Supplementary Web Appendix for: Robust estimation of encouragement-design intervention effects transported across sites

Kara E. Rudolph, Mark J. van der Laan

1 Identification proofs

1.1 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^a, Y^a), 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.

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

By assumption 1,

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

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

$$\begin{aligned}
\equiv\; & E_0[E_0\{E_0(Y^1 \mid S = 0, W, Z^1) \mid W, S = 0\} \\
& -E_0\{E_0(Y^1 \mid S = 0, W, Z^1) \mid W, S = 0\} \mid S = 0] \\
\equiv\; & E_0(Y^1 - Y^0 \mid S = 0)
\end{aligned}$$

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

1.2 CACE

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 $(Z^a, Y^{a)}$, 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 and $E_0(Z^1 - Z^0 | S = 0, W) \neq 0$. This first part 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 second part means that the nontransported average effect of the instrument on the exposure of participants with $S = 0$ does not equal 0.

Proof.

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

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. □

2 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.

2.1 ITTATE

2.1.1 Efficient influence curve

The following result provides the conditions under which an efficient influence curve-based estimator is consistent. Let $Q_W(w|s) = P(W \leq w \mid S = s), \bar{Q} = \{Q_W, \bar{Q}(1, W, A, Z)\}$, and $q_W(w|s) = \frac{dQ_W}{d\mu}(w)$ (i.e., density with respect to the appropriate dominating measure μ). Let $G_Z\bar{Q}(W) = E\{\bar{Q}(1, W, a, Z) \mid A = a, W, S = 0\}$. Recall that a subscript of 0 added to any of the above notation denotes values under the true data distribution, P_0.

Result 1. *We have*

$$P_0 D^a_{\Psi_1}(P) = \Psi^a_1(P_0) - \Psi^a_1(P) + R^a_2(P, P_0),$$

where

$$R^a_2(P, P_0) = \sum_{j=1}^{4} R^a_{2j}(P, P_0),$$

and

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

Inspection of the second order term R^a_2 gives three scenarios under which an estimator $\Psi^a_1(P_n)$ that solves the efficient influence equation $P_n D^a_{\Psi_1}(P_n)$ 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* $P_0 D^a_{\Psi_1}(P) = 0$ *and one of three above scenarios holds, then* $\Psi^a_1(P) = \Psi^a_1(P_0)$.

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 $\Psi^a_1(P_n)$ is consistent for $\Psi^a_1(P_0)$.

2.2 EACE

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

Result 2.

$$P_0 D^z_{\Psi_2}(P) \;=\; \Psi^z_2(P_0) - \Psi^z_2(P) + R^z_2(P, P_0),$$

where

$$R^z_2(P, P_0) = R^z_{21}(P, P_0) + R^z_{22}(P, P_0) + R^z_{23}(P, P_0),$$

and

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

Considering the case that $P(S=1) = P_0(S=1)$ so that $R^z_{23} = 0$, 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* $P_0 D^z_{\Psi_2}(P) = 0$ *and one of two above scenarios holds, then* $\Psi^z_2(P) = \Psi^z_2(P_0)$.

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 $\Psi^z_2(P_n)$ is consistent for $\Psi^z_2(P_0)$.

2.3 CACE

2.3.1 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 statistical estimand of the average effect of the instrument on the treatment.

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

3 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

$$S\{Z \mid Pa(Z)\} = [S\{1 \mid Pa(Z)\} - S\{0 \mid Pa(Z)\}][Z - E\{Z \mid Pa(Z)\}]$$

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

$$\begin{aligned} D^a_{Z,\Psi_1}(P) &= \frac{I(A=a, S=0)}{g_A(A=a \mid W, S=0)P(S=0)} \left\{\bar{Q}(1,W,a,Z=1) - \bar{Q}(1,W,a,Z=0)\right\} \\ &\quad \times\{Z - g_Z(Z=1 \mid S=0,W,a)\} \\ &\equiv C_Z(g_A,\bar{Q})\{Z - g_Z(Z=1 \mid S=0,W,a)\}. \end{aligned}$$

As in the TMLE shown in the main text, 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 offset, 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 $Z = z$ on W among the observations with $S_i = 0, A_i = a$, using predicted values $\text{Logit}\bar{g}^0_{Z,n}$ 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 $P_n D^a_{Z,\Psi_1}(g_{A,n}, g^K_{Z,n}, g_{S,n}, \bar{Q}^K_n, Q_{W,n|S=0}) = o_P(1/\sqrt{n})$. 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

$$P_n D^a_{Z,\Psi_1}(g_{S,n}, g_{A,n}, g^*_{Z,n}, \bar{Q}^*_n, Q_{W,n|S=0}) = 0 \text{ or } o_P(1/\sqrt{n}).$$

4 Alternative variance estimate of Ψ_3

Alternatively, $var(\Psi_3)$ can be estimated using the multivariate delta method.

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

5 Simulation results

Table 1: 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. N=500. The estimator standard error $\times\sqrt{n}$ should be compared to the efficiency bound, which is 0.47 for the ITTATE, 1.42 for the CACE, and 0.51 for the EACE.

Specification	%Bias	SE$\times\sqrt{n}$	95%CI Cov	MSE
bound, which is 1.49 for the ITTATE, 4.50 for the CACE, and 1.60 for the EACE.				
ITTATE				
All models correct	-0.09	1.53	94.75	0.0048
S model misspecified	0.03	1.39	94.91	0.0039
Z model misspecified	-0.09	1.51	94.40	0.0048
Y model misspecified	-0.03	1.56	94.85	0.0049
S,Z models misspecified	0.11	1.37	94.80	0.0038
S,Z,Y models misspecified	6.82	1.39	94.80	0.0039
EACE				
All models correct	-1.27	1.62	94.70	0.0052
S model misspecified	-0.83	1.46	93.60	0.0047
Z model misspecified	-1.27	1.49	92.82	0.0052
Y model misspecified	-1.17	1.65	95.40	0.0052
S,Z models misspecified	-0.88	1.35	93.09	0.0042
S,Z,Y models misspecified	13.98	1.37	94.03	0.0041
CACE				
All models correct	0.55	4.86	96.02	0.0495
S model misspecified	0.57	4.40	96.11	0.0397
Z model misspecified	0.55	4.81	95.93	0.0495
Y model misspecified	0.56	4.86	95.81	0.0503
S,Z models misspecified	0.78	4.35	96.27	0.0380
S,Z,Y models misspecified	7.16	4.41	96.34	0.0390

Table 2: 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. N=500. The estimator standard error $\times\sqrt{n}$ should be compared to the efficiency bound, which is 0.85 for the ITTATE, 3.58 for the CACE, and 1.29 for the EACE.

Specification	%Bias	SE$\times\sqrt{n}$	95%CI Cov	MSE
ITTATE				
All models correct	3.32	2.68	92.09	0.0165
S model misspecified	-0.06	1.40	95.18	0.0039
Z model misspecified	3.32	2.75	92.68	0.0165
Y model misspecified	7.72	2.91	93.84	0.0167
S,Z models misspecified	-0.25	1.40	95.42	0.0038
S,Z,Y models misspecified	18.87	1.42	94.97	0.0040
EACE				
All models correct	-11.13	3.21	87.31	0.0246
S model misspecified	-2.40	1.88	83.77	0.0122
Z model misspecified	-11.13	2.63	84.38	0.0246
Y model misspecified	-4.54	3.73	91.34	0.0242
S,Z models misspecified	0.55	1.38	78.59	0.0095
S,Z,Y models misspecified	-52.27	1.41	75.96	0.0102
CACE				
All models correct	8.15	12.41	94.84	0.3519
S model misspecified	2.89	6.40	96.94	0.0811
Z model misspecified	8.15	14.42	95.24	0.3519
Y model misspecified	12.93	12.41	94.76	0.3550
S,Z models misspecified	4.10	7.35	97.66	0.1225
S,Z,Y models misspecified	29.38	7.64	98.09	0.1337

Table 3: 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.

Truncation Level	%Bias	SE$\times\sqrt{n}$	95%CI Cov	MSE
N=5,000. 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 CACE, and 4.09 for the EACE.				
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
EACE				
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
CACE				
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
N=500. The estimator standard error $\times\sqrt{n}$ should be compared to the efficiency bound, which is 0.85 for the ITTATE, 3.58 for the CACE, and 1.29 for the EACE.				
ITTATE				
No modification	3.32	2.68	92.09	0.0165
Truncation at 0.01/100	3.37	2.67	92.11	0.0165
Truncation at 0.05/20	3.02	2.59	92.45	0.0158
Truncation at 0.1/10	3.53	2.00	87.86	0.0128
EACE				
No modification	-11.13	3.21	87.31	0.0246
Truncation at 0.01/100	-8.19	3.03	87.64	0.0237
Truncation at 0.05/20	-9.18	3.02	87.21	0.0234
Truncation at 0.1/10	-0.60	1.86	82.70	0.0139
CACE				
No modification	8.15	12.41	94.84	0.3519
Truncation at 0.01/100	8.20	12.38	94.84	0.3513
Truncation at 0.05/20	7.85	12.01	95.01	0.3354
Truncation at 0.1/10	7.84	9.34	91.35	0.2729

6 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.

7 R code

7.1 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)

  tmleest<-mean(plogis(q2[,3][site==0]))-mean(plogis(q2[,2][site==0]))

  ps0<-mean(I(site==0))

  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))

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

}

eatetmle<-function(a, z, y, site, w, nsitemodel, nzmodel, noutmodel){
  datw<-w
  n.dat<-nrow(w)

  #calculate components of clever covariate
  cps <- predict(glm(formula=nsitemodel, data=data.frame(cbind(site=site, datw)), family="binomial"), type="response")
  cpz<-predict(glm(formula=nzmodel, data=data.frame(cbind(a=a, z=z, datw)), family="binomial"), type="response")

  #calculate clever covariate
  g0w<-((1-cpz)*cps)/(1-cps)
  g1w<-(cpz*cps)/(1-cps)
  h0w<-((1-z)*I(site==1))/g0w
  h1w<-(z*I(site==1))/g1w

```

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

  data_new0<-data_new1<-datw
  data_new0$z<-0
  data_new1$z<-1
  #initial prediciton
  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))

  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)

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

  #get efficient influence curve values for everyone
  ps0<-mean(I(site==0))

  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
  n.dat<-nrow(datw)
  ps0<-mean(I(csite==0))
  camodel<-"a ~ 1"

  notransportate<-notransporttmle(a=ca, z=cz, site=csite, w=cw, ntamodel=camodel, ntzmodel=czmodel)
  ittate<-ittatetmle(a=ca, z=cz, y=cy, site=csite, w=cw, aamodel=camodel, asitemodel=csitemodel, azmodel=czmodel, aoutmodel=coutmodel, aq2model=cq2model)
  cate<-ittate$est/notransportate$est
  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))
  eic<-(ittate$eic/notransportate$est) - (ittate$est/(notransportate$est^2))*notransportate$eic

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

Functions.R

7.2 Code for example application

```
source("Functions.R")

  n<-5000

  site<-rbinom(n, 1, .5)
```

```
  race<-rbinom(n,1, .4 + (.2*site))

  crime<-rnorm(n, .1*site, 1)
  discrimination<-rnorm(n, 1+(.2*site), 1)

  #instrument
  voucher<-rbinom(n, 1, .5)

  #exposure
  move0<-rbinom(n,1,plogis( -log(1.6) - log(1.1)*crime -log(1.3)*discrimination))
  move1<-rbinom(n,1,plogis(  -log(1.6) +log(4) - log(1.1)*crime  -log(1.3)*discrimination))
  move<-ifelse(voucher==1, move1, move0)

  #outcomes
  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) )
  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) )
  inschoola<-ifelse(voucher==1, inschoola1, inschoola0)

  dat<-data.frame( w2=crime, w3=discrimination, w1=race, site=site, a=voucher, z=move, y=inschoola)

  wmat<-data.frame(w1=dat$w1, w2=dat$w2, w3=dat$w3)

  amodel<-"a ~ 1"
  sitemodel<-"site ~ w1 + w2 + w3 "
  zmodel<-"z ~ a + w2 + w3 "
  outmodel<-"y ~ z + w1 +w3 + z:w1"
  outmodelnoz<-"y ~ a + w1+w3+ a:w1"
  q2model<-"w1 + w2 + w3 "

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)$est
catetmletransportest<-catetmle(ca=dat$a, cz=dat$z, cy=dat$y, csite=dat$site, cw=wmat, csitemodel=sitemodel, czmodel=zmodel, coutmodel=outmodel, cq2model=q2model)$est
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

Bibliography

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.

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