Biostatistics (2022) **23**, 3, *pp.* 789–806
doi: 10.1093/biostatistics/kxaa057
Advance Access publication on February 2, 2021

Efficiently transporting causal direct and indirect effects to new populations under intermediate confounding and with multiple mediators

KARA E. RUDOLPH* and IVÁN DÍAZ

Department of Epidemiology, Mailman School of Public Health, Columbia University; and Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA
kr2854@cumc.columbia.edu

Summary

The same intervention can produce different effects in different sites. Existing transport mediation estimators can estimate the extent to which such differences can be explained by differences in compositional factors and the mechanisms by which mediating or intermediate variables are produced; however, they are limited to consider a single, binary mediator. We propose novel nonparametric estimators of transported interventional (in)direct effects that consider multiple, high-dimensional mediators and a single, binary intermediate variable. They are multiply robust, efficient, asymptotically normal, and can incorporate data-adaptive estimation of nuisance parameters. They can be applied to understand differences in treatment effects across sites and/or to predict treatment effects in a target site based on outcome data in source sites.

Keywords: Interventional indirect effect; Non-parametric methods; Mediation; Stochastic indirect effect; Targeted learning.

1. Introduction

The same intervention can produce different effects in different populations (e.g., Orr *and others*, 2003; Miller, 2015; Arnold *and others*, 2018). Different effects could arise from differences in (i) the distribution of compositional factors that modify aspects of the intervention's effectiveness (e.g., gender, age), (ii) probability take-up or degree of adherence to the intervention, (iii) the mechanism by which important mediating or intermediate variables are produced, and/or (iv) the mechanism by which the outcome is produced in different populations, including different population- or site-level contextual variables that are predictive of the outcome (Pearl and Bareinboim, 2018). Transportability has been defined by Pearl and Bareinboim (2018) as the "license to transfer causal information learned in experimental studies to a different environment." Previously, we proposed using the transport graphs of Pearl and Bareinboim (2018) coupled with a transport estimator that predicts effects "transported" to a target population as a tool for quantitatively examining the extent to which differences in effect estimates between sites could be explained by factors (i)–(iii) above (Rudolph *and others*, 2017, 2020). In this previous work, we

*To whom correspondence should be addressed.

developed an efficient and robust semi-parametric estimator of transported interventional (also called stochastic, see Rudolph *and others*, 2020) direct and indirect (what we refer to as (in)direct) effects in a target population. Although this previous estimator accounted for the presence of intermediate variables (those affected by treatment/exposure that could affect downstream mediator and outcome variables), it was limited in that it could only consider binary versions of a treatment/exposure variable, intermediate variable, and mediator variable and assumed that the distribution of the mediator was known (Rudolph *and others*, 2020). To our knowledge, it is currently the only available estimator for transporting (in)direct effects. However, many research questions involve continuous and/or multiple mediator variables. Thus, we address this methodologic gap by proposing novel nonparametric estimators of transported interventional (in)direct effects that allow for multiple, possibly high-dimensional mediators without constraints on their distributions.

To motivate this work, we consider an illustrative research question from the Moving to Opportunity study (MTO), a multi-site randomized controlled trial conducted by the US Department of Housing and Urban Development, where families living in high-rise public housing were randomized to receive a Section 8 housing voucher that they could use to move to a rental on the private market (Sanbonmatsu *and others*, 2011). Families were followed up at two subsequent time points with the final time point occurring 10–15 years after randomization. In this study, some unintended harmful effects on children's mental health, substance use, and risk behavior outcomes were documented (Sanbonmatsu *and others*, 2011), and these overall effects were partially mediated by aspects of the peer and school environments (Rudolph *and others*, 2018b). However, these unintended harmful effects and their indirect effect components were not universal across sites (Rudolph *and others*, 2018a, 2020). To illustrate our proposed methods, we use the transportability framework and our novel estimators to shed light on possible reasons why the intervention had harmful effects in some sites, particularly in Chicago, but not in others. For example, if we take Chicago as the site we would like to transport to, then we borrow information from the remaining sites to learn the outcome model, we can predict the effect for Chicago, standardizing based on the covariates, intermediate and mediating variables.

Putting the above in more general terms: our approach to estimate transported interventional (in)direct effects involves (i) borrowing information from the source population about the conditional distribution of the outcome given the mediating variables, intermediate confounding variables, treatment, and covariates, and (ii) using data from the target population for the distributions of the mediating variables, intermediate confounding variables, treatment, and covariates to get estimates using the outcome model that are essentially standardized to the target population. The utility of borrowing or transporting information across sites applies more broadly than the above MTO example. It applies to questions that seek to: (i) understand differences in treatment, policy, or intervention effects across sites in multi-site trials or cohort studies, or to (ii) predict treatment effects in a target site based on outcome data in source sites. This article is organized as follows. In Section 2, we introduce notation, define the structural causal models we consider, and define and identify the transported interventional (in)direct effects. In Section 3, we describe the efficient influence function (EIF), including a reparameterization that allows for estimation with multiple and/or continuously distributed mediators, and derive the robustness properties of the EIF. In Section 4, we describe two efficient estimators for the transported interventional (in)direct effects, based on the EIF derived in Section 3: an estimator that solves the EIF in one step and a targeted minimum loss-based estimator (TMLE). In Section 5, we present results from a simulation study in which we demonstrate the consistency, efficiency, and robustness of the two estimators across various scenarios. In Section 6, we apply the two estimators to estimate the transported indirect effects of housing voucher receipt on subsequent behavioral problems as adolescents among girls in Chicago, operating through aspects of the school environment, borrowing information from the other MTO sites. Section 7 concludes the manuscript.

2. Notation and definition of (in)direct effects

Let $O = (S, W, A, Z, M, SY)$ represent the observed data, where S denotes a binary variable indicating membership in the source population ($S = 1$) or target population ($S = 0$), W denotes a vector of observed pretreatment covariates, A denotes a categorical treatment variable, Z denotes an intermediate variable (a mediator-outcome confounder affected by treatment), M denotes a multivariate mediator, and Y denotes a continuous or binary outcome. Let $O_1, \ldots, O_n$ denote a sample of n i.i.d. observations of O. Note that the outcome is only observed for the source population/sites, $S = 1$, but we are interested in estimating effects for the target population/site, $S = 0$. We formalize the definition of our counterfactual variables using the following nonparametric structural equation model (NPSEM, Pearl, 2009) though equivalent methods may be developed by taking the counterfactual variables as primitives (Rubin, 1974). Assume the data-generating process satisfies:

$$S = f_S(U_S);\ W = f_W(S, U_W);\ A = f_A(S, W, U_A);\ Z = f_Z(S, W, A, U_Z);$$
$$M = f_M(S, W, A, Z, U_M);\ Y = f_Y(W, A, Z, M, U_Y). \quad (2.1)$$

Here, $U = (U_S, U_W, U_A, U_Z, U_M, U_Y)$ is a vector of exogenous factors, and the functions f are assumed deterministic but unknown. We use P to denote the distribution of O. We let P be an element of the nonparametric statistical model defined as all continuous densities on O with respect to some dominating measure ν. Let p denote the corresponding probability density function. We denote random variables with capital letters and realizations of those variables with lowercase letters. We define $\mathsf{P}f = \int f(o)\mathrm{d}\mathsf{P}(o)$ for a given function $f(o)$.

We use the following additional definitions. The function $\mathsf{c}(a, z, m, w)$ denotes $\mathsf{P}(S = 1 \mid A = a, Z = z, M = m, W = w)$, $\mathsf{g}(a \mid w, S = 0)$ denotes $\mathsf{P}(A = a \mid W = w, S = 0)$, $\mathsf{e}(a \mid m, w, S = 0)$ denotes $\mathsf{P}(A = a \mid M = M, W = w, S = 0)$, $\mathsf{q}(z \mid a, w, S = 0)$ denotes the density of Z conditional on $(A, W, S) = (a, w, 0)$, $\mathsf{r}(z \mid a, m, w, S = 0)$ denotes the density of Z conditional on $(A, M, W, S) = (a, m, w, 0)$, $\mathsf{b}(a, z, m, w, S = 1)$ denotes $\mathsf{E}(Y \mid A = a, Z = z, M = m, W = w, S = 1)$, and t denotes $\mathsf{P}(S = 0)$. Let X_a denote the counterfactual outcome observed in a hypothetical world in which $\mathsf{P}(A = a) = 1$. For example, we have $Z_a = f_Z(S, W, a, U_Z)$, $M_a = f_M(S, W, a, Z_a, U_M)$, and $Y_a = f_Y(W, a, Z_a, M_a, U_Y)$. Likewise, we let $Y_{a,m} = f_Y(W, a, Z_a, m, U_Y)$ denote the value of the outcome in a hypothetical world where $\mathsf{P}(A = a, M = m) = 1$.

2.1. *Transported interventional (in)direct effects*

We define the total effect of A on Y in the target population $S = 0$ in terms of a contrast between two user-given values $a', a^\star \in \mathcal{A}$ among those for whom $S = 0$. The total effect can be decomposed into the natural direct and indirect effects. However, natural direct and indirect effects are not generally identified in the presence of a mediator-outcome confounder affected by treatment (Z, using our notation above) (Avin *and others*, 2005; Tchetgen Tchetgen and VanderWeele, 2014). Direct and indirect effects may be alternatively defined considering a stochastic intervention on the mediator (Petersen *and others*, 2006; van der Laan and Petersen, 2008; Zheng and van der Laan, 2012; VanderWeele *and others*, 2014; Rudolph *and others*, 2017). Let G_a denote a random draw from the conditional distribution of M_a conditional on (S, W). The interventional indirect effect (also called randomized interventional indirect effect) among those for whom $S = 0$ can be written: $\mathsf{E}(Y_{a', G_{a'}} - Y_{a', G_{a^\star}} \mid S = 0)$. Generally speaking, this is the average effect of A on Y that operates through M in the target population. Specifically, it is the average difference in expected outcomes setting $A = a'$ and stochastically drawing M from the counterfactual joint distribution of mediator values, conditional on W, in a hypothetical world in which $A = a'$ versus drawing from the counterfactual joint distribution of mediator values, conditional on W, in which $A = a^\star$,

in the target population. The interventional direct effect among those for whom $S = 0$ can be similarly written: $\mathsf{E}(Y_{a',G_{a^\star}} - Y_{a^\star,G_{a^\star}} \mid S = 0)$, and, generally speaking, is the average effect of A on Y that does not operate through M in the target population. Specifically, it is the average difference in expected outcomes setting $A = a'$ versus $A = a^\star$ and stochastically drawing M from the counterfactual joint distribution of mediator values, conditional on W, in a hypothetical world in which $A = a^\star$, in the target population. We focus on identification and estimation of $\theta = \mathsf{E}(Y_{a',G_{a^\star}} \mid S = 0)$. Contrasts of θ under the values of a' and $a^\star$ given in the above definitions correspond to the transported interventional (in)direct effects. Under the assumptions

(1) $Y_{a',m} \perp\!\!\!\perp A \mid W, S = 0$,
(2) $M_{a^\star} \perp\!\!\!\perp A \mid W, S = 0$,
(3) $Y_{a',m} \perp\!\!\!\perp M \mid (Z, A, W, S = 0)$,
(4) $\mathsf{E}(Y \mid A = a', Z = z, M = m, W = w, S = 1) = \mathsf{E}(Y \mid A = a', Z = z, M = m, W = w, S = 0)$, and
(5) positivity:
- $P(w \mid S = 0) > 0$ implies $P(a \mid w, S = 0) > 0$ for $a \in \{a', a^\star\}$
- $P(m \mid a^\star, w, S = 0) > 0$ and $P(z \mid a', w, S = 0) > 0$ imply $P(m \mid a', z, w, S = 0) > 0$
- $P(a', z, m, w \mid S = 0) > 0$ implies $P(a', z, m, w \mid S = 1) > 0$,

θ is identified and is equal to

$$\theta = \int \mathsf{b}(a', z, m, w, S = 1)\mathsf{q}(z \mid a', w, S = 0)\mathsf{p}(m \mid a^\star, w, S = 0)\mathsf{p}(w \mid S = 0)\mathrm{d}\nu(w, z, m). \tag{2.2}$$

(The identification proof is in the Supplementary materials available at *Biostatistics* online.) Assumptions (1)–(3) are sequential randomization assumptions that involve the target population only. Assumption (1) states that, conditional on W, there is no unmeasured confounding of the relation between A and Y; assumption (2) states that conditional on W there is no unmeasured confounding of the relation between A and M; (3) states that conditional on (A, W, Z) there is no unmeasured confounding of the relation between M and Y. Assumption (4) is the transportability assumption and states that there is a common outcome model across source and target populations. It is this last assumption (4) that allows us to transport or borrow information on the outcome model from other sites. If an alternative data source is available where Y is observed among those for whom $S = 0$, then the null hypothesis of equivalence between $S = 0$ and $S = 1$ can be tested nonparametrically (Luedtke *and others*, 2019).

3. EFFICIENT INFLUENCE FUNCTION FOR θ

The *efficient influence function* (EIF) characterizes the asymptotic behavior of all regular and efficient estimators (Bickel *and others*, 1993; van der Vaart, 2002). In addition to being locally efficient, estimators constructed using the EIF have advantages of multiple robustness, which means that some components of the data distribution (i.e., nuisance parameters) can be inconsistently estimated while the estimator remains consistent. The multiple robustness property also allows the use data-adaptive machine learning algorithms in estimating nuisance parameters while retaining the ability to compute correct standard errors and confidence intervals. This is due to fact that the asymptotic analysis of the estimators yield second-order bias terms in differences of the nuisance parameters, and therefore allow slow convergence rates (e.g., $n^{-1/4}$) for estimating these nuisance parameters.

THEOREM 3.1 (Efficient influence function) For fixed a', $a^\star$ define

$$\begin{aligned}
\mathsf{h}(z,m,w,S=0) &= \frac{\mathsf{p}(m \mid a^\star, w, S=0)}{\mathsf{p}(m \mid a', z, w, S=0)} \\
\mathsf{u}(z,w,S=0) &= \int_{\mathcal{M}} \mathsf{b}(a',z,m,w,S=1)\mathsf{p}(m \mid a^\star, w, S=0)\mathrm{d}\nu(m) \\
\mathsf{v}(w,S=0) &= \int_{\mathcal{M}\times Z} \mathsf{b}(a',z,m,w,S=1)\mathsf{q}(z \mid a', w, S=0)\mathsf{p}(m \mid a^\star, w, S=0)\mathrm{d}\nu(m,z).
\end{aligned} \tag{3.3}$$

The efficient influence function for θ in the nonparametric model M is equal to

$$\begin{aligned}
D_{\mathsf{P},\theta}(o) =& D_{\mathsf{P},Y}(o) + D_{\mathsf{P},Z}(o) + D_{\mathsf{P},M}(o) + D_{\mathsf{P},W}(o), \text{ where} \\
D_{\mathsf{P},Y}(o) =& \frac{\mathbb{1}\{s=1, a=a'\}}{\mathsf{t}\times\mathsf{g}(a' \mid w, S=0)} \frac{1-\mathsf{c}(a',z,m,w)}{\mathsf{c}(a',z,m,w)} \mathsf{h}(z,m,w,S=0)\{y - \mathsf{b}(a',z,m,w,S=1)\} \\
D_{\mathsf{P},Z}(o) =& \frac{\mathbb{1}\{s=0, a=a'\}}{\mathsf{t}\times\mathsf{g}(a' \mid w, S=0)} \left\{\mathsf{u}(z,w,S=0) - \int_{\mathcal{Z}} \mathsf{u}(z,w,S=0)\mathsf{q}(z \mid a', w, S=0)\mathrm{d}\nu(z)\right\} \\
D_{\mathsf{P},M}(o) =& \frac{\mathbb{1}\{s=0, a=a^\star\}}{\mathsf{t}\times\mathsf{g}(a^\star \mid w, S=0)} \left\{\int_{\mathcal{Z}} \mathsf{b}(a',z,m,w,S=1)\mathsf{q}(z \mid a', w, S=0)\mathrm{d}\nu(z) - \mathsf{v}(w,S=0)\right\} \\
D_{\mathsf{P},\theta,W}(o) =& \frac{\mathbb{1}\{s=0\}}{\mathsf{t}}\{\mathsf{v}(w,S=0) - \theta\}.
\end{aligned} \tag{3.4}$$

This theorem makes two important contributions that advance the previous work deriving the EIF for a similar θ, but one that was limited in that it (i) assumed that the distribution of M conditional on (A, W, S) was known and (ii) could only consider a single binary M (Rudolph *and others*, 2020). First, the EIF we derive does not assume that that the distribution of M conditional on (A, W, S) is known, reflected in the $D_{\mathsf{P},M}(o)$ component of the EIF in Equation 3.4, above. Second, we can overcome the challenge of estimating multivariate or continuous densities on the mediator, M, and intermediate variable, Z, as well as integrals with respect to these densities, if either M or Z is low-dimensional (though it can be multivariate) by using an alternative parameterization of the densities that allows regression methods to be used in estimating the relevant quantities. In the remainder of this work, we assume Z is low-dimensional (e.g., binary, as in our MTO illustrative application), though similar parameterizations may be achieved if M is low-dimensional.

The EIF given in Theorem 3.1 may be represented in terms of the expressions given in Lemma 3.1 below, which does not depend on conditional densities or integrals on the mediating variables.

LEMMA 3.1 (Alternative representation of the EIF for univariate Z and multivariate M) The functions h, u, and v may be parameterized:

$$\mathsf{h}(z,m,w,S=0) = \frac{\mathsf{g}(a' \mid w, S=0)}{\mathsf{g}(a^\star \mid w, S=0)} \frac{\mathsf{q}(z \mid a', w, S=0)}{\mathsf{r}(z \mid a', m, w, S=0)} \frac{\mathsf{e}(a^\star \mid m, w, S=0)}{\mathsf{e}(a' \mid m, w, S=0)} \tag{3.5}$$

$$\mathsf{u}(z,w,S=0) = \mathsf{E}\left\{\mathsf{b}(a',Z,M,W,S=1)\mathsf{h}(Z,M,W,S=0), \middle| Z=z, a', W=w, S=0\right\}, \tag{3.6}$$

$$\mathsf{v}(w, S=0) = \mathsf{E}\left\{ \int_{\mathcal{Z}} \mathsf{b}(a', z, M, W, S=1)\mathsf{q}(z \mid a', W, S=0)\mathrm{d}\nu(z) \,\middle|\, A = a^{\star}, W = w, S = 0 \right\}. \tag{3.7}$$

In the remainder of the article, we denote $\eta = (\mathsf{c}, \mathsf{g}, \mathsf{e}, \mathsf{q}, \mathsf{r}, \mathsf{b}, \mathsf{u}, \mathsf{v})$ and $D_{\mathsf{P},\theta}(o) = D_{\eta,\theta}(o)$. We let $\hat{\eta}$ denote an estimator of η, and η_1 denotes the probability limit of $\hat{\eta}$, which may be different from the true value. We derive the robustness properties of $D_{\eta,\theta}(O)$ in the Supplementary materials available at *Biostatistics* online; they are given below in Lemma 3.2. The behavior of the term $\mathsf{P}D_{\eta_1,\theta}$ determines the robustness properties of the EIF as an estimating equation. Theorem 1 in the Supplementary materials available at *Biostatistics* online, together with the Cauchy–Schwarz inequality shows that $\mathsf{P}D_{\eta_1,\theta}$ yields a term of the order of:

$$\begin{aligned} R(\eta_1, \eta) = &\|\mathsf{v}_1 - \mathsf{v}\|\|\mathsf{g}_1 - \mathsf{g}\| \\ &+ \|\mathsf{u}_1 - \mathsf{u}\|\|\mathsf{q}_1 - \mathsf{q}\| \\ &+ \|\mathsf{b}_1 - \mathsf{b}\|\|\mathsf{q}_1 - \mathsf{q}\| \\ &+ \|\mathsf{b}_1 - \mathsf{b}\|\{\|\mathsf{c}_1 - \mathsf{c}\| + \|\mathsf{q}_1 - \mathsf{q}\| + \|\mathsf{r}_1 - \mathsf{r}\| + \|\mathsf{e}_1 - \mathsf{e}\|\} \end{aligned}$$

such that consistent estimation of θ is possible under consistent estimation of certain configurations of the parameters in η. The following lemma is a direct consequence.

LEMMA 3.2 (Multiple robustness of $D_{\eta,\theta}(O)$) Let $\eta_1 = (\mathsf{c}_1, \mathsf{g}_1, \mathsf{e}_1, \mathsf{q}_1, \mathsf{r}_1, \mathsf{b}_1, \mathsf{u}_1, \mathsf{v}_1)$ be such that one of the following conditions hold:

(1) $\mathsf{v}_1 = \mathsf{v}$ and either $(\mathsf{c}_1, \mathsf{q}_1, \mathsf{e}_1, \mathsf{r}_1) = (\mathsf{c}, \mathsf{q}, \mathsf{e}, \mathsf{r})$ or $(\mathsf{b}_1, \mathsf{q}_1) = (\mathsf{b}, \mathsf{q})$ or $(\mathsf{b}_1, \mathsf{u}_1) = (\mathsf{b}, \mathsf{u})$, or
(2) $\mathsf{g}_1 = \mathsf{g}$ and either $(\mathsf{c}_1, \mathsf{q}_1, \mathsf{e}_1, \mathsf{r}_1) = (\mathsf{c}, \mathsf{q}, \mathsf{e}, \mathsf{r})$ or $(\mathsf{b}_1, \mathsf{q}_1) = (\mathsf{b}, \mathsf{q})$ or $(\mathsf{b}_1, \mathsf{u}_1) = (\mathsf{b}, \mathsf{u})$.

Then $\mathsf{P}D_{\eta_1,\theta} = 0$ with $D_{\eta,\theta}$ defined as in Theorem 3.1.

We note that the cases $(\mathsf{b}_1, \mathsf{v}_1, \mathsf{u}_1) = (\mathsf{b}, \mathsf{v}, \mathsf{u})$ and $(\mathsf{b}_1, \mathsf{g}_1, \mathsf{u}_1) = (\mathsf{b}, \mathsf{g}, \mathsf{u})$ may be uninteresting if the re-parametrization in Lemma 3.1 is used to estimate the EIF, because in that case, consistent estimation of u and v will generally require consistent estimation of $(\mathsf{b}, \mathsf{c}, \mathsf{q}, \mathsf{r}, \mathsf{e})$ in addition to the outer conditional expectations in Equations (3.6) and (3.7).

4. ESTIMATORS

We describe two efficient, robust estimators of θ. In Section 4.1, we propose an estimator that solves the EIF estimating equation in one step (Pfanzagl and Wefelmeyer, 1985) (which we refer to as a one-step estimator), and in Section 4.2, we propose a targeted minimum loss-based estimator (TMLE, van der Laan and Rubin, 2006), which is a substitution estimator that also solves the EIF estimating equation, but does it through iterative de-biasing targeted updates to nuisance parameters. We provide the R code to implement the proposed estimators, freely available at https://github.com/kararudolph/transport.

Let $\hat{\theta}_{\text{os}}$ and $\hat{\theta}_{\text{tmle}}$ denote the estimators defined below in Sections 4.1 and 4.2. Per the theorem below, the two estimators are asymptotically normal and efficient.

THEOREM 4.1 (Asymptotic normality and efficiency) Assume

(1) Positivity, described as identification assumption (5) in Section 2.1, and

(2) The class of functions $\{D_{\eta,\theta} : |\theta - \theta_0| < \delta, ||\eta - \eta_1|| < \delta\}$ is Donsker for some $\delta > 0$ and such that $P_0(D_{\eta,\theta} - D_{\eta_1,\theta_0})^2 \to 0$ as $(\eta, \theta) \to (\eta_1, \theta_0)$, and
(3) The second-order term $R(\hat{\eta}, \eta)$ is $o_P(n^{-1/2})$.

Then, $\sqrt{n}(\hat{\theta}_{\text{os}} - \theta) \to N(0, \sigma^2)$, and $\sqrt{n}(\hat{\theta}_{\text{tmle}} - \theta) \to N(0, \sigma^2)$, where $\sigma^2 = \mathsf{Var}(D_\eta(O))$ is the nonparametric efficiency bound.

The proof of this theorem follows the general proof presented in Appendix 18 of van der Laan and Rose (2011). As a consequence, the variance of the estimators that follow can be estimated as the sample variance of the EIF, with $\hat{\theta}$ and the nuisance parameters estimated as described above. This variance estimate may be used to construct Wald-type confidence intervals.

The Donsker condition of Theorem 4.1 may be avoided by using cross-fitting (Klaassen, 1987; Zheng and van der Laan, 2011; Chernozhukov *and others*, 2019) in the estimation procedure. Let $\mathcal{V}_1, \ldots, \mathcal{V}_J$ denote a random partition of the index set $\{1, \ldots, n\}$ into J prediction sets of approximately the same size. That is, $\mathcal{V}_j \subset \{1, \ldots, n\}$; $\bigcup_{j=1}^{J} \mathcal{V}_j = \{1, \ldots, n\}$; and $\mathcal{V}_j \cap \mathcal{V}_{j'} = \emptyset$. In addition, for each j, the associated training sample is given by $\mathcal{T}_j = \{1, \ldots, n\} \setminus \mathcal{V}_j$. let $\hat{\eta}_j$ denote the estimator of η, obtained by training the corresponding prediction algorithm using only data in the sample $\mathcal{T}_j$. Further, we let $j(i)$ denote the index of the validation set which contains observation i. The one-step and TMLE estimators may be adapted to cross-fitting by substituting all occurrences of $\hat{\eta}(O_i)$ by $\hat{\eta}_{j(i)}(O_i)$ in the respective algorithms.

The third condition of Theorem 4.1 can be satisfied by many data-adaptive algorithms (e.g., lasso (Bickel *and others*, 2009), regression trees (Wager and Walther, 2015), neural networks (Chen and White, 1999), and highly adaptive lasso (HAL) (van der Laan, 2017)); we use HAL in the simulations that follow.

4.1. *One-step estimator*

The one-step estimate of θ is given by the solution to the EIF estimating equation:

$$\hat{\theta}_{\text{os}} = \frac{1}{n}\sum_{i=1}^{n}\{D_{\hat{\eta},Y}(O_i) + D_{\hat{\eta},Z}(O_i) + D_{\hat{\eta},M}(O_i)\} + \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}\{S_i = 0\}}{\hat{\mathsf{t}}}\hat{\mathsf{v}}(W_i, S = 0).$$

We first describe how to estimate $D_{\eta,Y}$. The regression $\mathsf{b}(a', z, m, w, S = 1)$ can be estimated by fitting a regression of Y on W, A, Z, M among observations with $S = 1$ and then predicting values of Y setting $A = a'$. The probability t is estimated as the empirical proportion of observations with $S = 0$ (i.e., in the target population). The regression function $\mathsf{c}(a', z, m, w)$ can be estimated by fitting a regression of S on W, A, Z, M and predicting the probability that $S = 1$ setting $A = a'$. The treatment mechanism $\mathsf{g}(a \mid w, S = 0)$ for $a \in \{a', a^\star\}$ can be estimated by fitting a regression of A on (S, W) and predicting the probability that $A = a$, setting $S = 0$. For the motivating example, we consider here in which assignment of A is randomized, these can be estimated as the empirical probabilities that $A = a'$ and $A = a^\star$ among those with $S = 0$. Under the reparameterization in Lemma 3.1 and in our motivating example, $\mathsf{q}(z \mid a', w, S = 0)$ can be estimated by fitting a regression of Z on S, A, W and predicting the probability that $Z = z$ setting $A = a', S = 0$. Likewise, $\mathsf{r}(z \mid a', m, w, S = 0)$ can be estimated by fitting a regression of Z on S, A, M, W and predicting the probability that $Z = z$ setting $A = a', S = 0$. The treatment probabilities $\mathsf{e}(a' \mid m, w, S = 0)$ and $\mathsf{e}(a^* \mid m, w, S = 0)$ can be estimated by fitting a regression of A on S, M, W and predicting the probability that $A = a'$ and $A = a^*$, respectively, setting $S = 0$.

We next describe how to estimate $D_{\eta,Z}$. For binary Z, the EIF simplifies to be

$$D_{\eta,Z}(o) = \frac{\mathbb{1}\{s=0, a=a'\}}{\mathsf{t} \times \mathsf{g}(a' \mid w, S=0)} \{\mathsf{u}(1,w,S=0) - \mathsf{u}(0,w,S=0)\} \left\{z - \mathsf{q}(1 \mid a', w, S=0)\right\}.$$

The parameters $\mathsf{t}, \mathsf{g}(a' \mid w, S=0)$ and $\mathsf{q}(z \mid a', w, S=0)$ can be estimated as described above. For each z, $\mathsf{u}(z,w,S=0)$ can be estimated by regressing the quantity $\mathsf{b}(a',Z,M,W,S=1) \times \mathsf{h}(Z,M,W,S=0)$ on S,Z,W and getting predicted values, setting $Z=z, S=0$.

To estimate $D_{\eta,M}$, we estimate $\mathsf{t}, \mathsf{g}(a^* \mid w, S=0), \mathsf{b}(a',z,m,w,S=1), \mathsf{q}(z \mid a', w, S=0)$ as described above. The function $\mathsf{v}(w, S=0)$ can be estimated by marginalizing out Z from $\mathsf{b}(a',z,M,W,S=1)$ using $\mathsf{q}(z \mid a', W, S=0)$ as predicted probabilities for each z, and then regressing the resulting quantity on A, W, S, and predicting values setting $A=a^*, S=0$.

4.2. *TML estimator*

We now describe how to compute a related TML estimator. As an overview, this estimator entails targeting $\mathsf{b}(a',z,m,w,S=1), \mathsf{q}(z|a',w,S=0)$, and $\mathsf{v}(w,S=0)$, which correspond to solving terms 1, 2, and 3, respectively, in (3.4). Plugging in $\hat{\theta}$ solves the last term in (3.4).

We assume Y can be bounded in $[0,1]$, as described previously (Gruber and van der Laan, 2010). Many of the steps are identical to those for the one-step estimator, the differences are in the targeting of $\mathsf{b}(a',z,m,w,S=1)$, $\mathsf{q}(z \mid a', w, S=0)$, and $\mathsf{v}(w,S=0)$.

Let $\hat{\mathsf{b}}(a',z,m,w,S=1)$ be an initial estimate of $\mathsf{b}(a',z,m,wS=1)$. We update this initial estimate using covariate

$$\hat{C}_{\mathsf{b}}(a',Z,M,W) = \frac{1-\hat{\mathsf{c}}(a',Z,M,W)}{\hat{\mathsf{c}}(a',Z,M,W)} \frac{\hat{\mathsf{h}}(Z,M,W,S=0)}{\hat{\mathsf{g}}(a' \mid W, S=0) \times \hat{\mathsf{t}}}$$

in a logistic regression of Y with logit $\hat{\mathsf{b}}(a',Z,M,W,S=1)$ as an offset, among the subset for which $S=1$. Let $\hat{\epsilon}_{\mathsf{b}}$ denote the maximum likelihood estimation (MLE) fitted coefficient associated with $\hat{C}_{\mathsf{b}}(a',Z,M,W)$. The targeted (i.e., updated) estimate is given by

$$\operatorname{logit} \tilde{\mathsf{b}}(a',z,m,w) = \operatorname{logit} \hat{\mathsf{b}}(a',z,m,w,S=1) + \hat{\epsilon}_{\mathsf{b}} \hat{C}_{\mathsf{b}}(a',z,m,w).$$

An alternative algorithm would use

$$\frac{\hat{\mathsf{e}}(a^\star|M,W,S=0)}{\hat{\mathsf{e}}(a' \mid M,W,S=0)} \frac{1}{\hat{\mathsf{g}}(a^\star \mid W,S=0) \times \hat{\mathsf{t}}}$$

as weights of what would become a weighted logistic regression model with covariate

$$\hat{C}_{\mathsf{b}}(a',Z,M,W) = \frac{1-\hat{\mathsf{c}}(a',Z,M,W)}{\hat{\mathsf{c}}(a',Z,M,W)} \frac{\hat{\mathsf{q}}(Z \mid a',W,S=0)}{\hat{\mathsf{r}}(Z \mid a',M,W,S=0)}.$$

Next, let $\hat{\mathsf{q}}(z \mid a', w, S=0)$ be an initial estimate of $\mathsf{q}(z \mid a', w, S=0)$. We update this initial estimate using covariate

$$\hat{C}_{\mathsf{q}}(a',W) = \frac{\hat{\mathsf{u}}(1,W,S=0) - \hat{\mathsf{u}}(0,W,S=0)}{\hat{\mathsf{g}}(a' \mid W,S=0) \times \hat{\mathsf{t}}}$$

in a logistic regression of Z with $\operatorname{logit}\hat{\mathsf{q}}(Z \mid a', W, S = 0)$ as an offset, among the subset for which $A = a', S = 0$. Let $\hat{\epsilon}_{\mathsf{q}}$ be the MLE fitted coefficient associated with $C_{\mathsf{q}}(A, W)$. The targeted estimate is given by

$$\operatorname{logit}\tilde{\mathsf{q}}(z \mid a', w, S = 0) = \operatorname{logit}\hat{\mathsf{q}}(z \mid a', w, S = 0) + \hat{\epsilon}_{\mathsf{q}}\hat{C}_{\mathsf{q}}(a', w).$$

To potentially improve performance in finite samples, we can move $\{\hat{\mathsf{g}}(a' \mid W, S = 0) \times \hat{\mathsf{t}}\}^{-1}$ into the weights of a weighted logistic regression model, leaving $\hat{\mathsf{u}}(1, W, S = 0) - \hat{\mathsf{u}}(0, W, S = 0)$ as $\hat{C}_{\mathsf{q}}(a', W)$.

Replacing $\hat{\mathsf{b}}$ and $\hat{\mathsf{q}}$ with $\tilde{\mathsf{b}}$ and $\tilde{\mathsf{q}}$, the above steps can be iterated until the score equation $n^{-1}\sum_i\{D_{\tilde{\eta},Y}(O_i) + D_{\tilde{\eta},Z}(O_i)\} = 0$ is solved up to a factor of $(\sqrt{n}\log(n))^{-1}$. This iterating process and stopping criterion ensures that the efficient influence function is solved up to $n^{-1/2}$ and mitigates risk of overfitting.

Next, we marginalize out Z from $\tilde{\mathsf{b}}(a', z, M, W)$ using $\tilde{\mathsf{q}}(z \mid a', W)$ as predicted probabilities for each z, and call the resulting quantity Q. This quantity is then regressed on (A, W) among units with $S = 0$ and $A = a^\star$ to obtain an estimator $\hat{\mathsf{v}}(W, S = 0)$. This estimate is updated using covariate

$$\hat{C}_{\mathsf{v}}(a^\star, W) = \frac{1}{\hat{\mathsf{g}}(a^\star \mid W, S = 0) \times \hat{\mathsf{t}}}$$

in a logistic regression of Q with $\operatorname{logit}\hat{\mathsf{v}}(W, S = 0)$ as an offset, among the subset for which $A = a^*, S = 0$. Let $\hat{\epsilon}_{\mathsf{v}}$ denote the MLE fitted coefficient on $C_{\mathsf{v}}(a^\star, W)$. The targeted estimate is given by

$$\operatorname{logit}\tilde{\mathsf{v}}(w, S = 0) = \operatorname{logit}\hat{\mathsf{v}}(w, S = 0) + \hat{\epsilon}_{\mathsf{v}}\hat{C}_{\mathsf{v}}(a^*, w).$$

To potentially improve finite sample performance, $\hat{C}_{\mathsf{v}}(a^\star, W)$ may be moved into the weights of a weighted logistic regression model with intercept only. The empirical mean of $\tilde{\mathsf{v}}(W_i, S = 0)$ among those for whom $S = 0$ is the TMLE estimate. Its variance can be estimated as the sample variance of the estimated EIF, given in (3.4).

5. Simulation

We conducted a limited simulation study to examine and compare finite sample performance of these two estimators. We consider the data-generating mechanism (DGM) as follows. All variables are Bernoulli distributed with probabilities given by

$$\begin{aligned}
P(W_1 = 1) &= 0.5\\
P(W_2 = 1 \mid W_1) &= 0.4 + 0.2W_1\\
P(\Delta = 1 \mid W) &= \operatorname{expit}(-1 + \log(4)W_1 + \log(4)W_2)\\
P(S = 1 \mid \Delta, W) &= \operatorname{expit}(\log(1.2)W_1 + \log(1.2)W_2 + \log(1.2)W_1W_2)\\
P(A = 1 \mid S, \Delta, W) &= 0.5\\
P(Z = 1 \mid A, S, \Delta, W) &= \operatorname{expit}(-\log(2) + \log(4)A + -\log(2)W_2 + \log(1.4)S + \log(1.43)A \times S)\\
P(M = 1 \mid Z, A, S, \Delta, W) &= \operatorname{expit}(-\log(2) + \log(4)Z - \log(1.4)W_2 + \log(1.4)S)\\
P(Y = 1 \mid M, Z, A, S, \Delta, W) &= \operatorname{expit}(-\log(5) + \log(8)Z + \log(4)M - \log(1.2)W_2 + \log(1.2)W_2Z.
\end{aligned}$$

This DGM is formulated to align with features of the MTO study we use for the illustrative example. For example, A is randomly assigned and adheres to the exclusion restriction (Angrist *and others*, 1996),

aligned with its role as an instrumental variable. In addition, we consider a modification of the observed data we have considered thus far: $\Delta \times O = \Delta \times (S, W, A, Z, M, SY)$, where Δ is an indicator of selection into the survey sample. We assume the survey sampling weights are known or can be estimated as

$$\hat{\Gamma}_i = \frac{1}{\Pi_i} \frac{\sum_{i=1}^{n}(1 - S_i)}{\sum_{i=1}^{n}(1 - S_i)\Pi_i^{-1}},$$

where $\Pi = \mathsf{P}(\Delta = 1 \mid X)$ and X represents unobserved variables used in the sampling design. Our previous identification result, which can alternatively be written as $\theta = \mathsf{E}[\mathsf{v}(W, S = 0)]$, then becomes

$$\theta_{\Delta=1} = \mathsf{E}\left[\Gamma\ \mathsf{v}(W, S = 0) \,\middle|\, \Delta = 1\right],$$

where we have added an index $\Delta = 1$ to emphasize that we are interested in parameters for the population from which the sample was drawn. The EIF is modified to be $D_{\mathsf{P},\Delta=1}(o) = \Gamma D_{\mathsf{P}}(o)$, and the estimators of the previous section can be applied by using the weights $\hat{\Gamma}_i$ for each subject in the sample.

We consider estimator performance in terms of absolute bias, absolute bias scaled by $\sqrt{n}$, influence curve-based standard error relative to the Monte Carlo-based standard error, standard deviation of the estimator relative to the efficiency bound scaled by $\sqrt{n}$, mean squared error relative to the efficiency bound scaled by n, and 95% confidence interval (CI) coverage. We run 1000 simulations for sample sizes $N = 1000$ and $N = 10\,000$. We also consider several model specifications. One in which all nuisance parameters in η are correctly specified, others that misspecify each nuisance parameter one at a time, another in which $\mathsf{g}(a' \mid w, S = 0), \mathsf{b}(a', z, m, w, S = 1), \mathsf{q}(z \mid a', w, S = 0)$ are correctly specified but the rest are not; and last, correctly specifying $\mathsf{b}(a', z, m, w, S = 1), \mathsf{q}(z \mid a', w, S = 0), \mathsf{v}(w, S = 0)$ but incorrectly specifying the rest. Under correct specification scenarios, we use HAL (Benkeser and van der Laan, 2016; van der Laan, 2017) to fit each nuisance parameter. For incorrect specification, we use an intercept-only model.

Table 1 shows simulation results for the transported interventional direct effect, and Table 2 shows simulation results for the transported interventional indirect effect comparing the one-step and TML estimators under correct specification of all nuisance parameters and various misspecifications. Given the robustness results in Lemma 3.2, we expect consistent estimates for all specifications in Tables 1 and 2 except when q is misspecified. We see this reflected in the results. We see that when the q model is misspecified, bias is more than an order of magnitude greater than any other specification for the transported interventional direct effect in Table 1, and also greater, though to a lesser extent for the transported interventional indirect effect in Table 2. 95% CI coverage using influence curve (IC)-based inference is close to 95% in the correctly specified scenario but is poor when q is misspecified for the transported interventional direct effect (Table 1), which is not unexpected given the biased estimates in this scenario. Coverage is less than 95% in other misspecified scenarios for both the transported direct and indirect effects (e.g., 68% when the b model is misspecified for the transported interventional indirect effect, Table 2). This is not unexpected; the IC may not provide accurate inference when the IC at the estimated distribution using misspecified models does not converge to the IC at the true distribution. For robustness to extend to IC-based inference, further targeting of the nuisance parameters would be necessary that would preserve asymptotic linearity with a known influence curve at the cost of some efficiency (van der Laan, 2014; Benkeser *and others*, 2016). We note that under the smaller sample size of $N = 1000$ we see some deterioration in performance, particularly for the indirect effect, which is expected given that the true indirect effect is over five times smaller than the direct effect.

We also give simulation results in Table 3 comparing performance of the transport one-step and TML estimators, assuming the outcome data are unobserved for $S = 0$, and the nontransported versions of

Table 1. *Simulation results for the transported interventional direct effect.*

Nuisance parameters misspecified	Estimator	|bias|	$\sqrt{n}$|bias|	relse	relsd	relrmse	95%CI Cov
Transported interventional direct effect							
$N = 10\,000$							
None	os	0.0005	0.0490	1.0200	0.9489	0.9488	0.9570
	tmle	0.0004	0.0415	1.0040	0.9610	0.9608	0.9530
c	os	0.0005	0.0519	1.0023	0.8226	0.8227	0.9570
	tmle	0.0003	0.0312	0.9577	0.8579	0.8576	0.9460
g	os	0.0005	0.0480	1.0213	0.9481	0.9480	0.9580
	tmle	0.0004	0.0408	1.0055	0.9600	0.9598	0.9520
e	os	0.0002	0.0156	1.0097	0.9431	0.9427	0.9520
	tmle	0.0003	0.0301	0.9878	0.9602	0.9599	0.9460
q	os	0.0885	8.8488	0.7750	1.4727	5.2339	0.0250
	tmle	0.0348	3.4814	1.0656	1.0154	2.2215	0.5580
r	os	0.0024	0.2382	1.0889	0.8724	0.8824	0.9640
	tmle	0.0021	0.2134	1.0809	0.8788	0.8867	0.9640
b	os	0.0047	0.4739	1.0460	0.9615	0.9979	0.9470
	tmle	0.0107	1.0661	0.9908	1.0007	1.1690	0.9070
u	os	0.0053	0.5285	0.9400	0.9405	0.9867	0.9230
	tmle	0.0053	0.5262	0.9249	0.9530	0.9983	0.9150
v	os	0.0005	0.0499	1.0213	0.9476	0.9476	0.9570
	tmle	0.0004	0.0421	1.0028	0.9621	0.9619	0.9520
c, e, r, u, v	os	0.0023	0.2293	0.8924	0.7159	0.7272	0.9140
	tmle	0.0019	0.1889	0.8519	0.7465	0.7538	0.9020
c, g, e, r, u	os	0.0023	0.2321	0.8914	0.7165	0.7281	0.9140
	tmle	0.0019	0.1904	0.8548	0.7438	0.7513	0.9030
$N = 1000$							
None	os	0.0014	0.0454	1.0200	0.8925	0.8921	0.9591
	tmle	0.0028	0.0880	0.9702	0.9309	0.9314	0.9414
c	os	0.0003	0.0104	1.0340	0.7691	0.7683	0.9600
	tmle	0.0021	0.0648	0.9648	0.8190	0.8190	0.9460
g	os	0.0019	0.0617	1.0167	0.8958	0.8957	0.9520
	tmle	0.0032	0.1009	0.9697	0.9317	0.9327	0.9424
e	os	0.0036	0.1134	1.0131	0.8855	0.8871	0.9520
	tmle	0.0049	0.1550	0.9611	0.9252	0.9286	0.9440
q	os	0.0672	2.1252	0.7993	1.3098	1.7796	0.7560
	tmle	0.0280	0.8862	1.0124	0.9771	1.0981	0.9300
r	os	0.0073	0.2303	1.1242	0.8149	0.8245	0.9620
	tmle	0.0070	0.2202	1.0994	0.8315	0.8400	0.9620
b	os	0.0047	0.1499	1.0460	0.3041	0.3156	0.9470
	tmle	0.0107	0.3371	0.9908	0.3164	0.3697	0.9070
u	os	0.0106	0.3362	0.9735	0.8614	0.8814	0.9420
	tmle	0.0101	0.3186	0.9234	0.8993	0.9164	0.9180
v	os	0.0009	0.0295	1.0304	0.8827	0.8819	0.9589
	tmle	0.0021	0.0668	0.9857	0.9152	0.9150	0.9498
c, e, r, u, v	os	0.0030	0.0949	0.9553	0.6643	0.6657	0.9315
	tmle	0.0013	0.0424	0.9141	0.6898	0.6895	0.9224
c, g, e, r, u	os	0.0019	0.0591	0.9533	0.6663	0.6664	0.9291
	tmle	0.0001	0.0034	0.9044	0.6981	0.6974	0.9222

Table 2. *Simulation results for the transported interventional indirect effect.*

Nuisance parameters misspecified	Estimator	|bias|	$\sqrt{n}$|bias|	relse	relsd	relrmse	95%CI Cov
Transported interventional indirect effect							
$N = 10\,000$							
None	os	0.0000	0.0030	0.9966	0.9760	0.9755	0.9420
	tmle	0.0001	0.0065	0.9895	0.9778	0.9774	0.9400
c	os	0.0003	0.0272	0.9864	0.9445	0.9456	0.9410
	tmle	0.0001	0.0147	0.9734	0.9530	0.9529	0.9370
g	os	0.0000	0.0004	0.9976	0.9749	0.9744	0.9430
	tmle	0.0000	0.0044	0.9907	0.9768	0.9764	0.9430
e	os	0.0006	0.0632	0.9610	0.8917	0.9003	0.9390
	tmle	0.0007	0.0668	0.9603	0.8887	0.8983	0.9380
q	os	0.0020	0.2035	0.9285	1.0709	1.1459	0.9070
	tmle	0.0030	0.2951	0.8061	1.2396	1.3737	0.8410
r	os	0.0003	0.0324	1.0081	1.0034	1.0051	0.9500
	tmle	0.0001	0.0075	1.0429	0.9652	0.9648	0.9590
b	os	0.0001	0.0067	0.4870	1.0879	1.0874	0.6850
	tmle	0.0001	0.0099	0.4747	1.1763	1.1759	0.6700
u	os	0.0000	0.0020	0.9997	0.9714	0.9709	0.9440
	tmle	0.0000	0.0009	0.9919	0.9744	0.9739	0.9440
v	os	0.0000	0.0026	1.0250	1.0114	1.0109	0.9550
	tmle	0.0001	0.0058	1.0015	1.0259	1.0254	0.9480
c, e, r, u, v	os	0.0006	0.0618	0.9375	0.9834	0.9907	0.9400
	tmle	0.0007	0.0696	0.9168	0.9947	1.0040	0.9310
c, g, e, r, u	os	0.0007	0.0676	0.9032	0.9399	0.9492	0.9150
	tmle	0.0008	0.0767	0.8996	0.9387	0.9508	0.9150
$N = 1000$							
None	os	0.0007	0.0229	0.9010	1.0200	1.0201	0.9041
	tmle	0.0006	0.0200	0.8900	1.0209	1.0207	0.8988
c	os	0.0013	0.0407	0.8891	0.9901	0.9924	0.8900
	tmle	0.0011	0.0337	0.8729	0.9970	0.9983	0.8880
g	os	0.0009	0.0293	0.9092	1.0185	1.0194	0.9072
	tmle	0.0008	0.0242	0.8974	1.0193	1.0197	0.8992
e	os	0.0026	0.0818	0.9025	0.9447	0.9582	0.8992
	tmle	0.0025	0.0797	0.8953	0.9415	0.9543	0.8976
q	os	0.0017	0.0541	0.8405	1.0920	1.0963	0.8700
	tmle	0.0017	0.0525	0.7671	1.1978	1.2012	0.8440
r	os	0.0004	0.0120	0.8955	1.0476	1.0468	0.9080
	tmle	0.0008	0.0263	0.9121	1.0109	1.0112	0.8980
b	os	0.0001	0.0021	0.4870	0.3440	0.3439	0.6850
	tmle	0.0001	0.0031	0.4747	0.3720	0.3719	0.6700
u	os	0.0008	0.0264	0.8872	1.0328	1.0331	0.8900
	tmle	0.0007	0.0215	0.8655	1.0424	1.0423	0.8800
v	os	0.0004	0.0114	0.9605	1.0274	1.0265	0.9247
	tmle	0.0003	0.0086	0.9279	1.0450	1.0440	0.9110
c, e, r, u, v	os	0.0023	0.0733	0.8522	1.0237	1.0331	0.8973
	tmle	0.0026	0.0831	0.8284	1.0286	1.0410	0.8881
c, g, e, r, u	os	0.0020	0.0618	0.8595	0.9285	0.9358	0.8741
	tmle	0.0022	0.0708	0.8492	0.9229	0.9328	0.8741

Table 3. *Simulation results comparing one-step (os) and TML (tmle) estimators for the transported interventional indirect and direct effects with those for the nontransported interventional indirect and direct effects.*

Effect type	Estimator	|bias|	$\sqrt{n}$|bias|	relse	relsd	relrmse	95%CI Cov
N = 10 000							
Transported, indirect effect	os	0.0000	0.0030	0.9966	0.9760	0.9755	0.9420
Transported, indirect effect	tmle	0.0001	0.0065	0.9895	0.9778	0.9774	0.9400
Nontransported, indirect effect	os	0.0000	0.0022	0.9413	1.1722	1.1717	0.9420
Nontransported, indirect effect	tmle	0.0000	0.0035	0.9412	1.1690	1.1685	0.9420
Transported, direct effect	os	0.0005	0.0490	1.0200	0.9489	0.9488	0.9570
Transported, direct effect	tmle	0.0004	0.0415	1.0040	0.9610	0.9608	0.9530
Nontransported, direct effect	os	0.0013	0.1344	1.0178	1.1003	1.1074	0.9620
Nontransported, direct effect	tmle	0.0014	0.1397	1.0150	1.1033	1.1110	0.9640
N = 1000							
Transported, indirect effect	os	0.0007	0.0229	0.9010	1.0200	1.0201	0.9041
Transported, indirect effect	tmle	0.0006	0.0200	0.8900	1.0209	1.0207	0.8988
Nontransported, indirect effect	os	0.0001	0.0045	0.9134	1.1626	1.1622	0.9340
Nontransported, indirect effect	tmle	0.0001	0.0042	0.9170	1.1495	1.1490	0.9300
Transported, direct effect	os	0.0014	0.0454	1.0200	0.8925	0.8921	0.9591
Transported, direct effect	tmle	0.0028	0.0880	0.9702	0.9309	0.9314	0.9414
Nontransported, direct effect	os	0.0025	0.0802	1.0212	1.0854	1.0876	0.9540
Nontransported, direct effect	tmle	0.0031	0.0968	1.0119	1.0949	1.0984	0.9540

the one-step and TML estimators developed previously (Díaz *and others*, 2020). These nontransported estimators are approximately 3 times more efficient than their transported counterparts (e.g., the efficiency bound of the transported TMLE of the indirect effect is 3 times greater than the efficiency bound of the nontransported TMLE of the indirect effect), reflecting the advantage of observing the outcome data in the target population.

6. Illustrative example

We apply the one-step and TML estimators proposed in Section 4 to estimate interventional indirect effects transported across MTO sites, as described in Section 1. Specifically, we are interested in the extent to which differences in: (i) the distribution of individual-level compositional factors between the sites, (ii) take-up of the intervention (i.e., using the housing voucher to move), and (iii) distribution of school environment mediating variables can explain the difference in the indirect effect estimates between MTO sites.

For this example, we consider the indirect effect of randomized receipt of a Section 8 housing voucher (A) and subsequent use (Z) on behavioral problems (Y) (Zill, 1990) through aspects of the school environment (M), (i) rank of the schools attended, (ii) whether ever attended a school in the top 50% of rankings, (iii) number of schools attended, (iv) number of moves since baseline, (v) average proportion of students receiving free or reduced lunch, (vi) ratio of students to teachers, (vii) proportion of schools attended that were Title I, and (viii) whether or not the most recent school attended was in the same district as the baseline school) among girls, comparing the Los Angeles (LA) and New York City (NYC) sites ($S = 1$, $N = 1000$) to the Chicago site ($S = 0$, $N = 600$) (rounded sample sizes per Census Bureau requirements). We do this in order to illustrate our methods: the outcomes in Chicago were actually observed, so we

can compared the transported estimate with estimates obtained using Chicago outcome data. Variables W and A were measured at baseline, when the children were 0–10 years old. Mediating variables were measured during the interval between baseline and the final follow-up timepoint 10–15 years later. The outcome was measured at the final follow-up timepoint. We account for a large number of covariates at the child and family levels: child age, race/ethnicity, history of behavioral problems, and gifted/talented status; parental education, marital status, whether or not the parent was under 18 at the birth of the child, employment, receipt of other public benefits, household size, feeling like the neighborhood was unsafe at night, feeling very dissatisfied with the neighborhood, whether or not the family had previously moved more than three times, wanting to move for better schools, whether or not the family had received a Section 8 voucher before, and poverty level of the baseline neighborhood. For this research question, randomization to receive a Section 8 housing voucher is an instrumental variable that affects M and Y through the intermediate variable of using the voucher to move out of public housing and into a rental on the private market (Z). We use the MTO sampling weights as described in Section 5. These weights account for sampling of children within families, changing randomization ratios, and loss to follow-up (Sanbonmatsu *and others*, 2011). We use data-adaptive methods for fitting the nuisance parameters, using a cross-validated ensemble of machine learning algorithms (Van der Laan *and others*, 2007), that includes generalized linear models, intercept-only models, and lasso (Tibshirani, 1996) that included all first and second-order predictors. To estimate the observed, nontransported interventional indirect effects, we use nontransported versions of the one-step and TML estimators (Díaz *and others*, 2020). Standard errors are estimated using the sample variance of the influence curve.

Figure 1 shows the transported and observed indirect effect estimates and their 95% CIs. Looking at the observed estimates, the indirect pathway from housing voucher receipt and use through the school environment to behavioral problems is protective for girls in LA and NYC, resulting in a reduction in behavioral problems at the final time point. However, the same pathway appears harmful for girls in Chicago, resulting in an increase of behavioral problems. Comparing the transported interventional indirect effect estimate (one-step estimator: 0.0043, 95% CI −0.0150 to 0.0237, risk difference scale; TMLE: 0.0153, 95% CI −0.0150 to 0.0420) to the observed estimate for girls in Chicago (one-step estimator: 0.0062, 95% CI 0.0027–0.0097; TMLE: 0.0089, 95% CI 0.0007–0.0171), we see that the two are similar even though the outcome data from Chicago was not used in the transported estimates. Thus, by taking the outcome model for LA and NYC and standardizing based on W, A, Z, M in Chicago, the predicted effect for Chicago is close to the observed. In contrast if they were not close to each other, this would suggest that the identification assumptions were not met.

In the context of MTO, identification assumption (iv) of a common outcome model (i.e., a common relation of the voucher, moving, and mediators on behavioral problems among girls across the MTO sites) is arguably the most tenuous. This assumption would not hold in the presence of any contextual-level effects on the outcome model, such as the local economy, housing market conditions, segregation, etc. In the presence of contextual-level effects on the conditional outcome distribution, we would be extrapolating from the source population to the target population using an inaccurate outcome model. Although we do not assume a common relation of voucher and moving on the mediators among girls across sites, we do assume that all $\{m, z, a', w\}$ observed in the target population are also observed in the source population.

7. CONCLUSIONS

We proposed estimators for transported interventional direct and indirect effects under intermediate confounding and allowing for multiple, possibly related mediating variables arising from a true, unknown joint distribution. These estimators solve the efficient influence function; one that does so in one step and the other that is a substitution estimator that incorporates a series of targeting steps to optimize the

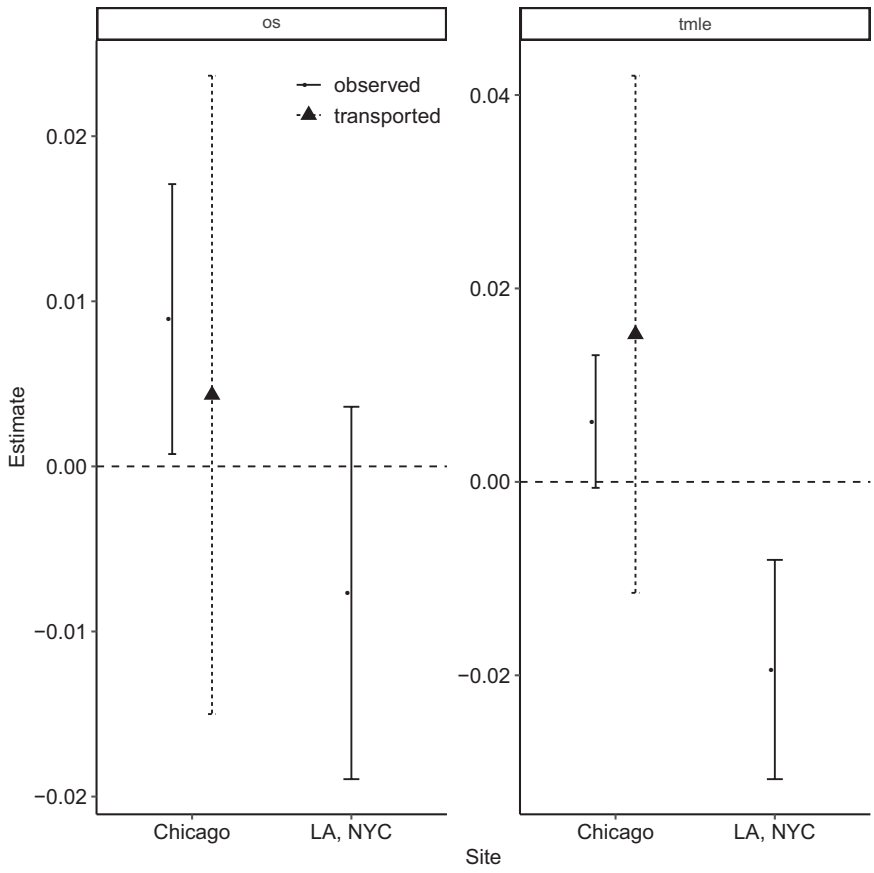

Fig. 1. Interventional indirect effects estimates of being randomized to the Section 8 voucher group on behavioral problems score in adolescence, 10–15 years later, mediated through features of the school environment, among girls. Estimates and 95% CIs for observed (nontransported) and transported predicted effects. All results were approved for release by the U.S. Census Bureau, authorization numbers CBDRB-FY20-ERD002-023 and CBDRB-FY20-ERD002-024.

bias-variance trade-off. We derived their multiple robustness properties and examined finite sample performance in a simulation study. Lastly, we applied our proposed estimators to better understand why a particular pathway from a housing intervention through changes in the school environment resulted in an unintended harmful effect on behavioral problems among girls in Chicago,when it led to improvements in behavioral problems among girls in other cities. However, in this illustrative example, the outcome data were, in fact, measured in the target population. Our proposed approach is arguably more useful in the scenario where outcome data are unobserved in the target population. An example of this could be a vaccine trial conducted in one or multiple countries where data consists of treatment (assignment to vaccine vs. placebo), intermediate variable (completion of treatment), mediators/surrogate outcomes (antibody titers), and long-term outcome (viral illness). One could use our proposed methods to predict the long-term effect of the same vaccine in a new target country where data on the treatment, intermediate variable, and mediators/surrogate outcomes have already been collected, but the long-term outcome has not yet had the opportunity to be observed.

8. SOFTWARE

We provide the R code to implement the proposed estimators, freely available at https://github.com/kararudolph/transport.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

This research was conducted as a part of the U.S. Census Bureau's Evidence Building Project Series. The U.S. Census Bureau has not reviewed the article for accuracy or reliability and does not endorse its contents. Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results were approved for release by the U.S. Census Bureau, authorization numbers CBDRB-FY20-ERD002-023 and CBDRB-FY20-ERD002-024.

Conflict of Interest: None declared.

FUNDING

KER's time was funded by the National Institute on Drug Abuse (R00DA042127).

REFERENCES

ANGRIST, J. D., IMBENS, G. W. AND RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.

ARNOLD, B. F., NULL, C., LUBY, S. P. AND COLFORD, J. M. (2018). Implications of wash benefits trials for water and sanitation—authors' reply. *The Lancet Global Health* **6**, e616–e617.

AVIN, C., SHPITSER, I. AND PEARL, J. (2005). Identifiability of path-specific effects. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. Morgan Kaufman, San Francisco, CA. MR2192340. pp. 357–363.

BENKESER, D., CARONE, M., VAN DER LAAN, M. J. AND GILBERT, P. (2016). Doubly-robust nonparametric inference on the average treatment effect. *Technical Report* 356, U.C. Berkeley Division of Biostatistics Working Paper Series.

BENKESER, D. AND VAN DER LAAN, M. (2016). The highly adaptive lasso estimator. In: *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. pp. 689–696.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. AND WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.

BICKEL, P. J., RITOV, Y., TSYBAKOV, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics* **37**, 1705–1732.

CHEN, X. AND WHITE, H. (1999). Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Transactions on Information Theory* **45**, 682–691.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C. Newey, W., Robins, J. (2019). Double machine learning for treatment and causal parameters. *Econometrics Journal* **21**, C1–C68.

DÍAZ, I., HEJAZI, N. S., RUDOLPH, K. E. AND VAN DER LAAN, M. J. (2020). Non-parametric efficient causal mediation with intermediate confounders. *Biometrika*, In Press. DOI: 10.1093/biomet/asaa085.

GRUBER, S. AND VAN DER LAAN, M. J. (2010). A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. *The International Journal of Biostatistics* **6**.

KLAASSEN, C. A. J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *The Annals of Statistics* **15**, 1548–1562.

LUEDTKE, A., CARONE, M. AND VAN DER LAAN, M. J. (2019). An omnibus non-parametric test of equality in distribution for unknown functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**, 75–99.

MILLER, T. R. (2015). Projected outcomes of nurse-family partnership home visitation during 1996–2013, USA. *Prevention Science* **16**, 765–777.

ORR, L., FEINS, J., JACOB, R., BEECROFT, E., SANBONMATSU, L., KATZ, L. F., LIEBMAN, J. B. AND KLING, J. R. (2003). *Moving to Opportunity: Interim Impacts Evaluation*. Washington DC: US Department of Housing and Urban Development, Office of Policy Development and Research.

PEARL, J. (2009). Myth, Confusion, and Science in Causal Analysis. *Technical Report* R-348, Cognitive Systems Laboratory, Computer Science Department University of California, Los Angeles, Los Angeles, CA.

PEARL, J. AND BAREINBOIM, E. (2018). Transportability across studies: a formal approach. Technical Report R-372, Cognitive Systems Laboratory, Dept. Computer Science, University of California, Los Angeles.

PETERSEN, M. L., SINISI, S. E. AND VAN DER LAAN, M. J. (2006). Estimation of direct causal effects. *Epidemiology* **17**, 276–284.

PFANZAGL, J. AND WEFELMEYER, W. (1985). Contributions to a general asymptotic statistical theory. *Statistics & Risk Modeling* **3**, 379–388.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized & nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

RUDOLPH, K. E., LEVY, J., SCHMIDT, N. M., STUART, E. A. AND AHERN, J. (2020). Using transportability to understand differences in mediation mechanisms across trial sites: applying a novel estimation approach to a large-scale housing voucher experiment. *Epidemiology* **31**, 523–533.

RUDOLPH, K. E., LEVY, J. AND VAN DER LAAN, M. J. (2020). Transporting stochastic direct and indirect effects to new populations. *Biometrics*, In Press. doi.org/10.1111/biom.13274.

RUDOLPH, K. E., SCHMIDT, N. M., GLYMOUR, M. M., CROWDER, R., GALIN, J., AHERN, J. AND OSYPUK, T. L. (2018a). Composition or context: using transportability to understand drivers of site differences in a large-scale housing experiment. *Epidemiology (Cambridge, MA)* **29**, 199–206.

RUDOLPH, K. E., SOFRYGIN, O., SCHMIDT, N. M., CROWDER, R., GLYMOUR, M. M., AHERN, J. AND OSYPUK, T. L. (2018b). Mediation of neighborhood effects on adolescent substance use by the school and peer environments. *Epidemiology* **29**, 590–598.

RUDOLPH, K. E., SOFRYGIN, O., ZHENG, W. AND VAN DER LAAN, M. J. (2017). Robust and flexible estimation of stochastic mediation effects: a proposed method and example in a randomized trial setting. *Epidemiologic Methods* **7**.

SANBONMATSU, L., KATZ, L. F., LUDWIG, J., GENNETIAN, L. A., DUNCAN, G. J., KESSLER, R. C., ADAM, E. K., MCDADE, T. AND LINDAU, S. T. (2011). Moving to opportunity for fair housing demonstration program: final impacts evaluation. Washington, DC: US Department of Housing and Urban Development, Office of Policy Development and Research.

TCHETGEN TCHETGEN, E. J. AND VANDERWEELE, T. J. (2014). On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology (Cambridge, MA)* **25**, 282.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288.

VAN DER LAAN, M. (2017). A generally efficient targeted minimum loss based estimator based on the highly adaptive lasso. *The International Journal of Biostatistics* **13**.

VAN DER LAAN, M. J. (2014). Targeted estimation of nuisance parameters to obtain valid statistical inference. *The International Journal of Biostatistics* **10**, 29–57.

VAN DER LAAN, M. J. AND PETERSEN, M. L. (2008). Direct effect models. *The International Journal of Biostatistics* **4**.

VAN DER LAAN, M. J., POLLEY, E. C. AND HUBBARD, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology* **6**.

VAN DER LAAN, M. J. AND ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York, NY: Springer.

VAN DER LAAN, M. J. AND RUBIN, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics* **2**.

VAN DER VAART, A. (2002). Semiparameric statistics. *Lectures on Probability Theory and Statistics*. Ecole d/Ete de Probabilities de Saint-Flour XXIX-1999. Lecture Notes in Math. 1781, 331-457. New York, NY: Springer. MR1385671.

VANDERWEELE, T. J., VANSTEELANDT, S. AND ROBINS, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology (Cambridge, MA)* **25**, 300.

WAGER, S. AND WALTHER, G. (2015). Adaptive concentration of regression trees, with application to random forests. *arXiv preprint arXiv:1503.06388*.

ZHENG, W. AND VAN DER LAAN, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In: *Targeted Learning*. Editors: van der Laan, MJ and Rose, S. New York, NY: Springer, pp. 459–474.

ZHENG, W. AND VAN DER LAAN, M. J. (2012). Targeted maximum likelihood estimation of natural direct effects. *The International Journal of Biostatistics* **8**, 1–40.

ZILL, N. (1990). *Behavior Problems Index Based on Parent Report*. Washington DC: Child Trends.

[*Received July 10, 2020; revised November 16, 2020; accepted for publication November 28, 2020*]