

Web Appendix for  
 “Generalizing trial evidence to target populations in non-nested designs:  
 Applications to AIDS clinical trials”

Fan Li, Ashley L. Buchanan and Stephen R. Cole

## 1 Efficient Influence Function

The verification of the efficient influence function follows the general approach outlined in Bickel et al. (1993), and includes two main steps: characterization of the tangent space and verification of pathwise differentiability in the target parameter. We first write out the observed data density as

$$f(d, s, x, y, z) = \{e f_1(y|z)\}^x \{(1-e) f_0(y|z)\}^{1-x} w(z)]^{ds} [\{1-w(z)\} \pi_0]^{d(1-s)} [f(z)]^d,$$

where  $e = \mathbb{P}(X = 1|S = 1)$  is the treatment propensity score,  $w(z)$  is the sampling score,  $f_1(y|z) = dF_1(y|z)$  and  $f_0(y|z) = dF_0(y|z)$  are densities of  $Y^1$  and  $Y^0$  given  $Z = z$ ,  $\pi_0 = \mathbb{P}(D = 1|S = 0) \leq 1$  is the inclusion probability for the cohort sample. Now consider a regular parametric submodel for the joint distribution of  $(D, S, X, Y, Z)$

$$\{e f_1(y|z; \theta)\}^x \{(1-e) f_0(y|z; \theta)\}^{1-x} w(z; \theta)]^{ds} [\{1-w(z; \theta)\} \pi_0]^{d(1-s)} [f(z; \theta)]^d$$

which equals  $f(d, s, x, y, z)$  when  $\theta = \theta_0$ , the true parameter value. The score function corresponding to the observed data density can be written as

$$H(d, s, x, y, z; \theta_0) = H_y(d, s, x, y, z) + H_s(d, s, z) + H_x(x),$$

where

$$H_y(d, s, x, y, z) = dsx \{ \partial \log f_1(y|z; \theta_0) / \partial \theta \} + ds(1-x) \{ \partial \log f_0(y|z; \theta_0) / \partial \theta \},$$

$$H_s(d, s, z) = \frac{d\{s - w(z; \theta_0)\}}{w(z; \theta_0) \{1 - w(z; \theta_0)\}} \frac{\partial w(z; \theta_0)}{\partial \theta}, \quad H_z(d, z) = d \{ \partial \log f(z; \theta_0) / \partial \theta \}.$$

Define  $\mathbb{G}^2(F)$  as the usual Hilbert space of mean-zero and square-integrable measurable functions with respect to distribution  $F$ . We can represent the tangent space by  $\mathbb{H} = \mathbb{H}_y + \mathbb{H}_s + \mathbb{H}_z$ , where

$$\mathbb{H}_y = \{ dsx l_1(y, z) + ds(1-x) l_0(y, z) \mid l_1(y, z) \in \mathbb{G}^2(F_1), l_0(y, z) \in \mathbb{G}^2(F_0) \}$$

$$\mathbb{H}_s = \{ dl_s(s, z) \mid l_s(s, z) \in \mathbb{G}^2(F_{S|Z}) \}, \quad \mathbb{H}_z = \{ dl_z(z) \mid l_z(z) \in \mathbb{G}^2(F_Z) \}$$

The moment condition for  $\Delta$  based on the potential outcomes is  $U(Y^1, Y^0; \Delta) = U(\Delta) = (Y^1 - Y^0) - \Delta$ , where  $\mathbb{E}[U(\Delta(\theta_0))] = 0$ . Under certain regularity conditions, we can interchange differentiation and integration and have  $0 = \frac{\partial}{\partial \theta} \mathbb{E}[U(\Delta(\theta_0))] = \frac{\partial}{\partial \Delta} \mathbb{E}[U(\Delta)] \frac{\partial}{\partial \theta} \Delta(\theta_0) + \frac{\partial}{\partial \theta} \mathbb{E}[U(\Delta)]$ . Since  $\frac{\partial}{\partial \Delta} \mathbb{E}[U(\Delta)] = -1$ , we obtain

$$\begin{aligned} \frac{\partial}{\partial \theta} \Delta(\theta_0) = \frac{\partial}{\partial \theta} \mathbb{E}[U(\Delta)] &= \int (y - \mu_1) f_1(y|z; \theta_0) f(z; \theta_0) [l_1(y, z) + l_z(z)] dy dz \\ &\quad - \int (y - \mu_0) f_0(y|z; \theta_0) f(z; \theta_0) [l_0(y, z) + l_z(z)] dy dz \\ &= \mathbb{E}[(Y^1 - \mu_1) l_1(Y^1, Z)] - \mathbb{E}[(Y^0 - \mu_0) l_0(Y^0, Z)] \\ &\quad + \mathbb{E}[(m_1(Z) - m_0(Z) - \Delta) l_z(Z)]. \end{aligned}$$

The mapping  $\Delta(\theta)$  is pathwise differentiable if there exists a function  $\mathcal{I}_{\text{eff}}(D, S, X, Y, Z) \in \mathbb{H}$  such that  $\frac{\partial}{\partial \theta} \Delta(\theta_0) = \mathbb{E}[\mathcal{I}_{\text{eff}}(D, S, X, Y, Z) H(D, S, X, Y, Z; \theta_0)]$  for all regular parametric submodels. Indeed, we can verify that following function

$$\begin{aligned} \mathcal{I}_{\text{eff}}(D, S, X, Y, Z) &= \frac{DSX}{we} (Y - m_1(Z)) - \frac{DS(1-X)}{w(1-e)} (Y - m_0(Z)) \\ &\quad + D\{S + \pi_0^{-1}(1-S)\} (m_1(Z) - m_0(Z) - \Delta), \end{aligned}$$

is an element in  $\mathbb{H}$  and satisfies the above condition, because

$$\begin{aligned} \mathbb{E} \left[ \frac{DSX}{we} (Y - m_1(Z)) H(D, S, X, Y, Z; \theta_0) \right] &= \mathbb{E}[(Y^1 - m_1(Z)) l_1(Y^1, Z)] \\ \mathbb{E} \left[ \frac{DS(1-X)}{w(1-e)} (Y - m_0(Z)) H(D, S, X, Y, Z; \theta_0) \right] &= \mathbb{E}[(Y^0 - m_0(Z)) l_0(Y^0, Z)] \\ \mathbb{E}[D\{S + \pi_0^{-1}(1-S)\} (m_1(Z) - m_0(Z) - \Delta) H(D, S, X, Y, Z; \theta_0)] \\ &= \mathbb{E}[(m_1(Z) - m_0(Z) - \Delta) l_z(Z)], \end{aligned}$$

and  $\mathbb{E}[(m_1(Z) - \mu_1) l_1(Y^1, Z)] = \mathbb{E}[(m_0(Z) - \mu_0) l_0(Y^0, Z)] = 0$ . This efficient influence function then motivates the first DR estimator, as mentioned in the main text.

## 2 Proof of Double Robustness

### 2.1 The DR1 Estimator

Recall the treatment propensity score model is expected to be correctly modeled because the data arise from a randomized trial. We will consider the probability limit of  $\widehat{\Delta}_{\text{DR1}}$  when either the outcome model or the sampling score model is correctly specified, but not necessarily both. Let the (weighted) maximum likelihood estimators for model parameters be  $\widehat{\alpha}_1$ ,  $\widehat{\alpha}_0$  and  $\widehat{\gamma}$ . Suppose that the outcome model  $m_1$  and  $m_0$  are correct but the sampling score model may be subject to misspecification. In this case, the probability limit of  $\widehat{\gamma}$  is  $\gamma^*$ , which may be different from the parameter value in the true sampling score model. In fact, according to the results of White (1982),  $\gamma^*$  is the least favorable value that minimizes the Kullback-Leibler distance between the probability distribution based on the postulated sampling score model and the true sampling score model. We write  $w^* = w(Z; \gamma^*)$ , which does not necessarily equal to  $w(Z; \gamma)$ , where  $\gamma$  is the true parameter value. Of note, our asymptotic analysis requires the target population size  $N$  to approach infinity, and as  $N \rightarrow \infty$ , the inclusion probability approaches a constant such that  $\Pi_0 = m/(N - n) \rightarrow \pi_0 = \mathbb{P}(D = 1|S = 0)$ . Because  $\Pi_1 = 1 = \pi_1 = \mathbb{P}(D = 1|S = 1)$ , we observe that

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \left\{ \frac{D_i S_i X_i}{\widehat{w}_i \widehat{e}_i} (Y_i - \widehat{m}_{1i}) + c_i \widehat{m}_{1i} \right\} \\ & \xrightarrow{p} \mathbb{E} \left[ \frac{DSX}{w^* e} (Y^1 - m_1) + D \{S + \pi_0^{-1}(1 - S)\} m_1 \right] \\ & = \mathbb{E} \left[ \frac{w}{w^*} \{E[Y^1|Z] - m_1\} \right] + \mathbb{E}[S m_1] + \mathbb{P}(S = 0) \pi_0^{-1} \mathbb{E}[D m_1 | S = 0] \\ & = \mathbb{E}[S m_1] + \mathbb{E}[(1 - S) m_1] = \mathbb{E}[m_1] = \mu_1. \end{aligned}$$

Similarly, we can show that

$$\frac{1}{N} \sum_{i=1}^N \left\{ \frac{D_i S_i (1 - X_i)}{\widehat{w}_i \widehat{e}_i} (Y_i - \widehat{m}_{0i}) + c_i \widehat{m}_{0i} \right\} \xrightarrow{p} \mu_0.$$

This proves  $\widehat{\Delta}_{\text{DR1}} \xrightarrow{p} \Delta$ .

Suppose that the sampling score model is correct but the outcome model may be misspecified. In this case,  $\widehat{\gamma}$  converges in probability to  $\gamma$ , the parameter value in the true sampling score model. However,

the probability limits of  $\widehat{\alpha}_1, \widehat{\alpha}_0$  are  $\alpha_1^*, \alpha_0^*$ , which could be different from the parameter values in the true outcome models. Denote  $m_1^* = m_1(Z; \alpha_1^*)$  and  $m_0^* = m_0(Z; \alpha_0^*)$ . We can write

$$\widehat{\Delta}_{\text{DR1}} = \widehat{\Delta}_{\text{IPSW1}} + \frac{1}{N} \sum_{i=1}^N \left\{ c_i - \frac{D_i S_i X_i}{\widehat{w}_i \widehat{e}_i} \right\} \widehat{m}_{1i} - \frac{1}{N} \sum_{i=1}^N \left\{ c_i - \frac{D_i S_i (1 - X_i)}{\widehat{w}_i (1 - \widehat{e}_i)} \right\} \widehat{m}_{0i}.$$

It is straightforward to show that  $\widehat{\Delta}_{\text{IPSW1}} \xrightarrow{p} \Delta$  when the sampling score model is correct. Further from the Weak Law of Large Numbers and the Law of Iterated Expectations, we obtain the expression

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \left\{ c_i - \frac{D_i S_i X_i}{\widehat{w}_i \widehat{e}_i} \right\} \widehat{m}_{1i} \\ & \xrightarrow{p} \mathbb{E} \left[ D \{ S + \pi_0^{-1} (1 - S) \} m_1^* - \frac{DSX}{we} m_1^* \right] \\ & = \mathbb{E}[S m_1^*] + P(S = 0) \pi_0^{-1} \mathbb{E}[D m_1^* | S = 0] - \mathbb{E} \left[ \frac{\mathbb{E}[SX|Z]}{we} m_1^* \right] \\ & = \mathbb{E}[m_1^*] - \mathbb{E}[m_1^*] = 0. \end{aligned}$$

Similarly

$$\frac{1}{N} \sum_{i=1}^N \left\{ c_i - \frac{D_i S_i (1 - X_i)}{\widehat{w}_i (1 - \widehat{e}_i)} \right\} \widehat{m}_{0i} \xrightarrow{p} \mathbb{E} \left[ D \{ S + \pi_0^{-1} (1 - S) \} m_0^* - \frac{DS(1 - X)}{w(1 - e)} m_0^* \right] = 0,$$

this gives  $\widehat{\Delta}_{\text{DR1}} \xrightarrow{p} \Delta$ ; therefore the double robustness property holds for the DR1 estimator.

## 2.2 The DR2 Estimator

To show the double robustness property of the DR2 estimator, suppose that the outcome models  $m_1$  and  $m_0$  are correct but the sampling score model may be misspecified. Notice that we can write the DR2 estimator as

$$\begin{aligned} \widehat{\Delta}_{\text{DR2}} &= \frac{N^{-1} \sum_{i=1}^N D_i S_i X_i (Y_i - \widehat{m}_{1i}) / \widehat{w}_i \widehat{e}_i}{N^{-1} \sum_{i=1}^N D_i S_i X_i / \widehat{w}_i \widehat{e}_i} - \frac{N^{-1} \sum_{i=1}^N D_i S_i (1 - X_i) (Y_i - \widehat{m}_{0i}) / \widehat{w}_i (1 - \widehat{e}_i)}{N^{-1} \sum_{i=1}^N D_i S_i (1 - X_i) / \widehat{w}_i (1 - \widehat{e}_i)} \\ & \quad + \frac{1}{N} \sum_{i=1}^N c_i (\widehat{m}_{1i} - \widehat{m}_{0i}). \end{aligned}$$

We observe

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i X_i}{\widehat{w}_i \widehat{e}_i} (Y_i - \widehat{m}_{1i}) &\xrightarrow{p} \mathbb{E} \left[ \frac{DSX}{w^* e} (Y^1 - m_1) \right] = \mathbb{E} \left[ \frac{w}{w^*} \{E[Y^1|Z] - m_1\} \right] = 0 \\ \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i (1 - X_i)}{\widehat{w}_i (1 - \widehat{e}_i)} (Y_i - \widehat{m}_{0i}) &\xrightarrow{p} \mathbb{E} \left[ \frac{DS(1-X)}{w^* (1-e)} (Y^0 - m_0) \right] = \mathbb{E} \left[ \frac{w}{w^*} \{E[Y^0|Z] - m_0\} \right] = 0. \end{aligned}$$

Further, similar to the results in Appendix 2.1

$$\frac{1}{N} \sum_{i=1}^N c_i (\widehat{m}_{1i} - \widehat{m}_{0i}) \xrightarrow{p} \mathbb{E}[D\{S + \pi_0^{-1}(1-S)\}(m_1 - m_0)] = \mathbb{E}[m_1] - \mathbb{E}[m_0] = \Delta$$

and hence  $\widehat{\Delta}_{\text{DR2}} \xrightarrow{p} \Delta$ .

Suppose that the sampling score model is correct but the outcome model may be misspecified. In this case, we can show

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i X_i}{\widehat{w}_i \widehat{e}_i} (Y_i - \widehat{m}_{1i}) &\xrightarrow{p} \mathbb{E} \left[ \frac{DSX}{we} (Y^1 - m_1^*) \right] = \mu_1 - \mathbb{E}[m_1^*] \\ \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i X_i}{\widehat{w}_i \widehat{e}_i} &\xrightarrow{p} \mathbb{E} \left[ \frac{DSX}{we} \right] = 1, \end{aligned}$$

which then gives

$$\frac{N^{-1} \sum_{i=1}^N D_i S_i X_i (Y_i - \widehat{m}_{1i}) / \widehat{w}_i \widehat{e}_i}{N^{-1} \sum_{i=1}^N D_i S_i X_i / \widehat{w}_i \widehat{e}_i} \xrightarrow{p} \mu_1 - \mathbb{E}[m_1^*].$$

Similarly, we must have

$$\frac{N^{-1} \sum_{i=1}^N D_i S_i X_i (Y_i - \widehat{m}_{0i}) / \widehat{w}_i (1 - \widehat{e}_i)}{N^{-1} \sum_{i=1}^N D_i S_i X_i / \widehat{w}_i (1 - \widehat{e}_i)} \xrightarrow{p} \mu_0 - \mathbb{E}[m_0^*].$$

Because

$$\frac{1}{N} \sum_{i=1}^N c_i (\widehat{m}_{1i} - \widehat{m}_{0i}) \xrightarrow{p} \mathbb{E}[D\{S + \pi_0^{-1}(1-S)\}(m_1^* - m_0^*)] = \mathbb{E}[m_1^*] - \mathbb{E}[m_0^*],$$

we have

$$\widehat{\Delta}_{\text{DR2}} \xrightarrow{p} (\mu_1 - \mathbb{E}[m_1^*]) - (\mu_0 - \mathbb{E}[m_0^*]) + \mathbb{E}[m_1^*] - \mathbb{E}[m_0^*] = \Delta.$$

Therefore  $\widehat{\Delta}_{\text{DR2}}$  is doubly-robust consistent for estimating  $\Delta$ .

### 3 Large-Sample Variance and Efficiency Considerations (Proof of Proposition 1)

We show that when the treatment propensity score is estimated from trial data, all five estimators (IPSW1, IPSW2, REG, DR1 and DR2) are asymptotically at least as efficient as when the true propensity score is used. Let  $g_i(\gamma) = \partial w(Z_i; \gamma) / \partial \gamma$ , recall that the weighted estimating equations for  $\hat{\gamma}$  is

$$\sum_{i=1}^N \psi_{\gamma}(D_i, S_i, Z_i; \gamma) = \sum_{i=1}^N D_i \Pi_{S_i}^{-1} \frac{S_i - w_i}{w_i(1 - w_i)} g_i(\gamma) = 0.$$

Because the model for the sampling score may not necessarily be correctly specified, we define  $\gamma^*$  as the probability limit of  $\hat{\gamma}$ . According to White (1982),  $\gamma^*$  is the least favorable value that minimizes the Kullback-Leibler distance between the distributions implied from the assumed sampling score and that corresponding to the true data generating process. Importantly, it satisfies  $\mathbb{E}[\psi_{\gamma}(D_i, S_i, Z_i; \gamma^*)] = 0$ . We similar define  $\alpha_1^*$  and  $\alpha_0^*$  as the probability limits of  $\hat{\alpha}_0$  and  $\hat{\alpha}_1$  in the outcome models. Further, let  $d_i(\beta) = \partial e(W_i; \beta) / \partial \beta$ , we have the estimating equation for  $\hat{\beta}$  as

$$\sum_{i=1}^N \psi_{\beta}(D_i, S_i, X_i, W_i; \beta) = \sum_{i=1}^N \frac{D_i S_i (X_i - e_i)}{e_i(1 - e_i)} d_i(\beta) = 0.$$

The information matrix of  $\hat{\beta}$  is  $E_{\beta\beta}^{-1}$ , where  $E_{\beta\beta} = \mathbb{E} [D_i S_i \{d_i(\beta) d_i^T(\beta)\} / (e_i(1 - e_i))]$  is a positive definite matrix. Finally, we note that  $\partial e^{-1} / \partial \beta = -d_i(\beta) / e_i^2$  and  $\partial(1 - e)^{-1} / \partial \beta = d_i(\beta) / (1 - e_i)^2$ , which will be used repeatedly in the following presentations.

#### 3.1 The IPSW1 Estimator

First consider the case when the true propensity score is used in the IPSW1 estimator. Let  $\tilde{\theta} = (\tilde{\mu}_1, \tilde{\mu}_0, \tilde{\gamma}^T)^T$ ,  $\theta^* = (\mu_1, \mu_0, \gamma^{*T})^T$  and note that  $\tilde{\theta}$  is the solution to  $\sum_{i=1}^N \Psi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \tilde{\theta}) = 0$ , where

$$\Psi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta) = \begin{pmatrix} D_i S_i X_i Y_i / (w_i e_i) - \mu_1 \\ D_i S_i (1 - X_i) Y_i / (w_i (1 - e_i)) - \mu_0 \\ \psi_{\gamma}(D_i, S_i, Z_i; \gamma) \end{pmatrix}.$$

Define the matrices

$$\begin{aligned} A(\theta^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \theta^T} \Psi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \right\}, \\ B(\theta^*) &= \mathbb{V} \{ \Psi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \}, \end{aligned}$$

where the expectation and covariance operators are defined with respect to the marginal distribution of the observed data vector. Under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\tilde{\theta} - \theta^*)$  converges in distribution to  $N(0, \Omega_{\theta^*})$ , where

$$\Omega_{\theta^*} = A(\theta^*)^{-1} B(\theta^*) A(\theta^*)^{-T}.$$

Thus, the asymptotic variance of  $\tilde{\Delta}_{\text{IPSW1}}$ , when true  $e_i$  is used, can be expressed as  $\sigma_{\text{IPSW1}}^2 = \lambda^T \Omega_{\theta^*} \lambda$ , where  $\lambda = (1, -1, 0_{1 \times p})^T$ .

Next consider the case when the propensity score is estimated from the trial data, and let  $\hat{\varpi} = (\hat{\mu}_1, \hat{\mu}_0, \hat{\gamma}^T, \hat{\beta}^T)^T$  and  $\varpi^* = (\mu_1, \mu_0, \gamma^{*T}, \beta^T)^T$  and note that  $\hat{\varpi}$  is the solution to the  $(2 + p + q) \times 1$  vector of estimating equations  $\sum_{i=1}^N \Phi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) = 0$ , where

$$\Phi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \begin{pmatrix} \Psi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) \\ \psi_{\beta}(D_i, S_i, X_i, W_i; \beta) \end{pmatrix}.$$

Here,  $\Psi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \Psi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta)$  whenever  $\beta$  is chosen such that  $e(W_i; \beta) = e_i$ , the true propensity score. Define matrices

$$\begin{aligned} C(\varpi^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \right\} \\ D(\varpi^*) &= \mathbb{V} \{ \Phi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \}. \end{aligned}$$

Then under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\hat{\varpi} - \varpi^*)$  converges in distribution to  $N(0, \Omega_{\varpi^*})$ , where

$$\Omega_{\varpi^*} = C(\varpi^*)^{-1} D(\varpi^*) C(\varpi^*)^{-T}.$$

Observe that  $\mathbb{E}[\psi_{\beta}(D_i, S_i, X_i, W_i; \beta) \psi_{\gamma}^T(D_i, S_i, Z_i; \gamma^*)] = 0_{q \times p}$ , and further define

$$\begin{aligned} G_1 &= \mathbb{E} [D_i S_i Y_i^1 d_i(\beta) / (w_i e_i)] \\ G_2 &= \mathbb{E} [-D_i S_i Y_i^0 d_i(\beta) / (w_i (1 - e_i))], \end{aligned}$$

and let  $G = (G_1, G_2, 0_{q \times p})$ . Then using block matrix notation, we can write

$$C(\varpi^*) = \begin{pmatrix} A(\theta^*) & -G^T \\ 0_{q \times (2+p)} & -E_{\beta\beta} \end{pmatrix} \quad \text{and} \quad D(\varpi^*) = \begin{pmatrix} B(\theta^*) & G^T \\ G & E_{\beta\beta} \end{pmatrix}.$$

Further,

$$C(\varpi^*)^{-1} = \begin{pmatrix} A(\theta^*)^{-1} & -A(\theta^*)^{-1}G^T E_{\beta\beta}^{-1} \\ 0_{q \times (2+p)} & -E_{\beta\beta}^{-1} \end{pmatrix}$$

It follows that the upper-left block of  $\Omega_{\varpi^*}$  can be written as

$$\Omega_{\theta^*} - A(\theta^*)^{-1}G^T E_{\beta\beta}^{-1}GA(\theta^*)^{-T}.$$

Therefore the asymptotic variance of  $\widehat{\Delta}_{\text{IPSW1}}$ , when the treatment propensity score is estimated from trial data, is

$$\tau_{\text{IPSW1}}^2 = \sigma_{\text{IPSW1}}^2 - \{\lambda^T A(\theta^*)G^T\}E_{\beta\beta}^{-1}\{\lambda^T A(\theta^*)G^T\}^T. \quad (3.1)$$

Because  $E_{\beta\beta}^{-1}$  is positive definite, we must have  $\tau_{\text{IPSW1}}^2 \leq \sigma_{\text{IPSW1}}^2$ .

To estimate the variance of  $\widehat{\Delta}_{\text{IPSW1}}$ , we can use the consistent sandwich estimator. Define the following matrices

$$C(\widehat{\varpi}) = \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi})$$

$$D(\widehat{\varpi}) = \frac{1}{N} \sum_{i=1}^N \Phi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \Phi_{\Delta_{\text{IPSW1}}}^T(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi})$$

Therefore a consistent estimator for the variance of  $\widehat{\Delta}_{\text{IPSW1}}$  is  $N^{-1}\eta^T C(\widehat{\varpi})^{-1}D(\widehat{\varpi})C(\widehat{\varpi})^{-T}\eta$ , where  $\eta = (\lambda^T, 0_{1 \times q})^T$ .

### 3.2 The IPSW2 Estimator

First consider the case when the true propensity score is used in the IPSW2 estimator. Let  $\tilde{\theta} = (\tilde{\mu}_1, \tilde{\mu}_0, \tilde{\gamma}^T)^T$ ,  $\theta^* = (\mu_1, \mu_0, \gamma^{*T})^T$  and note that  $\tilde{\theta}$  is the solution to  $\sum_{i=1}^N \Psi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \tilde{\theta}) =$



0, where

$$\Psi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta) = \begin{pmatrix} \{D_i S_i X_i (Y_i - \mu_1)\} / (w_i e_i) \\ \{D_i S_i (1 - X_i) (Y_i - \mu_0)\} / (w_i (1 - e_i)) \\ \psi_\gamma(D_i, S_i, Z_i; \gamma) \end{pmatrix}.$$

Define the matrices

$$\begin{aligned} A(\theta^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \theta^T} \Psi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \right\}, \\ B(\theta^*) &= \mathbb{V} \{ \Psi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \}. \end{aligned}$$

Under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\tilde{\theta} - \theta^*)$  converges in distribution to  $N(0, \Omega_{\theta^*})$ , where

$$\Omega_{\theta^*} = A(\theta^*)^{-1} B(\theta^*) A(\theta^*)^{-T}.$$

Thus, the asymptotic variance of  $\tilde{\Delta}_{\text{IPSW2}}$ , when true  $e_i$  is used, can be expressed as  $\sigma_{\text{IPSW2}}^2 = \lambda^T \Omega_{\theta^*} \tau$ , where  $\lambda = (1, -1, 0_{1 \times p})^T$ .

Next consider the case when the propensity score is estimated from the trial data, and let  $\hat{\varpi} = (\hat{\mu}_1, \hat{\mu}_0, \hat{\gamma}^T, \hat{\beta}^T)^T$  and  $\varpi^* = (\mu_1, \mu_0, \gamma^{*T}, \beta^T)^T$  and note that  $\hat{\varpi}$  is the solution to the  $(2 + p + q) \times 1$  vector of estimating equations  $\sum_{i=1}^N \Phi_{\Delta_{\text{IPSW1}}}(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) = 0$ , where

$$\Phi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \begin{pmatrix} \Psi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) \\ \psi_\beta(D_i, S_i, X_i, W_i; \beta) \end{pmatrix}.$$

Here,  $\Psi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \Psi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta)$  whenever  $\beta$  is chosen such that  $e(W_i; \beta) = e_i$ , the true propensity score. Define matrices

$$\begin{aligned} C(\varpi^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \right\} \\ D(\varpi^*) &= \mathbb{V} \{ \Phi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \}. \end{aligned}$$

Then under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\hat{\varpi} - \varpi^*)$  converges in distribution to  $N(0, \Omega_{\varpi^*})$ , where

$$\Omega_{\varpi^*} = C(\varpi^*)^{-1} D(\varpi^*) C(\varpi^*)^{-T}.$$

Further define

$$\begin{aligned} G_1 &= \mathbb{E} [D_i S_i (Y_i^1 - \mu_1) d_i(\beta) / (w_i e_i)] \\ G_2 &= \mathbb{E} [-D_i S_i (Y_i^0 - \mu_0) d_i(\beta) / (w_i (1 - e_i))], \end{aligned}$$

and let  $G = (G_1, G_2, 0_{q \times p})$ . Then using block matrix notation, we can write

$$C(\varpi^*) = \begin{pmatrix} A(\theta^*) & -G^T \\ 0_{q \times (2+p)} & -E_{\beta\beta} \end{pmatrix} \quad \text{and} \quad D(\varpi^*) = \begin{pmatrix} B(\theta^*) & G^T \\ G & E_{\beta\beta} \end{pmatrix}.$$

Similar to the results in Appendix 3.1, the upper-left block of  $\Omega_{\varpi^*}$  can be written as  $\Omega_{\theta^*} - A(\theta^*)^{-1} G^T E_{\beta\beta}^{-1} G A(\theta^*)^{-T}$ .

Therefore the asymptotic variance of  $\hat{\Delta}_{\text{IPSW2}}$ , when  $e_i$  is estimated from trial data, is

$$\tau_{\text{IPSW2}}^2 = \sigma_{\text{IPSW2}}^2 - \{\lambda^T A(\theta^*) G^T\} E_{\beta\beta}^{-1} \{\lambda^T A(\theta^*) G^T\}^T. \quad (3.2)$$

Because  $E_{\beta\beta}^{-1}$  is positive definite, we must have  $\tau_{\text{IPSW2}}^2 \leq \sigma_{\text{IPSW2}}^2$ .

To estimate the variance of  $\hat{\Delta}_{\text{IPSW2}}$ , we can use the consistent sandwich estimator. Define the following matrices

$$\begin{aligned} C(\hat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) \\ D(\hat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \Phi_{\Delta_{\text{IPSW2}}}(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) \Phi_{\Delta_{\text{IPSW2}}}^T(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) \end{aligned}$$

Therefore a consistent estimator for the variance of  $\hat{\Delta}_{\text{IPSW2}}$  is  $N^{-1} \eta^T C(\hat{\varpi})^{-1} D(\hat{\varpi}) C(\hat{\varpi})^{-T} \eta$ , where  $\eta = (\lambda^T, 0_{1 \times q})^T$ .

### 3.3 The REG Estimator

First consider the case when the true propensity score is used in the outcome regression estimator. Let

$\tilde{\theta} = (\tilde{\Delta}, \tilde{\alpha}_1^T, \tilde{\alpha}_0^T)^T$ ,  $\theta = (\Delta, \alpha_0^{*T}, \alpha_1^{*T})^T$  and note that  $\tilde{\theta}$  is the solution to  $\sum_{i=1}^N \Psi_{\Delta_{\text{REG}}}(Y_i, D_i, S_i, X_i, Z_i; \tilde{\theta}) =$

0

$$\Psi_{\Delta_{\text{REG}}}(Y_i, D_i, S_i, X_i, Z_i; \theta) = \begin{pmatrix} c_i m_{1i} - c_i m_{0i} - \Delta \\ D_i S_i X_i \psi_{\alpha_1}(Y_i, Z_i; \alpha_1) / e_i \\ D_i S_i (1 - X_i) \psi_{\alpha_0}(Y_i, Z_i; \alpha_0) / (1 - e_i) \end{pmatrix}.$$

Usually in a randomized trial, the true propensity score  $e_i = 1/2$  for all participants. In this case, the above estimating functions are equivalent to

$$\Psi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \theta) = \begin{pmatrix} c_i m_{1i} - c_i m_{0i} - \Delta \\ D_i S_i X_i \psi_{\alpha_1}(Y_i, Z_i; \alpha_1) \\ D_i S_i (1 - X_i) \psi_{\alpha_0}(Y_i, Z_i; \alpha_0) \end{pmatrix},$$

which does not involve the propensity scores. We include the propensity score in the estimating functions for technical convenience (and better connect with the case with the estimated propensity scores).

Define the matrices

$$\begin{aligned} A(\theta^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \theta^T} \Psi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \right\}, \\ B(\theta^*) &= \mathbb{V} \{ \Psi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \}. \end{aligned}$$

Under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\tilde{\theta} - \theta^*)$  converges in distribution to  $N(0, \Omega_{\theta^*})$ , where

$$\Omega_{\theta^*} = A(\theta^*)^{-1} B(\theta^*) A(\theta^*)^{-T}.$$

Thus, the asymptotic variance of  $\tilde{\Delta}_{\text{REG}}$ , when true  $e_i$  is used, can be expressed as  $\sigma_{\text{REG}}^2 = \lambda^T \Omega_{\theta^*} \lambda$ , where  $\lambda = (1, 0_{1 \times l_1}, 0_{1 \times l_0})^T$ .

Now consider the case when propensity score is estimated from the trial data, and let  $\hat{\varpi} = (\hat{\Delta}, \hat{\alpha}_1^T, \hat{\alpha}_0^T, \hat{\beta}^T)^T$ ,  $\theta^* = (\Delta, \alpha_0^{*T}, \alpha_1^{*T}, \beta^T)^T$  and note that  $\hat{\theta}$  is the solution to the  $(1 + l_1 + l_0 + q) \times 1$  vector of estimating equation  $\sum_{i=1}^N \Phi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) = 0$ , where

$$\Phi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \begin{pmatrix} \Psi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) \\ \psi_{\beta}(D_i, S_i, X_i, W_i; \beta) \end{pmatrix}.$$

Here,  $\Psi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \Psi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \theta)$  whenever  $\beta$  is chosen such that  $e(W_i; \beta) = e_i$ , the true propensity score. Define matrices

$$\begin{aligned} C(\varpi^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \varpi^T} \Phi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \right\} \\ D(\varpi^*) &= \mathbb{V} \{ \Phi_{\Delta\text{REG}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \}. \end{aligned}$$

Then under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\widehat{\varpi} - \varpi)$  converges in distribution to  $N(0, \Omega_{\varpi^*})$ , where

$$\Omega_{\varpi^*} = C(\varpi^*)^{-1}D(\varpi^*)C(\varpi^*)^{-T}.$$

Notice that  $\mathbb{E}[\psi_{\beta}(D_i, S_i, X_i, W_i; \beta)(c_i m_{1i} - c_i m_{0i} - \Delta)] = 0_{q \times 1}$ , and further define

$$\begin{aligned} G_3 &= \mathbb{E} [D_i S_i d_i(\beta) \psi_{\alpha_1}^T(Y_i, Z_i; \alpha_1^*) / e_i] \\ G_4 &= \mathbb{E} [-D_i S_i d_i(\beta) \psi_{\alpha_0}^T(Y_i, Z_i; \alpha_0^*) / (1 - e_i)], \end{aligned}$$

and let  $G = (0_{q \times 1}, G_3, G_4)$ . Then using block matrix notation, we can write

$$C(\varpi^*) = \begin{pmatrix} A(\theta^*) & -G^T \\ 0_{q \times 1} & -E_{\beta\beta} \end{pmatrix} \quad \text{and} \quad D(\varpi^*) = \begin{pmatrix} B(\theta^*) & G^T \\ G & E_{\beta\beta} \end{pmatrix}.$$

The upper-left block of  $\Omega_{\varpi^*}$  can be written as  $\Omega_{\theta^*} - A(\theta^*)^{-1}G^T E_{\beta\beta}^{-1}GA(\theta^*)^{-T}$ . Therefore the asymptotic variance of  $\widehat{\Delta}_{\text{REG}}$ , when  $e_i$  is estimated from trial data, is

$$\tau_{\text{REG}}^2 = \sigma_{\text{REG}}^2 - \{\lambda^T A(\theta^*)G^T\}E_{\beta\beta}^{-1}\{\lambda^T A(\theta^*)G^T\}^T. \quad (3.3)$$

Because  $E_{\beta\beta}^{-1}$  is positive definite, we must have  $\tau_{\text{REG}}^2 \leq \sigma_{\text{REG}}^2$ .

To estimate the variance of  $\widehat{\Delta}_{\text{REG}}$ , we can use the consistent sandwich estimator. Define the following matrices

$$\begin{aligned} C(\widehat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{REG}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \\ D(\widehat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \Phi_{\Delta_{\text{REG}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \Phi_{\Delta_{\text{REG}}}^T(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \end{aligned}$$

Therefore a consistent estimator for the variance of  $\widehat{\Delta}_{\text{REG}}$  is  $N^{-1}\eta^T C(\widehat{\varpi})^{-1}D(\widehat{\varpi})C(\widehat{\varpi})^{-T}\eta$ , where  $\eta = (\lambda^T, 0_{1 \times q})^T$ .

### 3.4 The DR1 Estimator

We first consider the case when the true treatment propensity score is used in the DR1 estimator. Recall that the true treatment propensity score  $e_i = r$ . Define

$$\begin{aligned}\tilde{\nu}_1 &= \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i X_i}{\widehat{w}_i e_i} (Y_i - \tilde{m}_{1i}) \\ \tilde{\nu}_2 &= \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i (1 - X_i)}{\widehat{w}_i (1 - e_i)} (Y_i - \tilde{m}_{0i}) \\ \tilde{\nu}_3 &= \frac{1}{N} \sum_{i=1}^N c_i (\tilde{m}_{1i} - \tilde{m}_{0i}),\end{aligned}$$

where  $\tilde{m}_{1i}, \tilde{m}_{0i}$  are defined in equation (3.5) in the main text. Then clearly we can write  $\tilde{\Delta}_{\text{DR1}} = \tilde{\nu}_1 - \tilde{\nu}_2 + \tilde{\nu}_3$ . Let  $\tilde{\theta} = (\tilde{\nu}_1, \tilde{\nu}_2, \tilde{\nu}_3, \tilde{\gamma}^T, \tilde{\alpha}_0^T, \tilde{\alpha}_1^T)^T$ , define  $\theta^* = (\nu_1, \nu_2, \nu_3, \gamma^{*T}, \alpha_0^{*T}, \alpha_1^{*T})^T$  as the limiting value of  $\tilde{\theta}$  and note that  $\tilde{\theta}$  is the solution to the estimating equation  $\sum_{i=1}^N \Psi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \tilde{\theta}) = 0$ , where

$$\Psi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta) = \begin{pmatrix} D_i S_i X_i (Y_i - m_{1i}) / (w_i e_i) - \nu_1 \\ D_i S_i (1 - X_i) (Y_i - m_{0i}) / (w_i (1 - e_i)) - \nu_2 \\ c_i m_{1i} - c_i m_{0i} - \nu_3 \\ \psi_\gamma(D_i, S_i, Z_i; \gamma) \\ D_i S_i X_i \psi_{\alpha_1}(Y_i, Z_i; \alpha_1) / e_i \\ D_i S_i (1 - X_i) \psi_{\alpha_0}(Y_i, Z_i; \alpha_0) / (1 - e_i) \end{pmatrix}.$$

Define the matrices

$$\begin{aligned}A(\theta^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \theta^T} \Psi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \right\}, \\ B(\theta^*) &= \mathbb{V} \{ \Psi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \}.\end{aligned}$$

Under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\tilde{\theta} - \theta^*)$  converges in distribution to  $N(0, \Omega_{\theta^*})$ , where

$$\Omega_{\theta^*} = A(\theta^*)^{-1} B(\theta^*) A(\theta^*)^{-T}.$$

Thus, the asymptotic variance of  $\tilde{\Delta}_{\text{DR1}}$ , when true  $e_i$  is used, can be expressed as  $\sigma_{\text{DR1}}^2 = \lambda^T \Omega_{\theta^*} \lambda$ , where  $\lambda = (1, -1, 1, 0_{1 \times p}, 0_{1 \times l_1}, 0_{1 \times l_0})^T$ .

Next consider the case when the treatment propensity score is estimated. Define

$$\begin{aligned}\hat{\nu}_1 &= \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i X_i}{\hat{w}_i \hat{e}_i} (Y_i - \hat{m}_{1i}) \\ \hat{\nu}_2 &= \frac{1}{N} \sum_{i=1}^N \frac{D_i S_i (1 - X_i)}{\hat{w}_i (1 - \hat{e}_i)} (Y_i - \hat{m}_{0i}) \\ \hat{\nu}_3 &= \frac{1}{N} \sum_{i=1}^N c_i (\hat{m}_{1i} - \hat{m}_{0i}),\end{aligned}$$

where  $\hat{m}_{1i}$ ,  $\hat{m}_{0i}$  are defined in equation (3.8) in the main text. Then clearly we can write  $\hat{\Delta}_{\text{DR1}} = \hat{\nu}_1 - \hat{\nu}_2 + \hat{\nu}_3$ . Let  $\hat{\varpi} = (\hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\gamma}^T, \hat{\alpha}_0^T, \hat{\alpha}_1^T, \hat{\beta})^T$  and define  $\varpi^* = (\nu_1, \nu_2, \nu_3, \gamma^{*T}, \alpha_0^{*T}, \alpha_1^{*T}, \beta^T)^T$  as the limiting value of  $\hat{\varpi}$ . Then  $\hat{\varpi}$  is the solution to the  $(3 + p + l_1 + l_0 + q) \times 1$  vector of estimating equation  $\sum_{i=1}^N \Phi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) = 0$ , where

$$\Phi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \begin{pmatrix} \Psi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) \\ \psi_{\beta}(D_i, S_i, X_i, W_i; \beta) \end{pmatrix}.$$

Here,  $\Psi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \Psi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \theta)$  whenever  $\beta$  is chosen such that  $e(W_i; \beta) = e_i$ , the true propensity score. Define matrices

$$\begin{aligned}C(\varpi^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \right\} \\ D(\varpi^*) &= \mathbb{V} \{ \Phi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \}.\end{aligned}$$

Then under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\hat{\varpi} - \varpi)$  converges in distribution to  $N(0, \Omega_{\varpi^*})$ , where

$$\Omega_{\varpi^*} = C(\varpi^*)^{-1} D(\varpi^*) C(\varpi^*)^{-T}.$$

Define

$$\begin{aligned}
G_1 &= \mathbb{E} [D_i S_i (Y_i^1 - m_{1i}) d_i(\beta) / (w_i e_i)] \\
G_2 &= \mathbb{E} [-D_i S_i (Y_i^0 - m_{0i}) d_i(\beta) / (w_i (1 - e_i))] \\
G_3 &= \mathbb{E} [D_i S_i d_i(\beta) \psi_{\alpha_1}^T (Y_i, Z_i; \alpha_1^*) / e_i] \\
G_4 &= \mathbb{E} [-D_i S_i d_i(\beta) \psi_{\alpha_0}^T (Y_i, Z_i; \alpha_0^*) / (1 - e_i)],
\end{aligned}$$

and let  $G = (G_1, G_2, 0_{q \times 1}, 0_{q \times p}, G_3, G_4)$ , then we can write

$$C(\varpi^*) = \begin{pmatrix} A(\theta^*) & -G^T \\ 0_{q \times (3+p+l_1+l_0)} & -E_{\beta\beta} \end{pmatrix} \quad \text{and} \quad D(\varpi^*) = \begin{pmatrix} B(\theta^*) & G^T \\ G & E_{\beta\beta} \end{pmatrix}.$$

Similar to the previous presentations, the upper-left block of  $\Omega_{\varpi^*}$  can be written as  $\Omega_{\theta^*} - A(\theta^*)^{-1} G^T E_{\beta\beta}^{-1} G A(\theta^*)^{-T}$ .

Therefore the asymptotic variance of  $\widehat{\Delta}_{\text{DR1}}$ , when  $e_i$  is estimated from trial data, is

$$\tau_{\text{DR1}}^2 = \sigma_{\text{DR1}}^2 - \{\lambda^T A(\theta^*) G^T\} E_{\beta\beta}^{-1} \{\lambda^T A(\theta^*) G^T\}^T. \quad (3.4)$$

Because  $E_{\beta\beta}^{-1}$  is positive definite, we must have  $\tau_{\text{DR1}}^2 \leq \sigma_{\text{DR1}}^2$ .

To estimate the variance of  $\widehat{\Delta}_{\text{DR1}}$ , we can use the consistent sandwich estimator. Define the following matrices

$$\begin{aligned}
C(\widehat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \\
D(\widehat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \Phi_{\Delta_{\text{DR1}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \Phi_{\Delta_{\text{DR1}}}^T(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi})
\end{aligned}$$

Therefore a consistent estimator for the variance of  $\widehat{\Delta}_{\text{DR1}}$  is  $N^{-1} \eta^T C(\widehat{\varpi})^{-1} D(\widehat{\varpi}) C(\widehat{\varpi})^{-T} \eta$ , where  $\eta = (\lambda^T, 0_{1 \times q})^T$ .

### 3.5 The DR2 Estimator

Again consider first the case when the true propensity score is used in constructing the DR2 estimator.

Recall that the true treatment propensity score  $e_i = r$ . Define

$$\begin{aligned}\tilde{\nu}_1 &= \frac{\sum_{i=1}^N D_i S_i X_i (Y_i - \tilde{m}_{1i}) / \hat{w}_i}{\sum_{i=1}^N D_i S_i X_i / \hat{w}_i} = \frac{\sum_{i=1}^N D_i S_i X_i (Y_i - \tilde{m}_{1i}) / \hat{w}_i e_i}{\sum_{i=1}^N D_i S_i X_i / \hat{w}_i e_i} \\ \tilde{\nu}_2 &= \frac{\sum_{i=1}^N D_i S_i (1 - X_i) (Y_i - \tilde{m}_{0i}) / \hat{w}_i}{\sum_{i=1}^N D_i S_i (1 - X_i) / \hat{w}_i} = \frac{\sum_{i=1}^N D_i S_i (1 - X_i) (Y_i - \tilde{m}_{0i}) / \hat{w}_i (1 - e_i)}{\sum_{i=1}^N D_i S_i (1 - X_i) / \hat{w}_i (1 - e_i)} \\ \tilde{\nu}_3 &= \frac{1}{N} \sum_{i=1}^N c_i (\tilde{m}_{1i} - \tilde{m}_{0i}),\end{aligned}$$

where  $\tilde{m}_{1i}, \tilde{m}_{0i}$  are defined in equation (3.5) in the main text. Then clearly we can write  $\tilde{\Delta}_{\text{DR2}} = \tilde{\nu}_1 - \tilde{\nu}_2 + \tilde{\nu}_3$ . Let  $\tilde{\theta} = (\tilde{\nu}_1, \tilde{\nu}_2, \tilde{\nu}_3, \tilde{\gamma}^T, \tilde{\alpha}_0^T, \tilde{\alpha}_1^T)^T$ , define  $\theta^* = (\nu_1, \nu_2, \nu_3, \gamma^{*T}, \alpha_0^{*T}, \alpha_1^{*T})^T$  as the limiting value of  $\tilde{\theta}$  and note that  $\tilde{\theta}$  is the solution to the estimating equation  $\sum_{i=1}^N \Psi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \tilde{\theta}) = 0$ , where

$$\Psi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta) = \begin{pmatrix} D_i S_i X_i (Y_i - m_{1i} - \nu_1) / (w_i e_i) \\ D_i S_i (1 - X_i) (Y_i - m_{0i} - \nu_2) / (w_i (1 - e_i)) \\ c_i m_{1i} - c_i m_{0i} - \nu_3 \\ \psi_\gamma(D_i, S_i, Z_i; \gamma) \\ D_i S_i X_i \psi_{\alpha_0}(Y_i, Z_i; \alpha_0) / e_i \\ D_i S_i (1 - X_i) \psi_{\alpha_1}(Y_i, Z_i; \alpha_1) / (1 - e_i) \end{pmatrix}.$$

Define the matrices

$$\begin{aligned}A(\theta^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \theta^T} \Psi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \right\}, \\ B(\theta^*) &= \mathbb{V} \{ \Psi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta^*) \}.\end{aligned}$$

Under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\tilde{\theta} - \theta^*)$  converges in distribution to  $N(0, \Omega_{\theta^*})$ , where

$$\Omega_{\theta^*} = A(\theta^*)^{-1} B(\theta^*) A(\theta^*)^{-T}.$$



Thus, the asymptotic variance of  $\tilde{\Delta}_{\text{DR2}}$ , when true  $e_i$  is used, can be expressed as  $\sigma_{\text{DR2}}^2 = \lambda^T \Omega_{\theta^*} \lambda$ , where  $\lambda = (1, -1, 1, 0_{1 \times p}, 0_{1 \times l_1}, 0_{1 \times l_0})^T$ .

Next consider the case when the treatment propensity score is estimated from trial data. Define

$$\begin{aligned}\tilde{\nu}_1 &= \frac{\sum_{i=1}^N D_i S_i X_i (Y_i - \hat{m}_{1i}) / \hat{w}_i \hat{e}_i}{\sum_{i=1}^N D_i S_i X_i / \hat{w}_i \hat{e}_i} \\ \tilde{\nu}_2 &= \frac{\sum_{i=1}^N D_i S_i (1 - X_i) (Y_i - \hat{m}_{0i}) / \hat{w}_i (1 - \hat{e}_i)}{\sum_{i=1}^N D_i S_i (1 - X_i) / \hat{w}_i (1 - \hat{e}_i)} \\ \tilde{\nu}_3 &= \frac{1}{N} \sum_{i=1}^N c_i (\hat{m}_{1i} - \hat{m}_{0i}),\end{aligned}$$

where  $\hat{m}_{1i}$ ,  $\hat{m}_{0i}$  are defined in equation (3.8) in the main text. Then clearly we can write  $\hat{\Delta}_{\text{DR2}} = \hat{\nu}_1 - \hat{\nu}_2 + \hat{\nu}_3$ . Let  $\hat{\theta} = (\hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3, \hat{\gamma}^T, \hat{\alpha}_0^T, \hat{\alpha}_1^T, \hat{\beta})^T$  define  $\theta = (\nu_1, \nu_2, \nu_3, \gamma^{*T}, \alpha_0^{*T}, \alpha_1^{*T}, \beta^T)^T$  as the limiting value of  $\hat{\theta}$  and note that  $\hat{\theta}$  is the solution to the  $(3 + p + l_1 + l_0 + q) \times 1$  vector of estimating equation  $\sum_{i=1}^N \Phi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \hat{\varpi}) = 0$ , where

$$\Phi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \begin{pmatrix} \Psi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) \\ \psi_{\beta}(D_i, S_i, X_i, W_i; \beta) \end{pmatrix}.$$

Here,  $\Psi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi) = \Psi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \theta)$  whenever  $\beta$  is chosen such that  $e(W_i; \beta) = e_i$ , the true propensity score. Define matrices

$$\begin{aligned}C(\varpi^*) &= \mathbb{E} \left\{ \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \right\} \\ D(\varpi^*) &= \mathbb{V} \{ \Phi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \varpi^*) \}.\end{aligned}$$

Then under suitable regularity conditions, as  $N \rightarrow \infty$ ,  $N^{1/2}(\hat{\varpi} - \varpi)$  converges in distribution to  $N(0, \Omega_{\varpi^*})$ , where

$$\Omega_{\varpi^*} = C(\varpi^*)^{-1} D(\varpi^*) C(\varpi^*)^{-T}.$$

Define

$$\begin{aligned}
G_1 &= \mathbb{E} [D_i S_i (Y_i^1 - m_{1i} - \nu_1) d_i(\beta) / (w_i e_i)] \\
G_2 &= \mathbb{E} [-D_i S_i (Y_i^0 - m_{0i} - \nu_2) d_i(\beta) / (w_i (1 - e_i))] \\
G_3 &= \mathbb{E} [D_i S_i d_i(\beta) \psi_{\alpha_1}^T (Y_i, Z_i; \alpha_1^*) / e_i] \\
G_4 &= \mathbb{E} [-D_i S_i d_i(\beta) \psi_{\alpha_0}^T (Y_i, Z_i; \alpha_0^*) / (1 - e_i)],
\end{aligned}$$

and let  $G = (G_1, G_2, 0_{q \times 1}, 0_{q \times p}, G_3, G_4)$ , then we can write

$$C(\varpi^*) = \begin{pmatrix} A(\theta^*) & -G^T \\ 0_{q \times (3+p+l_1+l_0)} & -E_{\beta\beta} \end{pmatrix} \quad \text{and} \quad D(\varpi^*) = \begin{pmatrix} B(\theta^*) & G^T \\ G & E_{\beta\beta} \end{pmatrix}.$$

Following the same arguments as in Appendix 3.4, the upper-left block of  $\Omega_{\varpi^*}$  can be written as  $\Omega_{\theta^*} - A(\theta^*)^{-1} G^T E_{\beta\beta}^{-1} G A(\theta^*)^{-T}$ . Therefore the asymptotic variance of  $\widehat{\Delta}_{\text{DR2}}$ , when  $e_i$  is estimated from trial data, is

$$\tau_{\text{DR2}}^2 = \sigma_{\text{DR2}}^2 - \{\lambda^T A(\theta^*) G^T\} E_{\beta\beta}^{-1} \{\lambda^T A(\theta^*) G^T\}^T. \quad (3.5)$$

Because  $E_{\beta\beta}^{-1}$  is positive definite, we must have  $\tau_{\text{DR2}}^2 \leq \sigma_{\text{DR2}}^2$ .

To estimate the variance of  $\widehat{\Delta}_{\text{DR2}}$ , we can use the consistent sandwich estimator. Define the following matrices

$$\begin{aligned}
C(\widehat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \varpi^T} \Phi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \\
D(\widehat{\varpi}) &= \frac{1}{N} \sum_{i=1}^N \Phi_{\Delta_{\text{DR2}}}(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi}) \Phi_{\Delta_{\text{DR2}}}^T(Y_i, D_i, S_i, X_i, Z_i; \widehat{\varpi})
\end{aligned}$$

Therefore a consistent estimator for the variance of  $\widehat{\Delta}_{\text{DR2}}$  is  $N^{-1} \eta^T C(\widehat{\varpi})^{-1} D(\widehat{\varpi}) C(\widehat{\varpi})^{-T} \eta$ , where  $\eta = (\lambda^T, 0_{1 \times q})^T$ .

## 4 Proof of Proposition 2

Define  $E_{\beta_1 \beta_1}^{-1}$  and  $E_{\beta_2 \beta_2}^{-1}$  as the information matrix of  $\widehat{\beta}_1$  and  $\widehat{\beta}_2$  (whose definition follow from the general definition of  $E_{\beta\beta}^{-1}$  in the beginning Appendix 3). By definition of nested models, we can write

$E_{\beta_2\beta_2}$  in blocks as

$$E_{\beta_2\beta_2} = \begin{pmatrix} E_{\beta_1\beta_1} & E_{\beta_2\beta_2}^{12} \\ E_{\beta_2\beta_2}^{21} & E_{\beta_2\beta_2}^{22} \end{pmatrix}.$$

where  $E_{\beta_2\beta_2}^{21} = \left(E_{\beta_2\beta_2}^{12}\right)^T$ . Block matrix inversion gives

$$E_{\beta_2\beta_2}^{-1} = \begin{pmatrix} F^{11} & F^{12} \\ F^{21} & F^{22} \end{pmatrix},$$

where

$$\begin{aligned} F^{11} &= E_{\beta_1\beta_1}^{-1} + E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12} \left(E_{\beta_2\beta_2}^{22} - E_{\beta_2\beta_2}^{21} E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12}\right)^{-1} E_{\beta_2\beta_2}^{21} E_{\beta_1\beta_1}^{-1} \\ F^{12} &= (F^{21})^T = -E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12} \left(E_{\beta_2\beta_2}^{22} - E_{\beta_2\beta_2}^{21} E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12}\right)^{-1} \\ F^{22} &= \left(E_{\beta_2\beta_2}^{22} - E_{\beta_2\beta_2}^{21} E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12}\right)^{-1} \end{aligned}$$

Now recall that equations (3.1)-(3.5) share the same form  $\tau_\chi^2 = \sigma_\chi^2 - M_\beta E_{\beta\beta}^{-1} M_\beta^T$ , where  $\chi \in \{\text{IPSW1, IPSW2, REG, DR1, DR2}\}$ , and the definition of  $M_\beta = \lambda^T A(\theta^*) G^T$  depends on the choice of estimator  $\chi$ , as elucidated in Appendix 3. However, the definition of  $E_{\beta\beta}^{-1}$  remains the same across estimators  $\chi$  but will vary according to the choice of the propensity score model. This means for each type of estimator  $\chi$ , the asymptotic variances of the estimators based on  $e(W_1; \hat{\beta}_1)$  and  $e(W_2; \hat{\beta}_2)$  are

$$\begin{aligned} \tau_\chi^2(\beta_1) &= \sigma_\chi^2 - M_{\beta_1} E_{\beta_1\beta_1}^{-1} M_{\beta_1}^T, \\ \tau_\chi^2(\beta_2) &= \sigma_\chi^2 - M_{\beta_2} E_{\beta_2\beta_2}^{-1} M_{\beta_2}^T, \end{aligned}$$

respectively. By definition of  $M_{\beta_2}$  in each type of estimator  $\chi$ , there exists some matrix  $M_{\beta_2}^{12}$  so that we can write  $M_{\beta_2}$  in block forms as

$$M_{\beta_2} = \begin{pmatrix} M_{\beta_1} & M_{\beta_2}^{12} \end{pmatrix}.$$

Based on this block form representation, we carry out some additional algebra to conclude that

$$\begin{aligned}
& \tau_{\chi}^2(\beta_1) - \tau_{\chi}^2(\beta_2) \\
&= M_{\beta_2} E_{\beta_2\beta_2}^{-1} M_{\beta_2}^T - M_{\beta_1} E_{\beta_1\beta_1}^{-1} M_{\beta_1}^T \\
&= \left( M_{\beta_1} F^{11} M_{\beta_1}^T - M_{\beta_1} E_{\beta_1\beta_1}^{-1} M_{\beta_1}^T \right) + M_{\beta_1} F^{12} (M_{\beta_2}^{12})^T + M_{\beta_2}^{12} F^{21} M_{\beta_1}^T + M_{\beta_2}^{12} F^{22} (M_{\beta_2}^{12})^T \\
&= \left( M_{\beta_1} E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12} - M_{\beta_2}^{12} \right) \left( E_{\beta_2\beta_2}^{22} - E_{\beta_2\beta_2}^{21} E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12} \right)^{-1} \left( M_{\beta_1} E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12} - M_{\beta_2}^{12} \right)^T \\
&\geq 0.
\end{aligned}$$

The last inequality follows because  $F^{22} = \left( E_{\beta_2\beta_2}^{22} - E_{\beta_2\beta_2}^{21} E_{\beta_1\beta_1}^{-1} E_{\beta_2\beta_2}^{12} \right)^{-1}$  is positive definite. This proves that the five PATE estimators constructed with  $e(W_2; \hat{\beta}_2)$  are asymptotically no less efficient than their counterparts constructed with  $e(W_1; \hat{\beta}_1)$ .

## 5 R Code for Generalizability Estimators and Simulations

To facilitate the implementation of the five estimators developed in this work, we have included the sample R functions used to estimate the PATE and their sandwich variances on the `github` page of the first author: [https://github.com/lifanfrank/RCode\\_generalizability](https://github.com/lifanfrank/RCode_generalizability). There are 8 R files, each of which is explained below.

- `IPSW_truePS.R`: implement the  $\tilde{\Delta}_{IPSW1}$  and  $\tilde{\Delta}_{IPSW2}$  estimators for PATE and the associated sandwich variance estimators, the true treatment propensity score ( $e_i = 1/2$ ) is used;
- `REG_truePS.R`: implement the  $\tilde{\Delta}_{REG}$  estimator for PATE and the associated sandwich variance estimator, the true treatment propensity score ( $e_i = 1/2$ ) is used;
- `DR_truePS.R`: implement the  $\tilde{\Delta}_{DR1}$  and  $\tilde{\Delta}_{DR2}$  estimators for PATE and the associated sandwich variance estimators, the true treatment propensity score ( $e_i = 1/2$ ) is used;
- `IPSW_estPS.R`: implement the  $\hat{\Delta}_{IPSW1}$  and  $\hat{\Delta}_{IPSW2}$  estimators for PATE and the associated sandwich variance estimators, the estimated treatment propensity score ( $\hat{e}_i$ ) is used;

- `REG_estPS.R`: implement the  $\widehat{\Delta}_{\text{REG}}$  estimator for PATE and the associated sandwich variance estimator, the estimated treatment propensity score ( $\widehat{e}_i$ ) is used;
- `DR_estPS.R`: implement the  $\widehat{\Delta}_{\text{DR1}}$  and  $\widehat{\Delta}_{\text{DR2}}$  estimators for PATE and the associated sandwich variance estimators, the estimated treatment propensity score ( $\widehat{e}_i$ ) is used;
- `simdata.R`: sample R code to generate a simulated data set for illustrative analysis;
- `example_analysis.R`: illustrative R code to perform the generalizability analysis using the above functions that implement point and variance estimators.

In addition, we have included in the same GitHub page the R code to carry out the simulation studies described in Section 6 of the main manuscript. Specifically, there are four folders:

- `Main Simulations`: This folder contains simulation code for producing Table 3, Web Tables 9, 10 and 11.
- `Web Table 12 and 13 - Moderate Misspecification of N`: This folder contains simulation code for producing Web Tables 12 and 13.
- `Web Table 14 - Severe Misspecification of N`: This folder contains simulation code for producing Web Table 14.
- `Web Table 15 to 16 - Small Sample Sizes`: This folder contains simulation code for producing Web Tables 15 and 16.

## 6 Web Tables and Figures

- Web Table 1-4 present the baseline characteristics of ACTG 320 participants, ACTG 320 women participants, ACTG A5202 participants, ACTG A5202 women participants by treatment groups. The within-trial standardized mean difference for each variable is calculated following Yang and Dalton (2012).

- Web Table 5 presents the baseline characteristics of the WIHS participants and CNICS participants (cohort samples) used to generalize ACTG 320 results.
- Web Table 6 presents the baseline characteristics of the WIHS participants and CNICS participants (cohort samples) used to generalize ACTG A5202 results.
- Web Table 7-8 present the sensitivity analyses of the generalizability results to smaller and larger values of the target population size  $N$ .
- Web Table 9-11 present simulation results referenced in Section 6.2 of the main text (Main Simulation Results).
- Web Table 12-16 present the additional simulation results referenced in Section 6.3 of the main text (Additional Simulations).
- Web Figure 1 presents the zoomed histograms of the estimated sampling scores for the generalizability analyses.
- Web Figure 2-5 present the forest plots of the SMDs by each covariate for the generalizability analyses.
- Web Figure 6 presents the histograms of the DR2 estimates across 5000 simulations under four typical scenarios, and is referenced in Section 6.2 of the main text.

Table 1 Characteristics of ACTG 320 participants at baseline with observed outcome (change in CD4 cell count from baseline to week 4). The treatment group refers the protease inhibitor (PI) group, and the control group refers to the non-protease inhibitor (non-PI) group.

Variable	Treatment ( $n_t = 530$ )	Control ( $n_c = 510$ )	Std Mean Diff
Male sex - %	82	85	0.06
Race or ethnic group - %			
White, non-Hispanic	53	52	0.06
Black, non-Hispanic	28	27	
Hispanic	17	18	
Asian/Other	2	2	
Median age - yr (Q1-Q3) <sup>a</sup>	38.32 (33.16, 44.93)	38.15 (33.06, 44.24)	0.02
Age group - %			0.11
[16, 30) yr	11	13	
[30, 40) yr	47	45	
[40, 50) yr	29	31	
[50, ·) yr	12	11	
Injection drug use - %	16	15	0.01
Median CD4 count (Q1-Q3)	78.25 (23.62, 137.62)	68.25 (22.0, 132.50)	0.06
Baseline CD4 count - %			0.11
(0, 50) cells/mm <sup>3</sup>	39	41	
[50, 100) cells/mm <sup>3</sup>	21	22	
[100, 200) cells/mm <sup>3</sup>	34	30	
[200, ·) cells/mm <sup>3</sup>	6	7	

<sup>a</sup> Q1 is the first quartile and Q3 is the third quartile.

Table 2 Characteristics of ACTG 320 participants at baseline among women only with observed outcome (change in CD4 cell count from baseline to week 4). The treatment group refers the protease inhibitor (PI) group, and the control group refers to the non-protease inhibitor (non-PI) group.

Variable	Treatment ( $n_t = 94$ )	Control ( $n_c = 79$ )	Std Mean Diff
Male sex - %	0	0	-
Race or ethnic group - %			0.17
White, non-Hispanic	26	38	
Black, non-Hispanic	49	42	
Hispanic	24	19	
Asian/Other	1	1	
Median age - yr (Q1-Q3) <sup>a</sup>	37.45 (31.20, 42.05)	35.37 (29.98, 43.07)	0.13
Age group - %			0.18
[16, 30) yr	18	25	
[30, 40) yr	46	43	
[40, 50) yr	27	27	
[50, ·) yr	10	5	
Injection drug use - %	22	10	0.34
Median CD4 count (Q1-Q3)	85.50 (27.50, 135.62)	58.00 (15.25, 125.25)	0.20
Baseline CD4 count - %			0.27
(0, 50) cells/mm <sup>3</sup>	35	43	
[50, 100) cells/mm <sup>3</sup>	22	28	
[100, 200) cells/mm <sup>3</sup>	40	30	
[200, ·) cells/mm <sup>3</sup>	2	4	

<sup>a</sup> Q1 is the first quartile and Q3 is the third quartile.



Table 3 Characteristics of ACTG A5202 participants at baseline with observed outcome (change in CD4 cell count from baseline to week 48). The treatment group refers the ABC-3TC group, and the control group refers to the TDF-FTC group.

Variable	Treatment ( $n_t = 721$ )	Control ( $n_c = 719$ )	Std Mean Diff
Male sex - %	82	82	0.002
Race or ethnic group - %			0.07
White, non-Hispanic	40	41	
Black, non-Hispanic	34	32	
Hispanic	23	24	
Asian/Other	4	4	
Median age - yr (Q1-Q3) <sup>a</sup>	38 (30, 45)	39 (31, 46)	0.05
Age group - no. %			0.14
[16, 30) yr	21	21	
[30, 40) yr	36	32	
[40, 50) yr	30	32	
[50, ·) yr	13	15	
Injection drug use - %	7	8	0.03
Hepatitis B/C - %	8	10	0.07
AIDS diagnosis - %	17	13	0.12
Median CD4 count (Q1-Q3)	244 (107, 350)	244 (116, 334)	0.03
Baseline CD4 count - %			0.15
(0, 50) cells/mm <sup>3</sup>	16	16	
[50, 100) cells/mm <sup>3</sup>	8	7	
[100, 200) cells/mm <sup>3</sup>	16	17	
[200, 350) cells/mm <sup>3</sup>	35	39	
[350, ·) cells/mm <sup>3</sup>	25	22	
Median log <sub>10</sub> viral load (Q1-Q3)	4.6 (4.3, 4.9)	4.6 (4.3, 4.8)	0.04
Viral load - %			
[0, 50, 000) cp/ml	58	60	0.08
[50, 000, 100, 000) cp/ml	20	21	
[100, 000, 300, 000) cp/ml	10	8	
[300, 000, 500, 000) cp/m	3	3	
[500, 000, ·) cp/ml	8	8	

<sup>a</sup> Q1 is the first quartile and Q3 is the third quartile.

Table 4 Characteristics of ACTG A5202 participants at baseline among women only with observed outcome (change in CD4 cell count from baseline to week 48). The treatment group refers the ABC-3TC group, and the control group refers to the TDF-FTC group.

Variable	Treatment ( $n_t = 128$ )	Control ( $n_c = 127$ )	Std Mean Diff
Male sex - %	0	0	-
Race or ethnic group - %			0.24
White, non-Hispanic	16	19	
Black, non-Hispanic	57	52	
Hispanic	23	27	
Asian/Other	4	2	
Median age - yr (Q1-Q3) <sup>a</sup>	38 (31, 46)	40 (32, 48)	0.05
Age group - no. %			0.35
[16, 30) yr	18	18	
[30, 40) yr	40	32	
[40, 50) yr	28	36	
[50, ·) yr	14	14	
Injection drug use - %	4	6	0.11
Hepatitis B/C - %	9	7	0.08
AIDS diagnosis - %	22	13	0.22
Baseline CD4 count - %	243 (91, 338)	250 (140, 315)	0.09
Median CD4 count (Q1-Q3)			0.27
(0, 50) cells/mm <sup>3</sup>	18	14	
[50, 100) cells/mm <sup>3</sup>	9	5	
[100, 200) cells/mm <sup>3</sup>	14	20	
[200, 350) cells/mm <sup>3</sup>	38	46	
[350, ·) cells/mm <sup>3</sup>	20	16	
Median log <sub>10</sub> viral load (Q1-Q3)	4.5 (4.0, 4.9)	4.5 (4.0, 4.8)	0.03
Viral load - %			0.06
[0, 50, 000) cp/ml	62	65	
[50, 000, 100, 000) cp/ml	20	18	
[100, 000, 300, 000) cp/ml	11	9	
[300, 000, 500, 000) cp/m	2	2	
[500, 000, ·) cp/ml	6	5	

<sup>a</sup> Q1 is the first quartile and Q3 is the third quartile.

Table 5 Baseline characteristics of WIHS participants and CNICS participants used to generalize ACTG 320 results.

Variable	WIHS ( <i>m</i> = 493)	CNICS ( <i>m</i> = 6, 158)
Calendar time of index visit - Min, Max	01/05/1995, 02/22/2012	01/02/1998, 11/06/2013
Male sex - %	0	80
Race or ethnic group - %		
White, non-Hispanic	18	40
Black, non-Hispanic	55	44
Hispanic	25	12
Asian/Other	2	5
Median age - yrs (Q1-Q3) <sup>a</sup>	40 (35-45)	41 (34-47)
Age group - no. %		
[16, 30) yr	7	12
[30, 40) yr	43	34
[40, 50) yr	40	38
[50, ·) yr	10	17
Injection drug use - %	37	20
Median CD4 count (Q1-Q3)	108 (41-172)	89 (27-172)
Baseline CD4 count - %		
(0, 50) cells/mm <sup>3</sup>	30	36
[50, 100) cells/mm <sup>3</sup>	17	17
[100, 200) cells/mm <sup>3</sup>	37	30
[200, ·) cells/mm <sup>3</sup>	16	17
Viral load <sup>b</sup> - %		
[0, 50, 000) cp/ml	52	39
[50, 000, 100, 000) cp/ml	14	15
[100, 000, 300, 000) cp/ml	20	24
[300, 000, 500, 000) cp/ml	5	8
[500, 000, ·) cp/ml	8	15
Median log <sub>10</sub> viral load (Q1-Q3)	4.65 (3.96-5.16)	4.93 (4.16-5.43)
Initiated any ART - %	81	17
Median time since ART initiation - yrs (Q1-Q3)	4.90 (1.19-5.62)	1.68 (0.56-4.11)

<sup>a</sup> Q1 is the first quartile and Q3 is the third quartile.

<sup>b</sup> In WIHS, 37 women were missing viral load; however, this variable is not used to generalize the results because it was not available in ACTG 320.

Table 6 Baseline characteristics of WIHS participants and CNICS participants used to generalize ACTG A5202 results.<sup>a</sup>

Variable	WIHS ( <i>m</i> = 1, 012)	CNICS ( <i>m</i> = 12, 302)
Calendar time of index visit - Min, Max	10/10/1994, 09/20/2012	01/09/1998, 11/19/2013
Male sex - %	0	82
Race or ethnic group - %		
White, non-Hispanic	17	45
Black, non-Hispanic	58	38
Hispanic	22	12
Asian/Other	3	5
Median age - yr (Q1-Q3) <sup>b</sup>	39 (33-44)	39 (31-46)
Age group - %		
[16, 30) yr	12	20
[30, 40) yr	43	34
[40, 50) yr	34	32
[50, ·] yr	11	14
Injection drug use - %	38	17
Hepatitis B/C - %	35	18
AIDS diagnosis - %	37	23
CD4 count - %		
(0, 50) cells/mm <sup>3</sup>	10	16
[50, 100) cells/mm <sup>3</sup>	6	7
[100, 200) cells/mm <sup>3</sup>	16	14
[200, 350) cells/mm <sup>3</sup>	29	27
[350, ·) cells/mm <sup>3</sup>	39	36
Median CD4 count (Q1-Q3)	290 (162-423)	271 (109, 427)
Viral load - %		
[0, 50, 000) cp/ml	55	52
[50, 000, 100, 000) cp/ml	14	15
[100, 000, 300, 000) cp/ml	19	18
[300, 000, 500, 000) cp/ml	5	6
[500, 000, ·) cp/ml	7	8
Median log <sub>10</sub> viral load (Q1-Q3)	4.61 (4.04-5.11)	4.64 (3.95-5.18)

<sup>a</sup> To generalize ACTG A5202 results, all participants in the cohort were ART naive.

<sup>b</sup> Q1 is the first quartile and Q3 is the third quartile.

Table 7 Sensitivity analyses of the target population average treatment effects (PATE) on change in CD4 cell count and corresponding 95% confidence intervals based on proposed sandwich variance estimators. The model specifications follow Section 5.2 of the main manuscript, but the population sizes are assumed as: all people living with HIV in the USA  $N = 0.7$  million and all women living with HIV in the USA  $N = 230,000$ .

Cohort	m	Trial	n	SATE	PATE IPSW1	PATE IPSW2	PATE REG	PATE DR1	PATE DR2
<i>True propensity scores</i>									
CNICS	6,158	320	1,040	19 (12, 25)	20 (11, 29)	18 (10, 25)	14 (5, 23)	15 (6, 24)	15 (6, 24)
CNICS	12,302	A5202	1,440	6 (-8, 20)	5 (-27, 36)	1 (-26, 28)	-2 (-27, 24)	-3 (-29, 24)	-3 (-29, 24)
WIHS	493	320	173	24 (7,41)	52 (19, 86)	42 (15, 69)	32 (-2, 65)	37 (1, 72)	37 (2, 72)
WIHS	1,012	A5202	255	1 (-35, 37)	106 (-44, 256)	31 (-44, 105)	0.04 (-103, 104)	4 (-84, 92)	4 (-85, 92)
<i>Estimated propensity scores with main-effects logistic model</i>									
CNICS	6,158	320	1,040	19 (12, 25)	18 (10, 26)	18 (10, 25)	14 (5, 23)	15 (7, 24)	15 (7, 24)
CNICS	12,302	A5202	1,440	6 (-8, 20)	-2 (-26, 21)	0.32 (-24, 25)	-0.86 (-26, 25)	-1 (-27, 25)	-1 (-27, 25)
WIHS	493	320	173	24 (7,41)	41 (13, 69)	47 (21, 73)	35 (3, 68)	40 (6, 74)	40 (5, 74)
WIHS	1,012	A5202	255	1 (-35, 37)	81 (-49, 211)	37 (-38, 111)	10 (-90, 110)	11 (-75, 96)	11 (-75, 96)
<i>Estimated propensity scores with full logistic model</i>									
CNICS	6,158	320	1,040	19 (12, 25)	18 (10, 25)	18 (10, 25)	14 (5, 23)	15 (6, 24)	15 (6, 24)
CNICS	12,302	A5202	1,440	6 (-8, 20)	0.24 (-21, 21)	-0.82 (-24, 22)	-3 (-27, 22)	-4 (-29, 22)	-3 (-29, 22)
WIHS	493	320	173	24 (7,41)	38 (16, 60)	39 (17, 61)	25 (-6, 55)	28 (-3, 60)	28 (-3, 60)
WIHS	1,012	A5202	255	1 (-35, 37)	53 (-39, 144)	23 (-28, 75)	-9 (-75, 57)	-8 (-75, 58)	-8 (-75, 59)

Table 8 Sensitivity analyses of the target population average treatment effects (PATE) on change in CD4 cell count and corresponding 95% confidence intervals based on proposed sandwich variance estimators. The model specifications follow Section 5.2 of the main manuscript, but the population sizes are assumed as: all people living with HIV in the USA  $N = 1.5$  million and all women living with HIV in the USA  $N = 330,000$ .

Cohort	m	Trial	n	SATE	PATE IPSW1	PATE IPSW2	PATE REG	PATE DR1	PATE DR2
<i>True propensity scores</i>									
CNICS	6,158	320	1,040	19 (12, 25)	20 (11, 29)	18 (10, 25)	14 (5, 23)	15 (6, 24)	15 (6, 24)
CNICS	12,302	A5202	1,440	6 (-8, 20)	5 (-27, 36)	1 (-26, 28)	-2 (-27, 24)	-2 (-29, 24)	-2 (-29, 24)
WIHS	493	320	173	24 (7,41)	52 (19, 86)	42 (15, 69)	32 (-2, 65)	37 (1, 72)	37 (2, 72)
WIHS	1,012	A5202	255	1 (-35, 37)	106 (-44, 257)	31 (-44, 105)	0.04 (-104, 104)	4 (-84, 92)	4 (-85, 92)
<i>Estimated propensity scores with main-effects logistic model</i>									
CNICS	6,158	320	1,040	19 (12, 25)	18 (10, 26)	18 (10, 25)	14 (5, 23)	15 (7, 24)	15 (7, 24)
CNICS	12,302	A5202	1,440	6 (-8, 20)	-2 (-26, 21)	0.33 (-24, 25)	-0.87 (-26, 25)	-1 (-27, 25)	-1 (-27, 25)
WIHS	493	320	173	24 (7,41)	41 (13, 69)	47 (21, 73)	35 (3, 68)	40 (6, 74)	39 (6, 74)
WIHS	1,012	A5202	255	1 (-35, 37)	81 (-49, 212)	37 (-38, 111)	10 (-90, 110)	11 (-75, 96)	11 (-75, 96)
<i>Estimated propensity scores with full logistic model</i>									
CNICS	6,158	320	1,040	19 (12, 25)	18 (10, 25)	18 (10, 25)	14 (5, 23)	15 (6, 24)	15 (6, 24)
CNICS	12,302	A5202	1,440	6 (-8, 20)	0.24 (-21, 21)	-0.81 (-24, 22)	-3 (-27, 22)	-4 (-29, 21)	-3 (-29, 22)
WIHS	493	320	173	24 (7,41)	38 (16, 60)	39 (17, 61)	25 (-5, 55)	28 (-3, 60)	28 (-3, 60)
WIHS	1,012	A5202	255	1 (-35, 37)	53 (-39, 144)	23 (-28, 75)	-9 (-75, 57)	-8 (-75, 58)	-8 (-75, 59)

Table 9 Comparison of performance of five different estimators for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.6, 0.6, 0.6)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (1, 1, 1)$  in the main simulation. The true PATE  $\Delta = 2.4$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>True treatment propensity scores used</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.01	21.0	20.1	94.9
$\tilde{\Delta}_{IPSW1}$	×	–	0.16	36.0	34.4	90.7
$\tilde{\Delta}_{IPSW2}$	✓	–	0.00	17.8	16.3	92.6
$\tilde{\Delta}_{IPSW2}$	×	–	-0.31	22.9	21.0	71.1
$\tilde{\Delta}_{REG}$	–	✓	0.00	9.6	10.2	96.1
$\tilde{\Delta}_{REG}$	–	×	-0.09	10.6	11.1	88.9
$\tilde{\Delta}_{DR1}$	✓	✓	0.00	10.8	11.1	95.8
$\tilde{\Delta}_{DR1}$	✓	×	0.00	12.7	12.7	95.4
$\tilde{\Delta}_{DR1}$	×	✓	0.00	13.3	13.1	95.4
$\tilde{\Delta}_{DR1}$	×	×	0.41	22.7	21.1	48.6
$\tilde{\Delta}_{DR2}$	✓	✓	0.00	10.7	11.0	95.7
$\tilde{\Delta}_{DR2}$	✓	×	0.00	12.6	12.5	95.1
$\tilde{\Delta}_{DR2}$	×	✓	0.00	12.0	11.9	95.3
$\tilde{\Delta}_{DR2}$	×	×	0.31	17.0	15.9	48.7
<i>Estimated treatment propensity scores used</i>						
$\hat{\Delta}_{IPSW1}$	✓	–	0.01	18.2	17.0	94.9
$\hat{\Delta}_{IPSW1}$	×	–	0.16	32.1	29.8	89.5
$\hat{\Delta}_{IPSW2}$	✓	–	0.00	17.6	15.7	92.7
$\hat{\Delta}_{IPSW2}$	×	–	-0.31	22.5	20.5	70.4
$\hat{\Delta}_{REG}$	–	✓	0.00	9.6	10.2	96.1
$\hat{\Delta}_{REG}$	–	×	-0.09	10.5	11.0	88.9
$\hat{\Delta}_{DR1}$	✓	✓	0.00	10.8	11.1	95.8
$\hat{\Delta}_{DR1}$	✓	×	0.00	12.6	12.6	95.5
$\hat{\Delta}_{DR1}$	×	✓	0.00	13.4	13.1	95.3
$\hat{\Delta}_{DR1}$	×	×	0.41	22.6	20.8	46.8
$\hat{\Delta}_{DR2}$	✓	✓	0.00	10.7	11.0	95.7
$\hat{\Delta}_{DR2}$	✓	×	0.00	12.5	12.5	95.3
$\hat{\Delta}_{DR2}$	×	✓	0.00	12.0	11.9	95.2
$\hat{\Delta}_{DR2}$	×	×	0.31	17.0	15.9	48.4

Table 10 Comparison of performance of five different estimators for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.3, 0.3, 0.3)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 2)$  in the main simulation. The true PATE  $\Delta = 2.8$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>True treatment propensity scores used</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.00	12.7	13.4	96.2
$\tilde{\Delta}_{IPSW1}$	×	–	0.01	12.8	13.6	95.9
$\tilde{\Delta}_{IPSW2}$	✓	–	0.00	11.6	11.6	95.1
$\tilde{\Delta}_{IPSW2}$	×	–	-0.03	11.6	11.6	94.7
$\tilde{\Delta}_{REG}$	–	✓	0.00	8.6	9.6	97.0
$\tilde{\Delta}_{REG}$	–	×	0.05	9.8	10.6	93.7
$\tilde{\Delta}_{DR1}$	✓	✓	0.00	8.7	9.7	97.0
$\tilde{\Delta}_{DR1}$	✓	×	0.00	9.1	10.1	96.9
$\tilde{\Delta}_{DR1}$	×	✓	0.00	8.7	9.7	96.8
$\tilde{\Delta}_{DR1}$	×	×	0.18	9.7	10.7	61.1
$\tilde{\Delta}_{DR2}$	✓	✓	0.00	8.7	9.7	97.0
$\tilde{\Delta}_{DR2}$	✓	×	0.00	9.1	10.1	97.0
$\tilde{\Delta}_{DR2}$	×	✓	0.00	8.7	9.7	96.8
$\tilde{\Delta}_{DR2}$	×	×	0.18	9.7	10.6	61.1
<i>Estimated treatment propensity scores used</i>						
$\hat{\Delta}_{IPSW1}$	✓	–	0.00	10.7	11.5	96.5
$\hat{\Delta}_{IPSW1}$	×	–	0.01	10.6	11.5	96.3
$\hat{\Delta}_{IPSW2}$	✓	–	0.00	11.6	11.6	94.8
$\hat{\Delta}_{IPSW2}$	×	–	-0.04	11.6	11.6	94.4
$\hat{\Delta}_{REG}$	–	✓	0.00	8.6	9.6	97.0
$\hat{\Delta}_{REG}$	–	×	0.05	9.8	10.6	93.6
$\hat{\Delta}_{DR1}$	✓	✓	0.00	8.7	9.7	97.0
$\hat{\Delta}_{DR1}$	✓	×	0.00	9.1	10.1	96.9
$\hat{\Delta}_{DR1}$	×	✓	0.00	8.7	9.7	96.8
$\hat{\Delta}_{DR1}$	×	×	0.18	9.7	10.7	61.0
$\hat{\Delta}_{DR2}$	✓	✓	0.00	8.7	9.7	97.0
$\hat{\Delta}_{DR2}$	✓	×	0.00	9.1	10.1	97.1
$\hat{\Delta}_{DR2}$	×	✓	0.00	8.7	9.7	96.8
$\hat{\Delta}_{DR2}$	×	×	0.18	9.7	10.6	61.1



Table 11 Comparison of performance of five different estimators for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.6, 0.6, 0.6)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 2)$  in the main simulation. The true PATE  $\Delta = 2.8$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>True treatment propensity scores used</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.01	20.4	19.5	94.5
$\hat{\Delta}_{IPSW1}$	×	–	-0.15	31.3	29.7	95.5
$\tilde{\Delta}_{IPSW2}$	✓	–	0.01	23.2	20.9	92.6
$\hat{\Delta}_{IPSW2}$	×	–	-0.62	29.7	27.6	37.9
$\tilde{\Delta}_{REG}$	–	✓	0.00	10.5	11.3	96.4
$\hat{\Delta}_{REG}$	–	×	-0.17	12.7	13.2	76.0
$\tilde{\Delta}_{DR1}$	✓	✓	0.00	11.6	12.1	96.0
$\hat{\Delta}_{DR1}$	✓	×	-0.01	14.3	14.3	95.5
$\tilde{\Delta}_{DR1}$	×	✓	0.00	14.0	13.9	95.7
$\hat{\Delta}_{DR1}$	×	×	0.83	29.1	27.1	5.9
$\tilde{\Delta}_{DR2}$	✓	✓	0.00	11.5	12.0	96.0
$\hat{\Delta}_{DR2}$	✓	×	-0.01	14.3	14.2	95.0
$\tilde{\Delta}_{DR2}$	×	✓	0.00	12.7	12.8	95.6
$\hat{\Delta}_{DR2}$	×	×	0.63	20.7	19.2	6.8
<i>Estimated treatment propensity scores used</i>						
$\hat{\Delta}_{IPSW1}$	✓	–	0.01	18.7	17.2	94.2
$\tilde{\Delta}_{IPSW1}$	×	–	-0.15	28.7	26.6	95.6
$\hat{\Delta}_{IPSW2}$	✓	–	0.00	23.3	20.9	92.8
$\tilde{\Delta}_{IPSW2}$	×	–	-0.62	29.7	27.6	37.3
$\hat{\Delta}_{REG}$	–	✓	0.00	10.5	11.3	96.4
$\tilde{\Delta}_{REG}$	–	×	-0.17	12.7	13.2	76.0
$\hat{\Delta}_{DR1}$	✓	✓	0.00	11.6	12.1	96.1
$\tilde{\Delta}_{DR1}$	✓	×	-0.01	14.3	14.4	95.5
$\hat{\Delta}_{DR1}$	×	✓	0.00	14.0	13.9	95.8
$\tilde{\Delta}_{DR1}$	×	×	0.83	29.4	27.1	5.8
$\hat{\Delta}_{DR2}$	✓	✓	0.00	11.5	12.0	96.0
$\tilde{\Delta}_{DR2}$	✓	×	-0.01	14.3	14.2	95.2
$\hat{\Delta}_{DR2}$	×	✓	0.00	12.7	12.8	95.6
$\tilde{\Delta}_{DR2}$	×	×	0.63	20.7	19.2	6.7

Table 12 Comparison of performance of five different estimators coupled with estimated propensity scores for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.6, 0.6, 0.6)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 2)$  when the population size  $N$  is underestimated. The true PATE  $\Delta = 2.8$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>N misspecified to be 0.8 million</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.01	18.6	17.2	94.2
$\hat{\Delta}_{IPSW1}$	×	–	-0.14	28.7	26.5	95.6
$\tilde{\Delta}_{IPSW2}$	✓	–	0.00	23.3	20.9	92.9
$\hat{\Delta}_{IPSW2}$	×	–	-0.62	29.7	27.5	37.5
$\tilde{\Delta}_{REG}$	–	✓	0.00	10.5	11.3	96.3
$\hat{\Delta}_{REG}$	–	×	-0.17	12.7	13.2	76.3
$\tilde{\Delta}_{DR1}$	✓	✓	0.00	11.6	12.1	96.1
$\hat{\Delta}_{DR1}$	✓	×	-0.01	14.4	14.3	95.5
$\tilde{\Delta}_{DR1}$	×	✓	0.00	14.0	13.9	95.9
$\hat{\Delta}_{DR1}$	×	×	0.83	29.3	27.1	5.8
$\tilde{\Delta}_{DR2}$	✓	✓	0.00	11.5	12.0	95.9
$\hat{\Delta}_{DR2}$	✓	×	0.00	14.3	14.2	95.2
$\tilde{\Delta}_{DR2}$	×	✓	0.00	12.7	12.8	95.6
$\hat{\Delta}_{DR2}$	×	×	0.63	20.7	19.2	6.6
<i>N misspecified to be 0.5 million</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.01	18.5	17.1	93.9
$\hat{\Delta}_{IPSW1}$	×	–	-0.14	28.6	26.4	95.6
$\tilde{\Delta}_{IPSW2}$	✓	–	0.01	23.2	20.8	92.6
$\hat{\Delta}_{IPSW2}$	×	–	-0.61	29.6	27.4	38.2
$\tilde{\Delta}_{REG}$	–	✓	0.01	10.5	11.3	96.2
$\hat{\Delta}_{REG}$	–	×	-0.17	12.7	13.2	77.2
$\tilde{\Delta}_{DR1}$	✓	✓	0.01	11.6	12.1	96.1
$\hat{\Delta}_{DR1}$	✓	×	-0.00	14.3	14.3	95.5
$\tilde{\Delta}_{DR1}$	×	✓	0.01	14.0	13.9	95.8
$\hat{\Delta}_{DR1}$	×	×	0.83	29.2	26.9	5.5
$\tilde{\Delta}_{DR2}$	✓	✓	0.01	11.5	12.0	96.1
$\hat{\Delta}_{DR2}$	✓	×	0.00	14.3	14.1	95.1
$\tilde{\Delta}_{DR2}$	×	✓	0.01	12.7	12.8	95.5
$\hat{\Delta}_{DR2}$	×	×	0.63	20.7	19.2	6.4

Table 13 Comparison of performance of five different estimators coupled with estimated propensity scores for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.6, 0.6, 0.6)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 2)$  when the population size  $N$  is overestimated. The true PATE  $\Delta = 2.8$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>N misspecified to be 1.2 million</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.00	18.7	17.2	94.3
$\hat{\Delta}_{IPSW1}$	×	–	-0.15	28.8	26.6	95.5
$\tilde{\Delta}_{IPSW2}$	✓	–	0.00	23.4	20.9	92.9
$\hat{\Delta}_{IPSW2}$	×	–	-0.62	29.8	27.6	37.1
$\tilde{\Delta}_{REG}$	–	✓	0.00	10.5	11.3	96.4
$\hat{\Delta}_{REG}$	–	×	-0.17	12.7	13.2	75.9
$\tilde{\Delta}_{DR1}$	✓	✓	0.00	11.6	12.1	96.1
$\hat{\Delta}_{DR1}$	✓	×	-0.01	14.4	14.4	95.5
$\tilde{\Delta}_{DR1}$	×	✓	0.00	14.1	14.0	95.8
$\hat{\Delta}_{DR1}$	×	×	0.83	29.5	27.2	5.9
$\tilde{\Delta}_{DR2}$	✓	✓	0.00	11.5	12.0	96.0
$\hat{\Delta}_{DR2}$	✓	×	-0.01	14.4	14.2	95.1
$\tilde{\Delta}_{DR2}$	×	✓	0.00	12.7	12.8	95.6
$\hat{\Delta}_{DR2}$	×	×	0.63	20.7	19.2	6.8
<i>N misspecified to be 1.5 million</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.00	18.7	17.2	94.3
$\hat{\Delta}_{IPSW1}$	×	–	-0.15	28.8	26.6	95.5
$\tilde{\Delta}_{IPSW2}$	✓	–	0.00	23.5	20.9	92.9
$\hat{\Delta}_{IPSW2}$	×	–	-0.62	29.8	27.6	36.9
$\tilde{\Delta}_{REG}$	–	✓	0.00	10.5	11.3	96.4
$\hat{\Delta}_{REG}$	–	×	-0.17	12.7	13.2	75.7
$\tilde{\Delta}_{DR1}$	✓	✓	0.00	11.6	12.1	96.1
$\hat{\Delta}_{DR1}$	✓	×	-0.01	14.4	14.4	95.5
$\tilde{\Delta}_{DR1}$	×	✓	0.00	14.1	14.0	95.8
$\hat{\Delta}_{DR1}$	×	×	0.83	29.5	27.2	5.9
$\tilde{\Delta}_{DR2}$	✓	✓	0.00	11.5	12.0	96.0
$\hat{\Delta}_{DR2}$	✓	×	-0.01	14.4	14.2	95.1
$\tilde{\Delta}_{DR2}$	×	✓	0.00	12.7	12.8	95.7
$\hat{\Delta}_{DR2}$	×	×	0.63	20.8	19.2	6.9

Table 14 Comparison of performance of five different estimators coupled with estimated propensity scores for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.6, 0.6, 0.6)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 2)$  when the population size  $N$  is severely overestimated (two extreme scenarios). The true PATE  $\Delta = 2.8$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>N misspecified to be 0.1 million</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.04	18.1	16.8	91.4
$\hat{\Delta}_{IPSW1}$	×	–	-0.10	27.4	25.5	95.4
$\tilde{\Delta}_{IPSW2}$	✓	–	0.05	22.4	20.2	89.9
$\hat{\Delta}_{IPSW2}$	×	–	-0.55	28.5	26.5	45.4
$\tilde{\Delta}_{REG}$	–	✓	0.04	10.4	11.2	95.2
$\hat{\Delta}_{REG}$	–	×	-0.13	12.6	13.1	83.6
$\tilde{\Delta}_{DR1}$	✓	✓	0.04	11.5	12.0	94.8
$\hat{\Delta}_{DR1}$	✓	×	0.03	14.0	14.0	94.7
$\tilde{\Delta}_{DR1}$	×	✓	0.04	13.7	13.6	94.7
$\hat{\Delta}_{DR1}$	×	×	0.83	27.7	25.7	4.1
$\tilde{\Delta}_{DR2}$	✓	✓	0.04	11.4	11.9	94.7
$\hat{\Delta}_{DR2}$	✓	×	0.04	13.9	13.9	94.2
$\tilde{\Delta}_{DR2}$	×	✓	0.04	12.6	12.7	94.4
$\hat{\Delta}_{DR2}$	×	×	0.65	20.2	18.8	4.7
<i>N misspecified to be 0.05 million</i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.08	17.8	16.5	86.5
$\hat{\Delta}_{IPSW1}$	×	–	-0.04	26.4	24.6	94.4
$\tilde{\Delta}_{IPSW2}$	✓	–	0.09	21.9	19.7	86.4
$\hat{\Delta}_{IPSW2}$	×	–	-0.49	27.7	25.7	54.9
$\tilde{\Delta}_{REG}$	–	✓	0.08	10.4	11.1	90.4
$\hat{\Delta}_{REG}$	–	×	-0.09	12.5	12.9	90.5
$\tilde{\Delta}_{DR1}$	✓	✓	0.08	11.4	11.8	91.0
$\hat{\Delta}_{DR1}$	✓	×	0.08	13.7	13.8	91.5
$\tilde{\Delta}_{DR1}$	×	✓	0.08	13.4	13.4	90.8
$\hat{\Delta}_{DR1}$	×	×	0.84	26.4	24.6	2.6
$\tilde{\Delta}_{DR2}$	✓	✓	0.08	11.3	11.8	90.8
$\hat{\Delta}_{DR2}$	✓	×	0.08	13.7	13.6	90.9
$\tilde{\Delta}_{DR2}$	×	✓	0.08	12.4	12.5	90.5
$\hat{\Delta}_{DR2}$	×	×	0.67	19.8	18.4	2.8

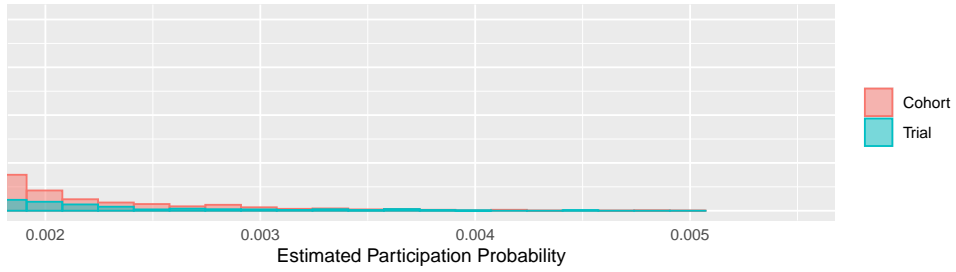
Table 15 Comparison of performance of five different estimators coupled with estimated propensity scores for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.6, 0.6, 0.6)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 2)$  under two levels of trial sample sizes. Cohort size is fixed at  $m = 4000$ , and the true PATE  $\Delta = 2.8$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>Trial sample size <math>n \approx 1000</math>, cohort size <math>m = 4000</math></i>						
$\hat{\Delta}_{IPSW1}$	✓	–	0.01	18.7	17.2	94.2
$\hat{\Delta}_{IPSW1}$	×	–	-0.15	28.7	26.6	95.6
$\hat{\Delta}_{IPSW2}$	✓	–	0.00	23.3	20.9	92.8
$\hat{\Delta}_{IPSW2}$	×	–	-0.62	29.7	27.6	37.3
$\hat{\Delta}_{REG}$	–	✓	0.00	10.5	11.3	96.4
$\hat{\Delta}_{REG}$	–	×	-0.17	12.7	13.2	76.0
$\hat{\Delta}_{DR1}$	✓	✓	0.00	11.6	12.1	96.1
$\hat{\Delta}_{DR1}$	✓	×	-0.01	14.3	14.4	95.5
$\hat{\Delta}_{DR1}$	×	✓	0.00	14.0	13.9	95.8
$\hat{\Delta}_{DR1}$	×	×	0.83	29.4	27.1	5.8
$\hat{\Delta}_{DR2}$	✓	✓	0.00	11.5	12.0	96.0
$\hat{\Delta}_{DR2}$	✓	×	-0.01	14.3	14.2	95.2
$\hat{\Delta}_{DR2}$	×	✓	0.00	12.7	12.8	95.6
$\hat{\Delta}_{DR2}$	×	×	0.63	20.7	19.2	6.7
<i>Trial sample size <math>n \approx 200</math>, cohort size <math>m = 4000</math></i>						
$\tilde{\Delta}_{IPSW1}$	✓	–	0.01	40.2	31.9	91.2
$\tilde{\Delta}_{IPSW1}$	×	–	-0.15	65.9	52.1	93.8
$\tilde{\Delta}_{IPSW2}$	✓	–	0.06	44.5	36.0	87.8
$\tilde{\Delta}_{IPSW2}$	×	–	-0.55	56.4	46.1	80.7
$\tilde{\Delta}_{REG}$	–	✓	0.00	21.8	21.3	94.3
$\tilde{\Delta}_{REG}$	–	×	-0.19	25.9	25.5	88.2
$\tilde{\Delta}_{DR1}$	✓	✓	0.00	24.0	22.7	93.6
$\tilde{\Delta}_{DR1}$	✓	×	-0.02	29.5	26.9	93.3
$\tilde{\Delta}_{DR1}$	×	✓	0.00	28.9	25.6	92.9
$\tilde{\Delta}_{DR1}$	×	×	0.79	62.6	48.9	67.9
$\tilde{\Delta}_{DR2}$	✓	✓	0.00	23.6	22.2	93.4
$\tilde{\Delta}_{DR2}$	✓	×	-0.02	28.4	25.9	92.6
$\tilde{\Delta}_{DR2}$	×	✓	0.00	25.4	22.8	92.0
$\tilde{\Delta}_{DR2}$	×	×	0.56	40.9	33.4	61.4

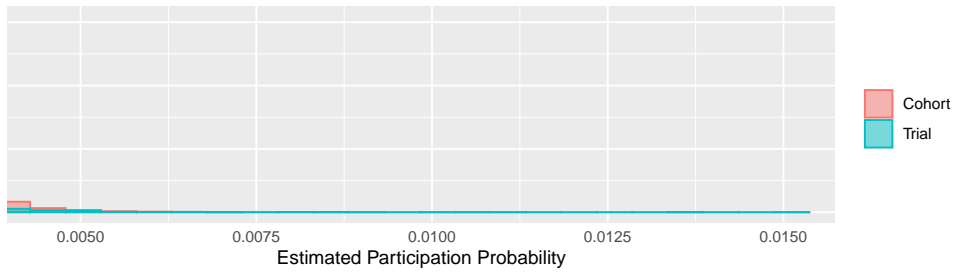
Table 16 Comparison of performance of five different estimators coupled with estimated propensity scores for estimating PATE with 5000 simulated data replications with  $(\gamma_1, \gamma_2, \gamma_3) = (0.6, 0.6, 0.6)$  and  $(\alpha_1, \alpha_2, \alpha_3) = (2, 2, 2)$  under two levels of trial sample sizes. Cohort size is fixed at  $m = 800$ , and the true PATE  $\Delta = 2.8$ . ESE: Empirical standard error; ASE: Average of the estimated standard errors.

Estimator	Correct $w(Z_i; \gamma)$	Correct $m_x(Z_i; \alpha_x)$	Bias	ESE ( $\times 100$ )	ASE ( $\times 100$ )	Coverage ( $\times 100$ )
<i>Trial sample size <math>n \approx 1000</math>, cohort size <math>m = 800</math></i>						
$\hat{\Delta}_{\text{IPSW1}}$	✓	–	-0.01	30.1	25.4	97.3
$\hat{\Delta}_{\text{IPSW1}}$	×	–	-0.17	34.7	33.3	97.9
$\hat{\Delta}_{\text{IPSW2}}$	✓	–	-0.06	34.9	29.2	93.8
$\hat{\Delta}_{\text{IPSW2}}$	×	–	-0.66	40.5	37.1	59.4
$\hat{\Delta}_{\text{REG}}$	–	✓	0.00	14.5	17.4	98.3
$\hat{\Delta}_{\text{REG}}$	–	×	-0.17	17.4	19.6	88.7
$\hat{\Delta}_{\text{DR1}}$	✓	✓	0.00	16.4	18.5	98.2
$\hat{\Delta}_{\text{DR1}}$	✓	×	-0.03	22.1	22.0	97.4
$\hat{\Delta}_{\text{DR1}}$	×	✓	0.00	17.6	19.9	97.6
$\hat{\Delta}_{\text{DR1}}$	×	×	0.88	39.7	37.6	25.8
$\hat{\Delta}_{\text{DR2}}$	✓	✓	0.00	15.5	18.2	98.1
$\hat{\Delta}_{\text{DR2}}$	✓	×	-0.03	20.7	21.0	96.3
$\hat{\Delta}_{\text{DR2}}$	×	✓	0.00	16.2	18.6	97.6
$\hat{\Delta}_{\text{DR2}}$	×	×	0.64	24.4	24.5	22.0
<i>Trial sample size <math>n \approx 200</math>, cohort size <math>m = 800</math></i>						
$\tilde{\Delta}_{\text{IPSW1}}$	✓	–	0.01	53.0	39.8	93.5
$\tilde{\Delta}_{\text{IPSW1}}$	×	–	-0.16	81.3	59.9	95.2
$\tilde{\Delta}_{\text{IPSW2}}$	✓	–	0.01	52.9	41.4	88.8
$\tilde{\Delta}_{\text{IPSW2}}$	×	–	-0.58	63.6	52.1	83.0
$\tilde{\Delta}_{\text{REG}}$	–	✓	0.00	24.1	25.2	96.0
$\tilde{\Delta}_{\text{REG}}$	–	×	-0.18	29.0	29.4	91.0
$\tilde{\Delta}_{\text{DR1}}$	✓	✓	0.00	27.8	27.6	95.6
$\tilde{\Delta}_{\text{DR1}}$	✓	×	-0.04	37.9	33.2	95.2
$\tilde{\Delta}_{\text{DR1}}$	×	✓	0.00	34.0	30.6	95.4
$\tilde{\Delta}_{\text{DR1}}$	×	×	0.86	77.1	58.6	75.7
$\tilde{\Delta}_{\text{DR2}}$	✓	✓	0.00	26.4	26.4	95.2
$\tilde{\Delta}_{\text{DR2}}$	✓	×	-0.03	33.2	30.5	94.0
$\tilde{\Delta}_{\text{DR2}}$	×	✓	0.00	28.0	26.8	94.5
$\tilde{\Delta}_{\text{DR2}}$	×	×	0.58	43.9	36.7	64.0

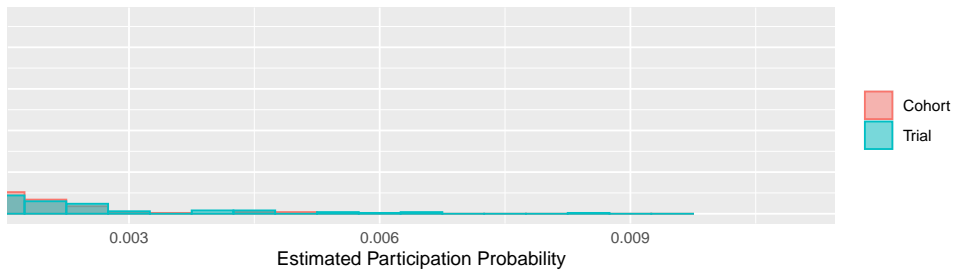
(a) Trial (320) and cohort (CNICS)



(b) Trial (A5202) and cohort (CNICS)



(c) Trial (320) and cohort (WIHS)



(d) Trial (A5202) and cohort (WIHS)

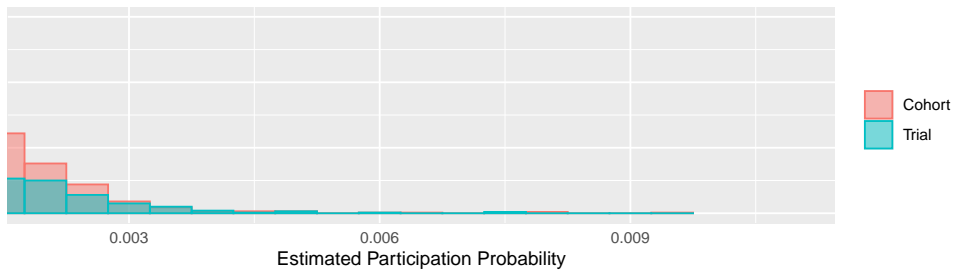


Figure 1 Zoomed histograms (right tails) of estimated sampling scores for each of the four generalizability analyses.

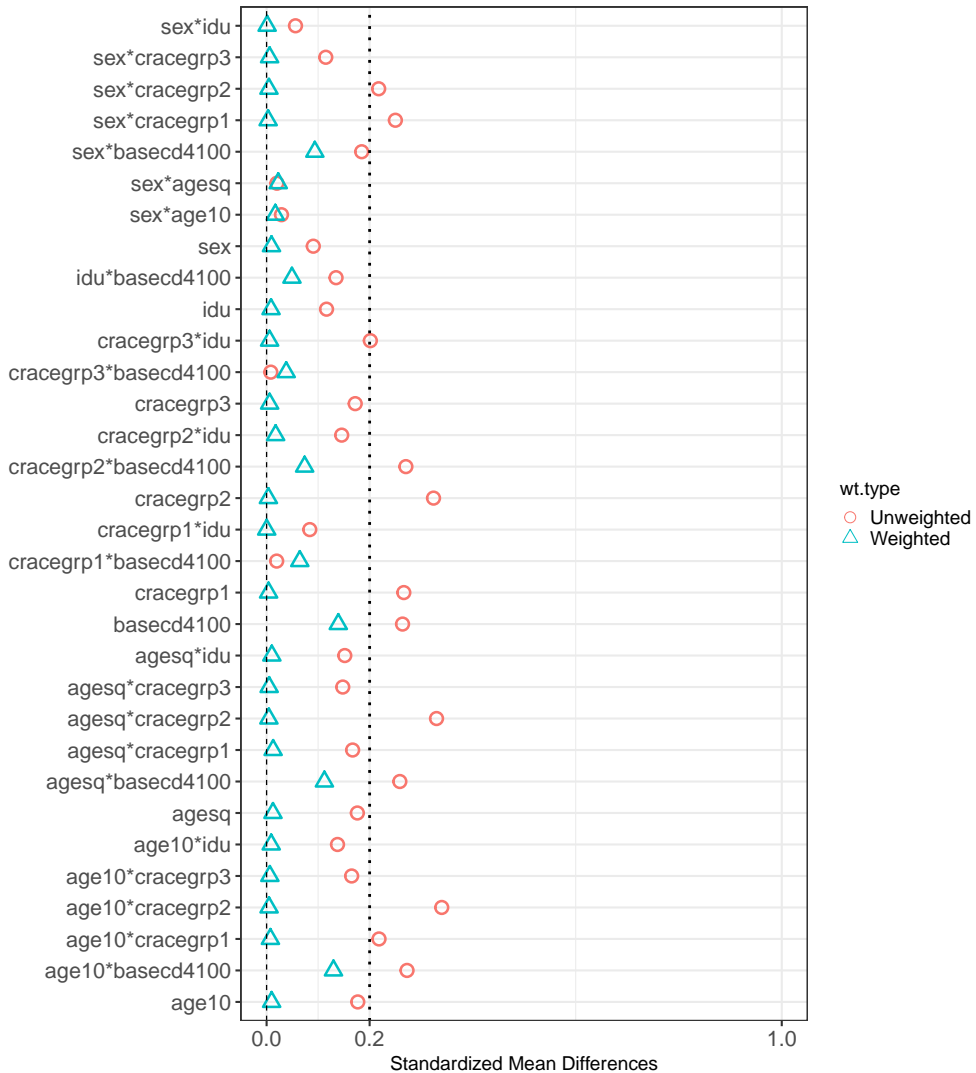


Figure 2 Forest plot of the standard mean differences of all covariates (and interaction terms) for the unweighted trial sample and inverse probability of participation weighted trial sample, when generalizing ACTG 320 to all people living with HIV in the USA based on the CNICS cohort.

<sup>a</sup> “wt.type” indicates the type of estimator (unweighted or sampling score weighted); “sex” is biological sex, “age10” is age in deciles, “agesq” is age in deciles squared, “cracegrp1” is an indicator for White, non-Hispanic race, “cracegrp2” is an indicator for Black, non-Hispanic race, “cracegrp3” is an indicator for Hispanic, “idu” is an indicator for injection drug use, “basecd4100” is CD4 cell count at baseline/100.



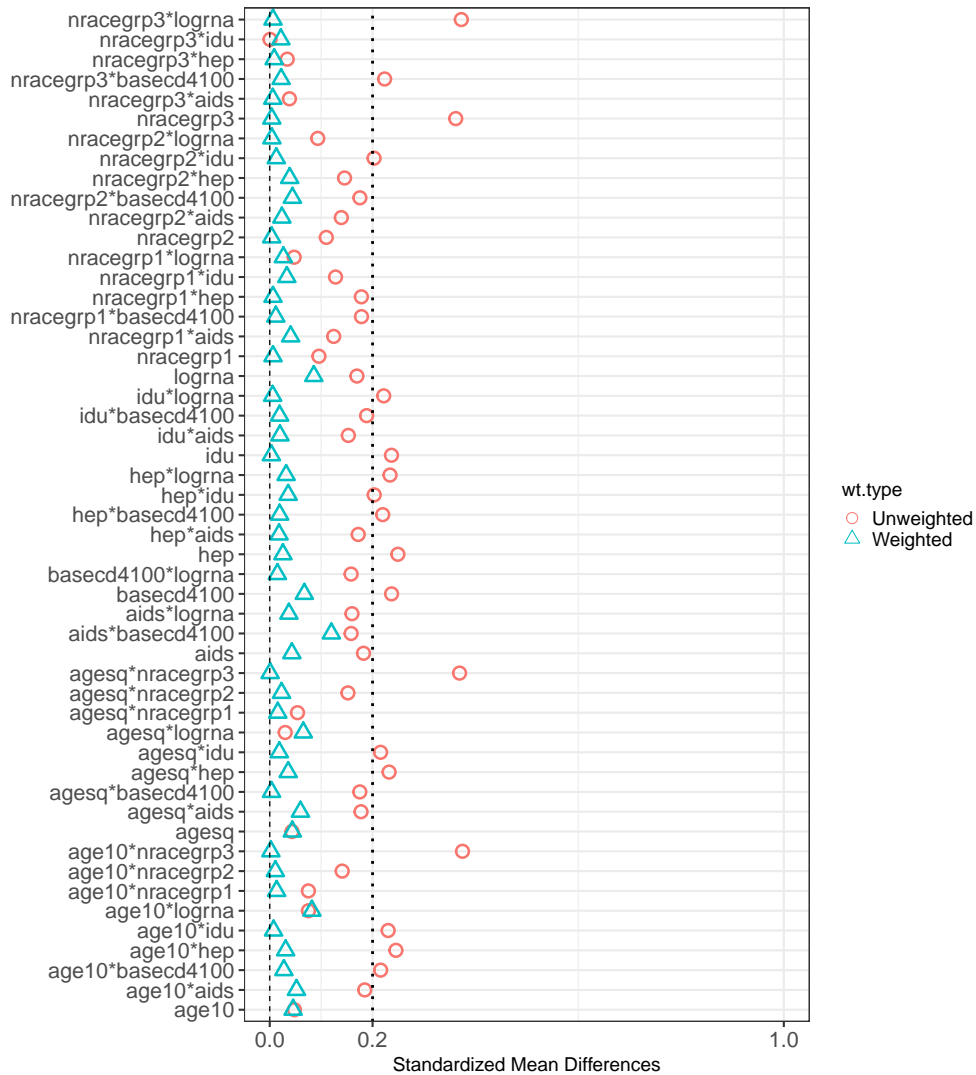


Figure 3 Forest plot of the standard mean differences of all covariates (and interaction terms) for the unweighted trial sample and inverse probability of participation weighted trial sample, when generalizing ACTG A5202 to all people living with HIV in the USA based on the CNICS cohort.

<sup>a</sup> “wt.type” indicates the type of estimator (unweighted or sampling score weighted); “age10” is age in deciles, “agesq” is age in deciles squared, “nracegrp1” is an indicator for White, non-Hispanic race, “nracegrp2” is an indicator for Black, non-Hispanic race, “nracegrp3” is an indicator for Hispanic, “hep” is an indicator for Hepatitis B/C, “idu” is an indicator for injection drug use, “aids” is an indicator for AIDS diagnosis, “basecd4100” is CD4 cell count at baseline/100, “logrna” is log10 viral load.

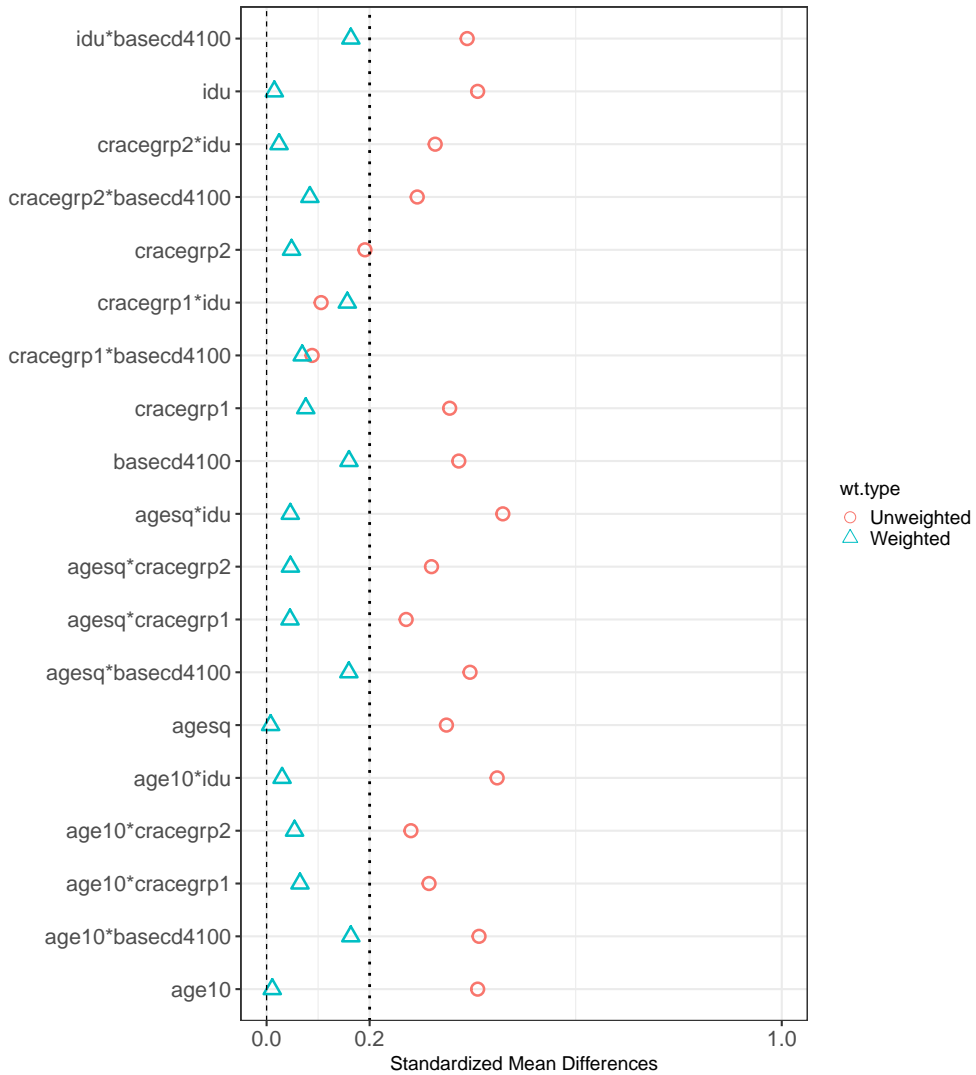


Figure 4 Forest plot of the standard mean differences of all covariates (and interaction terms) for the unweighted trial sample and inverse probability of participation weighted trial sample, when generalizing ACTG 320 to all women living with HIV in the USA based on the WIHS cohort.

<sup>a</sup> “wt.type” indicates the type of estimator (unweighted or sampling score weighted); “age10” is age in deciles, “agesq” is age in deciles squared, “cracegrp1” is an indicator for White, non-Hispanic race, “cracegrp2” is an indicator for Black, non-Hispanic race, “idu” is an indicator for injection drug use, “basecd4100” is CD4 cell count at baseline/100

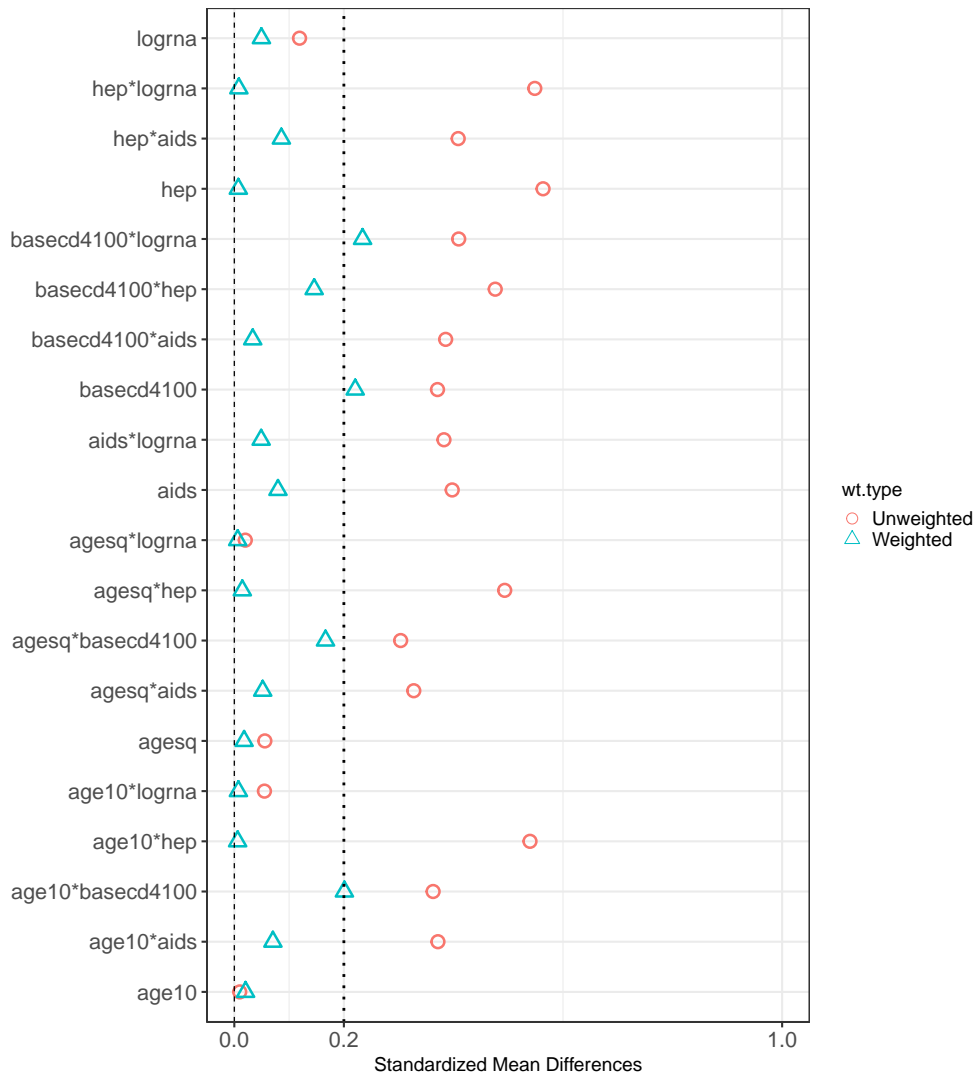


Figure 5 Forest plot of the standard mean differences of all covariates (and interaction terms) for the unweighted trial sample and inverse probability of participation weighted trial sample, when generalizing ACTG A5202 to all women living with HIV in the USA based on the WIHS cohort.

<sup>a</sup> “wt.type” indicates the type of estimator (unweighted or sampling score weighted); “age10” is age in deciles, “agesq” is age in deciles squared, “hep” is an indicator for Hepatitis B/C, “aids” is an indicator for AIDS diagnosis, “logrna” is log10 viral load.

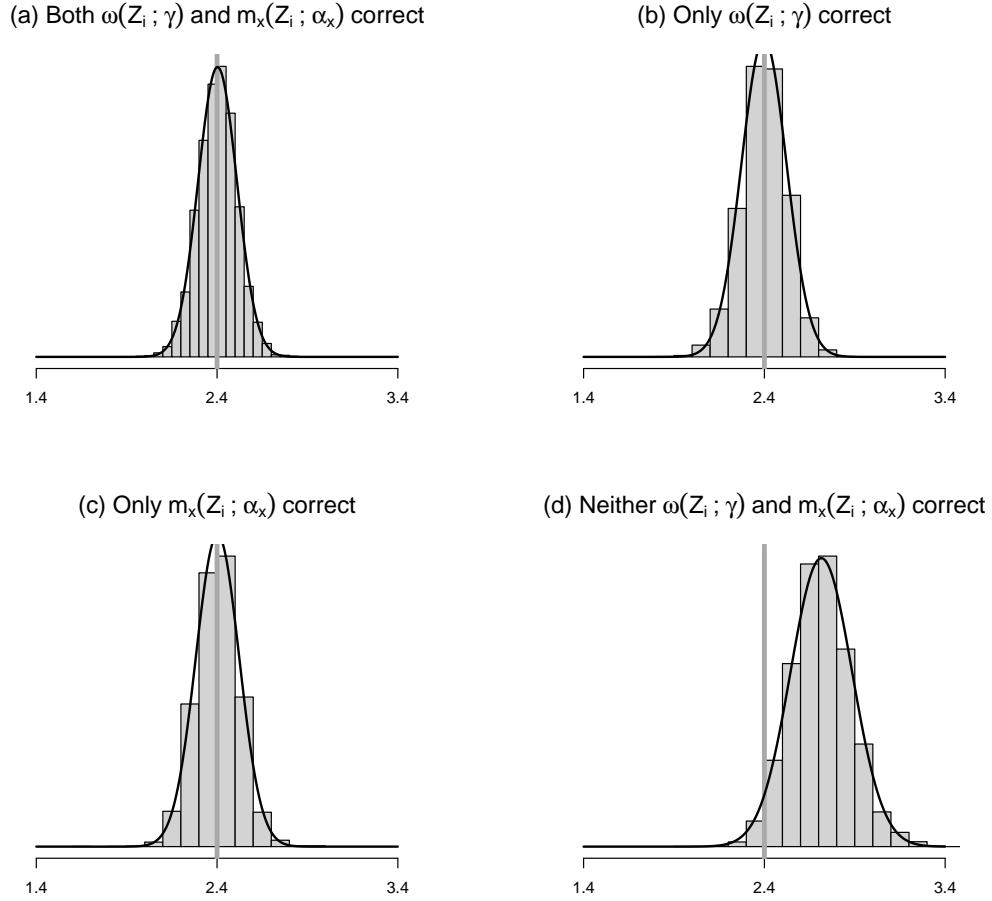


Figure 6 Empirical distributions (based on 5,000 simulations) of  $\widehat{\Delta}_{DR2}$  when (a) all models are correctly specified, (b) only trial participation model is correctly specified, (c) only outcome models are correctly specified and (d) all models are incorrectly specified, and under moderate selection effect,  $\gamma = (-7.698, 0.6, 0.6, 0.6)^T$  and moderate effect modification  $\zeta = (1, 1, 1)^T$ . The propensity scores were estimated from the trial data. The true PATE  $\Delta = 2.4$  is indicated by the gray vertical line. The histograms are overlaid with the normal density curves with the corresponding empirical variance.

## References

- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, New York: Springer.
- White, H. (1982), “Maximul Likelihood Estimation of misspecified models,” *Econometrica*, 50, 1–25.
- Yang, D. and Dalton, J. E. (2012), “A unified approach to measuring the effect size between two groups using SAS®,” in *SAS Global Forum*, volume 335, 1–6.