Statistica Sinica

Supplementary Material for "Identification and estimation of treatment effects on long-term outcomes in clinical trials with external observational data"

Wenjie Hu[a], Xiao-Hua Zhou[a,b] and Peng Wu[c*]

[a]Peking University, [b]Pazhou Lab, [c]Beijing Technology and Business University

This Supplementary Material consists of Sections 1–6, where Sections 1-4 give the technical proofs of Propositions 1–2 and Theorems 1–3, Sections 6 and 7 contain additional numerical results from the simulation study and empirical application.

∗correspond to: pengwu@btbu.edu.cn.

1. Proof of Propositions 1 and 2

Proof of Proposition 1. Let $e(X) = \mathrm{pr}(T = 1 \mid X, G = 1)$ and $h(X, S, T) = E[Y|X, S, T, G = 1]$, then we have

$$
\begin{aligned}
\tau &= E\left\{\frac{YT}{e(X)} - \frac{Y(1-T)}{1-e(X)} \mid G = 1\right\} \\
&= E\left\{\frac{h(X,S,T)T}{e(X)} - \frac{h(X,S,T)(1-T)}{1-e(X)} \mid G = 1\right\} \\
&= E\left[E\left\{\frac{h(X,S,T)T}{e(X)} - \frac{h(X,S,T)(1-T)}{1-e(X)} \mid X, T, G = 1\right\} \mid G = 1\right] \\
&= E\left[E\left\{h(X,S,T)|X, T = 1, G = 1\right\} - E\left\{h(X,S,T)|X, T = 0, G = 1\right\}|G = 1\right].
\end{aligned}
$$

By Assumption 3, $h(X, S, T)$ can be identified by the $E[Y|X, S, T, G = 0]$ in observational data, so τ can be identified.

$\square$

Proof of Proposition 2. For $t = 0$ or 1, we have

$$
\begin{aligned}
& E(Y \mid X, S, T = t, G = 1) \\
= & E(Y(t) \mid X, S(t), T = t, G = 1) \\
= & E(Y(t) \mid X, S(t), G = 1) && \text{Assumption 1} \\
= & E(Y(t) \mid X, S(t), G = 0) && \text{Assumption 4} \\
= & E(Y(t) \mid X, S(t), T = t, G = 0) && \text{Assumption 5} \\
= & E(Y|X, S, T = t, G = 0).
\end{aligned}
$$

□

2. Proof of Theorem 1

Proof of Theorem 1. Denote the density of X by $f(x)$, the conditional density $\mathrm{pr}(Y(t) = y|S(t) = s, X = x, G = 1)$ by $f_t(y|s, x)$ for $t = 0, 1$, $\mathrm{pr}(S(t) = s|X = x, G = 1)$ by $f_t(s|x)$, $\mathrm{pr}(Y = y|S = s, X = x, T = t, G = 0)$ by $f(y|s, x, t)$, $\mathrm{pr}(S = s|X = x, T = t, G = 0)$ by $f(s|x, t)$, $\mathrm{pr}(T = 1|X = x, G = 0)$ by $p(x)$ and $\mathrm{pr}(G = 1|X = x)$ by $\pi(x)$, then the full observed data distribution is

$$f(t,x,s,y,g) = f(x)\left[\{f_1(s|x)e(x)\}^t\{f_0(s|x)(1-e(x))\}^{1-t}\pi(x)\right]^g$$
$$\times\left[\{f(y|s,x,1)f(s|x,1)p(x)\}^t\{f(y|s,x,0)f(s|x,0)(1-p(x))\}^{1-t}(1-\pi(x))\right]^{1-g}.$$

Under Assumptions 1-3 in the manuscript, a parametric submodel indexed by θ is given as

$$f(x;\theta)\left[\{f_1(s|x;\theta)e(x;\theta)\}^t\{f_0(s|x;\theta)(1-e(x;\theta))\}^{1-t}\pi(x;\theta)\right]^g$$
$$\times\left[\{f(y|s,x,1;\theta)f(s|x,1;\theta)p(x;\theta)\}^t\{f(y|s,x,0;\theta)f(s|x,0;\theta)(1-p(x;\theta))\}^{1-t}(1-\pi(x;\theta))\right]^{1-g},$$

which equals $f(t,x,s,y,g)$ when $\theta=\theta_0$ and satisfies $\int yf_t(y|s,x)dy = \int yf(y|x,s,t)dy$ for $\forall s,x,t$.

The score $S(t,x,s,y,g;\theta)$ is

$$\begin{aligned}
&S(t,x,s,y,g;\theta)\\
=&gtS_1(s|x;\theta)+g(1-t)S_0(s|x;\theta)\\
&+g\frac{t-e(x;\theta)}{e(x;\theta)\{1-e(x;\theta)\}}\dot{e}(x;\theta)+\frac{g-\pi(x;\theta)}{\pi(x;\theta)\{1-\pi(x;\theta)\}}\dot{\pi}(x;\theta)\\
&+(1-g)t\{S(y|s,x,1;\theta)+S(s|x,1;\theta)\}+(1-g)(1-t)\{S(y|s,x,0;\theta)+S(s|x,0;\theta)\}\\
&+(1-g)\frac{t-p(x;\theta)}{p(x;\theta)\{1-p(x;\theta)\}}\dot{p}(x;\theta)+S_f(x;\theta),
\end{aligned}$$

where $S_t(s|x;\theta)=\partial\log f_t(s|x;\theta)/\partial\theta$ for $t=0,1$, $S(y|s,x,t;\theta)=\partial\log f(y|s,x,t;\theta)/\partial\theta$,

$S(s|x,t;\theta) = \partial \log f(s|x,t;\theta)/\partial\theta$, and $\dot{e}(x;\theta), \dot{p}(x;\theta), \dot{\pi}(x;\theta)$ are pathwise derivative with respect to θ for $e(x;\theta), p(x;\theta), \pi(x;\theta)$.

The tangent space $\mathcal{T}$ is

$$\begin{aligned}\mathcal{T} = \Big\{ & gtS_1(s|x) + g\{1-t\}S_0(s|x) + g\{t-e(x)\}a(x) + \{(g-\pi(x)\}b(x) \\ & + (1-g)t\left\{S(y|s,x,1) + S(s|x,1)\right\} + (1-g)(1-t)\left\{S(y|s,x,0) + S(s|x,0)\right\} \\ & + (1-g)\{t-p(x)\}c(x) + S(x)\Big\}\end{aligned}$$

where $S_t(s|x)$ satisfies $\int S_t(s|x)f_t(s|x)ds = 0$ for $\forall x, t = 0,1$, $S(y|s,x,t)$ satisfies

$\int S(y|s,x,t)f(y|s,x,t)dy = 0$ and $\int yS_t(y|s,x)f_t(y|s,x)dy = \int yS(y|s,x,t)f(y|s,x,t)dy$ for $\forall s,x,t = 0,1$, $S(s|x,t)$ satisfies $\int S(s|x,t)f(s|x,t)ds = 0$ for $\forall x, t = 0,1$, $S(x)$ satisfies $\int S(x)f(x)dx = 0$, and $a(x), b(x), c(x)$ are arbitrary square-integrable measurable functions.

Under the parametric submodel, the parameter of interest τ can be represented as

$$\begin{aligned}\tau(\theta) &= \frac{\int\{\int yf_1(y|s,x;\theta)f_1(s|x;\theta)dyds\}\pi(x;\theta)f(x;\theta)dx}{\int \pi(x;\theta)f(x;\theta)dx} \\ &\quad - \frac{\int\{\int yf_0(y|s,x;\theta)f_0(s|x;\theta)dyds\}\pi(x;\theta)f(x;\theta)dx}{\int \pi(x;\theta)f(x;\theta)dx} \\ &= \tau_1(\theta) - \tau_0(\theta)\end{aligned}$$

where $\tau_1 = E(Y(1)|G=1) = \tau_1(\theta_0)$ and $\tau_0 = E(Y(0)|G=1) = \tau_0(\theta_0)$ and $\tau = \tau(\theta_0)$. For notation convenience, we denote $\dot{e}(x) = \dot{e}(x;\theta_0), \dot{p}(x) = \dot{p}(x;\theta_0), \dot{\pi}(x) = \dot{\pi}(x;\theta_0)$, $S_f(x) = S_f(x;\theta_0)$, $q = \int \pi(x;\theta)f(x;\theta)dx = \mathrm{pr}(G = 1)$ and

$\mu_t(x) = E[Y(t)|X, G=1] = \int y f_t(y|s,x;\theta)f_t(s|x;\theta)dyds$. Then the path-wise derivative of $\tau_1(\theta)$ is

$$
\begin{aligned}
\left.\frac{\partial \tau_1(\theta)}{\partial \theta}\right|_{\theta=\theta_0} =& \frac{\int \left\{\int y(S_1(y|s,x;\theta_0) + S_1(s|x;\theta_0))f_1(y|s,x)f_1(s|x)dyds\right\}\pi(x)f(x)dx}{q} \\
&+ \frac{\int \{\mu_1(x) - \tau_1\}\{\dot{\pi}(x) + \pi(x)S_f(x)\} f(x)dx}{q} \\
=& \frac{E\big(\pi(X)E[E\{Y \cdot S(Y|S,X,1)|S,X,T=1,G=0\}|X,T=1,G=1]\big)}{q} \\
&+ \frac{E[\pi(X)E\{\mu_1(S,X) \cdot S_1(S|X) \mid X, T=1, G=1\}]}{q} \\
&+ \frac{E[\{\mu_1(X) - \tau_1\}\{\dot{\pi}(X) + \pi(X)S_f(X)\}]}{q}.
\end{aligned}
$$

We let

$$
\begin{aligned}
\phi_1 =& \frac{G}{q}\left[\{\mu_1(X) - \tau_1\} + \frac{T\{\mu_1(S,X) - \mu_1(X)\}}{e(X)}\right] \\
&+ \frac{1-G}{q}\frac{\pi(X)}{1-\pi(X)}\frac{T\{Y - \mu_1(S,X)\}}{r(S,X)}\frac{pr(S|X,T=1,G=1)}{pr(S|X,G=0)} \\
=& \frac{G}{q}\left[\{\mu_1(X) - \tau_1\} + \frac{T\{\mu_1(S,X) - \mu_1(X)\}}{e(X)}\right] \\
&+ \frac{1-G}{q}\frac{g_1(S,X)T\{Y - \mu_1(S,X)\}}{e(X)\{1 - g_1(S,X)\}}
\end{aligned}
$$

where $\pi(x) = \mathrm{pr}(G = 1|X = x)$, $r(s, x) = \mathrm{pr}(T = 1|S = s, X = s, G = 0)$ and $g_1(s, x) = \mathrm{pr}(G = 1|S = s, X = x, T = 1)$.

Pathwise differentiability of τ_1 can be verified by

$$\left.\frac{\partial \tau_1(\theta)}{\partial \theta}\right|_{\theta=\theta_0} = E\left\{\phi_1 \cdot S(T, X, S, Y, G; \theta_0)\right\} \tag{S.1}$$

Now we give a detailed proof of (S.1).

$$\begin{aligned}
&E\left\{\phi_1 \cdot S(T, X, S, Y, G; \theta_0)\right\} \\
=&\frac{1}{q}E\Bigg[G\{\mu_1(X) - \tau_1\}S(T, X, S, Y, G; \theta_0) \\
&+ \frac{GT\{\mu_1(S, X) - \mu_1(X)\}}{e(X)}S(T, X, S, Y, G; \theta_0) \\
&+ \frac{(1 - G)\pi(X)}{1 - \pi(X)}\frac{T\{Y - \mu_1(S, X)\}}{r(S, X)}\frac{pr(S|X, T = 1, G = 1)}{pr(S|X, G = 0)}S(T, X, S, Y, G; \theta_0)\Bigg],
\end{aligned}$$

where

$$\begin{aligned}
A =&E\left[G\{\mu_1(X) - \tau_1\}S(T, X, S, Y, G; \theta_0)\right] \\
=&E\left\{G\{\mu_1(X) - \tau_1\}\left[TS_1(S|X) + \frac{\{T - e(X)\}\dot{e}(X)}{e(X)\{1 - e(X)\}} + \frac{(G - \pi(X))\dot{\pi}(X)}{\pi(X)\{1 - \pi(X)\}} + S_f(X)\right]\right\} \\
=&E\Bigg[\{\mu_1(X) - \tau_1\}\{\dot{\pi}(X) + \pi(X)S_f(X)\}\Bigg],
\end{aligned}$$

$$
\begin{aligned}
B =& E\left[\frac{GT(\{\mu_1(S,X)-\mu_1(X)\}}{e(X)}S(T,X,S,Y,G;\theta_0)\right]\\
=& E\left[\frac{GT\{\mu_1(S,X)-\mu_1(X)\}}{e(X)}S_1(S|X)\right]\\
=& E\biggl(\pi(X)E\left[\{\mu_1(S,X)-\mu_1(X)\}S_1(S|X)\mid X,T=1,G=1\right]\biggr)\\
=& E\biggl\{\pi(X)E\left[\mu_1(S,X)S_1(S|X)\mid X,T=1,G=1\right]\biggr\},\\
C =& E\left[(1-G)\frac{\pi(X)}{1-\pi(X)}\frac{T\{Y-\mu_1(S,X)\}}{r(S,X)}\frac{pr(S|X,T=1,G=1)}{pr(S|X,G=0)}S(T,X,S,Y,G;\theta_0)\right]\\
=& E\left[(1-G)\frac{\pi(X)}{1-\pi(X)}\frac{T\{Y-\mu_1(S,X)\}}{r(S,X)}\frac{pr(S|X,T=1,G=1)}{pr(S|X,G=0)}S(Y|S,X,1)\right]\\
=& E\biggl(\pi(X)E\biggl\{E\bigl[\{Y-\mu_1(S,X)\}S(Y|S,X,1)\frac{pr(S|X,T=1,G=1)}{pr(S|X,G=0)}\\
& |S,X,T=1,G=0\bigr]|X,G=0\biggr\}\biggr)\\
=& E\biggl\{\pi(X)E(E\left[YS(Y|S,X,1)|S,X,T=1,G=0\right]|X,T=1,G=1)\biggr\}
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
& E\left\{\phi_1 S(T,X,S,Y,G;\theta_0)\right\}\\
& = \frac{1}{q}(A+B+C)\\
& = \left.\frac{\partial\tau_1(\theta)}{\partial\theta}\right|_{\theta=\theta_0}
\end{aligned}
$$

In addition, ϕ_1 can be represented as an element in $\mathcal{T}$. Let

$$\begin{aligned} S_1(s|x) &= \frac{1}{q}\frac{\mu_1(s,x)-\mu_1(x)}{e(x)}, \\ S(y|s,x,1) &= \frac{1}{q}\frac{\pi(x)}{1-\pi(x)}\frac{\{y-\mu_1(s,x)\}}{r(s,x)}\frac{pr(S=s|X=x,T=1,G=1)}{pr(S=s|X=x,G=0)}, \\ a(x) &= c(x) = 0, \\ b(x) &= \frac{1}{q}\{\mu_1(x)-\tau_1\}, \\ S(x) &= \frac{1}{q}\pi(x)\{\mu_1(x)-\tau_1\}. \end{aligned}$$

Then ϕ_1 can be decomposed as $\phi_1(y,s,x,t,g) = gtS_1(s|x)+(g-\pi(x))b(x)+(1-g)tS(y|s,x,1)+S(x)$. And the functions satisfy

$$\begin{aligned} \int S_1(s|x)f_1(s|x)ds &= 0, \\ \int S(y|s,x,1)f(y|s,x,1)dy &= 0, \\ \int S(x)f(x)dx &= 0, \\ \int yS_1(y|s,x)f_1(y|s,x)dy &= \int yS(y|s,x,1)f(y|s,x,1)dy, \end{aligned}$$

if we choose $S_1(y|s,x) = S(y|s,x,1)f(y|s,x,1)/f_1(y|s,x)$. So ϕ_1 is in the tangent space. For τ_0, we can get similar procedure and

$$
\begin{aligned}
\phi_0 =& \frac{G}{q}\left[\{\mu_0(X)-\tau_0\}+\frac{(1-T)\{\mu_0(S,X)-\mu_0(X)\}}{1-e(X)}\right] \\
&+\frac{1-G}{q}\frac{\pi(X)}{1-\pi(X)}\frac{(1-T)\{Y-\mu_0(S,X)\}}{1-r(S,X)}\frac{pr(S|X,T=0,G=1)}{pr(S|X,G=0)} \\
=& \frac{G}{q}\left[\{\mu_0(X)-\tau_0\}+\frac{(1-T)\{\mu_0(S,X)-\mu_0(X)\}}{1-e(X)}\right] \\
&+\frac{1-G}{q}\frac{g_0(S,X)(1-T)\{Y-\mu_0(S,X)\}}{\{1-e(X)\}\{1-g_0(S,X)\}}
\end{aligned}
$$

where $g_0(s,x) = f(G=1|S=s,X=x,T=0)$. So we have the efficient influence functions for τ is

$$
\begin{aligned}
\phi =& \phi_1-\phi_0 \\
=& \frac{G}{q}\left\{\frac{T\{\mu_1(S,X)-\mu_1(X)\}}{e(X)}-\frac{(1-T)\{\mu_0(S,X)-\mu_0(X)\}}{1-e(X)}+\{\mu_1(X)-\mu_0(X)\}-\tau\right\} \\
&+\frac{1-G}{q}\left\{\frac{g_1(S,X)T\{Y-\mu_1(S,X)\}}{e(X)\{1-g_1(S,X)\}}-\frac{g_0(S,X)(1-T)\{Y-\mu_0(S,X)\}}{\{1-e(X)\}\{1-g_0(S,X)\}}\right\}
\end{aligned}
$$

As τ can be represented as a functional of part of the likelihood, which is irrelevant with the propensity score models $e(X)$ and $r(S,X)$, so the efficient influence function is the same whether the propensity score models are known or not.

To show that in the nonparametric model, ϕ is the only influence function, we need only to show that $\mathcal{T}$ contains all the mean zero functions of the observed data. For $h(T,X,S,Y,G)$ that satisfies $E(h(T,X,S,Y,G)) = 0$, we can decompose it as

$$
\begin{aligned}
& h(T,X,S,Y,G) \\
=& GTh_1(X,S) + G(1-T)h_2(X,S) \\
& + (1-G)Th_3(X,S,Y) + (1-G)(1-T)h_4(X,S,Y) \\
=& GT\big[h_1(X,S) - E\{h_1(X,S)|X\}\big] + G(1-T)\big[h_2(X,S) - E\{h_2(X,S)|X\}\big] \\
& + G\{T - e(X)\}\big[E\{h_1(X,S)|X\} - E\{h_2(X,S)|X\}\big] \\
& + (1-G)T\big[h_3(X,S,Y) - E\{h_3(X,S,Y)|X,S\} + E\{h_3(X,S,Y)|X,S\} - E\{h_3(X,S,Y)|X\}\big] \\
& + (1-G)(1-T)\big[h_4(X,S,Y) - E\{h_4(X,S,Y)|X,S\} + E\{h_4(X,S,Y)|X,S\} - E\{h_4(X,S,Y)|X\}\big] \\
& + (1-G)\{T - p(X)\}\big[E\{h_3(X,S,Y)|X\} - E\{h_4(X,S,Y)|X\}\big] \\
& + \{G - \pi(X)\}\bigg(e(X)\big[E\{h_1(X,S)|X\} - E\{h_2(X,S)|X\}\big] + E\{h_2(X,S)|X\} \\
& - p(X)\big[E\{h_3(X,S,Y)|X\} - E\{h_4(X,S,Y)|X\}\big] - E\{h_4(X,S,Y)|X\}\bigg) \\
& + \pi(X)\bigg(e(X)\big[E\{h_1(X,S)|X\} - E\{h_2(X,S)|X\}\big] + E\{h_2(X,S)|X\} \\
& - p(X)\big[E\{h_3(X,S,Y)|X\} - E\{h_4(X,S,Y)|X\}\big] - E\{h_4(X,S,Y)|X\}\bigg)
\end{aligned}
$$

therefore we let

$$
\begin{aligned}
S_1(S|X) &= h_1(X,S) - E\{h_1(X,S)|X\} \\
S_0(S|X) &= h_2(X,S) - E\{h_2(X,S)|X\} \\
a(X) &= E\{h_1(X,S)|X\} - E\{h_2(X,S)|X\} \\
b(X) &= \Big(e(X)\big[E\{h_1(X,S)|X\} - E\{h_2(X,S)|X\}\big] + E\{h_2(X,S)|X\} \\
&\quad - p(X)\big[E\{h_3(X,S,Y)|X\} - E\{h_4(X,S,Y)|X\}\big] - E\{h_4(X,S,Y)|X\}\Big) \\
S(Y|S,X,1) &= h_3(X,S,Y) - E\{h_3(X,S,Y)|X,S\} \\
S(S|X,1) &= E\{h_3(X,S,Y)|X,S\} - E\{h_3(X,S,Y)|X\} \\
S(Y|S,X,0) &= h_4(X,S,Y) - E\{h_4(X,S,Y)|X,S\} \\
S(S|X,0) &= E\{h_4(X,S,Y)|X,S\} - E\{h_4(X,S,Y)|X\} \\
c(X) &= E\{h_3(X,S,Y)|X\} - E\{h_4(X,S,Y)|X\} \\
S(X) &= \pi(X)\Big(e(X)\big[E\{h_1(X,S)|X\} - E\{h_2(X,S)|X\}\big] + E\{h_2(X,S)|X\} \\
&\quad - p(X)\big[E\{h_3(X,S,Y)|X\} - E\{h_4(X,S,Y)|X\}\big] - E\{h_4(X,S,Y)|X\}\Big)
\end{aligned}
$$

So any function of observed data $h(T,X,S,Y,G)$ that has mean zero can be represented as an element of $\mathcal{T}$, so ϕ is the unique influence function for τ in the nonparametric model.

□

3. Proof of Theorem 2

Proof of Theorem 2. Following the proof of Vermeulen and Vansteelandt (2015), we first give the influence function for $\hat{\tau}_{dr}$, then the double robustness and asymptotically normal follow in a straightforward way.

Denote $Z = (T, X, S, Y, G)$ and

$$
\begin{aligned}
&\psi(Z; p, \alpha, \beta, \gamma, \eta) \\
=&\frac{G}{q}\left[\frac{T\{\mu_1(S,X;\alpha_1) - \mu_1(X;\beta_1)\}}{e(X;\gamma)} - \frac{(1-T)\{\mu_0(S,X;\alpha_0) - \mu_0(X;\beta_0)\}}{1-e(X;\gamma)} + \mu_1(X;\beta_1) - \mu_0(X;\beta_0) - \tau\right] \\
&+ \frac{1-G}{q}\left[\frac{g_1(S,X;\eta_1)T\{Y - \mu_1(S,X;\alpha_1)\}}{e(X;\gamma)\{1-g_1(S,X;\eta_1)\}} - \frac{g_0(S,X;\eta_0)(1-T)\{Y-\mu_0(S,X;\alpha_0)\}}{\{1-e(X;\gamma)\}\{1-g_0(S,X;\eta_0)\}}\right],
\end{aligned}
$$

where $\alpha = (\alpha_1^\top, \alpha_0^\top)^\top, \beta = (\beta_1^\top, \beta_0^\top)^\top$ and $\eta = (\eta_1^\top, \eta_0^\top)^\top$. We denote $\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\eta}$ as the probability limits of estimators $\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\eta}$. Besides, we denote $\alpha^*, \beta^*, \gamma^*, \eta^*$ as the true parameter when the working models are correctly specified, respectively. Under suitable regularity conditions (Vermeulen and Vansteelandt, 2015), we have that

$$
\begin{aligned}
n^{1/2}(\hat{\tau}_{dr} - \tau) =& n^{1/2}\hat{E}\{\psi(Z; p, \bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\eta}\} + E\{\frac{\partial \psi}{\partial p}\}n^{1/2}(\hat{p} - p) \\
&+ E\{\frac{\partial \psi}{\partial \alpha}\}n^{1/2}(\hat{\alpha} - \bar{\alpha}) + E\{\frac{\partial \psi}{\partial \beta}\}n^{1/2}(\hat{\beta} - \bar{\beta}) \\
&+ E\{\frac{\partial \psi}{\partial \gamma}\}n^{1/2}(\hat{\gamma} - \bar{\gamma}) + E\{\frac{\partial \psi}{\partial \eta}\}n^{1/2}(\hat{\eta} - \bar{\eta}) + o_p(1),
\end{aligned}
$$

So the consistency of $\hat{\tau}_{dr}$ relies on the mean zero property of $\psi(Z; p, \bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\eta})$.

Notice that

$$
\begin{aligned}
&E\left[\frac{T\{\mu_1(S, X; \alpha_1^*) - \mu_1(X; \beta_1^*)\}}{e(X; \bar{\gamma})}|X, G = 1\right] \\
=&E\left[\frac{T\{\mu_1(S(1), X; \alpha_1^*) - \mu_1(X; \beta_1^*)\}}{e(X; \bar{\gamma})}|X, G = 1\right] \\
=&E[\mu_1(S(1), X; \alpha_1^*) - \mu_1(X; \beta_1^*)|X, G = 1] = 0 \\
&E\left[\frac{g_1(S, X; \bar{\eta}_1)T\{Y - \mu_1(S, X; \alpha_1^*)\}}{e(X; \bar{\gamma})\{1 - g_1(S, X; \bar{\eta}_1)\}}|S, X, T, G = 0\right] \\
=&\frac{g_1(S, X; \bar{\eta}_1)TE\{Y|S, X, T, G = 0\}}{e(X; \bar{\gamma})\{1 - g_1(S, X; \bar{\eta}_1)\}} - \frac{g_1(S, X; \bar{\eta}_1)T\mu_1(S, X; \alpha_1^*)}{e(X; \bar{\gamma})\{1 - g_1(S, X; \bar{\eta}_1)\}} \\
=&\frac{g_1(S, X; \bar{\eta}_1)T\mu_1(S, X; \alpha_1^*)}{e(X; \bar{\gamma})\{1 - g_1(S, X; \bar{\eta}_1)\}} - \frac{g_1(S, X; \bar{\eta}_1)T\mu_1(S, X; \alpha_1^*)}{e(X; \bar{\gamma})\{1 - g_1(S, X; \bar{\eta}_1)\}} = 0
\end{aligned}
$$

When the condition (i) holds, that is the working regression models $\mu_i(S, X; \alpha_i)$ and $\mu_i(X; \beta_i)$ are correctly specified, thus $\bar{\alpha}, \bar{\beta}$ is the true parameter α^*, β^*.

So we have

$$
\begin{aligned}
&E\{\psi(Z;p,\alpha^*,\beta^*,\bar{\gamma},\bar{\eta})\}\\
=&E\Bigg[\frac{G}{q}\Bigg\{\frac{TE\{\mu_1(S,X;\alpha_1^*)-\mu_1(X;\beta_1^*)|T=1,X,G=1\}}{e(X;\bar{\gamma})}\\
&-\frac{(1-T)E\{\mu_0(S,X;\alpha_0^*)-\mu_0(X;\beta_0^*)|T=0,X,G=1\}}{1-e(X;\bar{\gamma})}\\
&+\mu_1(X;\beta_1^*)-\mu_0(X;\beta_0^*)-\tau\Bigg\}\\
&+\frac{1-G}{q}\Bigg\{\frac{g_1(S,X;\bar{\eta}_1)T\{Y-\mu_1(S,X;\alpha_1^*)\}}{e(X;\bar{\gamma})\{1-g_1(S,X;\bar{\eta}_1)\}}\\
&-\frac{g_0(S,X;\bar{\eta}_0)(1-T)\{Y-\mu_0(S,X;\alpha_0^*)\}}{\{1-e(X;\bar{\gamma})\}\{1-g_0(S,X;\bar{\eta}_0)\}}\Bigg\}\Bigg]\\
=&E\{\mu_1(X;\beta_1^*)-\mu_0(X;\beta_0^*)-\tau\}=0
\end{aligned}
$$

When condition (ii) holds, that is the working regression model $\mu_t(s,x;\alpha_t)$ and propensity score model $e(x;\gamma)$ are correctly specified, thus $\bar{\alpha},\bar{\gamma}$ are equal to α^*,γ^* respectively.

Then we have

$$
\begin{aligned}
&E\{\psi(Z;p,\alpha^*,\bar{\beta},\gamma^*,\bar{\eta})\}\\
=&E\Bigg[\frac{G}{q}\Bigg\{\frac{E[T\{\mu_1(S(1),X;\alpha_1^*)-\mu_1(X;\bar{\beta}_1)\}|X,G=1]}{e(X;\gamma^*)}\\
&-\frac{E[(1-T)\{\mu_0(S(0),X;\alpha_0^*)-\mu_0(X;\bar{\beta}_0)\}|X,G=1]}{1-e(X;\gamma^*)}\\
&+\mu_1(X;\bar{\beta}_1)-\mu_0(X;\bar{\beta}_0)-\tau\Bigg\}\\
&+\frac{1-G}{q}\Bigg\{\frac{g_1(S,X;\bar{\eta}_1)T\{Y-\mu_1(S,X;\alpha_1^*)\}}{e(X;\gamma^*)\{1-g_1(S,X;\bar{\eta}_1)\}}\\
&-\frac{g_0(S,X;\bar{\eta}_0)(1-T)\{Y-\mu_0(S,X;\alpha_0^*)\}}{\{1-e(X;\gamma^*)\}\{1-g_0(S,X;\bar{\eta}_0)\}}\Bigg\}\Bigg]\\
=&E\Bigg[E\{\mu_1(S(1),X;\alpha_1^*)-\mu_1(X;\bar{\beta}_1)|X,G=1\}\\
&-E\{\mu_0(S(0),X;\alpha_0^*)-\mu_0(X;\bar{\beta}_0)|X,G=1\}\\
&+E\{\mu_1(X;\bar{\beta}_1)-\mu_0(X;\bar{\beta}_0)|X,G=1\}-\tau\Bigg]\\
=&E\Bigg[E\{\mu_1(S(1),X;\alpha_1^*)|X,G=1\}-E\{\mu_0(S(0),X;\alpha_0^*)|X,G=1\}-\tau\Bigg]=0
\end{aligned}
$$

Therefore we have proved that $E\{\psi(Z;p,\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\eta})\}=0$ under condition (i) or (ii), thus $\hat{\tau}_{dr}$ has the doubly robust property. In particular, when all the working models are correctly specified, we have

$$
E\{\frac{\partial\psi}{\partial p}\}=E\{\frac{\partial\psi}{\partial\alpha}\}=E\{\frac{\partial\psi}{\partial\beta}\}=E\{\frac{\partial\psi}{\partial\gamma}\}=E\{\frac{\partial\psi}{\partial\eta}\}=0
$$

thus the influence function of $\hat{\tau}_{dr}$ is exactly the efficient influence function ϕ.

□

4. Proof of Theorem 3

Proof of Theorem 3(i). $\sqrt{n_1}(\hat{\tau}_{ipw} - \tau)$ can be decomposed as follows.

$$\sqrt{n_1}(\hat{\tau}_{ipw} - \tau) = \sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{(T_i - e_i)\cdot \hat{h}_i}{e_i(1-e_i)} - E[\frac{(T_i - e_i)\cdot h_i}{e_i(1-e_i)}|G_i = 1]\right\}$$
$$:= U_{1n} + U_{2n},$$

where

$$U_{1n} = \sqrt{\frac{n_1}{n_0}}\cdot\frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{(T_i - e_i)}{e_i(1-e_i)}\cdot\sqrt{n_0}(\hat{h}_i - h_i)\right\}$$
$$U_{2n} = \sqrt{n_1}\cdot\frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{(T_i - e_i)\cdot h_i}{e_i(1-e_i)} - E[\frac{(T_i - e_i)\cdot h_i}{e_i(1-e_i)}|G_i = 1]\right\}.$$

We focus on analyzing U_{1n}. By a Taylor expansion, we obtain that for $i = 1, ..., n_1$,

$$\sqrt{n_0}(\hat{h}_i - h_i) = [h_i'(\kappa^*)]^T \cdot \sqrt{n_0}(\hat{\kappa} - \kappa^*) + o_p(1). \tag{S.2}$$

By the property of maximum likelihood estimation for generalized linear models,

$$\sqrt{n_0}(\hat{\kappa} - \kappa^*) = I^{-1}(\kappa^*) \frac{1}{\sqrt{n_0}} \sum_{i=n_1+1}^{n_1+n_0} (Y_i - h_i)\tilde{X}_i/\phi + o_p(1), \tag{S.3}$$

where $I(\kappa^*)$ is the Fisher information matrix of κ at κ^* in the observational data, ϕ is the scale parameter in the corresponding generalized linear model. Then applying equations (S.2) and (S.3) yield that

$$\begin{aligned}
U_{1n} &= \sqrt{\frac{n_1}{n_0}} \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ \frac{(T_i - e_i)}{e_i(1-e_i)} \cdot [h_i'(\kappa^*)]^T \cdot \sqrt{n_0}(\hat{\kappa} - \kappa^*) \right\} + o_p(1) \\
&= \sqrt{\frac{n_1}{n_0}} \cdot \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{(T_i - e_i)}{e_i(1-e_i)} \cdot h_i'(\kappa^*) \right\}^T \cdot I^{-1}(\kappa^*) \frac{1}{\sqrt{n_0}} \sum_{i=n_1+1}^{n_1+n_0} (Y_i - h_i)\tilde{X}_i/\phi + o_p(1) \\
&= \sqrt{\rho} \cdot B_1^T \cdot I^{-1}(\kappa^*) \frac{1}{\sqrt{n_0}} \sum_{i=n_1+1}^{n_1+n_0} (Y_i - h_i)\tilde{X}_i/\phi + o_p(1),
\end{aligned}$$

where

$$B_1 = E\left[\frac{(T_i - e_i)}{e_i(1-e_i)} \cdot h_i'(\kappa^*) \mid G_i = 1 \right].$$

Thus,

$$\begin{aligned}
\sqrt{n_1}(\hat{\tau}_{ipw} - \tau) = \sqrt{\rho} \cdot B_1^T \cdot I^{-1}(\kappa^*) \frac{1}{\sqrt{n_0}} \sum_{i=n_1+1}^{n_1+n_0} (Y_i - h_i)\tilde{X}_i/\phi + \\
\frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \left[\frac{(T_i - e_i) \cdot h_i}{e_i(1-e_i)} - E\left\{ \frac{(T_i - e_i) \cdot h_i}{e_i(1-e_i)} | G_i = 1 \right\} \right] + o_p(1).
\end{aligned} \tag{S.4}$$

Since the first two terms on the right-hand side of the above equation are independent, then Theorem 3(i) follows immediately from equation (S.4) the true that

$$var\{\sqrt{\rho}B_1^T \cdot I^{-1}(\kappa^*)\frac{1}{\sqrt{n_0}}\sum_{i=n_1+1}^{n_1+n_0}(Y_i - h_i)\tilde{X}_i/\phi\}$$
$$= \rho B_1^T \cdot I^{-1}(\kappa^*) \cdot var\{(Y_i - h_i)\tilde{X}_i/\phi\}\}I^{-1}(\kappa^*)B_1 = \rho B_1^T I^{-1}(\kappa^*)B_1.$$

□

Proof of Theorem 3(ii). We decompose $\sqrt{n_1}(\tilde{\tau}_{ipw} - \tau)$ as follows.

$$\begin{aligned}\sqrt{n_1}(\tilde{\tau}_{ipw} - \tau) &= \sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{(T_i - \hat{e}_i)\cdot\hat{h}_i}{\hat{e}_i(1-\hat{e}_i)} - \tau\right\} \\ &= \sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{(T_i - \hat{e}_i)\cdot\hat{h}_i}{\hat{e}_i(1-\hat{e}_i)} - \frac{(T_i - e_i)\cdot\hat{h}_i}{e_i(1-e_i)} + \frac{(T_i - e_i)\cdot\hat{h}_i}{e_i(1-e_i)} - \tau\right\} \\ &:= U_{3n} - U_{4n} + \sqrt{n_1}(\hat{\tau}_{ipw} - \tau),\end{aligned}$$

where

$$\begin{aligned}U_{3n} &= \sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{T_i\hat{h}_i}{\hat{e}_i} - \frac{T_i\hat{h}_i}{e_i}\right\}, \\ U_{4n} &= \sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{(1-T_i)\hat{h}_i}{1-\hat{e}_i} - \frac{(1-T_i)\hat{h}_i}{1-e_i}\right\},\end{aligned}$$

Firstly, we discuss U_{3n}. According to the properties of logistic regression,

$$\sqrt{n_1}(\hat{e}_i - e_i) = e_i(1-e_i)X_i^T \cdot \sqrt{n_1}(\hat{\gamma} - \gamma^*) + o_p(1),$$
$$= e_i(1-e_i)X_i^T \cdot I^{-1}(\gamma^*) \cdot \frac{1}{\sqrt{n_1}} \sum_{j=1}^{n_1} X_j(T_j - e_j) + o_p(1), \tag{S.5}$$

where $I(\gamma^*)$ is the Fisher information matrix of γ at γ^*. By equation (S.5),

$$\begin{aligned} U_{3n} =& -\sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{T_i\hat{h}_i(\hat{e}_i - e_i)}{e_i^2} + o_p(1) \\ =& -\sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{T_i h_i(\hat{e}_i - e_i)}{e_i^2} + o_p(1) \\ =& -\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{T_i h_i}{e_i^2}e_i(1-e_i)X_i\right\}^T \cdot I^{-1}(\gamma^*) \cdot \frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1}X_j(T_j - e_j) + o_p(1). \end{aligned} \tag{S.6}$$

Secondly, we consider U_{4n}. Through a similar argument to U_{3n},

$$\begin{aligned} U_{4n} &= \sqrt{n_1}\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{(1-T_i)h_i(\hat{e}_i - e_i)}{(1-e_i)^2} + o_p(1) \\ &= \left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{(1-T_i)h_i}{(1-e_i)^2}e_i(1-e_i)X_i\right\}^T \cdot I^{-1}(\gamma^*) \cdot \frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1}X_j(T_j - e_j) + o_p(1). \end{aligned} \tag{S.7}$$

Finally, let

$$B_2 = E[\frac{T_i h_i(1-e_i)X_i}{e_i}|G_i = 1] + E[\frac{(1-T_i)h_i e_i X_i}{1-e_i}|G_i = 1].$$

Combing (S.4), (S.6) and (S.7) yield that

$$\begin{aligned}
\sqrt{n_1}(\tilde{\tau}_{ipw} - \tau) = & -B_2^T \cdot I^{-1}(\gamma^*) \cdot \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} X_i(T_i - e_i) + \sqrt{n_1}(\hat{\tau}_{ipw} - \tau) + o_p(1) \\
= & -B_2^T \cdot I^{-1}(\gamma^*) \cdot \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} X_i(T_i - e_i) + \\
& \sqrt{\rho} \cdot B_1^T \cdot I^{-1}(\kappa^*) \frac{1}{\sqrt{n_0}} \sum_{i=n_1+1}^{n_1+n_0} (Y_i - h_i)\tilde{X}_i/\phi + \\
& \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \left[\frac{(T_i - e_i) \cdot h_i}{e_i(1-e_i)} - E\left\{ \frac{(T_i - e_i) \cdot h_i}{e_i(1-e_i)} |G_i = 1 \right\} \right] + o_p(1).
\end{aligned}$$

Note that

$$var\left\{ -B_2^T \cdot I^{-1}(\gamma^*) \cdot \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} X_i(T_i - e_i) \right\} = B_2^T \cdot I^{-1}(\gamma^*) \cdot B_2,$$

and

$$cov\left\{X_i(T_i - e_i), \frac{(T_i - e_i)\cdot h_i}{e_i(1-e_i)}\right\} = cov\left\{X_i(T_i - e_i), \frac{T_i h_i}{e_i} - \frac{(1-T_i)h_i}{1-e_i}\right\}$$
$$= E[X_i(T_i - e_i)\frac{T_i h_i}{e_i}] - E[X_i(T_i - e_i)\frac{(1-T_i)h_i}{1-e_i}]$$
$$= B_2,$$

then the variance of $\sqrt{n_1}(\tilde{\tau}_{ipw} - \tau)$ is

$$var\{\sqrt{n_1}(\tilde{\tau}_{ipw} - \tau)\} = B_2^T \cdot I^{-1}(\gamma^*) \cdot B_2 + var\{\sqrt{n_1}(\hat{\tau}_{ipw} - \tau)\}$$
$$- 2B_2^T \cdot I^{-1}(\gamma^*) \cdot cov\left\{X_i(T_i - e_i), \frac{(T_i - e_i)\cdot h_i}{e_i(1-e_i)}\right\} + o_p(1)$$
$$= B_2^T \cdot I^{-1}(\gamma^*) \cdot B_2 + var\{\sqrt{n_1}(\hat{\tau}_{ipw} - \tau)\} - 2B_2^T \cdot I^{-1}(\gamma^*) \cdot B_2 + o_p(1)$$
$$= var\{\sqrt{n_1}(\hat{\tau}_{ipw} - \tau)\} - B_2^T \cdot I^{-1}(\gamma^*) \cdot B_2 + o_p(1).$$

This completes the proof of Theorem 3(ii).

$\square$

5. Additional Results for Simulation Study

5.1 Data generation mechanisms for binary outcome

In contrast with cases (1)-(6), the data generating process of cases (7)-(12) are as follows:

Case (7). For RCT data , $S = U + 2(X_1 + X_2) + T + \varepsilon_S$, $\text{pr}(Y = 1 \mid T, X, S, U) = \text{expit}\{T + 3(X_1 + X_2) + S\}$, $X = (X_1, X_2)^T \sim N(0, I_2)$. For observational data, $X \sim N(1, 4I_2), U \sim N(0, 1)$, $\text{pr}(T = 1|X, U) = \text{expit}\{U + X_1 + X_2\}$, S and Y are generated the same as in RCT data .

Case (8). For RCT data , $S = U^2 + 2(X_1^2 + X_2^2) + T + \varepsilon_S$, $\text{pr}(Y = 1 \mid T, X, S, U) = \text{expit}\{T + 3(X_1 + X_2) + S\}$, $X = (X_1, X_2)^T \sim N(0, I_2)$. For observational data, $X \sim N(0, 4I_2), U \sim N(0, 1)$, $\text{pr}(T = 1|X, U) = \text{expit}\{U + X_1 + X_2\}$, S and Y are generated the same as in RCT data .

Case (9). The data generation mechanism is the same as in case (7), except for setting $U \sim N(1, 4)$ in observational data.

Case (10). The data generation mechanism is the same as in case (8), except for setting $U \sim N(1, 4)$ in observational data.

Cases (11) and (12). The data generating process of cases (11)-(12) are the same as cases (7)-(8), respectively, except that setting X and S as discrete variables. For cases (11)-(12), the covariates X consisting of two binary variables (X_1, X_2), X_1 and X_2 are independent and identically distributed from a Binomial distribution $B(1, 0.5)$ for both the RCT and observational data. The surrogate S is generated through a logistic regression with $\text{pr}(S = 1 \mid X, S, U, T) = \text{expit}\{U + 2(X_1 + X_2) + T\}$ for case (11) and $\text{pr}(S = 1 \mid X, S, U, T) = \text{expit}\{U^2 + 2(X_1^2 + X_2^2) + T\}$ for case (12).

5.2 Additional simulation: Assumption Violation

We further consider scenarios where the unmeasured confounder U affects both S and Y in observational data. In this case, Assumption 3 may not be satisfied. Four extra cases are set as follows:

Case (13). For RCT data , $S = U + 2(X_1 + X_2) + T + \varepsilon_S$, $Y = U + T + 3(X_1 + X_2) + S + \varepsilon_Y$, $X = (X_1, X_2)^T \sim N(0, I_2)$. For observational data, $X \sim N(1, 4I_2), U \sim N(0, 1)$, $\text{pr}(T = 1|X, U) = \text{expit}\{X_1 + X_2\}$, S and Y are generated the same as in RCT data .

Case (14). For RCT data , $S = U^2 + 2(X_1^2 + X_2^2) + T + \varepsilon_S$, $Y = T + 3(X_1 + X_2) + S + \varepsilon_Y$, $X = (X_1, X_2)^T \sim N(0, I_2)$. For observational data, $X \sim N(1, 4I_2), U \sim N(0, 1)$, $\text{pr}(T = 1|X, U) = \text{expit}\{X_1 + X_2\}$, S and Y are generated the same as in RCT data.

Case (15). For RCT data , $S = U + 2(X_1 + X_2) + T + \varepsilon_S$, $\text{pr}(Y = 1 \mid T, X, S, U) = \text{expit}\{U + T + 3(X_1 + X_2) + S\}$, $X = (X_1, X_2)^T \sim N(0, I_2)$.

For observational data, $X \sim N(1, 4I_2), U \sim N(0, 1)$, $\text{pr}(T = 1|X, U) = \text{expit}\{X_1 + X_2\}$, S and Y are generated the same as in RCT data .

Case (16). For RCT data , $S = U^2 + 2(X_1^2 + X_2^2) + T + \varepsilon_S$, $\text{pr}(Y = 1 \mid T, X, S, U) = \text{expit}\{U + T + 3(X_1 + X_2) + S\}$, $X = (X_1, X_2)^T \sim N(0, I_2)$. For observational data, $X \sim N(1, 4I_2), U \sim N(0, 1)$, $\text{pr}(T = 1|X, U) = \text{expit}\{X_1 + X_2\}$, S and Y are generated the same as in RCT data.

The cases (13)-(14) for continuous outcomes, cases (15)-(16) for binary outcomes. Tables S1 and S2 summarize the numeric results for cases (13)-(14). These results are similar to those in Tables 1 and 3 of the manuscript and indicate that the proposed methods also perform well under the scenario where U affects both S and Y.

6. Additional Results for Application

We report the real-data analysis results where the asymptotic standard errors are computed with the plug-in method, which are presented in Table S3.

References

Athey, S., R. Chetty, G. Imbens, and H. Kang (2019). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Working

Table S1: Comparison of various estimators for cases (13)-(14), continuous outcome.

	$n_1 = 50$					$n_1 = 100$					$n_1 = 200$				
Case	Bias (SD)	ESE	CP95	ESE.b	CP95.b	Bias (SD)	ESE	CP95	ESE.b	CP95.b	Bias (SD)	ESE	CP95	ESE.b	CP95.b
IPW Estimator ($\hat{\tau}_{ipw}$), with True Propensity Score															
(13)	1.7 (210.0)	209.5	94.9	206.5	94.4	5.6 (152.6)	148.8	94.3	147.8	93.2	5.9 (102.0)	105.6	94.6	104.6	95.2
(14)	-1.3 (253.1)	238.7	92.3	234.5	91.7	-7.7 (169.6)	170.0	95.2	169.1	94.6	-6.7 (124.6)	121.4	94.4	121.9	94.3
IPW Estimator ($\tilde{\tau}_{ipw}$), with Estimated Propensity Score															
(13)	6.9 (69.6)	72.4	96.0	105.5	98.5	5.0 (47.3)	47.0	95.1	52.5	95.9	4.8 (33.6)	33.0	94.5	34.9	95.0
(14)	-3.1 (140.5)	125.2	91.0	162.7	95.4	-1.3 (92.6)	88.9	93.9	97.6	94.9	-4.7 (69.9)	63.7	92.8	66.9	94.0
Doubly Robust Estimator ($\hat{\tau}_{dr}$)															
(13)	-2.6 (59.5)	66.0	97.0	57.8	94.1	0.2 (45.8)	48.7	96.6	44.2	94.8	3.0 (33.5)	37.6	97.1	34.1	94.6
(14)	-10 (118.6)	119.1	94.3	105.1	91.0	-4.4 (84.9)	87.7	95.0	81.3	92.5	-6.0 (67.8)	65.0	94.1	62.2	92.6
Athey et al. (2019)'s Method															
(13)	-46.0 (211.8)	214.0	94.8	208.2	93.9	-41.9 (153.7)	151.1	92.9	149.6	91.9	-44.3 (106.7)	109.2	94.7	110.7	93.2
(14)	-104.3 (256.1)	244.2	91.9	237.0	89.8	-114.2 (171.7)	172.7	89.8	170.6	88.1	-118.2 (124.0)	122.9	83.1	122.4	81.9

Note: All the values in this table have been magnified 100 times. Bias and SD are the Monte Carlo bias and standard deviation over the 1000 simulations of the points estimates. ESE and CP95 are the averages of estimated asymptotic standard error and coverage proportions of the 95% confidence intervals based on the plug-in method, respectively. ESE.b and CP95.b have the same meaning as ESE and CP95 but are derived from 200 bootstraps.

Table S2: Comparison of various estimators for cases (15)-(16), binary outcome.

	$n_1 = 50$					$n_1 = 100$					$n_1 = 200$				
Case	Bias (SD)	ESE	CP95	ESE.b	CP95.b	Bias (SD)	ESE	CP95	ESE.b	CP95.b	Bias (SD)	ESE	CP95	ESE.b	CP95.b
IPW Estimator ($\hat{\tau}_{ipw}$), with True Propensity Score															
(15)	-0.2 (20.5)	20.2	93.6	20.0	93.1	0.3 (14.2)	14.6	94.9	14.6	94.7	-0.2 (11.0)	10.6	94.8	10.6	94.3
(16)	0.1 (25.9)	25.7	93.7	25.5	93.6	1.0 (18.2)	18.6	95.1	18.6	94.4	-0.2 (14.1)	13.6	94.4	13.8	94.3
IPW Estimator ($\tilde{\tau}_{ipw}$), with Estimated Propensity Score															
(15)	0.1 (8.2)	8.3	95.3	10.0	96.9	0.2 (6.0)	6.2	95.5	6.5	95.7	-0.1 (5.1)	5.0	94.6	5.2	95.1
(16)	0.3 (9.4)	8.9	93.7	12.3	97.4	0.2 (7.2)	7.0	93.4	7.7	95.6	0.0 (6.7)	6.1	90.7	6.6	92.5
Doubly Robust Estimator ($\hat{\tau}_{dr}$)															
(15)	-0.3 (8.2)	7.3	92.7	8.5	95.2	0.1 (6.2)	5.4	91.5	6.7	96.0	-0.1 (5.3)	4.1	86.5	5.6	95.3
(16)	-0.8 (9.5)	8.3	91.9	9.8	94.8	-1.0 (8.4)	6.9	89.5	8.4	94.6	-1.0 (7.6)	6.0	87.8	7.6	92.1
Athey et al. (2019)'s Method															
(15)	-2.6 (20.2)	20.1	93.8	19.6	93.0	-2.1 (13.9)	14.2	95.3	14.2	94.3	-2.5 (10.5)	10.0	93.2	10.0	93.2
(16)	-6.3 (24.8)	25.0	94.8	24.6	93.4	-5.3 (17.6)	17.6	94.5	17.5	92.8	-6.2 (12.8)	12.6	91.2	12.5	90.3

Note: All the values in this table have been magnified 100 times. Bias and SD are the Monte Carlo bias and standard deviation over the 1000 simulations of the points estimates. ESE and CP95 are the averages of estimated asymptotic standard error and coverage proportions of the 95% confidence intervals based on the plug-in method, respectively. ESE.b and CP95.b have the same meaning as ESE and CP95 but are derived from 200 bootstraps.

Table S3: Estimated effects of HCQ on renal failure.

		proportion = 0.3		proportion = 0.4		proportion = 0.5	
End time		Endpoint 1	Endpoint 2	Endpoint 1	Endpoint 2	Endpoint 1	Endpoint 2
		IPW, with True Propensity Score					
3	Estimate (ESE)	-0.376 (0.100)	-0.202 (0.090)	-0.162 (0.062)	-0.202 (0.090)	-0.123 (0.068)	-0.202 (0.090)
	p-value	$< 10^{-3}$	0.013	0.005	0.013	0.035	0.013
4	Estimate (ESE)	-0.418 (0.107)	-0.157 (0.076)	-0.211 (0.070)	-0.157 (0.076)	-0.172 (0.074)	-0.157 (0.076)
	p-value	$< 10^{-3}$	0.020	0.002	0.020	0.010	0.020
5	Estimate (ESE)	-0.488 (0.113)	-0.192 (0.082)	-0.265 (0.075)	-0.192 (0.082)	-0.211 (0.077)	-0.192 (0.082)
	p-value	$< 10^{-3}$	0.010	$< 10^{-3}$	0.010	0.003	0.010
		IPW, with Estimated Propensity Score					
3	Estimate (ESE)	-0.318 (0.063)	-0.178 (0.074)	-0.138 (0.051)	-0.178 (0.074)	-0.102 (0.062)	-0.178 (0.074)
	p-value	$< 10^{-3}$	0.008	0.004	0.008	0.051	0.008
4	Estimate (ESE)	-0.360 (0.063)	-0.145 (0.065)	-0.182 (0.055)	-0.145 (0.065)	-0.155 (0.063)	-0.145 (0.065)
	p-value	$< 10^{-3}$	0.013	$< 10^{-3}$	0.013	0.007	0.013
5	Estimate (ESE)	-0.423 (0.062)	-0.178 (0.070)	-0.232 (0.055)	-0.178 (0.070)	-0.189 (0.064)	-0.178 (0.070)
	p-value	$< 10^{-3}$	0.006	$< 10^{-3}$	0.006	0.002	0.006
		Doubly Robust Estimator					
3	Estimate (ESE)	-0.256 (0.086)	-0.187 (0.063)	-0.073 (0.077)	-0.187 (0.063)	-0.034 (0.053)	-0.187 (0.063)
	p-value	0.001	0.001	0.173	0.001	0.258	0.001
4	Estimate (ESE)	-0.337 (0.088)	-0.148 (0.050)	-0.125 (0.085)	-0.148 (0.050)	-0.092 (0.059)	-0.148 (0.050)
	p-value	$< 10^{-3}$	0.002	0.071	0.002	0.058	0.002
5	Estimate (ESE)	-0.396 (0.091)	-0.195 (0.057)	-0.169 (0.093)	-0.195 (0.057)	-0.119 (0.064)	-0.195 (0.057)
	p-value	$< 10^{-3}$	$< 10^{-3}$	0.035	$< 10^{-3}$	0.032	$< 10^{-3}$
		Athey et al. (2019)'s Method					
3	Estimate (ESE)	-0.034 (0.138)	-0.050 (0.070)	-0.045 (0.075)	-0.050 (0.070)	-0.031 (0.053)	-0.050 (0.070)
	p-value	0.402	0.239	0.275	0.239	0.282	0.239
4	Estimate (ESE)	-0.039 (0.150)	-0.043 (0.063)	-0.071 (0.086)	-0.043 (0.063)	-0.046 (0.064)	-0.043 (0.063)
	p-value	0.397	0.247	0.205	0.247	0.235	0.247
5	Estimate (ESE)	-0.053 (0.159)	-0.023 (0.071)	-0.082 (0.099)	-0.023 (0.071)	-0.060 (0.075)	-0.023 (0.071)
	p-value	0.371	0.370	0.203	0.370	0.210	0.370

Note: ESE is estimated asymptotic standard error based on the plug-in method. The p-values are obtained by two-sided test, that is $H_0 : \tau = 0$ against $H_1 : \tau \neq 0$

Paper 26463, National Bureau of Economic Research.

Vermeulen, K. and S. Vansteelandt (2015). Bias-reduced doubly robust estimation. *Journal of the American Statistical Association 110*(511), 1024–1036.