Web Appendix for "Propensity score weighting for covariate adjustment in randomized clinical trials" by Zeng et al.

APPENDIX A: PROOFS OF THE PROPOSITIONS IN SECTION 3.1

We proceed under a set of standard regularity conditions such as the expectations $E(Y_i|X_i, Z_i)$, $E(Y_i^2|X_i, Z_i)$ are finite and well defined.

Proof for Proposition 1(a). Suppose the propensity score model $e_i = e(X_i;\theta)$ is a smooth function of θ, and the estimated parameter $\hat{\theta}$ is obtained by maximum likelihood, we derive the score function $S_{\theta,i}$ for each observation i, namely the first order derivative of the log likelihood with respect to θ,

$$S_{\theta,i} = \frac{\partial}{\partial\theta} l_i(\theta) = \frac{\partial}{\partial\theta}\{Z_i \log e(X_i;\theta) + (1-Z_i)\log(1-e(X_i;\theta))\} = \frac{Z_i - e(X_i;\theta)}{e(X_i;\theta)(1-e(X_i;\theta))}\frac{\partial e(X_i;\theta)}{\partial\theta},$$

where $\frac{\partial e(X_i;\theta)}{\partial\theta}$ is the derivative evaluated at θ. As the true probability of being treated is a constant r and the logistic model is always correctly specified as long as it includes an intercept, there exists θ^* such that $e(X_i;\theta^*) = r$. When $\theta = \theta^*$, the score function is,

$$S_{\theta^*,i} = \frac{Z_i - r}{r(1-r)}\frac{\partial e(X_i;\theta^*)}{\partial\theta}.$$

Let $I_{\theta\theta}$ be the information matrix evaluated at θ, whose exact form is,

$$I_{\theta\theta} = E\left\{\frac{\partial}{\partial\theta} l_i(\theta)\frac{\partial}{\partial\theta} l_i(\theta)^T\right\} = E\left\{\frac{(Z_i - e(X_i;\theta))^2}{(e(X_i;\theta)(1-e(X_i;\theta)))^2}\frac{\partial e(X_i;\theta)}{\partial\theta}\frac{\partial e(X_i;\theta)}{\partial\theta}^T\right\}.$$

When $\theta = \theta^*$,

$$I_{\theta^*\theta^*} = \frac{1}{r(1-r)} E\left\{\frac{\partial e(X_i;\theta^*)}{\partial\theta}\frac{\partial e(X_i;\theta^*)}{\partial\theta}^T\right\}.$$

Applying the Cramer-Rao theorem, assume the propensity score model $e(X_i;\theta)$ satisfies certain regularity conditions[1], the Taylor expansion $\hat{\theta}$ at true value is,

$$\sqrt{N}(\hat{\theta} - \theta^*) = I^{-1}_{\theta^*\theta^*}\frac{1}{\sqrt{N}}\sum_{i=1}^N S_{\theta^*,i} + o_p(1),$$

By the Weak Law of Large Numbers (WLLN), we can establish the consistency of $\hat{\theta}$,

$$\hat{\theta} - \theta^* \xrightarrow{p} I^{-1}_{\theta^*\theta^*} E(S_{\theta^*,i}) = I^{-1}_{\theta^*\theta^*}\frac{E(Z_i - r)}{r(1-r)} E\left\{\frac{\partial e(X_i;\theta^*)}{\partial\theta}\right\} = 0.$$

With the consistency of $\hat{\theta}$, we also have,

$$\frac{1}{N}\sum_{i=1}^N Z_i(1 - e(X_i;\hat{\theta})) \xrightarrow{p} r(1-r), \frac{1}{N}\sum_{i=1}^N (1-Z_i)e(X_i;\hat{\theta}) \xrightarrow{p} r(1-r).$$

Next, we investigate the influence function of $\hat{\mu}_1^{\text{ow}} - \hat{\mu}_0^{\text{ow}}$,

$$\begin{aligned}
\sqrt{N}(\hat{\mu}_1^{\text{ow}} - \hat{\mu}_0^{\text{ow}}) =& \sqrt{N}\left(\frac{\sum_{i=1}^N Z_i Y_i(1-e(X_i;\hat{\theta}))}{\sum_{i=1}^N Z_i(1-e(X_i;\hat{\theta}))} - \frac{\sum_{i=1}^N Z_i Y_i(1-e(X_i;\hat{\theta}))}{\sum_{i=1}^N Z_i(1-e(X_i;\hat{\theta}))}\right),\\
=& \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_i Y_i(1-e(X_i;\hat{\theta}))}{r(1-r)} - \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N (1-Z_i) Y_i e(X_i;\hat{\theta})}{r(1-r)} + o_p(1),\\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\frac{Z_i Y_i(1-e(X_i;\hat{\theta}))e(X_i;\hat{\theta})}{e(X_i;\hat{\theta})r(1-r)} - \frac{(1-Z_i) Y_i(1-e(X_i;\hat{\theta}))e(X_i;\hat{\theta})}{(1-e(X_i;\hat{\theta}))r(1-r)}\right\} + o_p(1).
\end{aligned}$$

We perform the Taylor expansion at the true value θ^*,

$$
\begin{aligned}
\sqrt{N}(\hat{\mu}_1^{\text{ow}} - \hat{\mu}_0^{\text{ow}}) =& \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i Y_i (1-e(X_i;\theta^*))e(X_i;\theta^*)}{e(X_i;\theta^*)r(1-r)} - \frac{(1-Z_i)Y_i(1-e(X_i;\theta^*))e(X_i;\theta^*)}{(1-e(X_i;\theta^*))r(1-r)} \right\} - \\
& \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i Y_i (1-e(X_i;\theta^*))e(X_i;\theta^*)}{e(X_i;\theta^*)r(1-r)} - \frac{Z_i Y_i (1-e(X_i;\theta^*))e(X_i;\theta^*)}{e(X_i;\theta^*)r(1-r)} \right\} S_{\theta^*,i}^T (\hat{\theta} - \theta^*) + o_p(1), \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} - \frac{1}{N} \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] \sqrt{N}(\hat{\theta} - \theta^*) + o_p(1), \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} - E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N S_{\theta^*,i} + o_p(1).
\end{aligned}
$$

After plugging in the value of $S_{\theta^*,i}$ and $I_{\theta^*\theta^*}$, we can show that,

$$
\begin{aligned}
\hat{\mu}_1^{\text{ow}} - \hat{\mu}_0^{\text{ow}} &= \frac{1}{N} \sum_{i=1}^N \left[\frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} - \frac{Z_i - r}{r(1-r)} \left\{ (1-r) g_1(X_i) + r g_0(X_i) \right\} \right] + o_p(N^{-1/2}) \\
g_1(X_i) &= E \left[Y_i \frac{\partial e(X_i;\theta^*)}{\partial \theta} \middle| Z_i = 1 \right] E \left\{ \frac{\partial e(X_i;\theta^*)}{\partial \theta} \frac{\partial e(X_i;\theta^*)}{\partial \theta}^T \right\}^{-1} \frac{\partial e(X_i;\theta^*)}{\partial \theta}, \\
g_0(X_i) &= E \left[Y_i \frac{\partial e(X_i;\theta^*)}{\partial \theta} \middle| Z_i = 0 \right] E \left\{ \frac{\partial e(X_i;\theta^*)}{\partial \theta} \frac{\partial e(X_i;\theta^*)}{\partial \theta}^T \right\}^{-1} \frac{\partial e(X_i;\theta^*)}{\partial \theta}.
\end{aligned}
$$

Therefore, $\hat{\tau}^{\text{ow}}$ belongs to the augmented IPW estimator class $\mathcal{I}$ in the main text, which completes the proof of Proportion 1 (a).

Proof for Proposition 1(b): First, we build the relationship between the asymptotic variance of $\hat{\tau}^{\text{ow}}$ with the correpsonding information matrix $I_{\theta^*\theta^*}$ and score function $S_{\theta^*,i}$ evaluated at true value. Based on the results in Proposition 1(a), the asymptotic variance of $\hat{\tau}^{\text{ow}}$ depends on the following terms:

$$
\begin{aligned}
\lim_{N\to\infty} N \text{Var}(\hat{\tau}^{\text{ow}}) =& \text{Var} \left(\frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} - E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} S_{\theta^*,i} \right), \\
=& \text{Var} \left(\frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right) + \text{Var} \left(E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} S_{\theta^*,i} \right), \\
& -2\text{Cov} \left(\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\}, E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} S_{\theta^*,i} \right).
\end{aligned}
$$

Notice the facts that

$$
E \left(\frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right) = 0, E(S_{\theta^*,i}) = 0,
$$

$$
E(S_{\theta^*,i} S_{\theta^*,i}^T) = E \left\{ \frac{(Z_i - r)^2}{r^2(1-r)^2} \right\} E \left\{ \frac{\partial e(X_i;\theta^*)}{\partial \theta} \frac{\partial e(X_i;\theta^*)}{\partial \theta}^T \right\} = \frac{1}{(1-r)r} E \left\{ \frac{\partial e(X_i;\theta^*)}{\partial \theta} \frac{\partial e(X_i;\theta^*)}{\partial \theta}^T \right\} = I_{\theta^*\theta^*},
$$

we have,

$$
\begin{aligned}
\text{Var} \left(E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} S_{\theta^*,i} \right) =& E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i} \right], \\
=& \text{Cov} \left(\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\}, E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} S_{\theta^*,i} \right).
\end{aligned}
$$

We can further reduce the asymptotic variance to,

$$
\lim_{N\to\infty} N \text{Var}(\hat{\tau}^{\text{ow}}) = \text{Var} \left(\frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right) - \text{Var} \left(E \left[\left\{ \frac{Z_i Y_i}{r} - \frac{(1-Z_i)Y_i}{1-r} \right\} S_{\theta^*,i}^T \right] I_{\theta^*\theta^*}^{-1} S_{\theta^*,i} \right).
$$

Recall that X^1 and X^2 denote two nested sets of covariates with $X^2 = (X^1, X^{*1})$, and $e(X_i^1;\theta_1)$, $e(X_i^2;\theta_2)$ are the nested smooth parametric propensity score models. Suppose $\hat{\tau}_1^{\text{ow}}$ and $\hat{\tau}_2^{\text{ow}}$ are two OW estimators derived from the fitted propensity score $e(X_i^1;\hat{\theta}_1)$ and $e(X_i^1;\hat{\theta}_2)$ respectively. Denote the true value of the nested propensity score models as θ^{*1},θ^{*2}, the score functions at true value as $S_{\theta^{*1},i}$,$S_{\theta^{*2},i}$ and the information matrix as $I_{\theta^{*1},\theta^{*1}}$ and $I_{\theta^{*2},\theta^{*2}}$. To prove $\lim_{N\to\infty} N\mathrm{Var}(\hat{\tau}_1^{\mathrm{ow}}) \geq \lim_{N\to\infty} N\mathrm{Var}(\hat{\tau}_2^{\mathrm{ow}})$, it is equivalent to establish the following inequality,

$$\mathrm{Var}\left(E\left[\left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S^T_{\theta^{*2},i}\right]I^{-1}_{\theta^{*2}\theta^{*2}}S_{\theta^{*2},i}\right) \geq \mathrm{Var}\left(E\left[\left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S^T_{\theta^{*}1,i}\right]I^{-1}_{\theta^{*1}\theta^{*1}}S_{\theta^{*1},i}\right).$$

Using the equivalent expression, this inequality becomes,

$$E\left[\left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S^T_{\theta^{*2},i}\right]I^{-1}_{\theta^{*2}\theta^{*2}}E\left[\left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S_{\theta^{*2},i}\right] \geq$$
$$E\left[\left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S^T_{\theta^{*1},i}\right]I^{-1}_{\theta^{*1}\theta^{*1}}E\left[\left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S_{\theta^{*1},i}\right].$$

Additionally, as the two models are nested,

$$I_{\theta^{*2}\theta^{*2}} = \begin{bmatrix} I_{\theta^{*1}\theta^{*1}} & I^{12}_{\theta^{*2}\theta^{*2}} \\ I^{21}_{\theta^{*2}\theta^{*2}} & I^{22}_{\theta^{*2}\theta^{*2}} \end{bmatrix} \triangleq \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}, E\left[\left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S_{\theta^{*2},i}\right] = E\begin{bmatrix} \left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S_{\theta^{*1},i} \\ \left\{\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r}\right\}S^2_{\theta^{*1},i} \end{bmatrix} \triangleq \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}.$$

The inverse of the information matrix for the larger model is

$$I^{-1}_{\theta^{*2}\theta^{*2}} = \begin{bmatrix} I_{11}^{-1} + I_{11}^{-1}I_{12}(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}I_{21}I_{11}^{-1} & -I_{11}^{-1}I_{12}(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1} \\ -(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}I_{21}I_{11}^{-1} & (I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1} \end{bmatrix}.$$

Hence we can calculate the difference of asymptotic variance,

$$\begin{aligned}\lim_{N\to\infty} N\mathrm{Var}(\hat{\tau}_1^{\mathrm{ow}}) - \lim_{N\to\infty} N\mathrm{Var}(\hat{\tau}_2^{\mathrm{ow}}) =& U_1^T\{I_{11}^{-1}I_{12}(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}I_{21}I_{11}^{-1}\}U_1 - U_1^TI_{11}^{-1}I_{12}(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}U_2 \\ &-U_2^T(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}I_{21}I_{11}^{-1}U_1 + U_2^T(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}U_2^T, \\ =& (I_{21}I_{11}^{-1}U_1 - U_2)^T(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}(I_{21}I_{11}^{-1}U_1 - U_2) \geq 0.\end{aligned}$$

The last inequality follows from the fact that $(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}$ is positive definite. Hence, we have proved the asymptotic variance of the $\hat{\tau}_2^{\mathrm{ow}}$ is no greater than the OW estimator $\hat{\tau}_1^{\mathrm{ow}}$ with fewer covariates, which completes the proof of Proposition 1(b).

Proof for Proposition 1(c): When we are using logistic regression to estimate the propensity score, we have $\frac{\partial e(X_i;\theta^*)}{\partial\theta} = r(1-r)\tilde{X}_i, \tilde{X}_i = (1, X_i^T)^T$. Plugging this quantity into the g_1, g_0, we have,

$$\begin{aligned} g_1(X_i) &= E(Y_i\tilde{X}_i|Z_i = 1)E(\tilde{X}_i\tilde{X}_i^T)^{-1}\tilde{X}_i, \\ &= E(Y_i\tilde{X}_i|Z_i = 1)E(\tilde{X}_i\tilde{X}_i^T|Z_i = 1)^{-1}\tilde{X}_i, \\ g_0(X_i) &= E(Y_i\tilde{X}_i|Z_i = 0)E(\tilde{X}_i\tilde{X}_i^T|Z_i = 0)^{-1}\tilde{X}_i, \end{aligned}$$

where g_0 and g_1 correspond to linear projection of Y_i into the space of X_i (including a constant) in two arms. If the true outcome surface $E(Y_i|X_i, Z_i = 1)$ and $E(Y_i|X_i, Z_i = 0)$ are indeed linear functions of X_i, then the $g_1(X_i) = E(Y_i|X_i, Z_i = 1)$,$g_0(X) = E(Y_i|X_i, Z_i = 0)$, $\hat{\tau}^{\mathrm{ow}} = \hat{\mu}_1{}^{\mathrm{ow}} - \hat{\mu}_0{}^{\mathrm{ow}}$ is semiparametric efficient. As such, we complete the proof of Proposition 1(c).

Proof for Proposition 2: Since we require $h(x)$ to be a function of the propensity score, we denote the tilting function and the resulting balancing weights as $h(X_i;\theta), w_1(X_i;\theta), w_0(X_i;\theta)$ corresponding to each observation i. Also, we make the following assumptions:

(i) (Nonzero tilting function) There exists $\epsilon > 0$ such that $P\{h(X_i;\theta^*) > \epsilon\} = 1$.

(ii) (Smoothness) the first and second order derivatives of balancing weights with respect to the propensity score $\frac{d}{de}w_1(X_i;\theta)\}$, $\frac{d}{de}w_0(X_i;\theta)$,$\frac{d^2}{de^2}w_1(X_i;\theta)\}$, $\frac{d^2}{de^2}w_0(X_i;\theta)$ exists and are continuous in e.

(iii) (Bounded derivative in the neighborhood of θ^*) For the true value θ^*, there exists $c > 0$ and $M_1 > 0, M_2 > 0$ such that

$$\left|\frac{d}{de}w_0(X_i;\theta^*)\right| \leq M_1, \left|\frac{d}{de}w_0(X_i;\theta^*)\right| \leq M_1$$
$$\left|\frac{d^2}{de^2}w_0(X_i;\theta)\right| \leq M_2, \left|\frac{d^2}{de^2}w_1(X_i;\theta)\right| \leq M_2,$$

almost surely for θ in the neighborhood of θ^*, i.e. $\theta \in \{\theta|\,||\theta - \theta^*||_1 \leq c\}$.

We investigate the influence function of $\hat{\tau}^h$ for a given $h(x)$ and do Taylor expansion at the true value θ^*,

$$
\begin{aligned}
\sqrt{N}(\hat{\mu}_1^h - \hat{\mu}_0^h) =& \sqrt{N}\left(\frac{\sum_{i=1}^N Z_iY_iw_1(X_i;\hat{\theta})}{\sum_{i=1}^N Z_iw_1(X_i;\hat{\theta})} - \frac{\sum_{i=1}^N Z_iY_iw_0(X_i;\hat{\theta})}{\sum_{i=1}^N (1-Z_i)w_0(X_i;\hat{\theta})}\right),\\
=& \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N Z_iY_iw_1(X_i;\hat{\theta})}{E\{h(X_i;\theta^*)\}} - \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N (1-Z_i)Y_iw_0(X_i;\hat{\theta})}{E\{h(X_i;\theta^*)\}} + o_p(1),\\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{ \frac{Z_iY_iw_1(X_i;\theta^*)}{E\{h(X_i;\theta^*)\}} - \frac{(1-Z_i)Y_iw_0(X_i;\theta^*)}{E\{h(X_i;\theta^*)\}} \right.\\
&+ \frac{\left\{Z_iY_i\frac{d}{de}w_1(X_i;\theta^*) - (1-Z_i)Y_i\frac{d}{de}w_0(X_i;\theta^*)\right\}\frac{\partial e(X_i;\theta^*)}{\partial\theta}^T(\hat{\theta}-\theta^*)}{E\{h(X_i;\theta^*)\}}\\
&\left. + \frac{\left\{Z_iY_i\left[\frac{d^2}{de^2}w_1(X_i;\tilde{\theta}) + \frac{d}{de}w_1(X_i;\tilde{\theta})\right] - (1-Z_i)Y_i\left[\frac{d^2}{de^2}w_0(X_i;\tilde{\theta}) + \frac{d}{de}w_0(X_i;\tilde{\theta})\right]\right\}(\hat{\theta}-\theta^*)^T\frac{\partial^2 e(X_i;\tilde{\theta})}{\partial\theta^2}(\hat{\theta}-\theta^*)}{E\{h(X_i;\theta^*)\}} \right\}\\
&+ o_p(1),
\end{aligned}
$$

where $\tilde{\theta}$ lies in the line between θ^* and $\hat{\theta}$, such that $\tilde{\theta} = \theta^* + t(\hat{\theta}-\theta^*), t \in (0,1)$ (Taylor expansion with Lagrange remainder term). To see that the third term converges to zero in probability, we have $\sqrt{N}(\hat{\theta}-\theta^*)$ is asymptotic normal distributed with Cramer-Rao theorem and the asymptotic covariance is proportional to $E\left\{\frac{\partial^2 e(X_i;\theta^*)}{\partial\theta^2}\right\}^{-1}$, which means $N(\hat{\theta}-\theta^*)^T E\left\{\frac{\partial^2 e(X_i;\theta^*)}{\partial\theta^2}\right\}(\hat{\theta}-\theta^*)$ is tight, or equivalently

$$
P\left\{N(\hat{\theta}-\theta^*)^T E\left\{\frac{\partial^2 e(X_i;\theta^*)}{\partial\theta^2}\right\}(\hat{\theta}-\theta^*) < \infty\right\} = 1.
$$

Secondly, as $\hat{\theta} \xrightarrow{p} \theta^*$, $\tilde{\theta} \xrightarrow{p} \theta^*$, when N is sufficiently large, $||\tilde{\theta}-\theta^*||_1 \le c$, the first and second order derivative is bounded almost surely, such that

$$
\left|\frac{d^2}{de^2}w_1(X_i;\tilde{\theta}) + \frac{d}{de}w_1(X_i;\tilde{\theta})\right| \le M_1 + M_2, \left|\frac{d^2}{de^2}w_0(X_i;\tilde{\theta}) + \frac{d}{de}w_0(X_i;\tilde{\theta})\right| \le M_1 + M_2.
$$

Therefore, by the WLLN,

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N \left\{Z_iY_i\left[\frac{d^2}{de^2}w_1(X_i;\tilde{\theta}) + \frac{d}{de}w_1(X_i;\tilde{\theta})\right] - (1-Z_i)Y_i\left[\frac{d^2}{de^2}w_0(X_i;\tilde{\theta}) + \frac{d}{de}w_0(X_i;\tilde{\theta})\right]\right\}\\
&\le (M_1+M_2)\frac{1}{N}\sum_{i=1}^N |Z_iY_i| + |(1-Z_i)Y_i)| \xrightarrow{p} E\{|Z_iY_i| + |(1-Z_i)Y_i)|\} < \infty.
\end{aligned}
$$

Also, as $\tilde{\theta} \xrightarrow{p} \theta^*$, and we assume $e(X_i;\theta)$ is smooth (so that $\frac{\partial^2 e(X_i;\tilde{\theta})}{\partial\theta^2}$ is continuous),

$$
\frac{1}{N}\sum_{i=1}^N \frac{\partial^2 e(X_i;\tilde{\theta})}{\partial\theta^2} \xrightarrow{p} E\left\{\frac{\partial^2 e(X_i;\theta^*)}{\partial\theta^2}\right\}.
$$

As such, we can conclude that the third term converges to zero in probability,

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{\left\{Z_iY_i\left[\frac{d^2}{de^2}w_1(X_i;\tilde{\theta}) + \frac{d}{de}w_1(X_i;\tilde{\theta})\right] - (1-Z_i)Y_i\left[\frac{d^2}{de^2}w_0(X_i;\tilde{\theta}) + \frac{d}{de}w_0(X_i;\tilde{\theta})\right]\right\}(\hat{\theta}-\theta^*)^T\frac{\partial^2 e(X_i;\tilde{\theta})}{\partial\theta^2}(\hat{\theta}-\theta^*)}{E\{h(X_i;\theta^*)\}}\\
=& O_p\left(\frac{1}{\sqrt{N}}\frac{E\{|Z_iY_i| + |(1-Z_i)Y_i)|\}N(\hat{\theta}-\theta^*)^T E\left\{\frac{\partial^2 e(X_i;\theta^*)}{\partial\theta^2}\right\}(\hat{\theta}-\theta^*)}{E\{h(X_i;\theta^*)\}}\right) \xrightarrow{p} 0.
\end{aligned}
$$

Hence, we have,

$$\sqrt{N}(\hat{\mu}_1^h - \hat{\mu}_0^h) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{Z_iY_iw_1(X_i;\theta^*)}{E\{h(X_i;\theta^*)\}} - \frac{(1-Z_i)Y_iw_0(X_i;\theta^*)}{E\{h(X_i;\theta^*)\}}\right.$$
$$\left. + \frac{\left\{Z_iY_i\frac{d}{de}w_1(X_i;\theta^*) - (1-Z_i)Y_i\frac{d}{de}w_0(X_i;\theta^*)\right\}\frac{\partial e(X_i;\theta^*)}{\partial\theta}(\hat{\theta}-\theta^*)}{E\{h(X_i;\theta^*)\}}\right\} + o_p(1),$$
$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\frac{Z_iY_ih(X_i;\theta^*)/r}{E\{h(X_i;\theta^*)\}} - \frac{(1-Z_i)Y_ih(X_i;\theta^*)/(1-r)}{E\{h(X_i;\theta^*)\}}\right.$$
$$\left. + \frac{E\left[\left\{Z_iY_i\frac{d}{de}w_1(X_i;\theta^*) - (1-Z_i)Y_i\frac{d}{de}w_0(X_i;\theta^*)\right\}\frac{\partial e(X_i;\theta^*)}{\partial\theta}\right]}{E\{h(X_i;\theta^*)\}} I_{\theta^*\theta^*}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}S_{\theta^*,i}\right\} + o_p(1).$$

Since $h(X_i;\theta)$ is a function of propensity score, $h(X_i;\theta^*)$ is a function of r, which means $E\{h(X_i;\theta^*)\} = h(X_i;\theta^*)$. Applying this property and plugging in the value of $S_{\theta^*,i}, I_{\theta^*\theta^*}$, we have,

$$\hat{\mu}_1^h - \hat{\mu}_0^h = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{Z_iY_i}{r} - \frac{(1-Z_i)Y_i}{1-r} - \frac{Z_i - r}{r(1-r)}\left\{(1-r)g_1^h(X_i) + rg_0^h(X_i)\right\}\right] + o_p(N^{-1/2}),$$
$$g_1^h(X_i) = -\frac{r}{h(X_i;\theta^*)}E\left\{Z_iY_i\frac{d}{de}w_1(X_i;\theta^*)\right\}E\left\{\frac{\partial e(X_i;\theta^*)}{\partial\theta}\frac{\partial e(X_i;\theta^*)}{\partial\theta}^T\right\}^{-1}\frac{\partial e(X_i;\theta^*)}{\partial\theta},$$
$$g_0^h(X_i) = \frac{1-r}{h(X_i;\theta^*)}E\left\{(1-Z_i)Y_i\frac{d}{de}w_0(X_i;\theta^*)\right\}E\left\{\frac{\partial e(X_i;\theta^*)}{\partial\theta}\frac{\partial e(X_i;\theta^*)}{\partial\theta}^T\right\}^{-1}\frac{\partial e(X_i;\theta^*)}{\partial\theta},$$

which completes the proof of Proposition 2.

References

1. Lehmann EL, Casella G. *Theory of point estimation*. Springer Science & Business Media . 2006.

APPENDIX B: DERIVATION OF THE ASYMPTOTIC VARIANCE AND ITS CONSISTENT ESTIMATOR IN SECTION 3.3

Asymptotic variance derivation. As we have shown in the main text (Section 3.3), the asymptotic variance of $\hat{\tau}^{\text{OW}}$ depends on the elements in the sandwich matrix $A^{-1}BA^{-T}$, where $A = -E(\partial U_i/\partial\lambda), B = E(U_iU_i^T)$ evaluated at the true parameter value (μ_1, μ_0, θ^*). The exact form of the matrices A and B are as follows:

$$A = \begin{bmatrix} a_{11} & 0 & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}, A^{-1} = \begin{bmatrix} a_{11}^{-1} & 0 & -a_{11}^{-1}a_{13}a_{33}^{-1}, \\ 0 & a_{22}^{-1} & -a_{22}^{-1}a_{23}a_{33}^{-1} \\ 0 & 0 & a_{33}^{-1} \end{bmatrix}, B = \begin{bmatrix} b_{11} & 0 & b_{13} \\ 0 & b_{22} & b_{23} \\ b_{13}^T & b_{23}^T & b_{33} \end{bmatrix},$$

$$a_{11} = E\{Z_i(1-e_i)\}, a_{13} = E\{\tilde{X}_i^T(Y_i-\mu_1)Z_ie_i(1-e_i)\}, a_{22} = E\{(1-Z_i)e_i\},$$
$$a_{23} = -E(\tilde{X}_i^T(Y_i-\mu_0)(1-Z_i)e_i(1-e_i), a_{33} = E(e_i(1-e_i)\tilde{X}\tilde{X}^T],$$
$$b_{11} = E\{(Y_i-\mu_1)^2Z_i(1-e_i)^2\}, b_{13} = E\{\tilde{X}_i^T(Y_i-\mu_1)Z_i(Z_i-e_i)(1-e_i), b_{23} = E\{\tilde{X}_i^T(Y_i-\mu_0)(1-Z_i)(Z_i-e_i)e_i\},$$
$$b_{22} = E\{(Y_i-\mu_0)^2(1-Z_i)e_i^2\}, b_{33} = E\{(Z_i-e_i)^2\tilde{X}_i\tilde{X}_i^T\}.$$

After multiplying $A^{-1}BA^{-T}$ and extracting the upper left 2×2 matrix, we have,

$$\Sigma_{11} = [A^{-1}BA^{-T}]_{1,1} = \frac{1}{a_{11}^{-2}}(b_{11} - 2a_{13}a_{33}^{-1}b_{13}^T + a_{13}a_{33}^{-1}b_{33}a_{33}^{-1}a_{13}^T),$$
$$\Sigma_{22} = [A^{-1}BA^{-T}]_{2,2} = \frac{1}{a_{22}^{-2}}(b_{22} - 2a_{23}a_{33}^{-1}b_{23}^T + a_{23}a_{33}^{-1}b_{33}a_{33}^{-1}a_{23}^T),$$
$$\Sigma_{12} = \Sigma_{21} = [A^{-1}BA^{-T}]_{1,2} = \frac{1}{a_{11}a_{22}}(-a_{13}a_{33}^{-1}b_{23}^T - a_{23}a_{33}^{-1}b_{13}^T + a_{13}a_{33}^{-1}b_{33}a_{33}^{-1}a_{23}^T).$$

With the delta method, we can express the asymptotic variance for $\hat{\tau}^{\text{OW}}_{\text{RD}}, \hat{\tau}^{\text{OW}}_{\text{RR}}, \hat{\tau}^{\text{OW}}_{\text{OR}}$,

$$\text{Var}(\hat{\tau}^{\text{OW}}_{\text{RD}}) = \frac{1}{N}\left(\Sigma_{11} + \Sigma_{22} - 2\Sigma_{12}\right),$$
$$\text{Var}(\hat{\tau}^{\text{OW}}_{\text{RR}}) = \frac{1}{N}\left(\frac{\Sigma_{11}}{\mu_1^2} + \frac{\Sigma_{22}}{\mu_0^2} - \frac{2\Sigma_{12}}{\mu_1\mu_0}\right),$$
$$\text{Var}(\hat{\tau}^{\text{OW}}_{\text{OR}}) = \frac{1}{N}\left\{\frac{\Sigma_{11}}{\mu_1^2(1-\mu_1)^2} + \frac{\Sigma_{22}}{\mu_0^2(1-\mu_0)^2} - \frac{2\Sigma_{12}}{\mu_1(1-\mu_1)\mu_0(1-\mu_0)}\right\}.$$

Specifically, we write out the exact form of large sample variance for the estimator on additive scale after exploiting the fact that $E(Z_i) = E(e_i) = r$,

$$NVar(\hat{\tau}^{\text{OW}}) \to \frac{\text{Var}(Y_i|Z_i=1)}{r} + \frac{\text{Var}(Y_i|Z_i=1)}{1-r} - \frac{\{rm_1 + (1-r)m_0\}E(\tilde{X}_i\tilde{X}_i^T)^{-1}\{(2-3r)m_1 + (3r-1)m_0\}}{r(1-r)},$$

where $m_1 = E(\tilde{X}_i(Y_i-\mu_1)|Z_i=1), m_0 = E(\tilde{X}_i(Y_i-\mu_1)|Z_i=1)$

Connection to R-squared: When $r = 0.5$, the large sample variance of $\hat{\tau}^{\text{OW}}$ is,

$$\begin{aligned} N\text{Var}(\hat{\tau}^{\text{OW}}) \to& 2\left\{\text{Var}(Y_i|Z_i=1) + \text{Var}(Y_i|Z_i=0)\right\} - 4(\frac{1}{2}m_1 + \frac{1}{2}m_0)E(\tilde{X}\tilde{X}^T)^{-1}(\frac{1}{2}m_1 + \frac{1}{2}m_0), \\ =& 2\left\{\text{Var}(Y_i|Z_i=1) + \text{Var}(Y_i|Z_i=0)\right\} - 4E(\tilde{X}_i\tilde{Y}_i)E(\tilde{X}_i\tilde{X}_i^T)^{-1}E(\tilde{X}_i\tilde{Y}_i), \\ =& 2\left\{\text{Var}(Y_i|Z_i=1) + \text{Var}(Y_i|Z_i=0)\right\} - 4R^2_{\tilde{Y}\sim X}\text{Var}(\tilde{Y}_i), \\ =& 2\left\{\text{Var}(Y_i|Z_i=1) + \text{Var}(Y_i|Z_i=0)\right\} - 2R^2_{\tilde{Y}\sim X}\left\{\text{Var}(Y_i|Z_i=1) + \text{Var}(Y_i|Z_i=0)\right\}, \\ =& 4(1 - R^2_{\tilde{Y}\sim X})\text{Var}(\tilde{Y}_i), \\ =& \lim_{N\to\infty}(1 - R^2_{\tilde{Y}\sim X})N\text{Var}(\hat{\tau}^{\text{UNADJ}}). \end{aligned}$$

where $\tilde{Y}_i = Z_i(Y_i-\mu_1) + (1-Z_i)(Y_i-\mu_0)$. In the derivation, we use the fact that,

$$\text{Var}(\tilde{Y}_i) = E(\tilde{Y}_i^2) - E(\tilde{Y}_i)^2 = \frac{1}{2}E\left((Y_i-\mu_1)^2|Z_i=1\right) + \frac{1}{2}E\left((Y_i-\mu_1)^2|Z_i=1\right) = \frac{1}{2}\left\{\text{Var}(Y_i|Z_i=1) + \text{Var}(Y_i|Z_i=0)\right\}.$$

The efficiency gain is irrelevant to whether our model is correctly specified or not. Additionally, if we augment the covariate space from $\tilde{X}_i$ to X_i^*, the $R^2_{\tilde{Y}\sim X}$ is non-decreasing with $R^2_{\tilde{Y}\sim X} \le R^2_{\tilde{Y}\sim X^*}$. Therefore, the asymptotic variance of OW estimator with additional covariates decreases, $\mathrm{Var}(\hat{\tau}^{\mathrm{OW}*}) \leq \mathrm{Var}(\hat{\tau}^{\mathrm{OW}})$. This provides an intuitive justification of Proposition 1(b) when $r = 0.5$.

Consistent variance estimator: We obtain the empirical estimator for the asymptotic variance by plugging in the finite sample estimate for the elements in the sandwich matrix $A^{-1}BA^{-T}$,

$$\hat{\Sigma}_{11} = \frac{1}{\hat{a}_{11}^2}(\hat{b}_{11} - 2\hat{a}_{13}\hat{a}_{33}^{-1}\hat{b}_{13}^T + \hat{a}_{13}\hat{a}_{33}^{-1}\hat{a}_{13}^T),$$

$$\hat{\Sigma}_{22} = \frac{1}{\hat{a}_{11}^2}(\hat{b}_{22} - 2\hat{a}_{23}\hat{a}_{33}^{-1}\hat{b}_{23}^T + \hat{a}_{23}\hat{a}_{33}^{-1}\hat{a}_{23}^T),$$

$$\hat{\Sigma}_{12} = -\frac{1}{\hat{a}_{11}^2}(\hat{a}_{13}\hat{a}_{33}^{-1}\hat{b}_{23}^T + \hat{a}_{23}\hat{a}_{33}^{-1}\hat{b}_{13}^T - \hat{a}_{13}\hat{a}_{33}^{-1}\hat{a}_{23}^T),$$

$$\hat{a}_{11} = \hat{a}_{22} = \frac{1}{N}\sum_{i=1}^{N}\{\hat{e}_i(1-\hat{e}_i)\}, \qquad \hat{a}_{33} = \hat{b}_{33} = \frac{1}{N}\sum_{i=1}^{N}\{\hat{e}_i(1-\hat{e}_i)\tilde{X}_i^T\tilde{X}_i\},$$

$$\hat{a}_{13} = \frac{1}{N_1}\sum_{i=1}^{N} Z_i\hat{e}_i^2(1-\hat{e}_i)(Y_i-\hat{\mu}_1)^2\tilde{X}_i, \qquad \hat{a}_{23} = \frac{1}{N_0}\sum_{i=1}^{N}(1-Z_i)\hat{e}_i(1-\hat{e}_i)^2(Y_i-\hat{\mu}_0)^2\tilde{X}_i,$$

$$\hat{b}_{11} = \frac{1}{N_1}\sum_{i=1}^{N} Z_i\hat{e}_i(1-\hat{e}_i)^2(Y_i-\hat{\mu}_1)^2, \qquad \hat{b}_{22} = \frac{1}{N_0}\sum_{i=1}^{N}(1-Z_i)\hat{e}_i^2(1-\hat{e}_i)(Y_i-\hat{\mu}_0)^2,$$

$$\hat{b}_{13} = \frac{1}{N_1}\sum_{i} Z_i\hat{e}_i(1-\hat{e}_i)^2(Y_i-\hat{\mu}_1)\tilde{X}_i, \qquad \hat{b}_{23} = \frac{1}{N_0}\sum_{i}(1-Z_i)\hat{e}_i^2(1-\hat{e}_i)(Y_i-\hat{\mu}_0)\tilde{X}_i.$$

Hence, we summarize the estimators for the asymptotic variance of $\hat{\tau}^{\mathrm{OW}}_{\mathrm{RD}}, \hat{\tau}^{\mathrm{OW}}_{\mathrm{RR}}, \hat{\tau}^{\mathrm{OW}}_{\mathrm{OR}}$ in the following equations,

$$\mathrm{Var}(\hat{\tau}^{\mathrm{OW}}) = \frac{1}{N}\left[\hat{V}^{\mathrm{UNADJ}} - \hat{v}_1^T\left\{\frac{1}{N}\sum_{i=1}^{N}\hat{e}_i(1-\hat{e}_i)\tilde{X}_i^T\tilde{X}_i\right\}^{-1}(2\hat{v}_1 - \hat{v}_2)\right], \tag{1}$$

where

$$\hat{V}^{\mathrm{UNADJ}} = \left\{\frac{1}{N}\sum_{i=1}^{N}\hat{e}_i(1-\hat{e}_i)\right\}^{-1}\left(\frac{\hat{E}_1^2}{N_1}\sum_{i=1}^{N}Z_i\hat{e}_i(1-\hat{e}_i)^2(Y_i-\hat{\mu}_1)^2 + \frac{\hat{E}_0^2}{N_0}\sum_{i=1}^{N}(1-Z_i)\hat{e}_i^2(1-\hat{e}_i)(Y_i-\hat{\mu}_0)^2\right),$$

$$\hat{v}_1 = \left\{\frac{1}{N}\sum_{i=1}^{N}\hat{e}_i(1-\hat{e}_i)\right\}^{-1}\left(\frac{\hat{E}_1}{N_1}\sum_{i=1}^{N}Z_i\hat{e}_i^2(1-\hat{e}_i)(Y_i-\hat{\mu}_1)^2\tilde{X}_i + \frac{\hat{E}_0}{N_0}\sum_{i=1}^{N}(1-Z_i)\hat{e}_i(1-\hat{e}_i)^2(Y_i-\hat{\mu}_0)^2\tilde{X}_i\right),$$

$$\hat{v}_2 = \left\{\frac{1}{N}\sum_{i=1}^{N}\hat{e}_i(1-\hat{e}_i)\right\}^{-1}\left(\frac{\hat{E}_1}{N_1}\sum_{i=1}^{N}Z_i\hat{e}_i(1-\hat{e}_i)^2(Y_i-\hat{\mu}_1)^2\tilde{X}_i + \frac{\hat{E}_0}{N_0}\sum_{i=1}^{N}(1-Z_i)\hat{e}_i^2(1-\hat{e}_i)(Y_i-\hat{\mu}_0)^2\tilde{X}_i\right),$$

and $\hat{E}_k$ depends on the estimands. For $\hat{\tau}^{\mathrm{OW}}_{\mathrm{RD}}$, we have $\hat{E}_k = 1$; for $\hat{\tau}^{\mathrm{OW}}_{\mathrm{RR}}$, we set $\hat{E}_k = \hat{\mu}_k^{-1}$; for $\hat{\tau}^{\mathrm{OW}}_{\mathrm{OR}}$, we use $\hat{E}_k = \hat{\mu}_k^{-1}(1-\hat{\mu}_k)^{-1}$ with $k = 0, 1$.

APPENDIX C: VARIANCE ESTIMATOR FOR $\hat{\tau}^{\text{AIPW}}$

In this appendix, we provide the details on how to derive the variance estimator for $\hat{\tau}^{\text{AIPW}}$ in the main text. Let $\mu_1(X_i;\alpha_1)$, $\mu_0(X_i;\alpha_0)$ be the outcome surface for treated and control samples respectively, with α_1, α_0 being the regression parameters. Suppose $\hat{\alpha}_1$, $\hat{\alpha}_0$ are the MLEs that solve the score functions $\sum_{i=1}^N Z_i S_1(Y_i, X_i;\alpha_1) = 0$ and $\sum_{i=1}^N (1-Z_i) S_0(Y_i, X_i;\alpha_0) = 0$. We resume our notation and let $e(X_i;\theta)$ be the propensity score, $\hat{\theta}$ be the parameters and $S_\theta(X_i;\theta)$ be the corresponding score function. Recall that $\hat{\tau}^{\text{AIPW}}$ takes the following form:

$$\hat{\tau}^{\text{AIPW}} = \hat{\mu}_1^{\text{AIPW}} - \hat{\mu}_0^{\text{AIPW}} = \frac{1}{N}\sum_{i=1}^N \left\{ \frac{Z_i Y_i}{\hat{e}_i} - \frac{(Z_i - \hat{e}_i)\hat{\mu}_1(X_i)}{\hat{e}_i} \right\} - \left\{ \frac{(1-Z_i)Y_i}{1-\hat{e}_i} + \frac{(Z_i - \hat{e}_i)\hat{\mu}_0(X_i)}{1-\hat{e}_i} \right\},$$

Let $\lambda = (\nu_1, \nu_0, \alpha_0, \alpha_1, \theta)$ and $\hat{\lambda} = (\hat{\nu}_1, \hat{\nu}_0, \hat{\alpha}_0, \hat{\alpha}_1, \hat{\theta})$. Note that $\hat{\lambda}$ is the solution for λ in the equations below:

$$\sum_{i=1}^N \Psi_i = \sum_{i=1}^N \begin{bmatrix} \nu_1 - \{Z_i Y_i - (Z_i - e_i)\mu_1(X_i;\alpha_1)\}/e_i \\ \nu_0 - \{(1-Z_i)Y_i + (Z_i - e_i)\mu_0(X_i;\alpha_0)\}/(1-e_i) \\ Z_i S_1(Y_i, X_i;\alpha_1) \\ (1-Z_i)S_0(Y_i, X_i;\alpha_0) \\ S_\theta(X_i;\theta) \end{bmatrix} = 0.$$

The asymptotic covariance of $\hat{\lambda}$ can be obtained via M-estimation theory, which equals $A^{-1}BA^T$, with $A = -E(\partial\Psi_i/\partial\lambda)$, $B = E(\Psi_i\Psi_i^T)$. In practice, we use plug-in method to estimate A, B. We can express $\hat{\tau}^{\text{AIPW}}$ with the solution $\hat{\lambda}$ as $\hat{\tau}^{\text{AIPW}} = \hat{\nu}_1 - \hat{\nu}_0$. Next, we can calculate the asymptotic variance of $\hat{\tau}^{\text{AIPW}}$ based on the asymptotic covariance of $\hat{\lambda}$ and the delta method. Similarly, it is straightforward to obtain the estimator for risk ratio estimator $\hat{\tau}_{\text{RR}}^{\text{AIPW}} = \log\left(\hat{\nu}_1/\hat{\nu}_0\right)$ and odds ratio estimator $\hat{\tau}_{\text{OR}}^{\text{AIPW}} = \log\left(\hat{\nu}_1/(1-\hat{\nu}_1)\right) - \log\left(\hat{\nu}_0/(1-\hat{\nu}_0)\right)$, as in Web Appendix B.

APPENDIX D: PROGRAMMING CODE

In this appendix, we include the details to reproduce all results within the paper. Please download the codebase from https://github.com/zengshx777/OWRCT_codes_package with

R Scripts

`Main_RCT_Continuous.R`	Run simulations with continuous outcome.
`Main_RCT_Binary.R`	Run simulations with binary outcome.
`real_data_application.R`	Analyze BestAir study data produce Table 3 in the main text.
`Crude.R`	Function implements $\hat{\tau}^{\text{UNADJ}}$.
`IPWC.R`	Function implements $\hat{\tau}^{\text{IPW}}$.
`LinearR.R`	Function implements $\hat{\tau}^{\text{LR}}$.
`PS_AIPW.R`	Function implements $\hat{\tau}^{\text{AIPW}}$.
`OW.R`	Function implements $\hat{\tau}^{\text{OW}}$.
`plot_cont.R`	Visualize continuous simulation results, produce Figure 1 in main text.
`plot_bin.R`	Visualize binary simulation results, produce Figure 2,3 in main text.
`table_produce.R`	Summarize all results, produce Table 1 in the main text, Table 1,2,3 in Web Appendix E.
`example.R`	Simple demo for running simulations.
`all_jobs.sh`	Bash script to run all simulations.

To replicate the simulation results in the paper, the simplest way is to run `all_jobs.sh` after setting the code package as the working directory. The results will be automatically saved in folders ‘cont’ and ‘bin’. For real data application, running `real_data_application.R` will reproduce the results.

APPENDIX E: ADDITIONAL SIMULATION RESULTS

We include the additional numerical results for the simulations under different scenarios in Table 1 and 2. For binary outcome, we consider the following scenarios,

1. $u = 0.5, r = 0.5, b_1 = 0$, model is correctly specified, corresponding to scenario (a) in Figure 2.
2. $u = 0.5, r = 0.5, b_1 = 0$, model is misspecified, corresponding to scenario (b) in Figure 2.
3. $u = 0.3, r = 0.5, b_1 = 0$, model is correctly specified, corresponding to scenario (c) in Figure 2.
4. $u = 0.3, r = 0.5, b_1 = 0$, model is misspecified, corresponding to scenario (d) in Figure 2.
5. $u = 0.5, r = 0.5, b_1 = 0.75$, model is correctly specified, corresponding to scenario (e) in Figure 3.
6. $u = 0.5, r = 0.7, b_1 = 0$, model is correctly specified, corresponding to scenario (f) in Figure 3.
7. $u = 0.2, r = 0.5, b_1 = 0$, model is correctly specified, corresponding to scenario (g) in Figure 3.
8. $u = 0.1, r = 0.5, b_1 = 0$, model is correctly specified, corresponding to scenario (h) in Figure 3.

For binary outcome, we also report in Table 3 the number of non-convergences for fitting logistic regression under different baseline outcome prevalence $u = 0.5, 0.3, 0.2, 0.1$.

$\square$

TABLE 1 The relative efficiency of each estimator compared to the unadjusted, the ratio between the average estimated variance ({Est Var}) over Monte Carlo variance ({MC Var}) and 95% coverage rate of IPW, LR, AIPW and OW estimators for binary outcomes. The scenarios correspond to Figure 2 in the main manuscript.

		Relative efficiency				{Est Var}/{MC Var}				95% Coverage			
	N	IPW	LR	AIPW	OW	IPW	LR	AIPW	OW	IPW	LR	AIPW	OW
					$u = 0.5$, $b_1 = 0$, $r = 0.5$, correct specification **(a)**								
	50	0.729	0.966	0.854	0.880	0.936	1.387	0.903	1.124	0.903	0.940	0.906	0.943
τ_{RD}	100	1.034	1.100	1.061	1.083	0.796	0.924	0.763	0.972	0.914	0.934	0.905	0.945
	200	1.152	1.159	1.149	1.158	0.985	1.049	0.967	1.164	0.944	0.953	0.945	0.961
	500	1.186	1.191	1.191	1.184	0.969	0.995	0.969	1.151	0.946	0.948	0.947	0.962
	50	0.690	0.976	0.832	0.860	0.910	1.372	0.870	1.097	0.924	0.966	0.926	0.964
τ_{RR}	100	1.038	1.104	1.062	1.090	0.803	0.927	0.766	0.979	0.922	0.942	0.915	0.953
	200	1.154	1.160	1.150	1.160	0.987	1.050	0.969	1.165	0.948	0.957	0.947	0.964
	500	1.189	1.193	1.194	1.186	0.971	0.996	0.970	1.152	0.950	0.952	0.949	0.965
	50	0.702	0.960	0.836	0.864	0.950	1.395	0.905	1.128	0.913	0.966	0.915	0.955
τ_{OR}	100	1.031	1.101	1.060	1.082	0.795	0.925	0.763	0.973	0.920	0.938	0.910	0.950
	200	1.153	1.160	1.150	1.159	0.985	1.050	0.968	1.164	0.946	0.954	0.946	0.963
	500	1.187	1.191	1.192	1.184	0.969	0.994	0.968	1.150	0.948	0.951	0.948	0.964
					$u = 0.5$, $b_1 = 0$, $r = 0.5$, misspecification **(b)**								
	50	0.742	0.942	0.848	0.827	0.888	1.225	0.825	0.996	0.887	0.943	0.902	0.921
τ_{RD}	100	0.971	1.057	1.002	1.033	0.813	0.996	0.799	0.976	0.913	0.945	0.911	0.937
	200	1.074	1.086	1.076	1.082	0.921	0.993	0.912	1.039	0.936	0.943	0.936	0.950
	500	1.100	1.106	1.105	1.100	0.962	0.993	0.963	1.088	0.948	0.950	0.948	0.957
	50	0.697	0.944	0.824	0.811	0.869	1.244	0.834	1.000	0.909	0.943	0.914	0.948
τ_{RR}	100	0.968	1.072	1.013	1.036	0.806	0.992	0.797	0.966	0.925	0.956	0.924	0.947
	200	1.071	1.084	1.075	1.078	0.913	0.983	0.903	1.029	0.940	0.948	0.940	0.955
	500	1.103	1.110	1.109	1.103	0.966	0.997	0.967	1.092	0.949	0.952	0.948	0.958
	50	0.714	0.936	0.831	0.808	0.890	1.231	0.826	0.997	0.902	0.950	0.909	0.943
τ_{OR}	100	0.966	1.058	1.001	1.031	0.810	0.995	0.797	0.973	0.919	0.951	0.920	0.944
	200	1.075	1.087	1.077	1.083	0.921	0.992	0.911	1.039	0.938	0.947	0.938	0.953
	500	1.100	1.107	1.106	1.101	0.962	0.993	0.963	1.088	0.949	0.951	0.948	0.958
					$u = 0.3$, $b_1 = 0$, $r = 0.5$, correct specification **(c)**								
	50	0.797	0.946	0.899	0.942	0.915	1.369	0.892	1.141	0.896	0.944	0.892	0.937
τ_{RD}	100	1.002	1.044	1.021	1.043	0.852	1.138	0.814	1.015	0.925	0.951	0.914	0.945
	200	1.123	1.124	1.116	1.130	0.976	1.154	0.952	1.131	0.942	0.960	0.940	0.957
	500	1.187	1.201	1.198	1.188	1.014	1.147	1.014	1.185	0.951	0.964	0.951	0.966
	50	0.758	0.034	0.004	0.938	1.004	0.051	0.004	1.241	0.919	0.964	0.917	0.971
τ_{RR}	100	1.010	1.070	1.041	1.043	0.859	1.173	0.818	1.019	0.936	0.965	0.929	0.956
	200	1.124	1.132	1.122	1.129	0.962	1.148	0.939	1.114	0.949	0.968	0.945	0.962
	500	1.189	1.204	1.201	1.189	1.007	1.141	1.007	1.176	0.954	0.966	0.955	0.968
	50	0.748	0.073	0.008	0.924	1.013	0.112	0.009	1.225	0.915	0.959	0.917	0.958
τ_{OR}	100	1.005	1.057	1.031	1.043	0.855	1.158	0.816	1.019	0.931	0.961	0.922	0.952
	200	1.124	1.129	1.120	1.130	0.968	1.152	0.945	1.123	0.946	0.965	0.942	0.960
	500	1.188	1.203	1.200	1.189	1.011	1.144	1.010	1.181	0.952	0.964	0.953	0.967
					$u = 0.3$, $b_1 = 0$, $r = 0.5$, misspecification **(d)**								
	50	0.667	0.921	0.687	0.858	0.924	1.471	0.889	1.204	0.883	0.976	0.943	0.926
τ_{RD}	100	0.950	1.021	0.977	0.989	0.859	1.196	0.837	1.019	0.918	0.958	0.912	0.948
	200	1.126	1.139	1.133	1.126	0.946	1.156	0.931	1.072	0.940	0.963	0.938	0.953
	500	1.116	1.137	1.132	1.118	1.031	1.209	1.029	1.183	0.951	0.966	0.952	0.962
	50	0.543	0.952	0.630	0.795	0.885	1.515	1.039	1.189	0.905	0.986	0.953	0.959
τ_{RR}	100	0.941	1.041	0.993	0.975	0.843	1.202	0.822	1.000	0.932	0.971	0.923	0.961
	200	1.127	1.147	1.142	1.123	0.949	1.170	0.934	1.074	0.946	0.969	0.939	0.958
	500	1.115	1.139	1.135	1.117	1.028	1.208	1.026	1.178	0.953	0.968	0.954	0.964
	50	0.583	0.928	0.634	0.818	0.917	1.498	0.999	1.196	0.900	0.981	0.953	0.945
τ_{OR}	100	0.944	1.031	0.985	0.981	0.851	1.201	0.829	1.010	0.926	0.965	0.920	0.953
	200	1.127	1.143	1.138	1.125	0.947	1.163	0.932	1.074	0.940	0.966	0.940	0.957
	500	1.116	1.138	1.134	1.118	1.029	1.209	1.027	1.181	0.952	0.967	0.954	0.963

TABLE 2 The relative efficiency of each estimator compared to the unadjusted, the ratio between the average estimated variance ({Est Var}) over Monte Carlo variance ({MC Var}) and 95% coverage rate of IPW, LR, AIPW and OW estimators for binary outcomes. The scenarios correspond to Figure 3 in the main manuscript.

		Relative efficiency				{Est Var}/{MC Var}				95% Coverage			
	N	IPW	LR	AIPW	OW	IPW	LR	AIPW	OW	IPW	LR	AIPW	OW
		$u = 0.5$, $b_1 = 0.75$, $r = 0.5$, correct specification **(e)**											
	50	1.046	1.217	1.129	1.181	0.905	1.151	0.707	1.066	0.895	0.857	0.829	0.944
τ_{RD}	100	1.248	1.294	1.281	1.305	0.945	1.028	0.855	1.298	0.931	0.939	0.921	0.968
	200	1.365	1.420	1.411	1.367	0.988	1.014	0.966	1.353	0.945	0.947	0.941	0.976
	500	1.329	1.381	1.380	1.328	0.899	0.871	0.897	1.246	0.940	0.934	0.938	0.973
	50	0.910	1.128	0.989	1.069	0.866	1.066	0.634	0.998	0.916	0.914	0.857	0.966
τ_{RR}	100	1.257	1.283	1.272	1.305	0.959	1.022	0.855	1.306	0.938	0.940	0.933	0.976
	200	1.358	1.416	1.408	1.361	0.986	1.012	0.966	1.347	0.946	0.951	0.950	0.981
	500	1.330	1.384	1.383	1.329	0.899	0.871	0.898	1.244	0.940	0.936	0.940	0.974
	50	1.009	1.191	1.107	1.168	0.912	1.136	0.704	1.089	0.909	0.857	0.857	0.957
τ_{OR}	100	1.246	1.291	1.276	1.305	0.944	1.027	0.851	1.295	0.938	0.946	0.924	0.973
	200	1.368	1.425	1.416	1.371	0.988	1.015	0.966	1.353	0.945	0.948	0.944	0.979
	500	1.330	1.383	1.381	1.329	0.900	0.871	0.898	1.246	0.942	0.935	0.940	0.974
		$u = 0.5$, $b_1 = 0$, $r = 0.7$, correct specification **(f)**											
	50	0.619	1.379	1.328	0.882	0.871	18.187	0.560	0.803	0.848	0.917	0.836	0.901
τ_{RD}	100	0.902	0.999	0.956	1.026	0.850	0.971	0.760	1.134	0.898	0.949	0.905	0.951
	200	1.017	1.047	1.033	1.081	0.849	0.898	0.808	1.122	0.920	0.935	0.913	0.960
	500	1.165	1.180	1.173	1.189	0.981	1.007	0.972	1.281	0.945	0.948	0.944	0.973
	50	0.447	1.547	1.472	0.791	0.806	10.114	0.546	0.702	0.877	0.911	0.859	0.935
τ_{RR}	100	0.872	0.987	0.938	1.025	0.843	0.963	0.757	1.136	0.916	0.954	0.922	0.961
	200	1.017	1.052	1.038	1.085	0.843	0.893	0.804	1.112	0.928	0.941	0.920	0.963
	500	1.166	1.180	1.174	1.190	0.977	1.002	0.968	1.274	0.952	0.952	0.949	0.974
	50	0.489	1.512	1.450	0.816	0.881	5.454	0.545	0.728	0.892	0.915	0.842	0.928
τ_{OR}	100	0.888	0.996	0.949	1.026	0.848	0.972	0.759	1.134	0.908	0.956	0.914	0.958
	200	1.015	1.046	1.032	1.081	0.848	0.897	0.807	1.120	0.929	0.941	0.919	0.962
	500	1.166	1.181	1.174	1.189	0.981	1.007	0.972	1.280	0.946	0.951	0.946	0.973
		$u = 0.2$, $b_1 = 0$, $r = 0.5$, correct specification **(g)**											
	50	0.755	0.806	0.758	0.807	0.738	1.093	0.689	0.863	0.887	0.915	0.851	0.917
τ_{RD}	100	0.904	0.968	0.952	0.938	0.869	1.485	0.863	1.008	0.916	0.965	0.920	0.933
	200	1.103	1.129	1.120	1.114	0.925	1.296	0.918	1.048	0.938	0.973	0.933	0.955
	500	1.103	1.108	1.108	1.102	0.988	1.256	0.979	1.123	0.949	0.971	0.948	0.960
	50	0.642	0.010	0.001	0.671	0.868	0.017	0.002	1.034	0.914	0.957	0.900	0.973
τ_{RR}	100	0.908	1.028	1.004	0.933	0.860	1.532	0.856	0.997	0.925	0.977	0.939	0.952
	200	1.102	1.147	1.137	1.110	0.899	1.283	0.895	1.017	0.946	0.978	0.944	0.962
	500	1.097	1.104	1.104	1.096	0.983	1.253	0.973	1.116	0.949	0.977	0.949	0.964
	50	0.649	0.020	0.003	0.698	0.861	0.033	0.003	1.030	0.906	0.957	0.900	0.960
τ_{OR}	100	0.906	1.009	0.987	0.934	0.863	1.522	0.858	1.002	0.923	0.974	0.930	0.949
	200	1.103	1.142	1.133	1.112	0.907	1.289	0.903	1.028	0.943	0.976	0.938	0.960
	500	1.099	1.105	1.106	1.098	0.985	1.255	0.975	1.118	0.949	0.976	0.948	0.962
		$u = 0.1$, $b_1 = 0$, $r = 0.5$, correct specification **(h)**											
	50	0.995	0.800	0.785	1.032	0.238	0.255	0.193	0.277	0.888	0.440	0.417	0.912
τ_{RD}	100	0.892	0.881	0.852	0.939	1.064	2.224	0.996	1.194	0.922	0.980	0.947	0.940
	200	1.038	1.056	1.044	1.054	0.958	1.878	0.948	1.042	0.938	0.991	0.942	0.947
	500	1.076	1.101	1.100	1.078	0.985	1.577	0.989	1.068	0.949	0.988	0.947	0.954
	50	0.570	0.001	0.000	1.057	0.608	0.001	0.000	1.201	0.939	0.375	1.000	0.991
τ_{RR}	100	0.868	0.979	0.940	0.893	1.089	2.348	1.024	1.232	0.944	0.994	0.952	0.972
	200	1.052	1.132	1.115	1.065	0.938	1.910	0.940	1.019	0.949	0.994	0.948	0.957
	500	1.073	1.101	1.098	1.074	0.976	1.565	0.975	1.058	0.951	0.990	0.951	0.960
	50	0.610	0.002	0.000	1.078	0.685	0.002	0.000	1.335	0.928	0.375	1.000	0.985
τ_{OR}	100	0.872	0.960	0.923	0.901	1.085	2.329	1.018	1.226	0.938	0.993	0.948	0.965
	200	1.050	1.121	1.105	1.063	0.941	1.909	0.941	1.024	0.948	0.993	0.945	0.954
	500	1.074	1.101	1.098	1.075	0.977	1.568	0.977	1.060	0.951	0.990	0.950	0.958

TABLE 3 Number of times that the logistic regression fails to converge given different outcome prevalence $u \in \{0.5, 0.3, 0.2, 0.1\}$ and sample sizes $N \in [50, 200]$.

N	$u = 0.5$	$u = 0.3$	$u = 0.2$	$u = 0.1$
50	1649	1802	1905	1975
60	1025	1320	1699	1947
70	525	823	1245	1829
80	207	433	834	1659
90	84	194	527	1393
100	34	89	307	1199
110	5	41	159	941
120	5	20	88	684
130	0	3	44	498
140	0	0	17	331
150	0	1	10	251
160	0	0	11	176
170	0	0	2	117
180	0	0	0	85
190	0	0	0	45
200	0	0	0	38