit solutions is potentially empirically verifiable.

2.3. *Alternative identification strategies*

In this section, we establish two alternative proximal identification results to the proximal mediation formula, which rely on alternative assumptions regarding completeness and the existence of relevant confounding bridge functions, which are given in the Supplementary Material.

Theorem 2. (i) *Suppose that there exist confounding bridge functions* $h_1(w, M, X)$ *that satisfy* (2) *and* $q_0(z, X)$ *that satisfies*

$$\frac{1}{f(A=0 \mid W, X)} = E\{q_0(Z, X) \mid W, A=0, X\}. \tag{7}$$

Under Assumptions 1–5 and 6(i) above and Assumption A.1.2 in the Supplementary Material, it follows that both

$$\frac{1}{f(A=0 \mid U, X)} = E\{q_0(Z, X) \mid U, A=0, X\} \tag{8}$$

and (4) *hold, and furthermore that* $E[Y\{1, M(0)\}]$ *is identified as*

$$\psi = \iint I(a=0) q_0(z, x) h_1(w, m, x)\, \mathrm{d}F(w, z, a, m \mid x)\, \mathrm{d}F(x). \tag{9}$$

(ii) *Alternatively, suppose that there exist confounding bridge functions* $q_0(z, X)$ *that satisfy* (7) *and* $q_1(z, M, X)$ *that satisfies*

$$E\{q_0(Z, X) \mid W, A=0, M, X\} \frac{f(A=0 \mid W, M, X)}{f(A=1 \mid W, M, X)} = E\{q_1(Z, M, X) \mid W, A=1, M, X\}. \tag{10}$$

Under Assumptions 1–5 and Assumptions A.1.1–A.1.2 in the Supplementary Material, it follows that both

$$E\{q_0(Z,X) \mid U, A=0, M, X\} \frac{f(A=0 \mid U,M,X)}{f(A=1 \mid U,M,X)} = E\{q_1(Z,M,X) \mid U, A=1, M, X\} \quad (11)$$

and (8) *hold, and that* $E[Y\{1, M(0)\}]$ *is identified as*

$$\psi = \iint I(a=1) q_1(z,m,x) y \, \mathrm{d}F(y,z,a,m \mid x) \, \mathrm{d}F(x). \quad (12)$$

We therefore have three results for proximal identification, each of which relies on two confounding bridge assumptions. The strategy given in part (i) of the above theorem relies on a combination of outcome- and treatment-inducing confounding bridge functions, whereas the final strategy relies on two nested treatment-inducing confounding bridge functions. Result (8) follows from Cui et al. (2023), and formal conditions for the existence of solutions to (7) and (10) are given in that paper.

3. Semiparametric inference

3.1. *The semiparametric efficiency bound*

In this section, we consider inference for ψ under the semiparametric model $\mathcal{M}_{\mathrm{sp}}$ that places no restrictions on the observed data distribution besides existence, but not necessarily uniqueness, of bridge functions h_1 and h_0 that solve (2) and (3). Assumed existence of the outcome bridge functions places restrictions on the tangent space. Under the additional regularity conditions described below, we can also obtain the semiparametric efficiency bound under $\mathcal{M}_{\mathrm{sp}}$.

Assumption 7 (*Regularity conditions*).

(i) Let $T_1 : L_2(W,M,X) \to L_2(Z, A=1, M, X)$ denote the operator, given by $T_1(g) \equiv E\{g(W,M,X) \mid Z, A=1, M, X\}$. At the true data generating mechanism, T_1 is surjective.

(ii) Let $T_0 : L_2(W,M,X) \to L_2(Z, A=0, X)$ denote the operator, given by $T_0(g) \equiv E\{g(W,M,X) \mid Z, A=0, X\}$. Then at the true data generating mechanism, T_0 is surjective.

As noted in Ying et al. (2022), this condition relies on the functions $L_2(W,M,X)$ being rich enough such that any element in $L_2(Z, A=1, M, X)$ and $L_2(Z, A=0, X)$ can be generated via the conditional expectation map. Then we arrive at the following result.

Theorem 3. *Assume that there exist bridge functions h_1 and h_0 at all data laws that belong to the semiparametric model $\mathcal{M}_{\mathrm{sp}}$. Furthermore, suppose that at the true data law there exists q_0 and q_1 that solve (7) and (10), and that Assumption 6 holds, such that ψ is unique. Then*

$$\begin{aligned} IF_\psi &= I(A=1) q_1(Z,M,X)\{Y - h_1(W,M,X)\} \\ &\quad + I(A=0) q_0(Z,X)\{h_1(W,M,X) - h_0(W,X)\} + h_0(W,X) - \psi \end{aligned}$$

is a valid influence function for ψ under $\mathcal{M}_{\mathrm{sp}}$. Furthermore, the efficiency bound at the submodel where Assumption 7 holds and all bridge functions are unique is $E(IF_\psi^2)$.

3.2. *Multiply robust estimation*

We consider the setting where L is high dimensional, and parametric working models for h_1, h_0, q_0 and q_1 may be useful as a form of dimension reduction. In that case, we show that an estimator of ψ based on the efficient influence function is multiply robust, in the sense that only certain combinations of these working models need to be correctly specified in order to yield an unbiased estimator. To make this more concrete, consider the following semiparametric models that impose certain restrictions on the observed data distribution:

$$\mathcal{M}_1: h_1(W, M, X) \text{ and } h_0(W, X) \text{ are assumed to be correctly specified,}$$

$$\mathcal{M}_2: h_1(W, M, X) \text{ and } q_0(Z, X) \text{ are assumed to be correctly specified,}$$

$$\mathcal{M}_3: q_1(Z, M, X) \text{ and } q_0(Z, X) \text{ are assumed to be correctly specified.}$$

Here, $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ are all submodels of $\mathcal{M}_{\text{sp}}$. The proposed approach will rely on models for the confounding bridge functions, but we show that only one of $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ needs to hold to ensure unbiased estimation of the target parameter.

We first consider how to obtain inference in the submodels $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$. Let $h_1(W, M, X; \beta_1)$ and $h_0(W, X; \beta_0)$ denote models for the respective bridge functions $h_1(W, M, X)$ and $h_0(W, X)$, indexed by finite-dimensional parameters β_1 and β_0. Likewise, $q_1(Z, M, X; \gamma_1)$ and $q_0(Z, X; \gamma_0)$ are models for the bridge functions $q_1(Z, M, X)$ and $q_0(Z, X)$ indexed by the finite-dimensional parameters γ_1 and γ_0, respectively. It follows from Cui et al. (2023) that one can obtain estimates $\hat{\beta}_1$, $\hat{\beta}_0$ and $\hat{\gamma}_0$ of β_1, β_0 and γ_0 as the solutions to the respective estimating equations:

$$\sum_{i=1}^{n} A_i\{Y_i - h_1(W_i, M_i, X_i; \beta_1)\}c_1(Z_i, M_i, X_i) = 0,$$

$$\sum_{i=1}^{n}(1 - A_i)\{h_1(W_i, M_i, X_i; \beta_1) - h_0(W_i, X_i; \beta_0)\}c_0(Z_i, X_i) = 0,$$

$$\sum_{i=1}^{n}\{(1 - A_i)q_0(Z_i, X_i; \gamma_0) - 1\}d_0(W_i, X_i) = 0.$$

The first two sets of equations can be solved sequentially. Here $c_1(Z_i, M_i, X_i)$ is a function of the same dimension as β_1, and $c_0(Z_i, X_i)$ and $d_0(W_i, X_i)$ are similarly defined. Although $\hat{\beta}_0$ and hence $h_0(W, X; \hat{\beta}_0)$ depends on $\hat{\beta}_1$, this dependence is suppressed to simplify notation. The above estimating equations can be implemented using software for generalized method of moments or, when models are linear, two-stage least squares. Interestingly, despite the fact that (10) suggests that estimation of γ_0 would require postulation of a model for $1/f(A = 0 \mid W, X)$, Cui et al. (2023) showed that this is not the case. The efficient choices of $c_1(Z_i, M_i, X_i)$, $c_0(Z_i, X_i)$ and $d_0(W_i, X_i)$ are all implied by results in the Appendix of Cui et al. (2023). Since the resulting efficiency gain compared to using the choices $c_1(Z_i, M_i, X_i) = (1, Z_i^{\mathrm{T}}, M_i, X_i^{\mathrm{T}})^{\mathrm{T}}$, $c_0(Z_i, X_i) = (1, Z_i^{\mathrm{T}}, X_i^{\mathrm{T}})^{\mathrm{T}}$ and $d_0(W_i, X_i) = (1, W_i^{\mathrm{T}}, X_i^{\mathrm{T}})^{\mathrm{T}}$ is likely to be modest in most situations (Stephens et al., 2014), we do not consider locally efficient estimation of the nuisance parameters any further.

Since q_1 involves solving an integral equation (10) involving the ratio of propensity score functions, the results from previous work do not extend to inference for γ_1. The following theorem then suggests how to obtain semiparametric inference under model $\mathcal{M}_3$, and is more generally relevant for estimation of the average treatment effect in the treated population.

THEOREM 4. *All influence functions of regular and asymptotically linear estimators of γ_1 under the semiparametric model $\mathcal{M}_3$ are of the form*

$$-V^{-1}\Bigg[\{Aq_1(Z,M,X;\gamma_1)-(1-A)q_0(Z,X;\gamma_0)\}d_1(W,M,X) - E\left\{\frac{\partial q_0(Z,X;\gamma_0)}{\partial\gamma_0}(1-A)d_1(W,M,X)\right\}\varphi(W,Z,A,X;\gamma_0)\Bigg]$$

for some function $d_1(W,M,X)$ that is of the same dimension as γ_1, where

$$V = E\left\{\frac{\partial q_1(Z,M,X;\gamma_1)}{\partial\gamma_1}Ad_1(W,M,X)\right\}$$

and $\varphi(W,Z,A,X;\gamma_0)$ is the influence function for an estimator of γ_0.

This theorem indicates that inference for γ_1 can be obtained without the need to model either $f(A=0 \mid W,M,X)$ or $f(A=1 \mid W,M,X)$, or their ratio. Indeed, it suggests an estimation strategy for γ_1; namely, after obtaining $\hat{\gamma}_0$ as previously described, one can obtain $\hat{\gamma}_1$ as the solution to the equations

$$\sum_{i=1}^{n}\{A_i q_1(Z_i,M_i,X_i;\gamma_1)-(1-A_i)q_0(Z_i,X_i;\hat{\gamma}_0)\}(1,W_i^{\mathrm{T}},M_i,X_i^{\mathrm{T}})^{\mathrm{T}}=0.$$

Consistent estimation of γ_1 nevertheless relies on consistent estimation of γ_0. Given that $q_1(Z,M,X;\gamma_1)$ and $q_0(Z,X;\gamma_0)$ are both confounding bridges for the treatment assignment mechanism, this raises the question of how to postulate models for the two bridge functions that are compatible. A brief discussion on model compatibility is given in §4 below, with more detailed results given in the Supplementary Material.

Once we have strategies for estimating nuisance parameters indexing the bridge functions, one can construct proximal outcome regression, hybrid and inverse probability weighting estimators of ψ:

$$\hat{\psi}_{\text{P-OR}} = \frac{1}{n}\sum_{i=1}^{n}h_0(W_i,X_i;\hat{\gamma}_0),$$

$$\hat{\psi}_{\text{P-hybrid}} = \frac{1}{n}\sum_{i=1}^{n}(1-A_i)q_0(Z_i,X_i;\hat{\gamma}_0)h_1(W_i,M_i,X_i;\hat{\beta}_1),$$

$$\hat{\psi}_{\text{P-IPW}} = \frac{1}{n}\sum_{i=1}^{n}A_i q_1(Z_i,M_i,X_i;\hat{\gamma}_1)Y_i.$$

Then $\hat{\psi}_{\text{P-OR}}$ is a consistent and asymptotically normal estimator under model $\mathcal{M}_1$, $\hat{\psi}_{\text{P-hybrid}}$ is consistent and asymptotically normal under model $\mathcal{M}_2$, and $\hat{\psi}_{\text{P-IPW}}$ is consistent and asymptotically normal under model $\mathcal{M}_3$. Correctly specifying models for the different bridge functions may be challenging, since they are defined as solutions to integral equations, rather than the conditional expectations or probabilities more common in causal inference, e.g., $E(Y \mid A=1, L)$ or $f(A=1 \mid L)$. The development of a proximal multiply robust estimator, which enables the relaxation of parametric modelling assumptions, is therefore of interest.

Theorem 5. *Under typical regularity conditions,*

$$\hat{\psi}_{\text{P-MR}} = \frac{1}{n}\sum_{i=1}^{n} A_i q_1(Z_i, M_i, X_i; \hat{\gamma}_1)\{Y_i - h_1(W_i, M_i, X_i; \hat{\beta}_1)\}$$
$$+ (1 - A_i) q_0(Z_i, X_i; \hat{\gamma}_0)\{h_1(W_i, M_i, X_i; \hat{\beta}_1) - h_0(W_i, X_i; \hat{\gamma}_0)\} + h_0(W_i, X_i; \hat{\gamma}_0)$$

is a consistent and asymptotically normal estimator of ψ *under the union model* $\mathcal{M}_{\text{union}} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$*. Furthermore,* $\hat{\psi}_{\text{P-MR}}$ *attains the semiparametric efficiency bound at the intersection submodel* $\mathcal{M}_1 \cap \mathcal{M}_2 \cap \mathcal{M}_3$ *where Assumption 7 also holds.*

Using standard M-estimation arguments, and the influence functions for the nuisance parameter estimators, one can construct a nonparametric sandwich estimator of the standard error for $\hat{\psi}_{\text{P-MR}}$ that is robust to potential model misspecification; an alternative option is the nonparametric bootstrap. A weakness of our estimator is that, when Y is binary, $\hat{\psi}_{\text{P-MR}}$ is not guaranteed to fall within the [0,1] interval. This is an important topic for future work and could be remedied, e.g., by adapting the proposal in § 5 of Tchetgen Tchetgen & Shpitser (2012).

4. Simulation studies

In order to evaluate the finite sample performance of the proposed estimators, we conducted a simulation study. Specifically letting we generated data (Y, A, M, X, W, Z) by $X, U \sim MVN\{(0.25, 0.25, 0)^{\mathrm{T}}, \Sigma\}$; $f(A=1 \mid X, U) = \text{expit}\{-(0.5, 0.5)^{\mathrm{T}}X - 0.4U\}$; $Z \mid A, X, U \sim \mathcal{N}\{0.2 - 0.52A + (0.2, 0.2)^{\mathrm{T}}X - U, 1\}$; $W \mid X, U \sim \mathcal{N}\{0.3 + (0.2, 0.2)^{\mathrm{T}}X - 0.6U, 1\}$ and $M \mid A, X, U \sim \mathcal{N}\{-0.3A - (0.5, 0.5)^{\mathrm{T}}X + 0.4U, 1\}$, where $X = (X_1, X_2)^{\mathrm{T}}$ and

$$\Sigma = \begin{pmatrix} \sigma^2_{x_1} & \sigma_{x_1x_2} & \sigma_{x_1u} \\ \sigma_{x_1x_2} & \sigma^2_{x_2} & \sigma_{x_2u} \\ \sigma_{x_1u} & \sigma_{x_2u} & \sigma^2_u \end{pmatrix} = \begin{pmatrix} 0.25 & 0 & 0.05 \\ 0 & 0.25 & 0.05 \\ 0.05 & 0.05 & 1 \end{pmatrix}.$$

Finally, $Y = 2 + 2A + M + 2W - (1,1)^{\mathrm{T}}X - U + 2\epsilon^*$, where $\epsilon^* \sim \mathcal{N}(0,1)$. Since W and M are linear in X, U and the exposure, it follows that this data generating mechanism is compatible with the following models for h_1, h_0:

$$h_1(W, M, X; \beta_1) = \beta_{1,0} + \beta_{1,w}W + \beta^{\mathrm{T}}_{1,x}X + \beta_{1,m}M,$$
$$h_0(W, X; \beta_0) = \beta_{0,0} + \beta_{0,w}W + \beta^{\mathrm{T}}_{0,x}X.$$

Furthermore, in the Supplementary Material, we also show that, when A follows a Bernoulli distribution, given U and X, M is normally distributed, given U, A and X and $Z \sim \mathcal{N}(\epsilon_0 + \epsilon_u U + \epsilon_a A + \epsilon_x X, \sigma^2_{z|u,a,x})$; then the choice of bridge functions

$$q_0(Z, X) = 1 + \exp\{-(\gamma_{0,0} + \gamma_{0,z}Z + \gamma_{0,x}X)\},$$

Table 1. *Simulation results from experiments 1–4*

Experiment	Estimator	Bias	MSE	Coverage	Mean length	Med. length
1	$\hat{\theta}_{\text{P-IPW}}$	0.00	0.02	0.95	0.51	0.50
	$\hat{\theta}_{\text{P-hybrid}}$	0.00	0.02	0.96	0.50	0.50
	$\hat{\theta}_{\text{P-OR}}$	0.00	0.02	0.96	0.50	0.50
	$\hat{\theta}_{\text{P-MR}}$	0.00	0.02	0.95	0.51	0.50
2	$\hat{\theta}_{\text{P-IPW}}$	0.19	0.05	0.71	0.52	0.51
	$\hat{\theta}_{\text{P-hybrid}}$	−0.15	0.04	0.82	0.52	0.52
	$\hat{\theta}_{\text{P-OR}}$	0.00	0.02	0.95	0.50	0.50
	$\hat{\theta}_{\text{P-MR}}$	0.00	0.02	0.95	0.50	0.50
3	$\hat{\theta}_{\text{P-IPW}}$	0.38	0.16	0.30	0.60	0.59
	$\hat{\theta}_{\text{P-hybrid}}$	0.00	0.02	0.95	0.50	0.50
	$\hat{\theta}_{\text{P-OR}}$	−0.14	0.04	0.81	0.52	0.51
	$\hat{\theta}_{\text{P-MR}}$	0.00	0.02	0.95	0.51	0.50
4	$\hat{\theta}_{\text{P-IPW}}$	−0.01	0.02	0.94	0.51	0.51
	$\hat{\theta}_{\text{P-hybrid}}$	0.21	0.06	0.66	0.53	0.53
	$\hat{\theta}_{\text{P-OR}}$	0.17	0.05	0.72	0.52	0.51
	$\hat{\theta}_{\text{P-MR}}$	−0.01	0.02	0.95	0.51	0.51

Bias, Monte Carlo bias; MSE, mean squared error; Coverage, 95% confidence interval coverage; Mean length, average 95% confidence interval length; Med. length, median 95% confidence interval length.

$$q_1(Z, M, X) = \exp(\gamma_{1,0} + \gamma_{1,z}Z + \gamma_{1,m}M + \gamma_{1,x}X) + \{q_0(Z, X) - 1\} \exp\{\gamma_{1,0} + \gamma_{0,z}\epsilon_a + \gamma_{1,z}(Z + \gamma_{0,z}\sigma^2_{z|u,a,x}) + \gamma_{1,m}M + \gamma_{1,x}X\}$$

satisfies (7) and (10). Under the additional constraint that $\epsilon_a = -\gamma_{1,z}\sigma^2_{z|u,a,x}$, which is enforced here, it follows that the expression for q_1 simplifies to

$$q_1(Z, M, X) = q_0(Z, X) \exp(\gamma_{1,0} + \gamma_{1,z}Z + \gamma_{1,z}M + \gamma_{1,x}X).$$

Let $\hat{\delta}_{\text{P-DR}}$ be the proximal doubly robust estimator of $E\{Y(0)\}$ considered in Cui et al. (2023). We considered four proximal mediation estimators of the natural direct effect $E\{Y(1, M(0))\} - E\{Y(0)\}$: $\hat{\theta}_{\text{P-OR}} = \hat{\psi}_{\text{P-OR}} - \hat{\delta}_{\text{P-DR}}$, $\hat{\theta}_{\text{P-hybrid}} = \hat{\psi}_{\text{P-hybrid}} - \hat{\delta}_{\text{P-DR}}$, $\hat{\theta}_{\text{P-IPW}} = \hat{\psi}_{\text{P-IPW}} - \hat{\delta}_{\text{P-DR}}$ and $\hat{\theta}_{\text{P-MR}} = \hat{\psi}_{\text{P-MR}} - \hat{\delta}_{\text{P-DR}}$. We evaluated the performance of the proposed settings in settings where either all bridge functions were correctly modelled (experiment 1), q_1 and q_0 were misspecified (experiment 2), q_1 and h_0 were misspecified (experiment 3) or h_1 and h_0 were misspecified (experiment 4). We misspecified the models by including $|X_1|^{1/2}$ and $|X_2|^{1/2}$ in the bridge function rather than X. We conducted simulations at $n = 2000$ and repeated each experiment 1000 times.

The results are given in Table 1. As a benchmark, we also considered a naïve nonproximal estimator $\hat{\theta}_{\text{OLS}}$ of the direct effect, based on linearly regressing Y on A, M, Z, W and X; its Monte Carlo bias was 0.5, with 95% confidence intervals that included the true value only 29% of the time. When all bridge functions were correctly specified, the different proximal estimators had comparable performance, with $\hat{\theta}_{\text{P-OR}}$ and $\hat{\theta}_{\text{P-MR}}$ exhibiting slightly greater efficiency compared with the other methods. Across the different mechanisms of misspecification considered, we see that only the multiply robust estimator continues to have low bias, with confidence intervals that possess approximately their advertised coverage.

In a further set of simulation studies, we considered sensitivity of the different proposals to changes in the confounding mechanism as well as violations of the structural assumptions that underpin the methods. First, in experiment 5 we changed the above data-generating mechanism such that $f(A = 1 \mid X, U)$, $f(M \mid A, X, U)$ and $f(Y \mid A, M, W, X, U)$ no longer depended on U, to check how the methods performed when there was no unmeasured confounding. In this specific setting, to construct the benchmark estimator $\hat{\theta}_{\text{OLS}}$, we adjusted for A, M, X and W, but not Z, to avoid collider bias induced via an association between A and U. In experiment 6, we considered violations of the exclusion restriction Assumption 5.2, by generating $Y = 2 + 2A + M + 2W - (1, 1)^{\mathrm{T}}X - U - 0.5Z + 2\epsilon^{*}$. In experiment 7, we considered violations of the exclusion restriction Assumption 5.4, by generating W from $W \mid X, U, A \sim \mathcal{N}(0.3 + (0.2, 0.2)^{\mathrm{T}}X - 0.6U + 0.2A, 1)$. In experiment 8, we looked at a near violation of the completeness conditions, where the coefficient for U in the model for $E(W \mid X, U)$ was reduced in strength to 0.05, such that W was only weakly U-relevant and the conditional associations between W and Z were also weak. In experiment 9, we reversed this such that Z was now only weakly U-relevant. Each of the models used to construct the bridge functions included X, rather than the transformations considered in experiments 2–4. Values of the parameters in the data-generating mechanism were adjusted where necessary, to ensure that all bridge functions were correctly specified. The results of experiments 5–9 are given in the Supplementary Material. We see that in settings where conditional exchangeability holds, the proximal estimators perform similarly in terms of bias compared with $\hat{\theta}_{\text{OLS}}$, but displayed a decrease in efficiency. When the exclusion restrictions are violated, the proximal estimators performed similarly or slightly worse than the naïve, biased ordinary least-squares estimator. When Z or W are not U-relevant, we see that the standard errors for the proximal estimators can dramatically inflate; this is unsurprising, given how common instrumental variable estimators perform when instruments are weak/irrelevant. The proximal estimators also typically displayed considerably larger bias than $\hat{\theta}_{\text{OLS}}$, but smaller median bias. The multiply robust estimator $\hat{\theta}_{\text{P-MR}}$ generally displayed smaller mean and median bias compared with the other methods in experiments 8 and 9.

5. Data analysis

In the Job Corps study, participants were randomized from November 1994 to February 1996 either to treatment, i.e., access to the Job Corps program, or to control, i.e., no access. However, since individuals could choose whether to participate in the program or not, we treat the exposure of interest as nonrandomized. The outcome of interest was the number of arrests in the fourth year after assignment, and the mediator of interest was the percentage of weeks employed in the second year. Although data on a rich set of covariates were measured at baseline, it was nevertheless possible that unmeasured confounding could lead to biased estimates of both direct and indirect effects. Investigators collected information on factors that were known to be associated with duration in the program, e.g., expectations of the program, interactions with recruiters. Huber et al. (2020) noted that such variables may be strongly correlated with motivation, considered as an important latent source of confounding. They therefore adjusted for these variables in the analysis in the same way as standard measured confounders. In contrast, we treated these variables as proxies of an unmeasured confounder, and changed the analysis accordingly. Similar to Tchetgen Tchetgen et al. (2020), we restricted consideration to four potential proxies strongly correlated with the expose and/or the outcome: time being spent spoken to by the recruiter, worried about the Job Corps program, expected improvement in social skills

Table 2. *Results from the analysis of the Job Corps study*

Estimand	S-MR	95% CI	P-MR	95% CI
$E[Y\{1, M(0)\}] - [Y\{0, M(0)\}]$	−0.0107	−0.0341, 0.0128	0.0057	−0.0699, 0.0813
$E[Y\{1, M(1)\}] - [Y\{0, M(1)\}]$	−0.0110	−0.0345, 0.0124	−0.0016	−0.0781, 0.0749
$E[Y\{1, M(1)\}] - [Y\{1, M(0)\}]$	0.0006	−0.0229, 0.0241	−0.0088	−0.0314, 0.0138
$E[Y\{0, M(1)\}] - [Y\{0, M(0)\}]$	0.0009	−0.0001, 0.0019	−0.0016	−0.0110, 0.0078

S-MR, standard multiply robust estimator; P-MR, multiply robust estimators of $E[Y\{1, M(0)\}]$ and $E[Y\{0, M(1)\}]$; CI, confidence interval.

and whether the first contact by the recruiter was in the office or not. Since there was no *a priori* understanding as to whether these were candidates for Z or W, we used the algorithm described by Tchetgen Tchetgen et al. (2020) to assign them. This led to the first two being used for Z, and the second two used for W.

Our sample consisted of 10 775 participants; further information on the sample is given in Huber et al. (2020). There were 257 participants with missing data on Z who were removed from the analysis. Linear and logistic models were postulated for the outcome and exposure bridge functions, respectively. We considered both proximal inverse probability weighting, hybrid, outcome regression and multiply robust estimators of $E[Y\{1, M(0)\}]$ and $E[Y\{0, M(1)\}]$. We contrasted our estimator with a standard multiply robust approach $\hat{\theta}_{\text{S-MR}}$ based on the efficient influence function derived by Tchetgen Tchetgen & Shpitser (2012), that is valid under conditional exchangeability-type assumptions; we again postulated linear models for $E(Y \mid A = a, M, X, Z, W)$ and $E\{E(Y \mid A = a, M, X, Z, W) \mid A = a^*, X, Z, W\}$, and a logistic model for the odds $P(A = 1 \mid M, X, Z, W)/P(A = 0 \mid M, X, Z, W)$ and for $P(A = 1 \mid X, Z, W)$. All outcome regression models for the confounding bridge functions and otherwise were fitted separately in control and treatment groups, to allow for treatment-mediator and treatment-covariate interactions. Since the set of covariates measured at baseline was relatively high in dimension, we excluded variables with amounts of missingness $> 50\%$, or which had some missing values and were highly correlated with variables that were fully observed. For all other variables with missingness, we used the missing indicator method as in Huber et al. (2020). Standard errors for all estimators were calculated using sandwich estimators. In the Supplementary Material, following the AGReMA statement on good practice for conducting and reporting mediation analysis (Lee et al., 2021), we provide further information on the study and data analysis.

Results for the standard and proximal multiply robust approaches can be seen in Table 2. The total effect estimate given by the standard approach was −0.01 (95% confidence interval: −0.022, 0.002) and for the proximal approach, it was −0.003 (95% confidence interval: −0.040, 0.033). The estimates of the direct and indirect effects yielded by both approaches are also close to the null, and all 95% confidence intervals contain the null. The direct effect estimates under the proximal approach tended to be closer to the null; and the indirect effects were slightly larger in magnitude, although still very small. The results of the other estimators can be found in the Supplementary Material; the multiply robust estimator $\hat{\theta}_{\text{P-MR}}$ tended to agree more closely with $\hat{\theta}_{\text{P-hybrid}}$, although there was not a large disparity between the point estimates.

6. Discussion

An advantage of doubly/multiply robust methods, used in combination with cross-fitting, is that data-adaptive methods can be used to estimate nuisance parameters, yet their potentially slow rates of convergence are not necessarily inherited by the estimator of the target parameter (Chernozhukov et al., 2018). A complication in proximal learning is that the nuisances are defined as the solutions to integral equations. Progress in this direction is described in Ghassami et al. (2022) and Kallus et al. (2022); it would thus be useful to extend these ideas to mediation analysis. Another avenue for future work would be to extend the results of identification and estimation to more general path-specific effects (Avin et al., 2005; Shpitser, 2013), which are relevant in particular in settings when confounders of the mediator-outcome relationship are affected by the exposure. In such cases the cross-world assumption fails to hold, and standard natural effects are no longer identified. Finally, an important topic more generally in proximal learning is the development of sensitivity analysis methods. A simple way to check how sensitive results are to categorization of the Z and W proxies is to permute the labels. The development of more advanced tools to assess deviations from specific key assumptions, e.g., the exclusion restrictions involving the proxies, is left to future work. Under the failure of certain assumptions, methods for partial identification such as nonparametric bounds may also be useful (Robins, 1989; Manski, 1990).

Acknowledgement

Dukes gratefully acknowledges support from the Ghent University Special Research Fund and the Research Foundation Flanders. Shpitser and Tchetgen Tchetgen gratefully acknowledge support from the National Institutes of Health.

Supplementary material

The Supplementary Material includes proofs of the theorems and additional simulation results.

References

Avin, C., Shpitser, I. & Pearl, J. (2005). Identifiability of path-specific effects. In *Proc. 19th Int. Joint Conf. Artif. Intel.*, pp. 357–63. San Francisco, CA: Morgan Kaufmann.

Cheng, L., Guo, R. & Liu, H. (2022). Causal mediation analysis with hidden confounders. *arXiv:* 2102.11724v3.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Economet. J.* **21**, C1–68.

Cui, Y., Pu, H., Shi, X., Miao, W. & Tchetgen Tchetgen, E. (2023). Semiparametric proximal causal inference. *arXiv:* 2011.08411v4.

Ghassami, A., Ying, A., Shpitser, I. & Tchetgen, E. T. (2022). Minimax kernel machine learning for a class of doubly robust functionals with application to proximal causal inference. In *Proc. 25th Int. Conf. Artif. Intel. Statist.*, vol. 151, pp. 7187–209. Proceedings of Machine Learning Research.

Huber, M., Hsu, Y.-C., Lee, Y.-Y. & Lettry, L. (2020). Direct and indirect effects of continuous treatments based on generalized propensity score weighting. *J. Appl. Economet.* **35**, 814–40.

Imai, K., Keele, L. & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychol. Meth.* **15**, 309.

Kallus, N., Mao, X. & Uehara, M. (2022). Causal inference under unmeasured confounding with negative controls: a minimax learning approach. *arXiv:* 2103.14029v4.

Kuroki, M. & Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika* **101**, 423–37.

Lee, H., Cashin, A. G., Lamb, S. E., Hopewell, S., Vansteelandt, S., VanderWeele, T. J., MacKinnon, D. P., Mansell, G., Collins, G. S., Golub, R. M. et al. (2021). A guideline for reporting mediation analyses of randomized trials and observational studies: the agrema statement. *J. Am. Med. Assoc.* **326**, 1045–56.

Lipsitch, M., Tchetgen Tchetgen, E. & Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* **21**, 383–8.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *Am. Econ. Rev.* **80**, 319–23.

Miao, W., Geng, Z. & Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika* **105**, 987–93.

MIAO, W., SHI, X. & TCHETGEN TCHETGEN, E. (2020). A confounding bridge approach for double negative control inference on causal effects. *arXiv:* 1808.04945v3.

NEWEY, W. K. & POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71**, 1565–78.

PEARL, J. (2001). Direct and indirect effects. In *Proc. 17th Conf. Uncertainty Artif. Intel.*, pp. 411–20. San Francisco, CA: Morgan Kaufmann.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Mod.* **7**, 1393–512.

ROBINS, J. M. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*, pp. 113–159. U.S. Department of Health and Human Services.

ROBINS, J. M. & GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3**, 143–55.

ROBINS, J. M. & RICHARDSON, T. S. (2010). Alternative graphical causal models and the identification of direct effects. In *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*, pp. 103–58. New York: American Psychopathological Association.

ROBINS, J. M., RICHARDSON, T. S. & SHPITSER, I. (2021). An interventionist approach to mediation analysis. *arXiv:* 2008.06019v2.

SCHOCHET, P. Z., BURGHARDT, J. & MCCONNELL, S. (2008). Does job corps work? impact findings from the national job corps study. *Am. Econ. Rev.* **98**, 1864–86.

SHI, X., MIAO, W. & TCHETGEN TCHETGEN, E. (2020). A selective review of negative control methods in epidemiology. *Curr. Epidemiol. Rep.* **7**, 190–202.

SHPITSER, I. (2013). Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cogn. Sci.* **37**, 1011–35.

STEPHENS, A., TCHETGEN TCHETGEN, E. & DE GRUTTOLA, V. (2014). Locally efficient estimation of marginal treatment effects when outcomes are correlated: is the prize worth the chase? *Int. J. Biostatist.* **10**, 59–75.

TCHETGEN TCHETGEN, E. J. & SHPITSER, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Ann. Statist.* **40**, 1816.

TCHETGEN TCHETGEN, E. J., YING, A., CUI, Y., SHI, X. & MIAO, W. (2020). An introduction to proximal causal learning. *arXiv:* 2009.10982v1.

VANDERWEELE, T. J. & VANSTEELANDT, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statist. Interface* **2**, 457–68.

YING, A., MIAO, W., SHI, X. & TCHETGEN, E. J. T. (2022). Proximal causal inference for complex longitudinal studies. *arXiv:* 2109.07030v5.

[*Received on* 1 *October* 2021. *Editorial decision on* 14 *February* 2023]