Supplementary material

S1 Additional details on identification assumptions

S1.1 Causal mechanism of the placebo sample in the EITC study

Figure S1: Causal mechanism of a placebo sample in the EITC study. A encodes whether an individual's state of residence has enacted EITC laws, M whether she claimed the EITC, and Y whether she died of death of despair. A's effect on Y is exclusively mediated by M. For individuals in the placebo sample, A does not affect M and thus has no effect on Y. A subset of state residents, e.g., college graduates, are unlikely to be eligible for the EITC and thus cannot claim the EITC with or without EITC laws, i.e., A has no effect on M (as illustrated in (b)). (a) DAG for the primary sample ($S = 1$) and (b) DAG for the placebo sample ($S = 0$).

S1.2 Further discussion of Assumption 2

Let U denote the set of unmeasured confounders such that $A \perp (Y^{(0)}, Y^{(1)})|U, X, S$. One set of sufficient (but not necessary) conditions for Assumption 2 to hold is (i) $P(U|S = 1, A, X) = P(U|S = 0, A, X)$, and (ii) $E\{Y^{(0)}|U, X, S = 1\} - E\{Y^{(0)}|U, X, S = 0\}$ is a function of X alone. In words, Assumption 2 holds when the distribution of U is the same for $S = 0$ and $S = 1$ within each level of (A, X), and there is no additive S-U interaction in $E\{Y^{(0)}|U, X, S\}$.

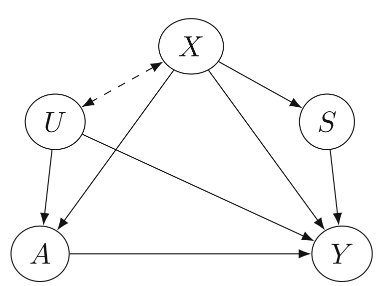

Figure S2: A DAG of a sufficient condition of Assumption 2.

We expect Assumption 2 to hold approximately in various applications. To give an example, in a study of the effect of intraoperative TEE on patients' post-surgery clinical outcomes, there is a concern of unmeasured confounding due to surgeons' experiences as a surgeon' preference for using intraoperative TEE may depend on her experience, and a surgeon's experience may affect complication management during the surgery and hence the clinical outcome [1,2]. Unfortunately, the disease registry data does not contain surgeon-level characteristics. The proposed placebo samples approach could help in this case. The healthiest cardiac surgery group, those under the age of 40 with an ejection fraction > 55% (normal range: 55-75%) are likely not to benefit (or benefit minimally) from the intraoperative TEE (see, e.g., empirical results from [3]). Assumption 1 is likely to hold for this placebo sample. On the other hand, Within the same hospital, it is largely random which surgeon operates on which patient, so that the unmeasured confounder U is likely to have similar distribution for patients in the placebo sample and other patients and Assumption 2 approximately holds. Note that Assumption 2 will not likely to hold without conditioning on the hospital indicator X = hospital indicator. Community hospitals tend to refer complex surgeries on higher-risk patients to university hospitals; therefore, it is likely that more experienced surgeons in university hospitals (compared to those in community hospitals) would operate on higher-risk patients; however, conditioning on the hospital would render U (surgeons' experience) independent of S (a patient being in the placebo sample). Of course, even though both Assumptions 1 and 2 are likely to approximately hold in this case, we cannot completely rule out the possibility of minor violation, and a sensitivity analysis could be helpful. In the following, we propose sensitivity analysis based on two different sensitivity models that may be useful in a broad range of applications.

S1.3 Estimators for sensitivity analysis under marginal sensitivity models

The bounds in the sensitivity analysis can be estimated using an IPW estimator, a regression-based estimator, and a doubly-robust estimator. We provide formulas for each below.

Using the IPW estimator, we have

$$\begin{aligned}
\hat{\theta}_{U,\mathrm{ipw}} = \hat{\theta}_{\mathrm{ipw}} + \Lambda - (\Gamma^{-1} - 1)\frac{1}{n_{11}}\sum_{i=1}^{n}\frac{\pi_S(X_i;\, \hat{\psi})\pi_A(X_i, 1;\, \hat{\alpha})}{(1 - \pi_S(X_i;\, \hat{\psi}))\pi_A(X_i, 0;\, \hat{\alpha})}(1 - S_i)A_i Y_i \\
+ (\Gamma - 1)\frac{1}{n_{11}}\sum_{i=1}^{n}\frac{\pi_S(X_i;\, \hat{\psi})\pi_A(X_i, 1;\, \hat{\alpha})}{(1 - \pi_S(X_i;\, \hat{\psi}))(1 - \pi_A(X_i, 0;\, \hat{\alpha}))}(1 - S_i)(1 - A_i)Y_i, \\
\hat{\theta}_{L,\mathrm{ipw}} = \hat{\theta}_{\mathrm{ipw}} - \Lambda - (\Gamma - 1)\frac{1}{n_{11}}\sum_{i=1}^{n}\frac{\pi_S(X_i;\, \hat{\psi})\pi_A(X_i, 1;\, \hat{\alpha})}{(1 - \pi_S(X_i;\, \hat{\psi}))\pi_A(X_i, 0;\, \hat{\alpha})}(1 - S_i)A_i Y_i \\
+ (\Gamma^{-1} - 1)\frac{1}{n_{11}}\sum_{i=1}^{n}\frac{\pi_S(X_i;\, \hat{\psi})\pi_A(X_i, 1;\, \hat{\alpha})}{(1 - \pi_S(X_i;\, \hat{\psi}))(1 - \pi_A(X_i, 0;\, \hat{\alpha}))}(1 - S_i)(1 - A_i)Y_i.
\end{aligned}$$

Using the regression-based estimator, we have

$$\begin{aligned}
\hat{\theta}_{U,\mathrm{reg}} = \hat{\theta}_{\mathrm{reg}} + \Lambda - (\Gamma^{-1} - 1)\frac{1}{n_{11}}\sum_{i:S_i=A_i=1}\mu_Y(0, 1, X_i;\, \hat{\beta}) + (\Gamma - 1)\frac{1}{n_{11}}\sum_{i:S_i=A_i=1}\mu_Y(0, 0, X_i;\, \hat{\beta}), \\
\hat{\theta}_{L,\mathrm{reg}} = \hat{\theta}_{\mathrm{reg}} - \Lambda - (\Gamma - 1)\frac{1}{n_{11}}\sum_{i:S_i=A_i=1}\mu_Y(0, 1, X_i;\, \hat{\beta}) + (\Gamma^{-1} - 1)\frac{1}{n_{11}}\sum_{i:S_i=A_i=1}\mu_Y(0, 0, X_i;\, \hat{\beta}).
\end{aligned}$$

Using the doubly-robust estimator, we have

$$\begin{aligned}
\hat{\theta}_{U,\mathrm{dr}} = \hat{\theta}_{\mathrm{dr}} + \Lambda - (\Gamma^{-1} - 1)\frac{1}{n_{11}}\sum_{i:S_i=A_i=1}\mu_Y(0, 1, X_i;\, \hat{\beta}) \\
- (\Gamma^{-1} - 1)\frac{1}{n_{11}}\sum_{i=1}^{n}\frac{\pi_S(X_i;\, \hat{\psi})\pi_A(X_i, 1;\, \hat{\alpha})}{(1 - \pi_S(X_i;\, \hat{\psi}))\pi_A(X_i, 0;\, \hat{\alpha})}(1 - S_i)A_i\{Y_i - \mu_Y(0, 1, X_i;\, \hat{\beta})\} \\
+ (\Gamma - 1)\frac{1}{n_{11}}\sum_{i:S_i=A_i=1}\mu_Y(0, 0, X_i;\, \hat{\beta}) \\
+ (\Gamma - 1)\frac{1}{n_{11}}\sum_{i=1}^{n}\frac{\pi_S(X_i;\, \hat{\psi})\pi_A(X_i, 1;\, \hat{\alpha})}{(1 - \pi_S(X_i;\, \hat{\psi}))(1 - \pi_A(X_i, 0;\, \hat{\alpha}))}(1 - S_i)(1 - A_i)\{Y_i - \mu_Y(0, 0, X_i;\, \hat{\beta})\}
\end{aligned}$$

$$
\begin{aligned}
\hat{\theta}_{L,\mathrm{dr}} = \ & \hat{\theta}_{\mathrm{dr}} - \Lambda \\
& - (\Gamma - 1)\frac{1}{n_{11}} \sum_{i:S_i=A_i=1} \mu_Y(0, 1, X_i;\ \hat{\beta}) \\
& - (\Gamma - 1)\frac{1}{n_{11}} \sum_{i=1}^{n} \frac{\pi_S(X_i;\ \hat{\psi})\pi_A(X_i, 1;\ \hat{\alpha})}{(1 - \pi_S(X_i;\ \hat{\psi}))\pi_A(X_i, 0;\ \hat{\alpha})}(1 - S_i)A_i\{Y_i - \mu_Y(0, 1, X_i;\ \hat{\beta})\} \\
& + (\Gamma^{-1} - 1)\frac{1}{n_{11}} \sum_{i:S_i=A_i=1} \mu_Y(0, 0, X_i;\ \hat{\beta}) \\
& + (\Gamma^{-1} - 1)\frac{1}{n_{11}} \sum_{i=1}^{n} \frac{\pi_S(X_i;\ \hat{\psi})\pi_A(X_i, 1;\ \hat{\alpha})}{(1 - \pi_S(X_i;\ \hat{\psi}))(1 - \pi_A(X_i, 0;\ \hat{\alpha}))}(1 - S_i)(1 - A_i)\{Y_i - \mu_Y(0, 0, X_i;\ \hat{\beta})\}.
\end{aligned}
$$

Under the Sensitivity Model 2.A and 2.B, and $Y_i \geq 0$ and for fixed Γ and Λ, the confidence interval for θ_0 is

$$
[\hat{\theta}_{L,*} - z_{\alpha/2}\hat{\sigma}_{L,*},\ \hat{\theta}_{U,*} + z_{\alpha/2}\hat{\sigma}_{U,*}], \tag{S1}
$$

where $z_{\alpha/2}$ is the $\alpha/2$-upper quantile of the standard normal distribution, and $\hat{\sigma}_{L,*}, \hat{\sigma}_{U,*}$ are respectively the standard error of $\hat{\theta}_{L,*}, \hat{\theta}_{U,*}$ for $*\in\{\mathrm{ipw}, \mathrm{reg}, \mathrm{dr}\}$.

S1.4 Stabilized estimators

The idea of stabilization is to normalize the weights to sum to one [4]. The stabilized estimators are asymptotically equivalent to their unstabilized counterparts but they have better finite-sample performance.

The stabilized IPW is

$$
\begin{aligned}
\hat{\theta}_{\mathrm{sipw}} = & \frac{1}{n_{11}} \sum_{i=1}^{n} S_i A_i Y_i \\
& - \left\{\sum_{i=1}^{n} \frac{S_i(1 - A_i)\pi_A(X_i, 1;\ \hat{\alpha})}{1 - \pi_A(X_i, 1;\ \hat{\alpha})}\right\}^{-1} \sum_{i=1}^{n} \frac{S_i(1 - A_i)\pi_A(X_i, 1;\ \hat{\alpha})}{1 - \pi_A(X_i, 1;\ \hat{\alpha})} Y_i \\
& - \left\{\sum_{i=1}^{n} \frac{(1 - S_i)A_i\pi_S(X_i;\ \hat{\psi})}{1 - \pi_S(X_i;\ \hat{\psi})} \frac{\pi_A(X_i, 1;\ \hat{\alpha})}{\pi_A(X_i, 0;\ \hat{\alpha})}\right\}^{-1} \sum_{i=1}^{n} \frac{(1 - S_i)A_i\pi_S(X_i;\ \hat{\psi})}{1 - \pi_S(X_i;\ \hat{\psi})} \frac{\pi_A(X_i, 1;\ \hat{\alpha})}{\pi_A(X_i, 0;\ \hat{\alpha})} Y_i \\
& + \left\{\sum_{i=1}^{n} \frac{(1 - S_i)(1 - A_i)\pi_S(X_i;\ \hat{\psi})}{1 - \pi_S(X_i;\ \hat{\psi})} \frac{\pi_A(X_i, 1;\ \hat{\alpha})}{1 - \pi_A(X_i, 0;\ \hat{\alpha})}\right\}^{-1} \sum_{i=1}^{n} \frac{(1 - S_i)(1 - A_i)\pi_S(X_i;\ \hat{\psi})}{1 - \pi_S(X_i;\ \hat{\psi})} \frac{\pi_A(X_i, 1;\ \hat{\alpha})}{1 - \pi_A(X_i, 0;\ \hat{\alpha})} Y_i.
\end{aligned}
$$

S2 Technical Proofs

S2.1 Proof of Proposition 1

The part (a) is directly from the assumptions. We only prove part (b). First note that

$$
\begin{aligned}
& E\left[\frac{S - \pi_S(X)}{\pi_S(X)\{1 - \pi_S(X)\}} \frac{A - \pi_A(X, S)}{\pi_A(X, S)\{1 - \pi_A(X, S)\}} Y | X\right] \\
& \quad = E\left[\frac{1}{\pi_S(X)} \frac{A - \pi_A(X, 1)}{\pi_A(X, 1)\{1 - \pi_A(X, 1)\}} Y | X, S = 1\right] P(S = 1 | X) \\
& \quad - E\left[\frac{1}{\{1 - \pi_S(X)\}} \frac{A - \pi_A(X, 0)}{\pi_A(X, 0)\{1 - \pi_A(X, 0)\}} Y | X, S = 0\right] P(S = 0 | X)
\end{aligned}
$$

$$
\begin{aligned}
&= E\left[\frac{A-\pi_A(X,1)}{\pi_A(X,1)\{1-\pi_A(X,1)\}}Y|X,S=1\right] - E\left[\frac{A-\pi_A(X,0)}{\pi_A(X,0)\{1-\pi_A(X,0)\}}Y|X,S=0\right]\\
&= E\left[\frac{A-\pi_A(X,1)}{\pi_A(X,1)\{1-\pi_A(X,1)\}}Y|X,S=1,A=1\right]P(A=1|X,S=1)\\
&\quad + E\left[\frac{A-\pi_A(X,1)}{\pi_A(X,1)\{1-\pi_A(X,1)\}}Y|X,S=1,A=0\right]P(A=0|X,S=1)\\
&\quad - E\left[\frac{A-\pi_A(X,0)}{\pi_A(X,0)\{1-\pi_A(X,0)\}}Y|X,S=0,A=1\right]P(A=1|X,S=0)\\
&\quad - E\left[\frac{A-\pi_A(X,0)}{\pi_A(X,0)\{1-\pi_A(X,0)\}}Y|X,S=0,A=0\right]P(A=0|X,S=0)\\
&= E[Y|X,S=1,A=1] - E[Y|X,S=1,A=0]\\
&\quad - E[Y|X,S=0,A=1] + E[Y|X,S=0,A=0]\\
&= E\{Y(1)-Y(0)|S=1,A=1,X\}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\theta_0 &= E\left(E\left[\frac{S-\pi_S(X)}{\pi_S(X)\{1-\pi_S(X)\}}\frac{A-\pi_A(X,S)}{\pi_A(X,S)\{1-\pi_A(X,S)\}}Y|X\right]\frac{P(A=1|S=1,X)P(S=1|X)}{P(S=1,A=1)}\right)\\
&= \frac{1}{P(S=1,A=1)}E\left[\frac{S-\pi_S(X)}{1-\pi_S(X)}\frac{\pi_A(X,1)\{A-\pi_A(X,S)\}}{\pi_A(X,S)\{1-\pi_A(X,S)\}}Y\right].
\end{aligned}
$$

S2.2 Proof of Theorem 1

In this section, we use subscripts to explicitly index quantities that depend on the distribution P, we use a zero subscript to denote a quantity evaluated at the true distribution $P = P_0$, we use a ε subscript to denote a quantity evaluated at the parametric submodel $P = P_\varepsilon$. We will show that $\varphi(O;\, \theta_P, \eta_P)$ is equal to the efficient influence function by showing that it is the canonical gradient of the pathwise derivative of θ_P, i.e., [5]

$$
\left.\frac{\partial \theta_\varepsilon}{\partial \varepsilon}\right|_{\varepsilon=0} = E_0\{\mathrm{EIF}(O;\, \theta_P, \eta_P)s_0(O)\}, \tag{S2}
$$

where $\theta_\varepsilon = \theta_{P_\varepsilon}$, $s_\varepsilon(O) = \partial \log dP_\varepsilon(O)/\partial\varepsilon$ denotes the parameter submodel score, which can be decomposed as $s_\varepsilon(Y|S,A,X) + s_\varepsilon(A|S,X) + s_\varepsilon(S|X) + s_\varepsilon(X)$ or $s_\varepsilon(X|Y,S,A) + s_\varepsilon(Y|S,A) + s_\varepsilon(S,A)$.

By the equality in Proposition 1 (a), we have

$$
\begin{aligned}
\left.\frac{\partial}{\partial\varepsilon}\theta_\varepsilon\right|_{\varepsilon=0} &= \frac{\partial}{\partial\varepsilon}E_\varepsilon(Y|S=1,A=1)|_{\varepsilon=0} \qquad\qquad (\mathrm{S3})\\
&\quad - \frac{\partial}{\partial\varepsilon}E_\varepsilon\{E_\varepsilon(Y|S=1,A=0,X)|S=1,A=1\}|_{\varepsilon=0}\\
&\quad - \frac{\partial}{\partial\varepsilon}E_\varepsilon\{E_\varepsilon(Y|S=0,A=1,X)|S=1,A=1\}|_{\varepsilon=0}\\
&\quad + \frac{\partial}{\partial\varepsilon}E_\varepsilon\{E_\varepsilon(Y|S=0,A=0,X)|S=1,A=1\}|_{\varepsilon=0}\\
&:= M_1 - M_2 - M_3 + M_4.
\end{aligned}
$$

In what follows, we use the following identities repeatedly:

$$
E_\varepsilon\{b(O_1)s_\varepsilon(O_2|O_1)\} = 0, \tag{S4}
$$

$$
E_\varepsilon\{b(O_1)(O_2 - E_\varepsilon(O_2|O_1))\} = 0, \tag{S5}
$$

for any b and any $(O_1, O_2) \subset O$.

Consider each term M_1, M_2, M_3, M_4 separately:

$$
\begin{aligned}
M_1 &= E_0\{Ys_0(Y|S,A)|S=1,A=1\}\\
&= E_0\left[\frac{SA}{E_0(SA)}Ys_0(Y|S,A)\right]\\
&= E_0\left[\frac{SA}{E_0(SA)}\{Y-E_0(Y|S,A)\}s_0(Y|S,A)\right] \quad \text{(S4)}\\
&= E_0\left[\frac{SA}{E_0(SA)}\{Y-E_0(Y|S,A)\}\{s_0(Y|S,A)+s_0(S,A)\}\right] \quad \text{(S5)}\\
&= E_0\left[\frac{SA}{E_0(SA)}\{Y-E_0(Y|S,A)\}\{s_0(Y|S,A)+s_0(S,A)+s_0(X|Y,S,A)\}\right] \quad \text{(S4)}\\
&= E_0\left[\frac{SA}{E_0(SA)}\{Y-E_0(Y|S,A)\}s_0(O)\right]
\end{aligned}
$$

$$
\begin{aligned}
M_2 &= E_0\{E_0(Ys_0(Y|S,A,X)|S=1,A=0,X)|S=1,A=1\}\\
&\quad + E_0\{E_0(Y|S=1,A=0,X)s_0(X|S,A)|S=1,A=1\}\\
&= E_0\left\{\frac{SA}{E_0(SA)}E_0\left(\frac{S(1-A)}{P_0(S=1,A=0|X)}Ys_0(Y|S,A,X)|X\right)\right\}\\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}E_0(Y|S=1,A=0,X)s_0(X|S,A)\right\}\\
&= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{S(1-A)}{E_0(S(1-A)|X)}Ys_0(Y|S,A,X)\right\}\\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}E_0(Y|S=1,A=0,X)s_0(X|S,A)\right\}\\
&= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{S(1-A)}{E_0(S(1-A)|X)}Ys_0(Y|S,A,X)\right\}\\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}E_0(Y|S=1,A=0,X)s_0(X|S,A)\right\}\\
&= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{S(1-A)}{E_0(S(1-A)|X)}\{Y-\mu_{Y0}(S,A,X)\}s_0(Y|S,A,X)\right\} \quad \text{(S4)}\\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}\{\mu_{Y0}(1,0,X)-E_0(\mu_{Y0}(1,0,X)|S,A)\}s_0(X|S,A)\right\} \quad \text{(S4)}\\
&= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{S(1-A)}{E_0(S(1-A)|X)}\{Y-\mu_{Y0}(S,A,X)\}\{s_0(Y|S,A,X)+s_0(S,A,X)\}\right\} \quad \text{(S5)}\\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}\{\mu_{Y0}(1,0,X)-E_0(\mu_{Y0}(1,0,X)|S,A)\}\{s_0(X|S,A)+s_0(S,A)\}\right\} \quad \text{(S5)}\\
&= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{S(1-A)}{E_0(S(1-A)|X)}\{Y-\mu_{Y0}(S,A,X)\}s_0(O)\right\}\\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}\{\mu_{Y0}(1,0,X)-E_0(\mu_{Y0}(1,0,X)|S,A)\}S_0(O)\right\} \quad \text{(S4)}
\end{aligned}
$$

$$\begin{aligned}
M_3 &= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{(1-S)A}{E((1-S)A|X)}Ys_0(Y|S,A,X)\right\} \\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}E_0(Y|S=0,A=1,X)s_0(X|S,A)\right\} \\
&= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{(1-S)A}{E_0((1-S)A|X)}\{Y-\mu_{Y0}(S,A,X)\}s_0(O)\right\} \\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}\{\mu_{Y0}(0,1,X)-E_0(\mu_{Y0}(0,1,X)|S,A)\}s_0(O)\right\} \\
M_4 &= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{(1-S)(1-A)}{E_0((1-S)(1-A)|X)}Ys_0(Y|S,A,X)\right\} \\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}E_0(Y|S=0,A=0,X)s_0(X|S,A)\right\} \\
&= E_0\left\{\frac{E_0(SA|X)}{E_0(SA)}\frac{(1-S)(1-A)}{E_0((1-S)(1-A)|X)}\{Y-\mu_{Y0}(S,A,X)\}s_0(O)\right\} \\
&\quad + E_0\left\{\frac{SA}{E_0(SA)}\{\mu_{Y0}(0,0,X)-E_0(\mu_{Y0}(0,0,X)|S,A)\}s_0(O)\right\}.
\end{aligned}$$

Combining the above derivations, we have

$$\left.\frac{\partial\theta_\varepsilon}{\partial\varepsilon}\right|_{\varepsilon=0} = E_0\{\mathrm{EIF}(O;\,\theta_P,\eta_P)s_0(O)\}, \tag{S6}$$

with $\mathrm{EIF}(O;\,\theta_P,\eta_P)$ being the efficient influence function.

S2.3 Proof of double robustness

Let $\bar{\eta} = (\bar{\mu}_Y, \bar{\pi}_S, \bar{\pi}_A)$ be the limit of $\hat{\eta}$. Here we will show that $E[\mathrm{EIF}(O,\theta_0;\,\bar{\eta})] = 0$ as long as either $\bar{\mu}_Y = \mu_{Y0}$ or $(\bar{\pi}_S, \bar{\pi}_A) = (\pi_{S0}, \pi_{A0})$. In this section, expectations $E = E_0$ are evaluated under P_0, but we drop the subscript for notational convenience.

When μ_Y is correctly specified, we have $\bar{\mu}_Y = \mu_{Y0}$. Then, by iterative expectation,

$$\begin{aligned}
E[\mathrm{EIF}(O,\theta_0;\,\bar{\eta})] &= E\Bigg[\frac{SA}{E(SA)}\{Y-\mu_{Y0}(1,0,X)-\mu_{Y0}(0,1,X)+\mu_{Y0}(0,0,X)-\theta_0\} \\
&\quad - \frac{S(1-A)\frac{\bar{\pi}_A(X,1)}{1-\bar{\pi}_A(X,1)}}{E\left[S(1-A)\frac{\bar{\pi}_A(X,1)}{1-\bar{\pi}_A(X,1)}\right]}\{Y-\mu_{Y0}(1,0,X)\} \\
&\quad - \frac{(1-S)A\frac{\bar{\pi}_A(X,1)}{\bar{\pi}_A(X,0)}\frac{\bar{\pi}_S(X)}{1-\bar{\pi}_S(X)}}{E\left[(1-S)A\frac{\bar{\pi}_A(X,1)}{\bar{\pi}_A(X,0)}\frac{\bar{\pi}_S(X)}{1-\bar{\pi}_S(X)}\right]}\{Y-\mu_{Y0}(0,1,X)\} \\
&\quad + \frac{(1-S)(1-A)\frac{\bar{\pi}_A(X,1)}{1-\bar{\pi}_A(X,0)}\frac{\bar{\pi}_S(X)}{1-\bar{\pi}_S(X)}}{E\left[(1-S)(1-A)\frac{\bar{\pi}_A(X,1)}{1-\bar{\pi}_A(X,0)}\frac{\bar{\pi}_S(X)}{1-\bar{\pi}_S(X)}\right]}\{Y-\mu_{Y0}(0,0,X)\}\Bigg] \\
&= E[Y-\mu_{Y0}(1,0,X)-\mu_{Y0}(0,1,X)+\mu_{Y0}(0,0,X)-\theta_0|S=1,A=1] \\
&= 0.
\end{aligned}$$

When π_A, π_S are correctly specified, we have $\bar{\pi}_S = \pi_{S0}$ and $\bar{\pi}_A = \pi_{A0}$. Then, by iterative expectation,

$$E[\mathrm{EIF}(O,\theta_0;\,\bar{\eta})] = E\Bigg[\frac{SA}{E(SA)}\{Y-\bar{\mu}_Y(1,0,X)-\bar{\mu}_Y(0,1,X)+\bar{\mu}_Y(0,0,X)-\theta_0\}$$

$$
\begin{aligned}
&- \frac{S(1-A)\frac{\pi_{A0}(X,1)}{1-\pi_{A0}(X,1)}}{E\left[S(1-A)\frac{\pi_{A0}(X,1)}{1-\pi_{A0}(X,1)}\right]}\{Y-\bar{\mu}_Y(1,0,X)\} \\
&- \frac{(1-S)A\frac{\pi_{A0}(X,1)}{\pi_{A0}(X,0)}\frac{\pi_{S0}(X)}{1-\pi_{S0}(X)}}{E\left[(1-S)A\frac{\pi_{A0}(X,1)}{\pi_{A0}(X,0)}\frac{\pi_{S0}(X)}{1-\pi_{S0}(X)}\right]}\{Y-\bar{\mu}_Y(0,1,X)\} \\
&+ \left.\frac{(1-S)(1-A)\frac{\pi_{A0}(X,1)}{1-\pi_{A0}(X,0)}\frac{\pi_{S0}(X)}{1-\pi_{S0}(X)}}{E\left[(1-S)(1-A)\frac{\pi_{A0}(X,1)}{1-\pi_{A0}(X,0)}\frac{\pi_{S0}(X)}{1-\pi_{S0}(X)}\right]}\{Y-\bar{\mu}_Y(0,0,X)\}\right] \\
= {} & E[Y|S=1,A=1] - E\{\bar{\mu}_Y(1,0,X)|S=1,A=1\} \\
&- E\{\bar{\mu}_Y(0,1,X)|S=1,A=1\} + E\{\bar{\mu}_Y(0,0,X)|S=1,A=1\} - \theta_0 \\
&- E\{E(Y|S=1,A=0,X) - \bar{\mu}_Y(1,0,X)|S=1,A=1\} \\
&- E\{E(Y|S=0,A=1,X) - \bar{\mu}_Y(0,1,X)|S=1,A=1\} \\
&+ E\{E(Y|S=0,A=0,X) - \bar{\mu}_Y(0,0,X)|S=1,A=1\} \\
= {} & 0.
\end{aligned}
$$

S2.4 Proof of Theorem 2

In what follows, we will use $P\{f(O)\} = \int f(O)dP$ to denote expectation treating the function f as fixed; thus $P\{f(O)\}$ is random when f is random, and is different from the fixed quantity $E\{f(O)\}$ which averages over randomness in both f and O.

Since $\hat{\theta}$ is a Z-estimator, using Theorem 5.31 of [6], we have that

$$\sqrt{n}(\hat{\theta}-\theta_0) = \sqrt{n}P\{\mathrm{EIF}(O;\,\theta_0,\hat{\eta})\} + n^{-1/2}\sum_{i=1}^{n}[\mathrm{EIF}(O_i;\,\theta_0,\bar{\eta}) - E\{\mathrm{EIF}(O_i;\,\theta_0,\bar{\eta})\}] + o_p(1+\sqrt{n}P\{\mathrm{EIF}(O;\,\theta_0,\hat{\eta})\}).$$

Using standard central limit theorem, the second term is asymptotically normal, and is $O_p(1)$. Hence, the consistency and rate of convergence of $\hat{\theta}$ depend on the property of the first term. We analyze $\sqrt{n}P\{\mathrm{EIF}(O;\,\theta_0,\hat{\eta})\}$ in the following.

Define $\eta = (\mu_Y, \pi_A, \pi_S, \lambda)$, $\hat{\eta} = (\hat{\mu}_Y, \hat{\pi}_A, \hat{\pi}_S, \hat{\lambda})$, $\tilde{\eta} = (\hat{\mu}_Y, \hat{\pi}_A, \hat{\pi}_S, \lambda_0)$, and $\eta_0 = (\pi_{S0}, \pi_{A0}, \mu_{Y0}, \lambda_0)$. From

$$\theta_0 = P\left[\frac{SA}{E(SA)}\{\mu_{Y0}(1,1,X) - \mu_{Y0}(1,0,X) - \mu_{Y0}(0,1,X) + \mu_{Y0}(0,0,X)\}\right],$$

we can write $P\{\mathrm{EIF}(O;\,\theta_0,\tilde{\eta})\}$ as

$$
\begin{aligned}
P\{\mathrm{EIF}(O;\,\theta_0,\tilde{\eta})\} = {} & P\left[\frac{SA}{E(SA)}\{Y - \hat{\mu}_Y(1,0,X) - \hat{\mu}_Y(0,1,X) + \hat{\mu}_Y(0,0,X) - \theta_0\}\right. \\
&- \frac{S(1-A)\frac{\hat{\pi}_A(X,1)}{1-\hat{\pi}_A(X,1)}}{E(SA)}\{Y-\hat{\mu}_Y(1,0,X)\} - \frac{(1-S)A\frac{\hat{\pi}_A(X,1)}{\hat{\pi}_A(X,0)}\frac{\hat{\pi}_S(X)}{1-\hat{\pi}_S(X)}}{E(SA)}\{Y-\hat{\mu}_Y(0,1,X)\} \\
&+ \left.\frac{(1-S)(1-A)\frac{\hat{\pi}_A(X,1)}{1-\hat{\pi}_A(X,0)}\frac{\hat{\pi}_S(X)}{1-\hat{\pi}_S(X)}}{E(SA)}\{Y-\hat{\mu}_Y(0,0,X)\}\right] \\
= {} & P\left[\frac{SA}{E(SA)}\{-\{\hat{\mu}_Y(1,0,X)-\mu_{Y0}(1,0,X)\} - \{\hat{\mu}_Y(0,1,X)-\mu_{Y0}(0,1,X)\}\}\right. \\
&+ \left.\frac{SA}{E(SA)}\{\hat{\mu}_Y(0,0,X)-\mu_{Y0}(0,0,X)\}\right] - \frac{S(1-A)\frac{\hat{\pi}_A(X,1)}{1-\hat{\pi}_A(X,1)}}{E(SA)}\{\mu_{Y0}(1,0,X)-\hat{\mu}_Y(1,0,X)\}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{(1-S)A\frac{\hat\pi_A(X,1)}{\hat\pi_A(X,0)}\frac{\hat\pi_S(X)}{1-\hat\pi_S(X)}}{E(SA)}\{\mu_{Y0}(0,1,X)-\hat\mu_Y(0,1,X)\} \\
&+ \left.\frac{(1-S)(1-A)\frac{\hat\pi_A(X,1)}{1-\hat\pi_A(X,0)}\frac{\hat\pi_S(X)}{1-\hat\pi_S(X)}}{E(SA)}\{\mu_{Y0}(0,0,X)-\hat\mu_Y(0,0,X)\}\right] \\
= {}& P\left[\left\{\frac{SA}{E(SA)} - \frac{S(1-A)\frac{\hat\pi_A(X,1)}{1-\hat\pi_A(X,1)}}{E(SA)}\right\}\{\mu_{Y0}(1,0,X)-\hat\mu_Y(1,0,X)\}\right] \\
&+ P\left[\left\{\frac{SA}{E(SA)} - \frac{(1-S)A\frac{\hat\pi_A(X,1)}{\hat\pi_A(X,0)}\frac{\hat\pi_S(X)}{1-\hat\pi_S(X)}}{E(SA)}\right\}\{\mu_{Y0}(0,1,X)-\hat\mu_Y(0,1,X)\}\right] \\
&+ P\left[\left\{-\frac{SA}{E(SA)} + \frac{(1-S)(1-A)\frac{\hat\pi_A(X,1)}{1-\hat\pi_A(X,0)}\frac{\hat\pi_S(X)}{1-\hat\pi_S(X)}}{E(SA)}\right\}\{\mu_{Y0}(0,0,X)-\hat\mu_Y(0,0,X)\}\right] \\
\coloneqq {}& T_1 + T_2 + T_3.
\end{aligned}
$$

From the facts that

$$
\begin{aligned}
&||\hat\pi_A(X,S) - \pi_{A0}(X,S)||_2^2 \\
&\quad = ||\{\hat\pi_A(X,1) - \pi_{A0}(X,1)\}\pi_{S0}(X)^{1/2}||_2^2 + ||\{\hat\pi_A(X,0) - \pi_{A0}(X,0)\}\{1-\pi_{S0}(X)\}^{1/2}||_2^2 \\
&||\hat\mu_Y(S,A,X) - \mu_{Y0}(S,A,X)||_2^2 \\
&\quad = ||\{\hat\mu_Y(1,1,X) - \mu_{Y0}(1,1,X)\}\pi_{S0}(X)^{1/2}\pi_{A0}(X,1)^{1/2}||_2^2 \\
&\qquad + ||\{\hat\mu_Y(1,0,X) - \mu_{Y0}(1,0,X)\}\pi_{S0}(X)^{1/2}(1-\pi_{A0}(X,1))^{1/2}||_2^2 \\
&\qquad + ||\{\hat\mu_Y(0,1,X) - \mu_{Y0}(0,1,X)\}(1-\pi_{S0}(X))^{1/2}\pi_{A0}(X,1)^{1/2}||_2^2 \\
&\qquad + ||\{\hat\mu_Y(0,0,X) - \mu_{Y0}(0,0,X)\}(1-\pi_{S0}(X))^{1/2}(1-\pi_{A0}(X,1))^{1/2}||_2^2,
\end{aligned}
$$

we have that every term on the right-hand side of the equations are bounded by the term on the left-hand side, e.g.,

$$
||\{\hat\pi_A(X,1) - \pi_{A0}(X,1)\}\pi_{S0}(X)^{1/2}||_2 \le ||\hat\pi_A(X,S) - \pi_{A0}(X,S)||_2.
$$

Hence, from Cauchy-Schwarz inequality, boundedness of $1/\hat\pi_A(X,0), 1/\{1-\hat\pi_A(X,0)\}, 1/\{1-\hat\pi_A(X,1)\}, 1/\{1-\hat\pi_S(X)\}$, and the triangle inequality, the first term satisfies

$$
\begin{aligned}
T_1 &= P\left[\frac{\pi_{S0}(X)^{1/2}\{\pi_{A0}(X,1) - \hat\pi_A(X,1)\}}{E(SA)\{1-\hat\pi_A(X,1)\}}\frac{\pi_{S0}(X)^{1/2}(1-\pi_{A0}(X,1))^{1/2}}{(1-\pi_{A0}(X,1))^{1/2}}\{\mu_{Y0}(1,0,X)-\hat\mu_Y(1,0,X)\}\right] \\
&\le C||\pi_{S0}(X)^{1/2}\{\pi_{A0}(X,1)-\hat\pi_A(X,1)\}||_2||\pi_{S0}(X)^{1/2}(1-\pi_{A0}(X,1))^{1/2}\{\mu_{Y0}(1,0,X)-\hat\mu_Y(1,0,X)\}||_2 \\
&= O_p(||\hat\pi_A(X,S) - \pi_{A0}(X,S)||_2||\hat\mu_Y(S,A,X) - \mu_{Y0}(S,A,X)||_2).
\end{aligned}
$$

Similarly, for the second and third terms, we have

$$
\begin{aligned}
T_2 &= P\left[\frac{\pi_{A0}(X,1)\pi_{S0}(X)\hat\pi_A(X,0)(1-\hat\pi_S(X)) - \pi_{A0}(X,0)(1-\pi_{S0}(X))\hat\pi_A(X,1)\hat\pi_S(X)}{E(SA)\hat\pi_A(X,0)(1-\hat\pi_S(X))}\{\mu_{Y0}(0,1,X)-\hat\mu_Y(0,1,X)\}\right] \\
&= O_p(\{||\hat\pi_S(X)-\pi_{S0}(X)||_2 + ||\hat\pi_A(X,S)-\pi_{A0}(X,S)||_2\}||\hat\mu_Y(S,A,X)-\mu_{Y0}(S,A,X)||_2) \\
T_3 &= -\frac{\pi_A(X,1)\pi_{S0}(X)(1-\hat\pi_A(X,0))(1-\hat\pi_S(X)) - (1-\pi_{A0}(X,0))(1-\pi_{S0}(X))\hat\pi_A(X,1)\hat\pi_S(X)}{E(SA)(1-\hat\pi_A(X,0))(1-\hat\pi_S(X))}\{\mu_{Y0}(0,0,X) \\
&\qquad \left. - \hat\mu_Y(0,0,X)\}\right] \\
&= O_p(\{||\hat\pi_S(X)-\pi_{S0}(X)||_2 + ||\hat\pi_A(X,S)-\pi_{A0}(X,S)||_2\}||\hat\mu_Y(S,A,X)-\mu_{Y0}(S,A,X)||_2).
\end{aligned}
$$

Therefore,

$$P\{\mathrm{EIF}(O;\, \theta_0, \tilde{\eta})\} = O_p(\{\|\hat{\pi}_S - \pi_{S0}\|_2 + \|\hat{\pi}_A - \pi_{A0}\|_2\}\|\hat{\mu}_Y - \mu_{Y0}\|_2)$$

Similarly, we can show that

$$P\{\varphi(O;\, \theta_0, \tilde{\eta})\} - P\{\varphi(O;\, \theta_0, \hat{\eta})\} = O_p\{(\|\hat{\pi}_S - \pi_{S0}\|_2 + \|\hat{\pi}_A - \pi_{A0}\|_2)\|\hat{\mu}_Y - \mu_{Y0}\|_2|\hat{\lambda} - \lambda_0|\}.$$

Again applying the triangle inequality and because $\hat{\lambda} - \lambda_0 = o_p(1)$, we arrive at

$$P\{\varphi(O;\, \theta_0, \hat{\eta})\} = O_p\{(\|\hat{\pi}_A - \pi_{A0}\|_2 + \|\hat{\pi}_S - \pi_{S0}\|_2)\|\hat{\mu}_Y - \mu_{Y0}\|_2\}.$$

S2.5 Proof of bounds on $V(X)$

Recall the definition $V(X) = E\{Y^{(0)}|S = 1, A = 1, X\} - E\{Y^{(0)}|S = 1, A = 0, X\} - [E\{Y^{(0)}|S = 0, A = 1, X\} - E\{Y^{(0)}|S = 0, A = 0, X\}]$.

The bounds on $V(X)$, that is,

$$\begin{aligned} &E[Y|S = 0, A = 1, X](\Gamma^{-1} - 1) - E[Y|S = 0, A = 0, X](\Gamma - 1) \\ \leq\ & V(X) \leq E[Y|S = 0, A = 1, X](\Gamma - 1) - E[Y|S = 0, A = 0, X](\Gamma^{-1} - 1), \end{aligned}$$

follows from the result that

$$V(X) = E[E\{Y|S = 0, U, X\}\{r(1, U, X) - 1\}|S = 0, A = 1, X] - E[E\{Y|S = 0, U, X\}\{r(0, U, X) - 1\}|S = 0, A = 0, X].$$

S3 Variance estimator

S3.1 Constructing a consistent variance estimator

A general recipe for constructing a consistent variance estimator under the M-estimation framework (see, e.g., tutorials by Stefanski and Boos [7] and Saul and Hudgens [8]) can be applied to our proposed estimators as follows. Let $\mathrm{IF}(O;\, \theta, \eta)$ denote the efficient influence function corresponding to $\hat{\theta}_{\mathrm{reg}}$, $\hat{\theta}_{\mathrm{ipw}}$, or $\hat{\theta}_{\mathrm{dr}}$. Nuisance parameters η are estimated by $\hat{\eta}$ via solving their corresponding estimating equations $\sum_{i=1}^n \phi(O_i;\, \hat{\eta}) = 0$, and the target parameter θ is estimated by $\hat{\theta}$ via solving $\sum_{i=1}^n \mathrm{IF}(O_i;\, \hat{\theta}, \hat{\eta}) = 0$. Write $\gamma = (\theta, \eta)$ for simplicity. To estimate γ, it suffices to stack estimating equations $\phi(O_i;\, \eta)$ and $\mathrm{IF}(O_i;\, \theta, \eta)$, and find $\gamma = \hat{\gamma}$ that solves

$$\sum_{i=1}^n \psi(O_i;\, \hat{\gamma}) = \begin{pmatrix} \sum_{i=1}^n \phi(O_i;\, \hat{\eta}) \\ \sum_{i=1}^n \mathrm{IF}(O_i;\, \hat{\theta}, \hat{\eta}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{S7}$$

Define $A(\gamma) = \mathbb{E}\{-\partial\psi(O;\, \gamma)/\partial\gamma\}$ and $B(\gamma) = \mathbb{E}\{\psi(O;\, \gamma)\psi(O;\, \gamma)^T\}$. A consistent estimator of the variance-covariance matrix of $\hat{\gamma}$ can then be obtained using

$$\hat{V}(\hat{\gamma}) = (1/n) \cdot A(\hat{\gamma})^{-1}B(\hat{\gamma})\{A(\hat{\gamma})^{-1}\}^T. \tag{S8}$$

A consistent variance estimator of the target parameter θ is then obtained as the $(1, 1)$-th element of $\hat{V}(\hat{\gamma})$.

S3.2 Implementation in R using the geex package

The R package geex is a general purpose software to find roots and compute the empirical sandwich estimator for a set of user-supplied, unbiased estimating equations ([8]). Below, we provide an implementation of the sandwich variance estimator via the geex package.

```
##############################################
# Estimating equation
dr_est_func <- function(data, models){
A = data$A
S = data$S
Y = data$Y
Xs <- grab_design_matrix(data = data,
rhs_formula = grab_fixed_formula(models$S_md))
Xa = grab_design_matrix(data = data,
rhs_formula = grab_fixed_formula(models$A_md))
Xmu <- grab_design_matrix(data = data,
rhs_formula = grab_fixed_formula(models$mu_md))
Xa_S1 = Xa
Xa_S1[, 'S'] = 1
Xa_S0 = Xa
Xa_S0[, 'S'] = 0
Xmu_S1_A0 = Xmu
Xmu_S1_A0[, 'S'] = 1
Xmu_S1_A0[, 'A'] = 0
Xmu_S1_A0[, 'S:A'] = 0
Xmu_S0_A1 = Xmu
Xmu_S0_A1[, 'S'] = 0
Xmu_S0_A1[, 'A'] = 1
Xmu_S0_A1[, 'S:A'] = 0
Xmu_S0_A0 = Xmu
Xmu_S0_A0[, 'S'] = 0
Xmu_S0_A0[, 'A'] = 0
Xmu_S0_A0[, 'S:A'] = 0
S_md_pos = 1:length(coef(models$S_md))
S_scores <- grab_psiFUN(models$S_md, data = data)
A_md_pos = (length(coef(models$S_md))+1):(length(coef(models$S_md)) +
length(coef(models$A_md)))
A_scores <- grab_psiFUN(models$A_md, data = data)
mu_md_pos = (length(coef(models$S_md)) + length(coef(models$A_md)) + 1):
(length(coef(models$S_md)) + length(coef(models$A_md)) +
length(coef(models$mu_md)))
mu_scores <- grab_psiFUN(models$mu_md, data = data)
function(theta, n, n_11) {
pi_s = plogis(Xs %*% theta [S_md_pos ])
pi_a_s_1 = plogis(Xa_S1 %*% theta [A_md_pos ])
pi_a_s_0 = plogis(Xa_S0 %*% theta [A_md_pos ])
mu_s_1_a_0 = Xmu_S1_A0 %*% theta [mu_md_pos ]
mu_s_0_a_1 = Xmu_S0_A1 %*% theta [mu_md_pos ]
mu_s_0_a_0 = Xmu_S0_A0 %*% theta [mu_md_pos ]
```

```
beta = S * A * (n/n_11) * (Y - mu_s_1_a_0 - mu_s_0_a_1 +
mu_s_0_a_0 - theta[length(theta)]) -
(n/n_11) * S * (1 - A) * (pi_a_s_1/(1-pi_a_s_1)) *
(Y - mu_s_1_a_0) - (n/n_11) * (1 - S) * A *
(pi_a_s_1/pi_a_s_0) * (pi_s/(1-pi_s)) *
(Y - mu_s_0_a_1) + (n/n_11) * (1 - S) * (1 - A) *
(pi_a_s_1/(1-pi_a_s_0)) * (pi_s/(1-pi_s)) * (Y - mu_s_0_a_0)
c(S_scores(theta[S_md_pos]),
A_scores(theta[A_md_pos]),
mu_scores(theta[mu_md_pos]),
beta)
}
}
```

```
##########################################################
# Solve the estimating equation and compute the empirical
# sandwich estimator. The function returns the point
# estimator and the estimated standard error.
estimate_aipw <- function(data, models){
n_11 = length(which(data$S == 1 & data$A == 1))
n = dim(data)[1]
res = m_estimate(estFUN = dr_est_func,
data = data,
root_control = setup_root_control(start =
c(unlist(lapply(models, function(x) coef(x))), 0)),
outer_args = list(models = models),
inner_args = list(n = n, n_11 = n_11))
nparam = length(roots(res))
point_est = roots(res)[nparam]
se = sqrt(vcov(res)[nparam, nparam])
return(c(point_est, se))
}
```

```
########################################################
# Run the code
res = estimate_aipw(data = data,
models = list(S_md = S_md,
A_md = A_md,
mu_md = mu_md))
```

S4 Additional simulation results

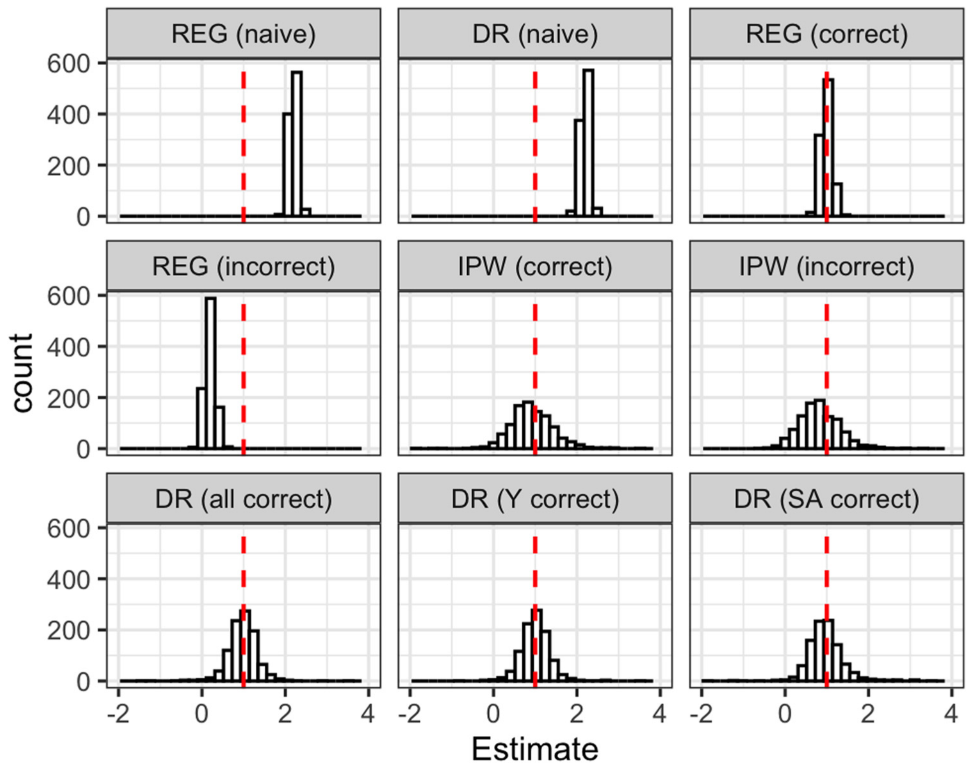

Figure S3: Sampling distribution of nine estimators under Scenario 1.

Table 1: Simulation results for four scenarios. Scenario 5: $b = (b1)$, $d = (d1)$, $e = (e)$, $f = (f1)$. Scenario 6: $b = (b1)$, $d = (d1)$, $e = (e)$, $f = (f2)$. Scenario 7: $b = (b1)$, $d = (d2)$, $e = (e)$, $f = (f1)$. Scenario 8: $b = (b1)$, $d = (d2)$, $e = (e)$, $f = (f2)$. $\theta_0 = 1$ in all four scenarios

Sample size	Estimator	Model Spec.	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)
			Scenario 5			Scenario 6			Scenario 7			Scenario 8		
1,000	$\hat{\theta}_{\text{reg,naive}}$		1.19	0.13	0.00	1.19	0.13	0.00	1.19	0.13	0.00	1.18	0.13	0.00
2,000	$\hat{\theta}_{\text{reg,naive}}$		1.19	0.09	0.00	1.19	0.09	0.00	1.19	0.09	0.00	1.19	0.09	0.00
5,000	$\hat{\theta}_{\text{reg,naive}}$		1.19	0.06	0.00	1.19	0.06	0.00	1.18	0.06	0.00	1.19	0.06	0.00
1,000	$\hat{\theta}_{\text{dr,naive}}$		1.19	0.14	0.20	1.19	0.14	0.20	1.20	0.13	0.20	1.20	0.13	0.40
2,000	$\hat{\theta}_{\text{dr,naive}}$		1.20	0.10	0.00	1.20	0.10	0.00	1.20	0.09	0.10	1.20	0.09	0.20
5,000	$\hat{\theta}_{\text{dr,naive}}$		1.19	0.07	0.00	1.19	0.07	0.00	1.19	0.06	0.00	1.20	0.06	0.00
1,000	$\hat{\theta}_{\text{reg}}$	μ_Y correct	−0.01	0.19	95.97	−0.02	0.20	93.85	−0.01	0.19	95.36	−0.02	0.19	94.56
2,000	$\hat{\theta}_{\text{reg}}$	μ_Y correct	−0.01	0.14	95.99	−0.01	0.14	96.29	−0.01	0.13	95.59	0.00	0.13	95.69
5,000	$\hat{\theta}_{\text{reg}}$	μ_Y correct	−0.01	0.09	95.06	−0.01	0.09	94.63	−0.01	0.08	93.45	−0.01	0.08	94.84
1,000	$\hat{\theta}_{\text{reg}}$	μ_Y incorrect	−0.81	0.18	0.50	−0.82	0.19	1.11	−0.83	0.18	0.40	−0.83	0.18	0.60
2,000	$\hat{\theta}_{\text{reg}}$	μ_Y incorrect	−0.81	0.13	0.00	−0.81	0.13	0.00	−0.82	0.13	0.00	−0.82	0.13	0.00
5,000	$\hat{\theta}_{\text{reg}}$	μ_Y incorrect	−0.81	0.08	0.00	−0.81	0.08	0.00	−0.83	0.08	0.00	−0.82	0.08	0.00
1,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) correct	−0.10	0.49	92.54	−0.12	0.48	88.61	−0.09	0.42	92.04	−0.08	0.42	92.04
2,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) correct	−0.07	0.38	90.97	−0.03	0.37	91.17	−0.06	0.33	89.97	−0.04	0.33	92.18
5,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) correct	−0.03	0.26	89.90	−0.06	0.26	89.15	−0.03	0.23	92.37	−0.04	0.23	92.27
1,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) incorrect	−0.21	0.49	89.21	−0.22	0.49	87.60	−0.18	0.43	89.42	−0.17	0.43	90.42
2,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) incorrect	−0.17	0.38	87.56	−0.13	0.38	87.56	−0.15	0.33	86.66	−0.13	0.33	89.87

(*Continued*)

Table 1: *Continued*

Sample size	Estimator	Model Spec.	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)
5,000	$\hat{\theta}_{\mathrm{ipw}}$	(π_S, π_A) incorrect	−0.14	0.26	82.92	−0.16	0.27	83.57	−0.13	0.23	86.14	−0.14	0.23	85.18
1,000	$\hat{\theta}_{\mathrm{dr}}$	All correct	−0.04	0.32	93.85	−0.03	0.33	93.65	−0.01	0.31	95.56	−0.03	0.31	95.06
2,000	$\hat{\theta}_{\mathrm{dr}}$	All correct	−0.01	0.25	94.88	0.01	0.25	94.38	0.00	0.23	94.68	0.01	0.23	94.08
5,000	$\hat{\theta}_{\mathrm{dr}}$	All correct	−0.01	0.17	93.34	−0.02	0.18	93.56	0.01	0.16	94.84	0.00	0.15	94.63
1,000	$\hat{\theta}_{\mathrm{dr}}$	μ_Y correct	−0.01	0.36	93.45	−0.02	0.36	93.75	−0.03	0.31	94.96	−0.04	0.32	93.65
2,000	$\hat{\theta}_{\mathrm{dr}}$	μ_Y correct	0.01	0.27	95.09	0.03	0.27	94.88	−0.02	0.24	94.98	0.01	0.24	94.98
5,000	$\hat{\theta}_{\mathrm{dr}}$	μ_Y correct	0.00	0.18	93.66	−0.01	0.19	92.59	−0.01	0.16	93.23	−0.01	0.16	93.56
1,000	$\hat{\theta}_{\mathrm{dr}}$	(π_S, π_A) correct	−0.03	0.36	90.83	−0.02	0.36	90.12	−0.01	0.34	92.64	−0.03	0.34	90.73
2,000	$\hat{\theta}_{\mathrm{dr}}$	(π_S, π_A) correct	−0.01	0.28	92.28	0.05	0.28	92.68	−0.02	0.26	91.88	0.00	0.26	91.98
5,000	$\hat{\theta}_{\mathrm{dr}}$	(π_S, π_A) correct	0.00	0.19	90.87	−0.02	0.19	92.70	0.02	0.18	92.16	−0.01	0.18	93.23

Table 2: Simulation results for four scenarios. Scenario 9: $b = (b2), d = (d1), e = (e), f = (f1)$. Scenario 10: $b = (b2), d = (d1), e = (e), f = (f2)$. Scenario 11: $b = (b2), d = (d2), e = (e), f = (f1)$. Scenario 12: $b = (b2), d = (d2), e = (e), f = (f2)$. $\theta_0 = 1$ in all four scenarios

Sample size	Estimator	Model Spec.	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	MedianEst. SE	95% CI Cov. (%)
			Scenario 9			Scenario 10			Scenario 11			Scenario 12		
1,000	$\hat{\theta}_{\text{reg,naive}}$		1.20	0.14	0.00	1.19	0.15	0.00	1.18	0.15	0.00	1.19	0.15	0.00
2,000	$\hat{\theta}_{\text{reg,naive}}$		1.20	0.10	0.00	1.20	0.10	0.00	1.19	0.10	0.00	1.18	0.10	0.00
5,000	$\hat{\theta}_{\text{reg,naive}}$		1.19	0.06	0.00	1.19	0.07	0.00	1.18	0.07	0.00	1.18	0.07	0.00
1,000	$\hat{\theta}_{\text{dr,naive}}$		1.20	0.15	0.20	1.19	0.16	0.20	1.19	0.14	0.10	1.20	0.14	0.10
2,000	$\hat{\theta}_{\text{dr,naive}}$		1.20	0.11	0.10	1.19	0.11	0.10	1.20	0.10	0.30	1.19	0.10	0.00
5,000	$\hat{\theta}_{\text{dr,naive}}$		1.19	0.07	0.32	1.19	0.07	0.00	1.19	0.06	0.00	1.19	0.07	0.00
1,000	$\hat{\theta}_{\text{reg}}$	μ_Y correct	−0.01	0.19	94.35	−0.02	0.19	94.76	−0.01	0.19	94.05	−0.01	0.19	94.66
2,000	$\hat{\theta}_{\text{reg}}$	μ_Y correct	0.00	0.14	95.89	0.00	0.14	94.08	−0.01	0.13	94.48	−0.01	0.13	95.29
5,000	$\hat{\theta}_{\text{reg}}$	μ_Y correct	−0.01	0.09	94.41	−0.01	0.09	95.38	−0.01	0.08	95.27	−0.01	0.08	95.60
1,000	$\hat{\theta}_{\text{reg}}$	μ_Y incorrect	−0.78	0.19	1.81	−0.78	0.19	1.21	−0.78	0.18	1.11	−0.78	0.18	0.60
2,000	$\hat{\theta}_{\text{reg}}$	μ_Y incorrect	−0.77	0.13	0.00	−0.77	0.13	0.00	−0.78	0.13	0.00	−0.78	0.13	0.00
5,000	$\hat{\theta}_{\text{reg}}$	μ_Y incorrect	−0.77	0.08	0.00	−0.77	0.08	0.00	−0.78	0.08	0.00	−0.78	0.08	0.00
1,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) correct	−0.07	0.48	91.83	−0.09	0.48	90.02	−0.05	0.42	90.62	−0.09	0.42	90.93
2,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) correct	−0.06	0.36	92.38	−0.05	0.37	92.58	−0.05	0.32	92.58	−0.05	0.32	91.57
5,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) correct	−0.05	0.26	89.47	−0.03	0.25	91.62	−0.04	0.23	90.01	−0.04	0.23	91.62
1,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) incorrect	−0.15	0.49	90.22	−0.19	0.49	89.31	−0.13	0.43	91.13	−0.15	0.43	89.82
2,000	$\hat{\theta}_{\text{ipw}}$	(π_S, π_A) incorrect	−0.15	0.38	89.37	−0.13	0.38	89.67	−0.10	0.33	89.77	−0.12	0.33	88.87

(*Continued*)

Table 2: *Continued*

Sample size	Estimator	Model Spec.	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	Median Est. SE	95% CI Cov. (%)	Bias	MedianEst. SE	95% CI Cov. (%)
5,000	$\hat{\theta}_{\mathrm{ipw}}$	(π_S, π_A) incorrect	−0.14	0.26	87.00	−0.11	0.26	88.08	−0.11	0.23	87.22	−0.11	0.23	87.86
1,000	$\hat{\theta}_{\mathrm{dr}}$	All correct	0.00	0.33	94.66	0.00	0.32	92.84	0.00	0.30	94.76	0.00	0.29	94.35
2,000	$\hat{\theta}_{\mathrm{dr}}$	All correct	0.00	0.24	95.09	0.01	0.25	94.18	−0.01	0.22	94.98	0.00	0.22	93.58
5,000	$\hat{\theta}_{\mathrm{dr}}$	All correct	−0.01	0.17	94.31	−0.01	0.17	95.06	−0.02	0.15	93.88	−0.01	0.15	95.06
1,000	$\hat{\theta}_{\mathrm{dr}}$	μ_Y correct	0.01	0.37	93.25	0.00	0.36	93.85	−0.01	0.33	94.96	−0.01	0.32	94.25
2,000	$\hat{\theta}_{\mathrm{dr}}$	μ_Y correct	0.01	0.28	93.78	0.01	0.28	94.68	−0.01	0.24	94.68	−0.02	0.24	93.28
5,000	$\hat{\theta}_{\mathrm{dr}}$	μ_Y correct	0.00	0.19	93.88	0.01	0.19	93.88	−0.03	0.16	93.56	−0.02	0.16	95.38
1,000	$\hat{\theta}_{\mathrm{dr}}$	(π_S, π_A) correct	0.01	0.36	91.43	0.00	0.36	92.04	0.00	0.33	92.14	−0.03	0.32	91.94
2,000	$\hat{\theta}_{\mathrm{dr}}$	(π_S, π_A) correct	0.02	0.27	93.38	0.01	0.27	93.28	0.00	0.24	92.88	0.00	0.25	92.78
5,000	$\hat{\theta}_{\mathrm{dr}}$	(π_S, π_A) correct	−0.01	0.19	93.02	0.00	0.18	93.13	−0.01	0.17	92.16	−0.01	0.17	93.77

References

[1] MacKay EJ, Zhang B, Heng S, Ye T, Neuman MD, Augoustides JG, et al. Association between transesophageal echocardiography and clinical outcomes after coronary artery bypass graft surgery. J Amer Soc Echocardiography. 2021;34(6):571–81.
[2] MacKay EJ, Zhang B, Augoustides JG, Groeneveld PW, Desai ND. Association of intraoperative transesophageal echocardiography and clinical outcomes after open cardiac valve or proximal aortic surgery. JAMA Netw Open. 2022 Feb;5(2):e2147820–e2147820. doi: 10.1001/jamanetworkopen.2021.47820.
[3] Metkus TS, Thibault D, Grant MC, Badhwar V, Jacobs JP, Lawton J, et al. Transesophageal echocardiography in patients undergoing coronary artery bypass graft surgery. J Amer College Cardiol. 2021;78(2):112–122.
[4] Robins J, Sued M, Lei-Gomez Q, Rotnitzky A. Comment: Performance of double-robust estimators when “inverse probability” weights are highly variable. Stat Sci. 2007;22(4):544–559.
[5] Newey WK. The asymptotic variance of semiparametric estimators. Econometrica. 1994;62(6):1349–82. http://www.jstor.org/stable/2951752.
[6] van der Vaart AW. Asymptotic statistics. Cambridge University Press; 2000.
[7] Stefanski LA, Boos DD. The calculus of M-estimation. Amer Stat. 2002;56(1):29–38.
[8] Saul BC, Hudgens MG. The calculus of M-estimation in R with geex. J Stat Softw. 2020;92(2).