University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2015 *Paper* 335

Double Robust Estimation of Encouragement-design Intervention Effects Transported Across Sites

Kara E. Rudolph* Mark J. van der Laan†

*Johns Hopkins Bloomberg School of Public Health, kara.rudolph@gmail.com

†University of California - Berkeley, laan@berkeley.edu

Double Robust Estimation of Encouragement-design Intervention Effects Transported Across Sites

Kara E. Rudolph and Mark J. van der Laan

Abstract

We develop double robust targeted maximum likelihood estimators (TMLE) for transporting intervention effects from one population to another. Specifically, we develop TMLE estimators for three transported estimands: intent-to-treat average treatment effect (ATE) and complier ATE, which are relevant for encouragement-design interventions and instrumental variable analyses, and the ATE of the exposure on the outcome, which is applicable to any randomized or observational study. We demonstrate finite sample performance of these TMLE estimators using simulation, including in the presence of practical violations of the positivity assumption. We then apply these methods to the Moving to Opportunity trial, a multi-site, encouragement-design intervention in which families in public housing were randomized to receive housing vouchers and logistical support to move to low-poverty neighborhoods. This application sheds light on whether effect differences across sites can be explained by differences in population composition.

1 Introduction

Multi-site interventions are common in public health, public policy, and economics. Do we expect an intervention effect in one site to be the same as the intervention effect in another site? In many cases, we would answer "no" for one of two reasons. First, there could be differences in site-level variables related to intervention design/implementation or contextual variables, like the economy, that would modify intervention effectiveness. Such variables suggest that the intervention either is not the same or does not work the same in the two sites. Second, there could be differences in person-level variables—population composition—across sites that also modify intervention effectiveness. This could cause intervention effects to differ across sites even if the interventions are structured and implemented in an identical fashion.

That intervention effects may differ for sites with different population composition motivates previous work on transportability (Pearl and Bareinboim, 2011). Transportability (which has been discussed as generalizability (Cole and Stuart, 2010) and external validity (Rothwell, 2005)) is the idea of applying the results of an experiment in one setting/population to a target setting/population based on the observed characteristics of that target population. Pearl and Bareinboim have formalized this goal by developing transport formulas and enumerating the necessary assumptions associated with each transport formula (Pearl and Bareinboim, 2011).

These transport formulas can be applied to predict the effect of an intervention in a target population based on the observed composition of that population and intervention results from the original population. This prediction can be useful for researchers wanting to estimate the potential long-term effects of an intervention in a new site based on long-term follow-up results in an original site. An example of this would be predicting effects from the expansion of home-visiting programs for low-income pregnant women under the Affordable Care Act based on long-term follow-up results from the Nurse Family Partnership trials (Eckenrode et al., 2010).

Transported predictions may also be useful in determining the extent to which differences in intervention effects across sites can be explained by differences in population composition. An example of this, which we use to motivate this work, is from the Moving to Opportunity (MTO) trial (Kling et al., 2007). MTO is a five-site, encouragement-design intervention in which families in public housing were randomized to receive housing vouchers and logistical support to move to low-poverty neighborhoods. To date, there has been no quantitative examination of the underlying reasons for differences in MTO's effects across sites (Orr et al., 2003).

We are not aware of any literature on the development of estimators incorporating transport formulas. However, there is a related literature on generalizing results from randomized controlled trials. The simplest of these methods is post-stratification or non-parametric direct standardization (Miettinen, 1972), but this method breaks down when there are many population characteristics to control for or if those characteristics are continuous. Previous model-based approaches have involved Horvitz-Thompson weighting, propensity score matching, and principle stratification (Stuart et al., 2011; Cole and

Stuart, 2010; Frangakis, 2009). These are important contributions but may be limited by their reliance on correct model specification. In addition, with the exception of principal stratification, we know of no extensions of these methods to encouragement-design interventions. Model-based approaches for such an intervention design would involve models relating 1) site (or population) to covariates, 2) instrument to exposure conditional on covariates and relevant effect modifiers, and 3) exposure to outcome conditional on covariates and relevant effect modifiers.

We address this research gap by first extending Pearl and Bareinboim's transport formulas to the case of an encouragement design intervention such as MTO. Next we develop and evaluate targeted maximum likelihood estimators (TMLEs) for transporting three estimands: the intent-to-treat average treatment effect (ITTATE), the complier ATE (CATE), and the ATE of the effect of the exposure on the outcome, ignoring the instrument (henceforth referred to as the EATE). This estimation approach has several advantages. First, it is robust to multiple model misspecification scenarios. Second, TMLE is efficient. Third, we target marginal population quantities, which are most relevant to policy and program leaders, while allowing for potential effect modification across a high-dimensional vector of covariates. Fourth, these estimators can easily incorporate machine learning algorithms, thereby reducing bias due to model misspecification.

The paper is organized as follows. In Section 2, we introduce notation and define the structural causal model. In Sections 3-5, we develop a TMLE for each of the three estimands of interest. The ITTATE is discussed in Section 3, the EATE is discussed in Section 4, and the CATE is discussed in Section 5. A reader who is interested in one of the three estimands can skip the other two sections without compromising understanding. For each estimand, we present the identification result, robustness properties, influence function-based inference, and steps for computing the TMLE. In Section 6, we present results from a simulation study in which we demonstrate consistency, efficiency and robustness of each TMLE estimate under different model specification scenarios and degrees of practical positivity violations. In Section 7, we apply these methods to the MTO example. Section 8 concludes.

2 Notation and Structural Causal Model

We observe the following vector of data for each of n participants: $O = (S, W, A, Z, S{\times}Y)$. S is an indicator of the site; $S = 1$ for the site for which we have long-term follow-up data and $S = 0$ for the site for which we do not have follow-up data. W is a vector of covariates, the distribution of which depends on S. A is a binary instrument, which is randomized. Z is the binary exposure of interest. Y is the outcome of interest, which we only observe for those in the site with long-term outcome data ($S = 1$).

We assume each participant's data vector O is an independent, random draw from the unknown distribution P_0 on O. We also assume the following causal relationships:

$$S = f_S(U_S); W = f_W(S, U_W); A = f_A(W, S, U_A); Z = f_Z(W, S, A, U_Z); Y = f_Y(W, Z, U_Y),$$

where U_S, U_W, U_A, U_Z, U_Y are exogenous random variables. Knowledge of these functions

and knowledge of U define a causal model, and in particular, a model for the distribution of the observed data, O.

The objective is to develop a TMLE to estimate each of three target parameters: ψ_1, ψ_2, ψ_3. TMLE is a semiparametric estimation approach that has been described previously (van der Laan and Rubin, 2006). It results in a substitution estimator for a particular parameter of interest. It is double robust and locally efficient under regular asymptotically linear conditions. Its consistency and efficiency properties derive from the fact that it solves the efficient influence curve estimating equation, which itself determines the linear approximation of any regular asymptotically linear and efficient estimator. Thus, to develop a TMLE, we first need to identify the parameter of interest and derive that parameter's efficient influence curve. We go through each step in Sections 3-5 that follow.

3 Intent-to-treat average effect of instrument

ψ_1 is the intent-to-treat average effect of the instrument on the outcome for participants in the site without follow-up data ($S = 0$), defined to be $\psi_1 = E(Y^1 - Y^0|S = 0)$, where for each $a \in \{0, 1\}, Y^a$ denotes the counterfactual outcome that would be observed if instrument $A = a$ were assigned and if Y were observed for participants with $S = 0$.

3.0.1 Identification

Under the assumptions given below, this parameter can be identified from the true data distribution P_0 (a subscript of 0 added to any notation denotes the truth):

$$\begin{aligned}\psi_1 = \Psi_1(P) &\equiv E(Y^1 - Y^0|S = 0) \\ &\equiv E(\{E(E(Y|S = 1, W, A = 1, Z)|S = 0, W, A = 1) \\ &- E(E(Y|S = 1, W, A = 0, Z)|S = 0, W, A = 0)\}|S = 0),\end{aligned} \tag{1}$$

where Ψ_1 is the statistical target parameter defined as a mapping $\Psi_1 : \mathcal{M} \to R$ that takes a probability distribution P in statistical model $\mathcal{M}$, and maps it to a real number. The true value is obtained by applying Ψ_1 to the true distribution, P_0. The statistical model, $\mathcal{M}$, is the collection of probability distributions of O that satisfy assumption 3 (below) and possibly put restrictions on $P(A|W, S = 0)$. The other two assumptions do not put restrictions on the data, so they do not affect the statistical model.

The assumptions needed for identifiability are:

1. $E_0(Y \mid S = 0, W, A, Z) = E_0(Y \mid S = 1, W, A, Z)$,
2. A is independent of (Z^0, Y^0, Z^1, Y^1), given $W, S = 0$, and
3. $P_0(S = 1, A = a \mid W, Z) > 0$ $P_{0,W,Z|A=a,S=0}$-a.e. This is the positivity assumption and means that every $P(S = 1, A = a|W, Z)$ that one could draw from the true joint distribution of W, Z given $A = a$ and $S = 0$ must be greater than 0.

The proof of this identifiability result is in the supplementary Web Appendix.

3.0.2 Efficient influence curve and robustness properties

Let $\Psi_1(P) \equiv \Psi_1^1(P) - \Psi_1^0(P) = \psi_1$, where for each $a \in \{0,1\}, \Psi_1^a(P)$ is defined as in Equation 1 and denotes the counterfactual mean outcome one would observe if instrument $A = a$ were assigned and if Y were observed for participants with $S = 0$. In addition, let $\bar{Q}(s = 1, W, A, Z) = E(Y|S = 1, W, A, Z)$, let $g_A(A \mid s, W) = P(A \mid S = s, W)$, and let $g_Z(Z \mid a, s, W) = P(Z \mid A = a, S = s, W)$.

Result 1. *The efficient influence curve of* $P \to \Psi_1^a(P) = E_P(Y^a \mid S = 0)$ *on a model* $\mathcal{M}$ *that makes at most assumptions about* $P(A = a \mid W, S = 0)$ *and positivity is given by:*

$$D^a(P) = D_Y^a(P) + D_Z^a(P) + D_W^a(P),$$

where

$$\begin{aligned}
D_W^a(P) &= \frac{I(S=0)}{P(S=0)}\{E_P(\bar{Q}(s=1,W,a,Z) \mid S=0,W,A=a) \\
&\quad -E_P(E_P(\bar{Q}(s=1,W,a,Z) \mid S=0,W,A=a) \mid S=0)\} \\
D_Z^a(P) &= \frac{I(A=a,S=0)}{g_A(a \mid W,S=0)P(S=0)} \\
&\quad \times \left\{\bar{Q}(s=1,W,a,Z) - E_P(\bar{Q}(s=1,W,a,Z) \mid S=0,W,A=a)\right\} \\
D_Y^a(P) &= \frac{I(S=1,A=a)}{g_A(a \mid W,S=1)P(S=1))} \frac{g_Z(Z \mid A=a,W,S=0)}{g_Z(Z \mid A=a,W,S=1)} \frac{P(W \mid S=0)}{P(W \mid S=1)} \\
&\quad \times (Y - \bar{Q}(s=1,W,a,Z)).
\end{aligned}$$

We note that

$$\frac{P(W \mid S=0)}{P(W \mid S=1)} = \frac{P(S=0 \mid W)P(S=1)}{P(S=1 \mid W)P(S=0)}.$$

There are three scenarios under which an estimator that solves the efficient influence equation will be consistent (robustness result). First, the Y model may be misspecified if all other models are correct. Second, the S and A models may be misspecified if the Y and Z models are correct. Third, the S and Z models may be misspecified if the Y and A models are correct.

We provide the derivation for the robustness properties in the supplementary Web Appendix.

3.0.3 Targeted maximum likelihood estimator

There are two TMLEs that can be computed to estimate the transported ITTATE. In this section, we describe how to compute one of them and describe how to compute the other in the supplementary Web Appendix. This TMLE can be computed in one-step and is particularly suitable when Z is high dimensional.

Let $\bar{Q}^0$ be an initial estimate of $\bar{Q}$. Let $\bar{Q}^0_n, g_{Z,n}, g_{S,n}$, and $g_{A,n}$ be estimators of $\bar{Q}^0$, $g_Z(Z \mid a, s, W)$, $g_S(S \mid W)$, and $g_A(A \mid s, W)$, respectively. Consider submodel $\text{Logit}\bar{Q}^0_n(\epsilon) = \text{Logit}\bar{Q}^0_n + \epsilon C_Y(g_{Z,n}, g_{S,n})$, where

$$C_Y(g_Z, g_S, g_A) = \frac{I(S=1, A=a)}{g_A(a \mid W, S=1)P(S=1)} \frac{g_Z(Z \mid A=a, W, S=0)}{g_Z(Z \mid A=a, W, S=1)} \frac{P(W \mid S=0)}{P(W \mid S=1)},$$

noting again that

$$\frac{P(W \mid S=0)}{P(W \mid S=1)} = \frac{P(S=1)}{P(S=0)} \frac{g_S(0 \mid W)}{g_S(1 \mid W)}.$$

C_Y is called the clever covariate. Its components can be calculated as follows. $\bar{Q}^0_n$ can be estimated by the predicted values of Y from a regression of Y on A, Z, W among participants with $S_i = 1$. Alternatively, one could use a machine learning approach, but we will use regression terminology for simplicity. Similarly, $g_A(a \mid W, S=1)$ can be estimated by the predicted probabilities from a logistic regression model of $A = a$ on W among participants with $S_i = 1$, and $g_S(s|W)$ can be estimated by the predicted probabilities from a logistic regression model of $S = s$ on W. For binary $Z, g_Z(z|a, s, W)$ can be estimated by the predicted probabilities setting $A = a$ from a logistic regression model of $Z = z$ on A and W among strata of observations with $S_i = s$.

Let ϵ^0_n be the fitted coefficient for C_Y in the univariate logistic regression model of Y on C_Y using $\text{Logit}\bar{Q}^0_n$ as an off-set, using the binary log-likelihood loss function multiplied by $I(S=1, A=a)$ (i.e., only using the observations with $S_i = 1, A_i = a$). If Y is not on the $[0, 1]$ scale, it can be bounded as previously recommended (Gruber and van der Laan, 2010). The updated estimator is denoted with $\bar{Q}^1_n = \bar{Q}^0_n(\epsilon^0_n)$.

Next, run a regression of $E_0(\bar{Q}^1_n(s=1, W, a, Z) \mid S=0, W, A=a)$ by regressing the predicted values $\bar{Q}^1_n(s=1, W, a, Z)$ on W among strata of observations with $A_i = a, S_i = 0$. Denote this estimator of $\bar{Q}_{0,Z} = E_0(\bar{Q}_0(s=1, W, a, Z) \mid S=0, W, A=a)$ with $\bar{Q}^0_{Z,n}$. Consider the submodel

$$\text{Logit}\bar{Q}^0_{Z,n}(\epsilon) = \text{Logit}\bar{Q}^0_{Z,n} + \epsilon C_Z(g_{A,n}),$$

where,

$$C_Z(g_A) = \frac{I(A=a, S=0)}{g_A(a \mid W, S=0)P(S=0)}.$$

Let $\epsilon_{1,n}$ be the fitted coefficient for this univariate logistic regression model that uses the binary log-likelihood loss function treating $\bar{Q}^1_n(s=1, W, a, Z)$ as the outcome and $\text{Logit}\bar{Q}^0_{Z,n}$ as an offset, restricted to the observations with $S_i = 0, A_i = a$. Denote this update with $\bar{Q}^1_{Z,n} = \bar{Q}^0_{Z,n}(\epsilon_{1,n})$.

The TMLE of ψ^a_1 is given by $\psi^a_{1,n} = Q_{W,n|S=0}\bar{Q}^{1,a}_{Z,n}$, where $Q_{W,n|S=0}$ is the empirical distribution of W_i among those with $S_i = 0$ — a function of the participant's covariates. In other words, the TMLE of ψ^a_1 is the empirical mean of $\bar{Q}^{1,a}_{Z,n}$ among the observations with $S_i = 0$ and setting $A = a$. So, our final estimator is $\psi_1 = \psi^1_1 - \psi^0_1 = Q_{W,n|S=0}(\bar{Q}^{1,a=1}_{Z,n} - \bar{Q}^{1,a=0}_{Z,n})$. This is the empirical mean of the difference in $\bar{Q}^1_{Z,n}$ setting $a = 1$ versus $a = 0$ among observations with $S_i = 0$. This TMLE solves

the efficient influence function $P_n D^a(g_{Z,n}, g_{S,n}, g_{A,n}, \bar{Q}_n^{1,a}, \bar{Q}_{Z,n}^1, Q_{W,n|S=0}) = 0$, where $P_n f = 1/n \sum_i f(O_i)$ (the empirical mean of function $f(O)$).

We can conservatively estimate the variance of the TMLE with $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (D_n^1(O_i) - D_n^0(O_i))^2$, which is the sample variance of the efficient influence curve, which was given in Result 1.

4 Average effect of exposure, ignoring instrument

ψ_2 is the average effect of the exposure on the outcome for participants in the site without long-term follow-up data ($S = 0$), defined to be $\psi_2 = E_0(Y^1 - Y^0 \mid S = 0)$, where for each $z \in \{0, 1\}, Y^z$ denotes the counterfactual outcome that would be observed if exposure $Z = z$ were assigned and if Y were observed for participants with $S = 0$.

4.0.4 Identification

Under the assumptions given below, this parameter can be identified:

$$\begin{aligned} \psi_2 = \Psi_2(P) &\equiv E(Y^1 - Y^0 | S = 0) \\ &\equiv E(\{E(Y \mid S = 1, W, Z = 1) - E(Y \mid S = 1, W, Z = 0)\} \mid S = 0), \end{aligned} \tag{2}$$

where Ψ_2 is the statistical target parameter defined as $\Psi_2 : \mathcal{M} \to R$. The true value is obtained by applying Ψ_2 to the true distribution, P_0. The statistical model, $\mathcal{M}$, is the collection of probability distributions of O that satisfy assumption 3 (below). The other two assumptions do not put restrictions on the data, so they do not affect the statistical model.

The assumptions needed for identifiability are:

1. $E_0(Y|S = 0, W, Z) = E_0(Y|S = 1, W, Z)$,
2. Z is independent of (Y^0, Y^1) given W, and
3. $P_0(Z = z|W, S = 1) > 0$ $P_{0W|S=0} - a.e.$ This is the positivity assumption and means that every P(Z = z|W,S=1) that one could draw from the true distribution of W given S = 0 must be greater than 0.

The proof of this identifiability result is trivial and known from the average treatment effect literature.

4.0.5 Efficient influence curve

Let $\Psi_2(P) \equiv \Psi_2^1(P) - \Psi_2^0(P) = \psi_2$, where for each $z \in \{0, 1\}$, $\Psi_2^z(P)$ is defined as in Equation 2 and denotes the counterfactual mean outcome estimated if exposure $Z = z$ were assigned and if Y were observed for participants with $S = 0$. Unless otherwise specified, we use the same notation as in Section 3.

Result 2. *The efficient influence curve of* $P \rightarrow \Psi_2^z(P) = E_P(Y^z \mid S = 0)$ *on a model* $\mathcal{M}$ *that at most makes assumptions about* $P(Z = z|W, S = 0)$ *and positivity is given by*

$$\begin{aligned} D^z(P)(O) &= \frac{I(S=1,Z=z)}{P(S=1,Z=z \mid W)} \frac{P(S=0 \mid W)}{P(S=0)}(Y - \bar{Q}(s=1,W,z)) \\ &\quad + \frac{I(S=0)}{P(S=0)} \left\{ \bar{Q}(s=1,W,z) - E_P(\bar{Q}(s=1,W,z) \mid S=0) \right\}. \end{aligned}$$

We have two scenarios under which an estimator that solves the efficient influence curve will be consistent (robustness result). First, the S and Z models may be misspecified if the Y model is correct. Second, the Y model may be misspecified if the S and Z models are correct. We provide the derivation of the robustness properties in the supplementary Web Appendix.

4.0.6 Targeted maximum likelihood estimator

Consider the submodel $\text{Logit}\bar{Q}_n^0(\epsilon) = \text{Logit}\bar{Q}_n^0 + \epsilon C_Y(g_{Z,n}, g_{S,n})$, where

$$\begin{aligned} C_y(g_Z, g_S) &\equiv \frac{I(S=1,Z=z)}{P(S=1,Z=z \mid W)} \frac{P(S=0 \mid W)}{P(S=0)} \\ &= \frac{I(S=1,Z=z)}{P(Z=z \mid S=1,W)P(S=1 \mid W)} \frac{P(S=0 \mid W)}{P(S=0)}. \end{aligned}$$

The components of C_Y can be calculated as described in Section 3. Let ϵ_n^0 be the fitted coefficient for this clever covariate C_Y in the univariate logistic regression model of Y on C_Y using $\text{Logit}\bar{Q}_n^0$ as an off-set, using the binary log-likelihood loss function multiplied by $I(S = 1, Z = z)$ (i.e., only using the observations with $S_i = 1, Z_i = z$). Again, if Y is not on the $[0, 1]$ scale, it can be bounded as recommended previously (Gruber and van der Laan, 2010). The updated estimator is denoted with $\bar{Q}_n^1 = \bar{Q}_n^0(\epsilon_n^0)$.

The TMLE of ψ_2^z is given by $Q_{W,n|S=0}\bar{Q}_n^{1,z}$, which is the empirical mean of $\bar{Q}_n^1$ among the observations with $S_i = 0$ setting $Z = z$. So, our final estimator, $\psi_2 = \psi_2^1 - \psi_2^0 = Q_{W,n|S=0}(\bar{Q}_n^{1,z=1} - \bar{Q}_n^{1,z=0})$, is the empirical mean of the difference in $\bar{Q}_n^1$ setting $z = 1$ versus $z = 0$ among the observations with $S_i = 0$. This TMLE solves the efficient influence function $P_n D^z(g_{Z,n}, g_{S,n}, \bar{Q}_n^{1,z}) = 0$.

Again, we can conservatively estimate the variance of the TMLE with $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(D_n^1(O_i) - D_n^0(O_i))^2$, which is the sample variance of the efficient influence curve, which was given in Result 2.

5 Complier average effect of exposure, using instrument

ψ_3 is the complier average effect of the exposure on the outcome in the site without long-term follow-up data, defined to be $\psi_3 = E(Y^1 - Y^0|Z^1 - Z^0 = 1, S = 0)$, where for each

$a \in \{0,1\}, Y^a$ denotes the counterfactual outcome that would be observed if instrument $A = a$ were assigned and if Y were observed for participants with $S = 0$, and Z^a denotes the counterfactual exposure that would be observed if instrument $A = a$ were assigned. The complier average causal effect is also called the instrumental variables (IV) estimand and the local average instrument effect (LATE), even in the case of a binary instrument and binary exposure (Angrist et al., 1996).

5.0.7 Identification

Under the assumptions given below, this parameter can be identified from the true data distribution P_0:

$$\psi_3 = \Psi_3(P) \equiv E(Y^1 - Y^0 | Z^1 - Z^0 = 1, S = 0) = \tag{3}$$
$$\frac{E(E(E(Y|S=1,W,Z)|S=0,W,A=1) - E(E(Y|S=1,W,Z)|S=0,W,A=0)|S=0)}{E(E(Z|S=0,W,A=1) - E(Z|S=0,W,A=0)|S=0)}$$

where Ψ_3 is the statistical target parameter defined as $\Psi_3 : \mathcal{M} \to R$. The true value is obtained by applying Ψ_3 to the true distribution, P_0. The statistical model, $\mathcal{M}$, is the collection of probability distributions of O that satisfy assumption 5. Note that the numerator of this identification is the same as for ψ_1 as specified in Equation 1.

Identification relies on five assumptions.

1. $E_0(Y \mid S = 0, W, A, Z) = E_0(Y \mid S = 1, W, A, Z)$,
2. $A = f_A(U_A)$ is independent of $(Z^{a=0}, Z^{a=1}, Y^{a=0}, Y^{a=1})$, given $W, S = 0$,
3. $Y^{az} = Y^z$, which is the exclusion restriction assumption, stating that the instrument A only affects the outcome Y through the exposure Z,
4. $Z^1 - Z^0 \geq 0$, which is the monotonicity assumption, meaning that the instrument A cannot decrease exposure, and
5. $P_0(S = 1, A = a \mid W, Z) > 0$ $P_{0,W,Z|A=a,S=0}$-a.e. This is the positivity assumption and means that every $P(S = 1, A = a|W, Z)$ that one could draw from the true joint distribution of W, Z given $A = a$ and $S = 0$ must be greater than 0.

5.0.8 Efficient Influence Curve

Let $\Psi_3(P) \equiv \Psi_3^1(P) - \Psi_3^0(P) = \psi_3$, where for each $a \in \{0,1\}, \Psi_3^a(P)$ is defined as in Equation 3 and denotes the counterfactual mean outcome estimated if instrument $A = a$ were assigned and if Y were observed for participants with $S = 0$. Unless otherwise specified, we use the same notation as in Section 3.

Result 3. *The efficient influence curve of $P \to \Psi_3^a(P) = E_P(Y^a \mid S = 0)$ on a model $\mathcal{M}$ that at most makes assumptions about $P(A = a|W, S = 0)$, $P(Z = z|W, S = 0)$, and positivity is given by*

$$D^a_{\Psi_1,\tilde{\Psi}}(P) = \frac{1}{\tilde{\Psi}(P)} D_{\Psi_1}(P) - \frac{\Psi_1(P)}{\tilde{\Psi}(P)^2} D_{\tilde{\Psi}}(P),$$

where $\tilde{\Psi}(P)$ is the ATE TMLE, which is the nontransported average effect of the instrument on the exposure for participants with $S = 0$. The TMLE for the ATE has been described previously (van der Laan and Rubin, 2006).

An estimator that solves the above efficient influence curve will be consistent (robustness result) if both the numerator and denominator are correct. So, applying the robustness results from the ITTATE TMLE and for the ATE TMLE (van der Laan and Rubin, 2006) there are three scenarios under which this will happen. These scenarios are the same as those for the ITTATE. First, the Y model may be misspecified if the S, Z, and A models are correct. Second, the S and A models may be misspecified if the Y and Z models are correct. Third, the S and Z models may be misspecified if the Y and A models are correct. We provide the derivation for the efficient influence curve in the supplementary Appendix.

5.0.9 Targeted maximum likelihood estimator

This TMLE is estimated as the ratio of two TMLEs: the TMLE detailed in Section 3.0.3 over the TMLE for the average effect of A on Z as detailed previously (van der Laan and Rubin, 2006).

We refer to Section 3.0.3 for the steps to estimate the TMLE in numerator. The TMLE in the denominator can be estimated as follows (van der Laan and Rubin, 2006).

Consider the submodel $\text{Logit}\bar{Q}_n^0(\epsilon) = \text{Logit}\bar{Q}_n^0 + \epsilon C_Z(g_{A,n})$, where

$$C_Z(g_A) \equiv \frac{I(S = 1, A = a)}{P(S = 1, A = a \mid W)}$$

The components of C_Z can be calculated as described in Section 3. Let ϵ_n^0 be the fitted coefficient for this clever covariate C_Z in the univariate logistic regression model of Z on C_Z using $\text{Logit}\bar{Q}_n^0$ as an off-set, using the binary log-likelihood loss function multiplied by $I(S = 0, A = a)$ (i.e., only using the observations with $S_i = 0, A_i = a$). The updated estimator is denoted with $\bar{Q}_n^1 = \bar{Q}_n^0(\epsilon_n^0)$.

The TMLE in the denominator is $Q_{W,n|S=0}(\bar{Q}_n^{1,a=1} - \bar{Q}_n^{1,a=0})$, which is the empirical mean of the difference in $\bar{Q}_n^1$ setting $a = 1$ versus $a = 0$ among the observations with $S_i = 0$. This TMLE solves the efficient influence function $P_n D^a_{\Psi_1,\tilde{\Psi}} = 0$.

We can conservatively estimate the variance of the TMLE with the sample variance of it's efficient influence curve, which was given in Result 3. Alternatively, the variance can be estimated using the multivariate delta method, which we show in the supplementary Web Appendix.

6 Simulation Study

6.1 Overview and set-up

We conduct a simulation study to examine finite sample performance of the TMLE estimators for ψ_1, ψ_2, and ψ_3. We consider two data-generating mechanisms from the

same structural causal model, shown in Table 1. The magnitude of several coefficients increases in the second data-generating mechanism compared to the first, which results in practical positivity violations. Comparing performance between the two data-generating mechanisms allows us to examine sensitivity to the positivity assumption.

Table 1: Simulation data generating mechanisms.

Data-generating mechanism 1: without positivity violations	Data-generating mechanism 2: with positivity violations
$S \sim Ber(0.5)$	$S \sim Ber(0.5)$
$W_1 \sim Ber(0.4 + 0.2S)$	$W_1 \sim Ber(0.3 + 0.5S)$
$W_2 \sim N(0.1S, 1)$	$W_2 \sim N(0.5S, 1)$
$W_3 \sim N(1 + 0.2S, 1)$	$W_2 \sim N(1 + S, 2)$
$A \sim Ber(0.5)$	$A \sim Ber(0.5)$
$Z \sim Ber(-log(1.6) + log(4)A - log(1.1)W_2 - log(1.3)W_3)$	$Z \sim Ber(-log(1.6)+log(4)A-log(2)W_2+ log(2)W_3)$
$Y \sim Ber(log(1.6) + log(1.9)Z - log(1.3)W_3 - log(1.2)W_1 + log(1.2)AW_1)$	$Y \sim Ber(log(1.6) + log(1.9)Z - log(1.3)W_3 - log(1.2)W_1 + log(1.2)AW_1)$

Table 2: Characteristics of the clever covariate from the first simulation iteration for data-generating mechanisms 1 and 2 and from the application to the Moving to Opportunity Study.

	$C_Y(A=1)$			$C_Y(A=0)$		
	Mean (SD)	Min	Max	Mean (SD)	Min	Max
ITTATE						
Data-generating mechanism 1	0.55(0.26)	0.14	2.12	0.60(0.28)	0.15	2.13
Data-generating mechanism 2	1.16(1.58)	0.84×10^{-2}	21.35	1.25(1.64)	1.06×10^{-2}	22.82
Application	1.09(1.54)	0.83×10^{-2}	8.77	2.69(3.89)	1.73×10^{-2}	21.77
EATE						
Data-generating mechanism 1	0.49(0.38)	0.05	2.46	0.55(0.31)	0.14	1.75
Data-generating mechanism 2	1.07(1.62)	0.15×10^{-2}	26.26	1.33(2.26)	0.04×10^{-2}	41.49
Application	2.05(2.76)	4.54×10^{-2}	13.11	1.13(1.82)	1.01×10^{-2}	10.69

For each of the ITTATE, CATE, and EATE, we show TMLE estimator performance in terms of mean percent bias, closeness to the efficiency bound (mean estimator standard error (SE) $\times$ the square root of the number of observations), 95% confidence interval coverage, and mean squared error (MSE) across 10,000 simulations for a sample size of

N=5,000. We evaluate performance under correct model specification and various model misspecifications where misspecification of the S and Z models involved specifying a null model and misspecification of the Y included a term for Z only.

6.2 Results

As seen in Table 3, the TMLE estimators are consistent under the robustness properties derived for each estimand. Specifically, the TMLE estimators have less than 1% bias for all model specifications except when all of the models (site, exposure, and outcome models) are misspecified. The 95% CI for the TMLE estimator results in coverage of about 95% for unbiased estimates.

Table 3: Results from data-generating mechanism 1 without positivity violations. TMLE estimator performance under correct and incorrect model specification across 10,000 simulations in terms of percent bias, estimator standard error $\times\sqrt{n}$, 95% confidence interval coverage, and mean squared error. The estimator standard error $\times\sqrt{n}$ should be compared to the efficiency bound, which is 1.49 for the ITTATE, 4.50 for the CATE, and 1.60 for the EATE.

Specification	%Bias	SE$\times\sqrt{n}$	95%CI Cov	MSE
ITTATE				
All models correct	-0.67	1.50	95.01	0.0004
S model misspecified	-0.49	1.37	95.34	0.0004
Z model misspecified	-0.67	1.49	95.00	0.0004
Y model misspecified	-0.71	1.52	95.36	0.0005
S,Z models misspecified	-0.49	1.37	95.29	0.0004
S,Z,Y models misspecified	6.05	1.38	94.84	0.0004
CATE				
All models correct	-0.13	4.54	95.17	0.0041
S model misspecified	0.04	4.15	95.50	0.0034
Z model misspecified	-0.13	4.53	95.20	0.0041
Y model misspecified	-0.17	4.54	95.00	0.0042
S,Z models misspecified	0.05	4.14	95.37	0.0034
S,Z,Y models misspecified	6.60	4.18	94.76	0.0036
EATE				
All models correct	-0.31	1.60	94.94	0.0005
S model misspecified	-0.38	1.46	93.68	0.0005
Z model misspecified	-0.31	1.48	93.01	0.0005
Y model misspecified	-0.29	1.62	95.09	0.0005
S,Z models misspecified	-0.43	1.36	92.95	0.0004
S,Z,Y models misspecified	14.46	1.37	76.27	0.0009

Performance of these estimators in the presence of practical positivity violations is of interest for several reasons. First, the sites involved may have very different covariate dis-

tributions, which could contribute to such violations. Second, the sites may differ in how the instrument, A, is related to the exposure of interest, Z, which could also contribute to the violations. Third, predicted probabilities from these two models are multiplied together in the clever covariate, which may compound positivity violations from the first two sources. When there are practical violations of the positivity assumption, theory no longer guarantees consistency of the estimators (Petersen et al., 2010).

As compared to the results in Table 3 without positivity violations, Table 4 shows that in the presence of such violations even the estimators using correctly specified models are slightly biased. MSE is particularly compromised by these practical positivity violations due to increased variability across the simulations. Coverage for the TMLE EATE is also compromised because of this variability. The standardized TMLE EATE estimates are slightly skewed with heavier tails. Calculating the 95% CI coverage using the percentile method from bootstrapping corrects this under-coverage (results not shown but available upon request).

The presence of these positivity violations exacerbates sensitivity to model misspecification of the TMLE estimator. This is largely due to increased variability in the estimates across the simulations and non-normally distributed standardized estimates—the consequences of which are seen in the lower coverage and greater MSE.

Weight truncation is a common and easy-to-implement strategy that may lessen sensitivity to practical positivity violations. Although truncation has the potential to improve both bias and variance due to positivity violations, it may also increase bias due to misspecification (Petersen et al., 2010; Cole and Hernán, 2008; Bembom and van der Laan, 2008). We repeated the simulations under data-generating mechanism 2 truncating the clever covariate at several different lower bound/upper bound truncation levels: 0.01/100, 0.05/20, and 0.1/10. We compare the untruncated results to the truncated results in Table 5. Truncation resulted in the expected improvements in terms of reduced variance and MSE but compromised confidence interval coverage for all estimands. Truncation also resulted in increased bias for the EATE.

7 Application

7.1 Overview and set-up

We now apply the transportability estimators to an example from the Moving to Opportunity trial (MTO). MTO is a large-scale social policy experiment that has been described in the Introduction and previously (Kling et al., 2007). In discussing potential differences in effects across sites, MTO researchers concluded:

> *Of course, if it had been possible to attribute differences in impacts across sites to differences in site characteristics, that would have been very valuable information. Unfortunately, that was not possible. With only five sites, which differ in innumerable potentially relevant ways, it was simply not possible to disentangle the underlying factors that cause impacts to vary across sites.* (Orr et al., 2003, p.B11)

Table 4: Results from data-generating mechanism 2 with practical positivity violations. TMLE estimator performance under correct and incorrect model specification across 10,000 simulations in terms of percent bias, estimator standard error $\times\sqrt{n}$, 95% confidence interval coverage, and mean squared error. The estimator standard error $\times\sqrt{n}$ should be compared to the efficiency bound, which is 2.68 for the ITTATE, 11.33 for the CATE, and 4.09 for the EATE.

Specification	%Bias	SE$\times\sqrt{n}$	95%CI Cov	MSE
ITTATE				
All models correct	-0.88	2.68	94.85	0.0015
S model misspecified	0.32	1.38	95.19	0.0004
Z model misspecified	-0.88	2.77	95.61	0.0015
Y model misspecified	-0.41	2.85	95.81	0.0015
S,Z models misspecified	0.34	1.39	95.28	0.0004
S,Z,Y models misspecified	18.34	1.42	94.06	0.0004
CATE				
All models correct	2.57	11.42	94.96	0.0265
S model misspecified	3.73	5.84	95.42	0.0068
Z model misspecified	2.57	11.85	95.86	0.0265
Y model misspecified	3.06	11.42	94.85	0.0270
S,Z models misspecified	3.98	5.93	95.60	0.0069
S,Z,Y models misspecified	23.00	6.12	94.23	0.0085
EATE				
All models correct	0.18	3.60	91.36	0.0029
S model misspecified	1.98	1.96	86.33	0.0012
Z model misspecified	0.18	2.67	82.93	0.0029
Y model misspecified	2.09	4.17	96.05	0.0027
S,Z models misspecified	2.18	1.38	79.27	0.0009
S,Z,Y models misspecified	-52.11	1.41	2.49	0.0065

We ask whether our transportability estimators can shed light on this previously intractable problem. Taking two MTO sites, Boston and Los Angeles (LA), we test the null hypothesis that the predicted effect of the intervention on school dropout for LA equals the true effect for LA, where the predicted effect borrows the conditional outcome model from Boston and makes use of differing distributions of population characteristics between the sites through transport formulas. If we fail to reject the null, this suggests that the intervention may be transportable based on the covariates included in the transport formula. If we reject the null, it suggests that the intervention is not transportable given our measured covariates. We consider two estimands that are typically reported for MTO data: the ITTATE and the CATE.

We use the same school dropout outcome as reported previously for adolescents aged 15-19 years (completed less than 12 years of school, did not receive high school diploma or GED, and is not enrolled in school) (Sanbonmatsu et al., 2011). We define a binary

Table 5: Results of truncation of the clever covariate to lessen sensitivity to practical positivity violations (from data-generating mechanism 2). Estimator performance under correct model specification across 10,000 simulations in terms of percent bias, estimator standard error $\times\sqrt{n}$, 95% confidence interval coverage, and mean squared error. The estimator standard error $\times\sqrt{n}$ should be compared to the efficiency bound, which is 2.68 for the ITTATE, 11.33 for the CATE, and 4.09 for the EATE.

Truncation Level	%Bias	SE$\times\sqrt{n}$	95%CI Cov	MSE
ITTATE				
No modification	-0.88	2.68	94.85	0.0015
Truncation at 0.01/100	-0.87	2.68	94.87	0.0015
Truncation at 0.05/20	-0.87	2.48	93.71	0.0014
Truncation at 0.1/10	-0.74	2.05	89.45	0.0012
CATE				
No modification	2.57	11.42	94.96	0.0265
Truncation at 0.01/100	2.58	11.41	94.96	0.0265
Truncation at 0.05/20	2.59	10.57	93.80	0.0250
Truncation at 0.1/10	2.75	8.72	89.74	0.0221
EATE				
No modification	0.18	3.60	91.36	0.0029
Truncation at 0.01/100	2.29	3.23	92.50	0.0024
Truncation at 0.05/20	2.71	2.40	89.78	0.0016
Truncation at 0.1/10	2.60	1.90	84.96	0.0013

instrument as has been done previously: randomized receipt of a voucher to move versus no voucher (Osypuk et al., 2012). The exposure of interest is defined as moving to a low-poverty neighborhood during follow-up. Neighborhoods were defined based on participant addresses geocoded to Census tracts. Neighborhood poverty was calculated as the percent of residents living at or below the federal poverty line based on the 2000 Census. A low-poverty neighborhood was defined as less than 25% of residents living below poverty based on theory and breakpoints in the site-specific distributions. Population composition characteristics included an extensive set of baseline characteristics spanning several domains: sociodemographic characteristics of the adolescent and adult family member, behavior and learning characteristics of the adolescent, neighborhood characteristics, and reasons for participation. A full list of the characteristics included is in the supplementary Web Appendix. We consider two sites for simplicity. For the purpose of this illustration, we ignore MTO study weights and only consider participants with non-missing data (n=260 adolescents in the Boston site; n=270 adolescents in the LA site). A more in-depth analysis of this and other MTO effects is the topic of a future paper.

Because we do not know the true models relating the instrument to exposure, the exposure to the outcome, and covariates to site, we use nonparametric methods instead of the standard parametric regression models. Specifically, we use the nonparametric,

ensemble machine learning method, Superlearner (van der Laan et al., 2007), to generate the predicted probabilities needed for the TMLE transportability estimators. Superlearner has been described previously (van der Laan et al., 2007). Briefly, it weights multiple machine learning algorithms to minimize the cross-validated mean squared prediction error. We conducted a simulation showing that incorporating Superlearner into the TMLE transportability estimators did not change estimator performance (results not shown but available from the first author upon request). The clever covariate for the ITTATE TMLE estimator ranges from 0.83×10^{-2} to 21.77 (Table 2). This suggests that practical positivity violations may be a minor problem.

7.2 Results

Figure 1 shows the ITTATE and CATE estimates for the average effect of 1) being randomized to a voucher group and 2) moving to a low-poverty neighborhood on probability of dropping out of high school, respectively. We see that the true site-specific effects for Boston and LA differ. For Boston, the ITTATE is statistically significant, which suggests that the MTO intervention was successful in reducing high school dropout. We see no effect of the intervention on high school dropout for the LA site.

Our goal is to determine if the differences in these effect estimates between the sites can be explained by population characteristics. Specifically, we transport the effects estimated for the Boston site to the LA site using the population characteristics in LA but no outcome data. Figure 1 compares the TMLE transported ITTATE and CATE estimates to the site-specific estimates. We see that the transported estimates for LA are similar to true LA estimates, which means that the difference in effects between Boston and LA can be largely explained by population composition.

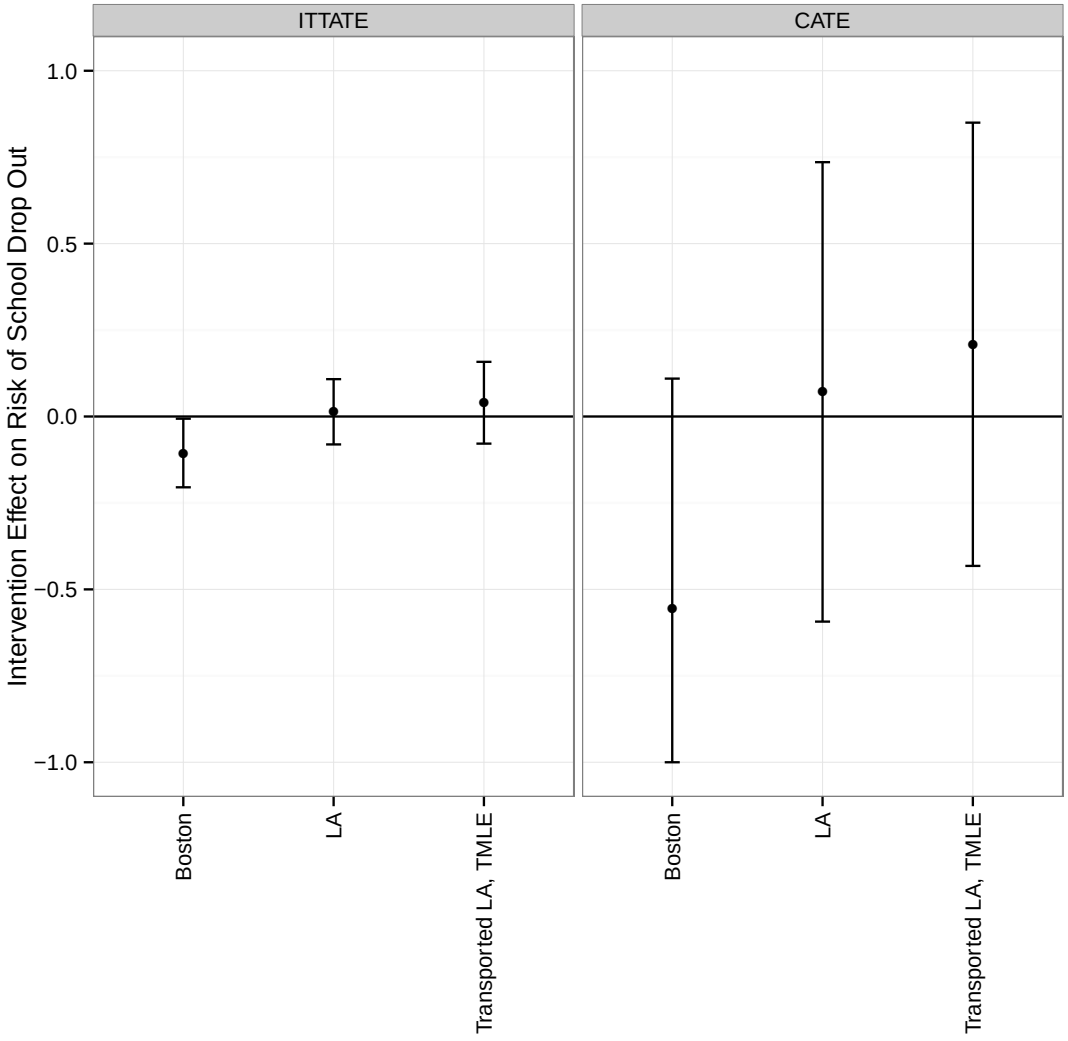
8 Conclusion

In this paper, we developed double robust TMLE estimators for transporting average treatment effects from a study population to a target population. This complements graphical work on the subject of transportability and fills the key gap in estimation strategies in this area (Pearl and Bareinboim, 2011). These transport estimators are applicable for encouragement design interventions as well as randomized experiments and observational studies.

Development of new estimators is useful insofar as they are practical and easy to implement. To facilitate the use of these estimators, we provide step-by-step instructions for implementing each transport TMLE in the article. In the supplementary Web Appendix, we provide R code for each estimator as well as sample code for application.

A limitation of these estimators is their sensitivity to practical violations of the positivity assumption. This limitation is not unique to these estimators, but applies to broad classes of estimators that rely on weights either exclusively or partially outside the $\bar{Q}$ model, e.g., TMLE estimators, inverse probability of treatment weighted (IPTW) estimators, and augmented IPTW (A-IPTW) estimators (Robins et al., 2007). Truncation

Figure 1: Application results: average effect estimates and 95% confidence intervals using data from the Moving to Opportunity Interim Follow-up. The ITTATE is interpreted as the effect of being randomized to one of the voucher groups on risk of dropping out of high school. The CATE is interpreted as the effect of moving to a low-poverty neighborhood on the risk of dropping out of high school.



of the clever covariate, which is related to the general strategy of weight truncation, is a common strategy to deal with this limitation (Petersen et al., 2010; Cole and Hernán, 2008; Bembom and van der Laan, 2008), but we found that it did not appreciably improve performance in our simulations. Although it slightly improved MSE, the trade-off was increased bias and reduced CI coverage. An area for future work is to optimize estimator performance in the presence of such practical positivity violations. We are currently pursuing two strategies. The first is to reduce instances of practical positivity violations by drawing on the screening and pruning strategies employed in collaborative TMLE (van der Laan and Gruber, 2010). The second is reduce the influence of practical positivity violations by moving part of the clever covariate into the $\bar{Q}$ model (Stitelman et al., 2012). This is a middle ground between TMLE and weighted G-computation, the latter of which has been shown to be robust to practical positivity assumptions (Kang and Schafer, 2007; Robins et al., 2007; Rudolph et al., 2014).

In an era of shrinking budgets, it is important to recognize that what works in one population may not work for another population so that resources can be targeted optimally. Applying these TMLE estimators to examine site differences in multi-site epidemiologic studies and large-scale policy or program interventions contribute to achieving that goal.

Acknowledgements

Kara Rudolph was supported by the Robert Wood Johnson Foundation Health & Society Scholars Program. Mark van der Laan was supported by 5 R01 AI074345-07.

A Identification proofs

A.1 ITTATE

ψ_1 is the intent-to-treat average effect of the instrument on the outcome for participants in the site without follow-up data ($S = 0$), defined to be $\psi_1 = E(Y^1 - Y^0|S = 0)$, where for each $a \in \{0, 1\}, Y^a$ denotes the counterfactual outcome that would be observed if instrument $A = a$ were assigned and if Y were observed for participants with $S = 0$.

The assumptions needed for identifiability are:

1. $E_0(Y \mid S = 0, W, A, Z) = E_0(Y \mid S = 1, W, A, Z)$,
2. A is independent of (Z^0, Y^0, Z^1, Y^1), given $W, S = 0$, and
3. $P_0(S = 1, A = a \mid W, Z) > 0$ $P_{0,W,Z|A=a,S=0}$-a.e. This is the positivity assumption and means that every $P(S = 1, A = a|W, Z)$ that one could draw from the true joint distribution of W, Z given $A = a$ and $S = 0$ must be greater than 0.

Proof.

$$\begin{aligned}
\psi_1 = \Psi_1(P) &\equiv E_0(\{E_0(E_0(Y|S=1,W,A=1,Z)|S=0,W,A=1) \\
&\quad -E_0(E_0(Y|S=1,W,A=0,Z)|S=0,W,A=0)\}|S=0)
\end{aligned}$$

By assumption 1,

$$\begin{aligned}
&\equiv E_0(\{E_0(E_0(Y|S=0,W,A=1,Z)|S=0,W,A=1) \\
&\quad -E_0(E_0(Y|S=0,W,A=0,Z)|S=0,W,A=0)\}|S=0)
\end{aligned}$$

By assumption 2, $P(Z=z \mid S=0,W,A=a) = P(Z^a=z \mid S=0,W)$, so

$$\begin{aligned}
&\equiv E_0(E_0(E_0(Y^a \mid S=0,W,Z^a) \mid W,S=0) \mid S=0) \\
&\equiv E(Y^a \mid S=0)
\end{aligned}$$

By assumption 3, we have that ψ_1 is defined. □

A.2 CATE

ψ_3 is the complier average effect of the exposure on the outcome in the site without long-term follow-up data, defined to be $\psi_3 = E(Y^1 - Y^0|Z^1 - Z^0 = 1, S = 0)$, where for each $a \in \{0,1\}, Y^a$ denotes the counterfactual outcome that would be observed if instrument $A = a$ were assigned and if Y were observed for participants with $S = 0$, and Z^a denotes the counterfactual exposure that would be observed if instrument $A = a$ were assigned.

The assumptions needed for identifiability are:

1. $E_0(Y \mid S=0,W,A,Z) = E_0(Y \mid S=1,W,A,Z)$,
2. $A = f_A(U_A)$ is independent of $(Z^{a=0}, Z^{a=1}, Y^{a=0}, Y^{a=1})$, given $W, S = 0$,
3. $Y^{az} = Y^z$, which is the exclusion restriction assumption, stating that the instrument A only affects the outcome Y through the exposure Z,
4. $Z^1 - Z^0 \geq 0$, which is the monotonicity assumption, meaning that the instrument A cannot decrease exposure, and
5. $P_0(S=1, A=a \mid W,Z) > 0$ $P_{0,W,Z|A=a,S=0}$-a.e. This is the positivity assumption and means that every $P(S=1, A=a|W,Z)$ that one could draw from the true joint distribution of W, Z given $A = a$ and $S = 0$ must be greater than 0.

Proof.

$$\begin{aligned}
\psi_3 = \Psi_3(P) &\equiv \frac{\Psi_1(P)}{\tilde{\Psi}(P)} \\
&\equiv \{E_0(\{E_0(E_0(Y|S=1,W,A=1,Z)|S=0,W,A=1) \\
&\quad -E_0(E_0(Y|S=1,W,A=0,Z)|S=0,W,A=0)\}|S=0)\} \\
&\quad /\{E_0(E_0(Z|S=0,W,A=1) - E_0(Z|S=0,W,A=0)|S=0)\}
\end{aligned}$$

By assumption 1,

$$\begin{aligned}
&\equiv \{E_0(\{E_0(E_0(Y|S=0,W,A=1,Z)|S=0,W,A=1) \\
&\quad -E_0(E_0(Y|S=0,W,A=0,Z)|S=0,W,A=0)\}|S=0)\} \\
&\quad /\{E_0(E_0(Z|S=0,W,A=1) - E_0(Z|S=0,W,A=0)|S=0)\}
\end{aligned}$$

By assumption 2, $P(Z = z \mid S = 0, W, A = a) = P(Z^a = z \mid S = 0, W)$, so

$$\begin{aligned}
&\equiv \{E_0(E_0(E_0(Y^1 \mid S=0,W,Z^1) \mid S=0,W) \\
&\quad -E_0(E_0(Y^0 \mid S=0,W,Z^0) \mid S=0,W) \mid S=0)\} \\
&\quad /\{E_0(E_0(Z^1|S=0,W) - E_0(Z^0|S=0,W|S=0)\} \\
&\equiv \frac{E_0(Y^1 - Y^0 \mid S=0)}{E_0(Z^1 - Z^0 \mid S=0)} \\
&\equiv \{E_0(Y^1 - Y^0|Z^1 - Z^0 = 1, S=0)P_0(Z^1 - Z^0 = 1|S=0) \\
&\quad +E_0(Y^1 - Y^0|Z^1 - Z^0 = 0, S=0)P_0(Z^1 - Z^0 = 0|S=0) \\
&\quad +E_0(Y^1 - Y^0|Z^1 - Z^0 = -1, S=0)P_0(Z^1 - Z^0 = -1|S=0)\} \\
&\quad /E_0(Z^1 - Z^0 \mid S=0)
\end{aligned}$$

By assumption 3,

$$\begin{aligned}
&\equiv \{E_0(Y^1 - Y^0|Z^1 - Z^0 = 1, S=0)P_0(Z^1 - Z^0 = 1|S=0) \\
&\quad +E_0(Y^1 - Y^0|Z^1 - Z^0 = -1, S=0)P_0(Z^1 - Z^0 = -1|S=0)\} \\
&\quad /E_0(Z^1 - Z^0 \mid S=0)
\end{aligned}$$

By assumption 4,

$$\equiv \frac{E_0(Y^1 - Y^0|Z^1 - Z^0 = 1, S=0)P_0(Z^1 - Z^0 = 1|S=0)}{E_0(Z^1 - Z^0 \mid S=0)}$$

By assumption 4, $Z^1 - Z^0$ can only have values in$\{0, 1\}$,
so the conditional expression is a conditional probability.

$$\begin{aligned}
&\equiv \frac{E_0(Y^1 - Y^0|Z^1 - Z^0 = 1, S=0)E(Z^1 - Z^0 \mid S=0)}{E_0(Z^1 - Z^0 \mid S=0)} \\
&\equiv E_0(Y^1 - Y^0|Z^1 - Z^0 = 1, S=0)
\end{aligned}$$

By assumption 5, we have that ψ_3 is defined. □

B Robustness results

We derived each efficient influence curve $D(P)(o)$ of Ψ on the nonparametric model by noting that it is given by the Gateaux derivative of Ψ at P in the direction $(\delta_o - P\delta_o)$, where δ_o is the probability distribution that puts mass 1 on o (Gill et al., 1989): $D(P)(o) = \lim_{\epsilon \to 0} \Psi(P + \epsilon(\delta_o - P(o)) - \Psi(P)/\epsilon$. In addition, we note that the efficient influence curve (EIC) for a parameter $\Psi(P) = \Psi(Q(P))$ is not affected by model assumptions on nuisance parameters whose tangent space is orthogonal to the tangent space of Q, so the EIC of the nonparametric model equals the EIC for the actual model.

B.1 ITTATE

B.1.1 Efficient influence curve

The following result provides the conditions under which an efficient influence curve-based estimator is consistent. Let $Q_W(w|s) = P(W \leq w \mid S = s), \bar{Q} = (Q_W, \bar{Q}(s = 1, W, A, Z))$, and $q_W(w|s) = \frac{dQ_W}{d\mu}(w)$ (i.e., density with respect to the appropriate dominating measure μ). Let $G_Z\bar{Q}(W) = E_P(\bar{Q}(s = 1, W, a, Z) \mid A = a, W, S = 0)$. A subscript of 0 added to any of the above notation denotes the truth. For example, $g_{Z,0}(Z \mid a, s, W)$ denotes the true exposure mechanism.

Result 4. *We have*

$$P_0 D^a(P) = \Psi^a(P_0) - \Psi^a(P) + R_2^a(P, P_0),$$

where

$$R_2^a(P, P_0) = \sum_{j=1}^{4} R_{2j}^a(P, P_0),$$

and

$$\begin{aligned}
R_{21}^a &= Q_{W,0|S=0} \left(\frac{g_{A,0|S=0}}{g_{A|S=0}} \frac{P_0(S=0)}{P(S=0)} - 1 \right) (G_{0,Z} - G_Z)\bar{Q} \\
R_{22}^a &= Q_{W|S=0} G_Z \left(\frac{g_{A,0|S=1}}{g_{A|S=1}} \frac{q_{W,0|S=1}}{q_{W|S=1}} \frac{P_0(S=0)}{P(S=0)} \frac{g_{Z,0|S=1}}{g_{Z|S=1}} - 1 \right) (\bar{Q}_0 - \bar{Q}) \\
R_{23}^a &= (Q_{W,0|S=0} - Q_{W|S=0}) G_Z (\bar{Q} - \bar{Q}_0) \\
R_{24}^a &= Q_{W,0|S=0} (G_{Z,0} - G_Z)(\bar{Q} - \bar{Q}_0).
\end{aligned}$$

Inspection of the second order term R_2^a gives three scenarios under which an estimator $\Psi_1^a(P_n)$ that solves the efficient influence equation $P_n D^a(P_n)$ will be consistent (robustness result). First, if $Q_W, G_Z, g_A, q_{W|S=1}$, and $P(S = 0)$ are correctly specified in the sense that $(Q_{W|S=0} - Q_{W,0|S=0}) G_Z (\bar{Q} - \bar{Q}_0) = 0$, $G_Z = G_{Z,0}$, $g_{A|S=1} = g_{A,0|S=1}$, $q_{W|S=1} = q_{W,0|S=1}$,and $P(S - 0) = P_0(S = 0)$. In other words, the Y model may be misspecified if all other models are correct. Second, if $\bar{Q}$ and G_Z are correctly specified in the sense that $\bar{Q} = \bar{Q}_0$, $G_Z = G_{Z,0}$. In other words, the S and A models may be

misspecified if the Y and Z models are correct. Third, if $\bar{Q}$ and g_A (and $P(S=0)$) are correctly specified in the sense that $\bar{Q} = \bar{Q}_0$, $g_{A|S=0} = g_{A,0|S=0}$, $P(S-0) = P_0(S=0)$. In other words, the S and Z models may be misspecified if the Y and A models are correct.

Corollary 1. *If $P_0 D^a(P) = 0$ and one of three above scenarios holds, then $\Psi_1^a(P) = \Psi_1^a(P_0)$.*

The implication of this corollary is that if P_n of P_0 converges to a P so that $P_0 D(P) = 0$ (which will be true for a TMLE since a TMLE solves $P_n D(P_n) = 0$) and one of the above three scenarios holds, then $\Psi_1^a(P_n)$ is consistent for $\Psi_1^a(P_0)$.

B.2 EATE

The following result provides the conditions for consistency based on the efficient influence curve.

Result 5.

$$P_0 D^z(P) \quad = \quad \Psi_2^z(P_0) - \Psi_2^z(P) + R_2^z(P, P_0),$$

where

$$R_2^z(P, P_0) = R_{21}^z(P, P_0) + R_{22}^z(P, P_0) + R_{23}^z(P, P_0),$$

and

$$\begin{aligned}
R_{21}^z(P, P_0) &= \int \frac{g_0(S=1, z \mid W) - g(S=1, z \mid W)}{g(S=1, z \mid W)} (\bar{Q}_0 - \bar{Q})(W) \frac{g_S(S=0 \mid W)}{P(S=0)} dP_0(W) \\
R_{22}^z(P, P_0) &= (Q_W - Q_{0,W})(\bar{Q}_0 - \bar{Q}) \\
&\quad + \int \frac{g_S(S=0 \mid W)}{P(S=0)} (\bar{Q}_0 - \bar{Q}) d(P - P_0)(W) \\
R_{23}^z(P, P_0) &= \frac{(P_0 - P)(S=0)}{P(S=0)} (Q_{0,W} - Q_W)\bar{Q}.
\end{aligned}$$

Considering the case that $P(S=1) = P_0(S=1)$ so that $R_{23}^z = 0$, we have two scenarios under which an estimator that solves the efficient influence curve will be consistent (robustness result). First, if $\bar{Q} = \bar{Q}_0$ (i.e., the Y model is correct). Second, if $P(S=1, Z=z \mid W) = P_0(S=1, Z=z \mid W)$ and $(Q_W, P_W) = (Q_{0,W}, P_{0,W})$ (i.e., the S and Z models are correct).

Corollary 2. *If $P_0 D^z(P) = 0$ and one of two above scenarios holds, then $\Psi_2^z(P) = \Psi_2^z(P_0)$.*

The implication of this corollary is that if P_n of P_0 converges to a P so that $P_0 D(P) = 0$ (which will be true for a TMLE since a TMLE solves $P_n D(P_n) = 0$) and one of the above two scenarios holds, then $\Psi_2^z(P_n)$ is consistent for $\Psi_2^z(P_0)$.

B.3 CATE

B.3.1 Efficient influence curve

We know the efficient influence curve of Ψ_1 and $\tilde{\Psi}$, so by the delta method, the efficient influence curve of Ψ_3 is given by the following ratio:

$\Psi_3(P) = \frac{\Psi_1(P)}{\tilde{\Psi}(P)}$, where $\tilde{\Psi}(P)$ is the TMLE of the estimand in the denominator.

$$\begin{aligned} D^a_{\Psi_1,\tilde{\Psi}}(P) &= \left(\frac{\Psi_1(P)}{\tilde{\Psi}(P)}\right)' \\ &= \frac{\tilde{\Psi}(P)\Psi_1(P)' - \Psi_1(P)\tilde{\Psi}(P)'}{\tilde{\Psi}(P)^2} \\ &= \frac{\tilde{\Psi}(P)D_{\Psi_1}(P) - \Psi_1(P)D_{\tilde{\Psi}}(P)}{\tilde{\Psi}(P)^2} \\ &= \frac{1}{\tilde{\Psi}(P)}D_{\Psi_1}(P) - \frac{\Psi_1(P)}{\tilde{\Psi}(P)^2}D_{\tilde{\Psi}}(P) \end{aligned}$$

C Alternative ITTATE TMLE

Let $Pa(Z)$ represent variables that are parents of Z. If Z is binary, then we can use that for any function $S(Z \mid Pa(Z))$ with conditional mean zero, given $Pa(Z)$, we have

$$S(Z \mid Pa(Z)) = (S(1 \mid Pa(Z)) - S(0 \mid Pa(Z))(Z - E(S \mid Pa(Z)))$$

(Van der Laan and Robins, 2003). Therefore, we can rewrite $D^a_Z(P)$ as follows:

$$\begin{aligned} D^a_Z(P) &= \frac{I(A=a,S=0)}{g_A(a \mid W, S=0)P(S=0)}\left\{\bar{Q}(s=1,W,a,z=1) - \bar{Q}(s=1,W,a,z=0)\right\} \\ &\quad \times(Z - g_Z(Z=1 \mid S=0,W,A=a)) \\ &\equiv C_Z(g_A,\bar{Q})(Z - g_Z(Z=1 \mid S=0,W,A=a)). \end{aligned}$$

As in TMLE I, consider submodel $\text{Logit}\bar{Q}^0_n(\epsilon) = \text{Logit}\bar{Q}^0_n + \epsilon C_Y(g^0_{Z,n}, g_{S,n})$, and let ϵ^0_n be the fitted coefficient for this clever covariate C_Y in the univariate logistic regression model using $\text{Logit}\bar{Q}^0_n$ as off-set, using the binary log-likelihood loss function multiplied with $I(S=1,A=a)$ (i.e., only using the observations with $S_i = 1, A_i = a$). The updated estimator is denoted with $\bar{Q}^1_n = \bar{Q}^0_n(\epsilon^0_n)$. Consider the submodel

$$\text{Logit}\bar{g}^0_{Z,n}(\epsilon) = \text{Logit}\bar{g}^0_{Z,n} + \epsilon C_Z(g_{A,n}, \bar{Q}^0_n).$$

Let ϵ^0_{1n} be the fitted coefficient using logistic regression of Z on W among the observations with $(S_i = 0, A_i = a)$, using predicted values $Logit\bar{g}^0_{Z,n}$ as an offset. This defines now $g^1_{Z,n} = g^0_{Z,n}(\epsilon^0_{1n})$. This process can be iterated: Let $k = 0$, set $\bar{Q}^{k+1}_n = \bar{Q}^k_n(\epsilon^k_n)$ and $g^{k+1}_{Z,n} = g^k_{Z,n}(\epsilon^k_{1n})$, set $k \leftarrow k+1$, and repeat until convergence. Assume that $\epsilon^k_n, \epsilon^k_{1n}$ converge to zero as $k \to \infty$ or that at a step K we have that $P_n D^a(g_{A,n}, g^K_{Z,n}, g_{S,n}, \bar{Q}^K_n, Q_{W,n|S=0}) =$

$o_P(1/\sqrt{n})$. Let $g^*_{Z,n}, \bar{Q}^*_n$ denote the resulting final fits. The TMLE of $\psi^a_{0,1}$ is defined by the substitution estimator $\psi^{a*}_{n,1} = \Psi^a_1(g^*_{Z,n}, \bar{Q}^*_n, Q_{W,n|S=0})$. This TLME solves

$$P_n D^a(g_{S,n}, g_{A,n}, g^*_{Z,n}, \bar{Q}^*_n, Q_{W,n|S=0}) = 0 \text{ or } o_P(1/\sqrt{n}).$$

D Alternative variance estimate of ψ_3

Alternatively, $var(\psi_3)$ can be estimated using the multivariate delta method.

$$\begin{aligned}
Var\left(\frac{\Psi_1}{\tilde{\Psi}}\right) &= \left(\nabla\frac{\Psi_1}{\tilde{\Psi}}\right)' Cov(\Psi_1, \tilde{\Psi})\left(\nabla\frac{\Psi_1}{\tilde{\Psi}}\right) \\
&= \left[\frac{1}{\tilde{\mu}}, \frac{-\mu_1}{\tilde{\mu}^2}\right]\begin{bmatrix} \sigma_1^2 & \sigma_1\tilde{\sigma} \\ \sigma_1\tilde{\sigma} & \tilde{\sigma}^2 \end{bmatrix}\begin{bmatrix} \frac{1}{\tilde{\mu}} \\ \frac{-\mu_1}{\tilde{\mu}^2} \end{bmatrix} \\
&= \left[\frac{\sigma_1^2}{\tilde{\mu}} - \frac{\mu_1\sigma_1\tilde{\sigma}}{\tilde{\mu}^2}, \frac{\sigma_1\tilde{\sigma}}{\tilde{\mu}} - \frac{\tilde{\sigma}^2\mu_1}{\tilde{\mu}^2}\right]\begin{bmatrix} \frac{1}{\tilde{\mu}} \\ \frac{-\mu_1}{\tilde{\mu}^2} \end{bmatrix} \\
&= \frac{\sigma_1^2}{\tilde{\mu}^2} - \frac{\mu_1\sigma_1\tilde{\sigma}}{\tilde{\mu}^2} - \frac{\mu_1\sigma_1\tilde{\sigma}}{\tilde{\mu}^3} + \frac{\tilde{\sigma}^2\mu_1^2}{\tilde{\mu}^4} \\
&= \frac{1}{\tilde{\mu}^2}\left(\sigma_1^2 - \frac{2\mu_1\sigma_1\tilde{\sigma}}{\tilde{\mu}} + \frac{\tilde{\sigma}^2\mu_1^2}{\tilde{\mu}^2}\right) \\
&= \frac{\mu_1^2}{\tilde{\mu}^2}\left(\frac{\sigma_1^2}{\mu_1^2} - \frac{2\sigma_1\tilde{\sigma}}{\mu_1\tilde{\mu}} + \frac{\tilde{\sigma}^2}{\tilde{\mu}^2}\right)
\end{aligned}$$

E Baseline covariates used in MTO application

An extensive set of baseline characteristics were included in applying our transport estimators to the MTO research question.

- Adolescent characteristics: age, gender, race, number of family members.
- Characteristics related to the child's behavior and learning: child was suspended or expelled from school during 2 years prior to baseline, child had gone to a special class or school or had gotten special help in school for behavioral or emotional problems during 2 years prior to baseline, child had gone to a special class or school or had gotten special help in school for a learning problem during 2 years prior to baseline, someone from school asked to discuss problems the child had with schoolwork or behavior during the 2 years prior to baseline, child enrolled in special class for gifted and talented students, child had problems that made it difficult to get to school or play active games/sports.
- Adult family member characteristics included: level of education, marital status, age at birth of the adolescent, work status, receipt of AFDC/TANF, car status, disability status.

- Neighborhood characteristics: family lived in neighborhood for at least 5 years; felt neighborhood streets were unsafe at night; household member had been assaulted, threatened with a knife or gun, or robbed during the 6 months prior to baseline; chat with a neighbor at least once per week; would likely tell neighbor if neighbor's child was getting into trouble; family living in neighborhood; friends in neighborhood; neighborhood satisfaction.
- Reported reasons for participating in MTO: to get away from drugs or gangs, to have access to better schools.
- Moving-related characteristics: confidence about finding an apartment in a different part of the city, moved more then 3 times during the 5 years prior to baseline, and previous application for Section 8 voucher.

F R code

F.1 Code for TMLE functions

```
## a variable needs to be named a and have values 0/1
## site variable needs to be named 'site' and needs to have value 0 for the site where the outcome data is not used and value 1 for the site where the outcome data is used
## z variable needs to be named z and have values 0/1
## y variable needs to be named y and have values 0/1
## w variables in a dataframe named w and with names w1:wx

ittatetmle<-function(a, z, y, site, w, aamodel, asitemodel, azmodel, aoutmodel, aq2model){
  datw<-w
  n.dat<-nrow(datw)

  #calculate components of clever covariate
  cpa <- predict(glm(formula=aamodel, family="binomial", data=data.frame(cbind(datw, a=a))), newdata=datw, type="response")
  cps <- predict(glm(formula=asitemodel, data=data.frame(cbind(site=site, datw)), family="binomial"), type="response")

  zmodels0 <- glm(formula=azmodel, data=data.frame(cbind(a=a, z=z, site=site, datw)), subset=site==0, family="binomial")
  zmodels1 <- glm(formula=azmodel, data=data.frame(cbind(a=a, z=z, site=site, datw)),subset=site==1, family="binomial")
  data_new0<-data_new1<-datw
  data_new0$a<-0
  data_new1$a<-1

  dga1s0<-dbinom(z, 1,prob=predict(zmodels0, newdata=data_new1, type="response"))
  dga1s1<-dbinom(z, 1,prob=predict(zmodels1, newdata=data_new1, type="response"))
```

```
dga0s0<-dbinom(z, 1,prob=predict(zmodels0, newdata=data_new0, type="response"))
dga0s1<-dbinom(z, 1,prob=predict(zmodels1, newdata=data_new0, type="response"))

#calculate clever covariate
g0w<-(1-cpa)*(dga0s1/dga0s0)*(cps/(1-cps))
g1w<-cpa*(dga1s1/dga1s0)*(cps/(1-cps))
h0w<-((1-a)*I(site==1))/g0w
h1w<-(a*I(site==1))/g1w

ymodel<-glm(formula=aoutmodel,family="binomial", data=data.frame(cbind(datw, a=a,z=z, site=site, y=y)), subset=site==1)

#initial prediciton
q<-cbind(predict(ymodel, type="link", newdata=data.frame(cbind(datw, a=a, z=z))), predict(ymodel, type="link", newdata=data.frame(cbind(datw, a=0,z=z))), predict(ymodel, type="link", newdata=data.frame(cbind(datw, a=1,z=z))))

epsilon<-coef(glm(y ~ -1 + offset(q[,1]) + h0w + h1w , family="binomial", subset=site==1 ))

#update initial prediction
 q1<- q + c((epsilon[1]*h0w + epsilon[2]*h1w), epsilon[1]/g0w, epsilon[2]/g1w)

predmodela0<-suppressWarnings(glm(formula=paste("plogis(q1)", aq2model, sep="~"), data=data.frame(cbind(w, a=a,site=site, q1=q1[,2])), subset=site==0 & a==0 , family="binomial"))
predmodela1<-suppressWarnings(glm(formula=paste("plogis(q1)", aq2model, sep="~"), data=data.frame(cbind(w,a=a, site=site, q1=q1[,3])), subset=site==0 & a==1 , family="binomial"))
predmodelaa<-suppressWarnings(glm(formula=paste("plogis(q1) ~", aq2model, "+a", sep=""), data=data.frame(cbind(w, site=site, q1=q1[,1], a=a)), subset=site==0, family="binomial"))

#get initial prediction for second regression model
q2pred<-cbind(predict(predmodelaa, type="link", newdata=data.frame(cbind(datw, a=a))), predict(predmodela0, type="link", newdata=datw), predict(predmodela1, type="link", newdata=datw))

cz<-cbind(ifelse(a==0,I(site==0)/(1-cpa), I(site==0)/cpa), I(site==0)/(1-cpa), I(site==0)/cpa)

epsilon2<-suppressWarnings(coef(glm(plogis(q1[,1]) ~ -1 + offset(q2pred[,1]) + cz[,2] + cz[,3], family="binomial", subset= site==0)))
  for(k in 1:2){
      epsilon2[k]<-ifelse(is.na(epsilon2[k]), 0, epsilon2[k])
  }

q2<- q2pred + c((epsilon2[1]*cz[,2] + epsilon2[2]*cz[,3]), epsilon2[1]/(1-cpa), epsilon2[2]/cpa)
```

```
57
58    tmleest<-mean(plogis(q2[,3][site==0]))-mean(plogis(q2[,2][site==0]))
59
60    ps0<-mean(I(site==0))
61
62    eic<-(((h1w/ps0) - (h0w/ps0))*(y - plogis(q[,1]))) +  (((a*cz[,3]/ps0) -
          ((1-a)*cz[,2]/ps0))* (plogis(q[,1]) - plogis(q2pred[,1]))) + ((I(site
          ==0)/ps0)*((plogis(q2pred[,3]) - plogis(q2pred[,2])) - tmleest))
63
64    return(list("est"=tmleest, "var"=var(eic)/n.dat, "eic"=eic))
65
66 }
67
68 eatetmle<-function(a, z, y, site, w, nsitemodel, nzmodel, noutmodel){
69    datw<-w
70    n.dat<-nrow(w)
71
72    #calculate components of clever covariate
73    cps <- predict(glm(formula=nsitemodel, data=data.frame(cbind(site=site,
          datw)), family="binomial"), type="response")
74    cpz<-predict(glm(formula=nzmodel, data=data.frame(cbind(a=a, z=z, datw)),
           family="binomial"), type="response")
75
76    #calculate clever covariate
77    g0w<-((1-cpz)*cps)/(1-cps)
78    g1w<-(cpz*cps)/(1-cps)
79    h0w<-((1-z)*I(site==1))/g0w
80    h1w<-(z*I(site==1))/g1w
81
82    ymodel<-glm(formula=noutmodel,family="binomial", data=data.frame(cbind(
          datw, a=a,z=z, site=site, y=y)), subset=site==1)
83
84    data_new0<-data_new1<-datw
85    data_new0$z<-0
86    data_new1$z<-1
87    #initial prediciton
88    q<-cbind(predict(ymodel, type="link", newdata=data.frame(cbind(datw, a=a,
          z=z))), predict(ymodel, type="link", newdata=data_new0), predict(
          ymodel, type="link", newdata=data_new1))
89
90    epsilon<-coef(glm(y ~ -1 + offset(q[,1]) + h0w + h1w , family="binomial",
           subset=site==1 ))
91
92    #update initial prediction
93    q1<- q + c((epsilon[1]*h0w + epsilon[2]*h1w), epsilon[1]/g0w, epsilon[2]/
          g1w)
94
95    tmleest<-mean(plogis(q1[,3][site==0]))-mean(plogis(q1[,2][site==0]))
96
97    #get efficient influence curve values for everyone
98    ps0<-mean(I(site==0))
99
```

```
  eic<-(((z*h1w/ps0) - ((1-z)*h0w/ps0))*(y - plogis(q[,1]))) + (I(site==0)/
      ps0*plogis(q1[,3])) - (I(site==0)/ps0*plogis(q1[,2])) - (tmleest/ps0)

  return(list("est"=tmleest, "var"=var(eic)/n.dat, "eic"=eic[site==0]))
}

notransporttmle<-function(a,z,w, site, ntamodel, ntzmodel){
  datw<-w

  n.dat<-nrow(datw)
  ps0<-mean(I(site==0))

  #calculate components of clever covariate
  cpa <- predict(glm(formula=ntamodel, data=data.frame(cbind(a=a, site=site
      , datw)), subset=site==0, family="binomial"), newdata=datw, type="
      response")

  g0w<-1-cpa
  g1w<-cpa

  #clever covariates
  h0w<-I(site==0)*(1- a)/(g0w*ps0)
  h1w<-I(site==0)*a/(g1w*ps0)

  zmodel<-glm(formula=ntzmodel,family="binomial", data=data.frame(cbind(a=a
      , z=z, site=site, datw)), subset=site==0)

  data_new0<-data_new1<-datw
  data_new0$a<-0
  data_new1$a<-1

  q<-cbind(predict(zmodel, type="link", newdata=data.frame(cbind(a=a, z=z,
      datw))), predict(zmodel, type="link", newdata=data_new0), predict(
      zmodel, type="link", newdata=data_new1))

  epsilon<-coef(glm(z ~ -1 + offset(q[,1]) + h0w + h1w , family="binomial",
       subset= site==0))

  q1<- q + c((epsilon[1]*h0w + epsilon[2]*h1w), I(site==0)*epsilon[1]/(g0w*
      ps0), I(site==0)*epsilon[2]/(g1w*ps0))

  tmleest<-mean(plogis(q1[,3][site==0]))-mean(plogis(q1[,2][site==0]))

  eic<-(((a*h1w) - ((1-a)*h0w))*(z - plogis(q[,1]))) + ((I(site==0)/ps0)*((
      plogis(q1[,3]) - plogis(q1[,2])) - tmleest))

  return(list("est"=tmleest, "var"=var(eic)/n.dat, "eic"=eic))
}


catetmle<-function(ca, cz, cy, csite, cw, czmodel, csitemodel, coutmodel,
    cq2model){
  datw<-cw
```

```
143   n.dat<-nrow(datw)
144   ps0<-mean(I(csite==0))
145   camodel<-"a ~ 1"
146
147   notransportate<-notransporttmle(a=ca, z=cz, site=csite, w=cw, ntamodel=
          camodel, ntzmodel=czmodel)
148   ittate<-ittatetmle(a=ca, z=cz, y=cy, site=csite, w=cw, aamodel=camodel,
          asitemodel=csitemodel, azmodel=czmodel, aoutmodel=coutmodel, aq2model
          =cq2model)
149   cate<-ittate$est/notransportate$est
150   varcate<-(ittate$est^2/notransportate$est^2)*(((ittate$var*n.dat)/ittate$
          est^2) - ((2*cov(cbind(ittate$eic, notransportate$eic))[1,2])/(ittate
          $est*notransportate$est)) + ((notransportate$var*n.dat)/
          notransportate$est^2))
151   eic<-(ittate$eic/notransportate$est) - (ittate$est/(notransportate$est^2)
          )*notransportate$eic
152
153   return(list("est"=cate, "var"=var(eic)/n.dat, "eic"=eic))
154 }
```

Functions.R

F.2 Code for example application

```
1  source("Functions.R")
2
3    n<-5000
4
5    site<-rbinom(n, 1, .5)
6
7    race<-rbinom(n,1, .4 + (.2*site))
8
9    crime<-rnorm(n, .1*site, 1)
10   discrimination<-rnorm(n, 1+(.2*site), 1)
11
12   #instrument
13   voucher<-rbinom(n, 1, .5)
14
15   #exposure
16   move0<-rbinom(n,1,plogis( -log(1.6) - log(1.1)*crime -log(1.3)*
         discrimination))
17   move1<-rbinom(n,1,plogis(  -log(1.6) +log(4) - log(1.1)*crime  -log(1.3)*
         discrimination))
18   move<-ifelse(voucher==1, move1, move0)
19
20   #outcomes
21   inschoola0<-rbinom(n,1, plogis(log(1.6)  + (log(1.9)*move0) -log(1.3)*
         discrimination - log(1.2)*race + log(1.2)*race*move0) )
22   inschoola1<-rbinom(n,1, plogis(log(1.6) + (log(1.9)*move1)  - log(1.3)*
         discrimination - log(1.2)*race+  log(1.2)*race*move1) )
23   inschoola<-ifelse(voucher==1, inschoola1, inschoola0)
24
```

```
25  dat<-data.frame( w2=crime, w3=discrimination, w1=race, site=site, a=
        voucher, z=move, y=inschoola)
26
27  wmat<-data.frame(w1=dat$w1, w2=dat$w2, w3=dat$w3)
28
29  amodel<-"a ~ 1"
30  sitemodel<-"site ~ w1 + w2 + w3 "
31  zmodel<-"z ~ a + w2 + w3 "
32  outmodel<-"y ~ z + w1 +w3 + z:w1"
33  outmodelnoz<-"y ~ a + w1+w3+ a:w1"
34  q2model<-"w1 + w2 + w3 "
35
36 ittatetmletransportest<-ittatetmle(a=dat$a, z=dat$z, y=dat$y, site=dat$site
       , w=wmat, aamodel=amodel, asitemodel=sitemodel, azmodel=zmodel,
       aoutmodel=outmodel, aq2model=q2model)$est
37 catetmletransportest<-catetmle(ca=dat$a, cz=dat$z, cy=dat$y, csite=dat$site
       , cw=wmat, csitemodel=sitemodel, czmodel=zmodel, coutmodel=outmodel,
       cq2model=q2model)$est
38 eatetmletransportest<-eatetmle(a=dat$a, z=dat$z, y =dat$y, site=dat$site, w
       =wmat, nsitemodel=sitemodel, nzmodel=zmodel, noutmodel=outmodel)$est
```

examp.R

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.

Bembom, O. and van der Laan, M. J. (2008) Data-adaptive selection of the truncation level for inverse-probability-of-treatment-weighted estimators.

Cole, S. R. and Hernán, M. A. (2008) Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, **168**, 656–664.

Cole, S. R. and Stuart, E. A. (2010) Generalizing evidence from randomized clinical trials to target populations the actg 320 trial. *American journal of epidemiology*, **172**, 107–115.

Eckenrode, J., Campa, M., Luckey, D. W., Henderson, C. R., Cole, R., Kitzman, H., Anson, E., Sidora-Arcoleo, K., Powers, J. and Olds, D. (2010) Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Archives of Pediatrics & Adolescent Medicine*, **164**, 9–15.

Frangakis, C. (2009) The calibration of treatment effects from clinical trials to target populations. *Clinical trials (London, England)*, **6**, 136.

Gill, R. D., Wellner, J. A. and Præstgaard, J. (1989) Non-and semi-parametric maximum likelihood estimators and the von mises method (part 1)[with discussion and reply]. *Scandinavian Journal of Statistics*, 97–128.

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**.

Kang, J. D. and Schafer, J. L. (2007) Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*, 523–539.

Kling, J. R., Liebman, J. B. and Katz, L. F. (2007) Experimental analysis of neighborhood effects. *Econometrica*, **75**, 83–119.

van der Laan, M. J. and Gruber, S. (2010) Collaborative double robust targeted maximum likelihood estimation. *The international journal of biostatistics*, **6**.

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 Robins, J. M. (2003) *Unified methods for censored longitudinal data and causality*. Springer.

van der Laan, M. J. and Rubin, D. (2006) Targeted maximum likelihood learning. *The International Journal of Biostatistics*, **2**.

Miettinen, O. S. (1972) Standardization of risk ratios. *American Journal of Epidemiology*, **96**, 383–388.

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.

Osypuk, T. L., Schmidt, N. M., Bates, L. M., Tchetgen-Tchetgen, E. J., Earls, F. J. and Glymour, M. M. (2012) Gender and crime victimization modify neighborhood effects on adolescent mental health. *Pediatrics*, **130**, 472–481.

Pearl, J. and Bareinboim, E. (2011) Transportability across studies: A formal approach. *Tech. rep.*, DTIC Document.

Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y. and van der Laan, M. J. (2010) Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, 0962280210386207.

Robins, J., Sued, M., Lei-Gomez, Q. and Rotnitzky, A. (2007) Comment: Performance of double-robust estimators when" inverse probability" weights are highly variable. *Statistical Science*, 544–559.

Rothwell, P. M. (2005) External validity of randomised controlled trials:"to whom do the results of this trial apply?". *The Lancet*, **365**, 82–93.

Rudolph, K. E., Díaz, I., Rosenblum, M. and Stuart, E. A. (2014) Estimating population treatment effects from a survey subsample. *American Journal of Epidemiology*, **180**, 737–748.

Sanbonmatsu, L., Ludwig, J., Katz, L. F., Gennetian, L. A., Duncan, G. J., Kessler, R. C., Adam, E., McDade, T. W. and Lindau, S. T. (2011) Moving to opportunity for fair housing demonstration program–final impacts evaluation.

Stitelman, O. M., De Gruttola, V. and van der Laan, M. J. (2012) A general implementation of tmle for longitudinal data applied to causal inference in survival analysis. *The international journal of biostatistics*, **8**.

Stuart, E. A., Cole, S. R., Bradshaw, C. P. and Leaf, P. J. (2011) The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **174**, 369–386.