# Supplement: Doubly Robust Transfer Learning Under Sub-group Shift for Cohort-Level Missing Indicator Covariates

Huali Zhao

*Department of Mathematical Sciences, Tsinghua University*

Tianying Wang

*Department of Statistics, Colorado State University*

### Supplementary Material

This supplement includes the potential nuisance models, the cross-fitted version of the proposed method, proofs of all technical details, and additional simulation and data analysis results.

## S1 Potential nuisance models and the cross-fitted version of DRTL-comb

In this section, we first provide the alternative working nuisance models for the density ratio $\mathbb{w}(y, \boldsymbol{z})$ and the imputation model $\mathbb{m}(y, \boldsymbol{z})$. Then we provide the cross-fitted version of the DRTL-comb method when these nuisance functions are estimated with flexible machine learning methods.

## S1.1 Potential nuisance models

To explore flexibility in modeling the nuisance models, we outline alternative approaches adaptable to the study's context. For the density ratio $\mathbb{w}(y, \boldsymbol{z})$, one can adopt a semi-parametric working model:

$$\omega(y, \boldsymbol{z}) = \exp\left\{\mathbf{b}(y, \boldsymbol{z})^\top \boldsymbol{\eta} + r(y, \boldsymbol{z})\right\},$$

where $\mathbf{b}(y, \boldsymbol{z}) \in \mathbb{R}^{1+p}$ denotes a vector of pre-specified basis functions capturing nonlinear effects, $\boldsymbol{\eta} \in \mathbb{R}^{1+p}$ represents the associated nuisance parameters, and $r(y, \boldsymbol{z})$ is an unknown nonparametric function, as proposed by Liu et al. (2023). The term $\mathbf{b}(y, \boldsymbol{z})^\top \boldsymbol{\eta}$ captures parametric components. Further, one can adopt the divergence-based method of Yan and Chen (2024) to estimate the density ratio function, which is amenable to machine learning algorithms including the deep learning. This method is more stable than the kernel smoothing or the classification-based methods to estimate the density ratio.

For the imputation model $\mathbb{m}(y, \boldsymbol{z})$, one can also use some machine learning algorithm (e.g., random forest or neural network) to learn the mean $E(X \mid Y, \boldsymbol{Z})$ directly. Adopting these approaches would require the cross-fitted version of our DRTL-comb method, and require re-deriving the asymptotic theoretical results. We will discuss these in the next section.

## S1.2 The cross-fitted version of DRTL-comb

In general, assume that $\widehat{\omega}(y, \boldsymbol{z})$ and $\widehat{m}(y, \boldsymbol{z})$ are estimators of $\omega(y, \boldsymbol{z})$ and $m(y, \boldsymbol{z})$ obtained via nonparametric or flexible machine learning algorithms. We now present a general cross-fitting framework of the DRTL-comb method.

Following the idea of Chernozhukov et al. (2018), we apply cross-fitting on the source sample to eliminate the dependence between the nuisance estimators and the samples on which they are evaluated. Specifically, we randomly split the source sample into $K$ disjoint subsets of equal size, each containing $n_{\mathcal{S}}/K$ observations, indexed by $\mathcal{I}_1, \ldots, \mathcal{I}_K$, with $\{1, \ldots, n_{\mathcal{S}}\} = \cup_{k=1}^{K} \mathcal{I}_k$, and denote $\mathcal{I}_{-k} = \{1, \ldots, n_{\mathcal{S}}\} \setminus \mathcal{I}_k$. For each $k \in [K]$, let $\widehat{\omega}^{[-k]}(y, \boldsymbol{z})$ and $\widehat{m}^{[-k]}(y, \boldsymbol{z})$ denote the estimators of $\omega(y, \boldsymbol{z})$ and $m(y, \boldsymbol{z})$ constructed using the observations in $\mathcal{I}_{-k} \cup \{n_{\mathcal{S}} + 1, \ldots, n_{\mathcal{S}} + n_{\mathcal{T}}\}$, and let $\widehat{m}(y, z) = K^{-1} \sum_{k=1}^{K} \widehat{m}^{[-k]}(y, \boldsymbol{z})$. The resulting cross-fitted estimating equation is then given by

$$
\begin{aligned}
\widehat{U}_{\mathrm{DML}}(\beta_x, \boldsymbol{\beta}_z) := &\frac{1}{n_{\mathcal{S}}} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \widehat{\omega}^{[-k]}(Y_i, \boldsymbol{Z}_i)\{X_i - \widehat{m}^{[-k]}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_x - \boldsymbol{Z}_i^{\top} \boldsymbol{\beta}_z) \\
&+ \frac{1}{n_{\mathcal{T}}} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \widehat{m}(Y_i, \boldsymbol{Z}_i)(Y_i - \beta_x - \boldsymbol{Z}_i^{\top} \boldsymbol{\beta}_z) = 0, \qquad \text{(S1.1)}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\boldsymbol{V}}_{\mathrm{DML}}(\beta_x, \boldsymbol{\beta}_z) := &\frac{1}{n_{\mathcal{S}}} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \widehat{\omega}^{[-k]}(Y_i, \boldsymbol{Z}_i) \boldsymbol{Z}_i \{\widehat{m}^{[-k]}(Y_i, \boldsymbol{Z}_i) - X_i\} \beta_x \\
&+ \frac{1}{n_{\mathcal{T}}} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{Z}_i \{Y_i - \widehat{m}(Y_i, \boldsymbol{Z}_i) \beta_x - \boldsymbol{Z}_i^{\top} \boldsymbol{\beta}_z\} = \boldsymbol{0}. \qquad \text{(S1.2)}
\end{aligned}
$$

Define the solution of Eqs (S1.1)-(S1.2) as $(\widehat{\beta}_{\mathrm{DML,x}}, \widehat{\boldsymbol{\beta}}_{\mathrm{DML,z}}^{\top})^{\top}$.

Let $\omega^*(y, \boldsymbol{z})$ and $m^*(y, \boldsymbol{z})$ denote the corresponding best approximations of $\widehat{\omega}(y, \boldsymbol{z})$ and $\widehat{m}(y, \boldsymbol{z})$. Assume that the nuisance estimators satisfy

$$\|\widehat{\omega}(y, \boldsymbol{z}) - \omega^*(y, \boldsymbol{z})\|_2 = o(n^{-1/4}), \quad \|\widehat{m}(y, \boldsymbol{z}) - m^*(y, \boldsymbol{z})\|_2 = o(n^{-1/4}). \tag{S1.3}$$

These regularity conditions on nuisance estimators are standard and widely adopted in double/debiased machine learning literature (Chernozhukov et al., 2018). Assume that the correct specification conditions are

$$\text{Either } \omega^*(y, \boldsymbol{z}) = \mathbb{w}(y, \boldsymbol{z}), \text{ or } m^*(y, \boldsymbol{z}) = \mathbb{m}(y, \boldsymbol{z}). \tag{S1.4}$$

This assumption is necessary to establish the double robustness property. Under Assumptions (S1.3)-(S1.4), one can follow the double machine learning framework of Chernozhukov et al. (2018) to derive the desirable double robustness of the DRTL-comb estimator. The root-$n$ consistency and the asymptotic results can be established similar to Liu et al. (2023). We leave the details to readers who are interested in developing machine-learning–based versions of the nuisance models.

# S2 Proof of Theorem 1

*Proof.* Let $\|\cdot\|_\infty$ represent the maximum norm of a vector or matrix. First, we derive the error rates for $\widehat{\beta}_x$ and $\widehat{\boldsymbol{\beta}}_z$. We expand the left side of the Eq (2.9) as

$$\widehat{U}_{\mathrm{DR}}(\beta_x, \boldsymbol{\beta}_z) = U(\beta_x, \boldsymbol{\beta}_z) + \Delta_{11} + \Delta_{12} + \Delta_{13}, \tag{S2.5}$$

where

$$\begin{aligned}
U(\beta_x, \boldsymbol{\beta}_z) &= \frac{1}{n_{\mathcal{S}}} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i)\{X_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_x - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_z) + \frac{1}{n_{\mathcal{T}}} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \bar{m}(Y_i, \boldsymbol{Z}_i)(Y_i - \beta_x - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_z), \\
\Delta_{11} &= \frac{1}{n_{\mathcal{S}}} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\}\{\bar{m}(Y_i, \boldsymbol{Z}_i) - \widehat{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_x - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_z), \\
\Delta_{12} &= \frac{1}{n_{\mathcal{S}}} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i)\{\bar{m}(Y_i, \boldsymbol{Z}_i) - \widehat{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_x - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_z) \\
&\quad + \frac{1}{n_{\mathcal{T}}} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_x - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_z), \\
\Delta_{13} &= \frac{1}{n_{\mathcal{S}}} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\}\{X_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_x - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_z).
\end{aligned}$$

Similarly, we expand the left side of the Eq (2.10) as

$$\widehat{\boldsymbol{V}}_{\mathrm{DR}}(\beta_x, \boldsymbol{\beta}_z) = \boldsymbol{V}(\beta_x, \boldsymbol{\beta}_z) + \boldsymbol{\Delta}_{21} + \boldsymbol{\Delta}_{22} + \boldsymbol{\Delta}_{23}, \tag{S2.6}$$

where

$$\begin{aligned}
\boldsymbol{V}(\beta_x, \boldsymbol{\beta}_z) &= \frac{1}{n_{\mathcal{S}}} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i)\boldsymbol{Z}_i\{\bar{m}(Y_i, \boldsymbol{Z}_i) - X_i\}\beta_x + \frac{1}{n_{\mathcal{T}}} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{Z}_i\{Y_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\beta_x - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_z\}, \\
\boldsymbol{\Delta}_{21} &= \frac{1}{n_{\mathcal{S}}} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\}\boldsymbol{Z}_i\{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}\beta_x,
\end{aligned}$$

$$\boldsymbol{\Delta}_{22} = \frac{1}{n_{\mathcal{S}}}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\bar{\omega}(Y_i,\boldsymbol{Z}_i)\boldsymbol{Z}_i\{\widehat{m}(Y_i,\boldsymbol{Z}_i)-\bar{m}(Y_i,\boldsymbol{Z}_i)\}\beta_x - \frac{1}{n_{\mathcal{T}}}\sum_{i\in\mathcal{I}_{\mathcal{T}}}\boldsymbol{Z}_i\{\widehat{m}(Y_i,\boldsymbol{Z}_i)-\bar{m}(Y_i,\boldsymbol{Z}_i)\}\beta_x,$$

$$\boldsymbol{\Delta}_{23} = \frac{1}{n_{\mathcal{S}}}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i)-\bar{\omega}(Y_i,\boldsymbol{Z}_i)\}\boldsymbol{Z}_i(\bar{m}(Y_i,\boldsymbol{Z}_i)-X_i)\beta_x.$$

Using Theorem 5.21 in Van der Vaart (2000), Chebyshev's inequality, and $n_{\mathcal{S}}/n_{\mathcal{T}} = O(1)$, we have

$$n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i)-\bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2 - E_{\mathcal{S}}\{\widehat{\omega}(Y,\boldsymbol{Z})-\bar{\omega}(Y,\boldsymbol{Z})\}^2 = o_p(n_{\mathcal{S}}^{-1/2}),$$

$$n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i)-\bar{m}(Y_i,\boldsymbol{Z}_i)\}^2 - E_{\mathcal{S}}\{\widehat{m}(Y,\boldsymbol{Z})-\bar{m}(Y,\boldsymbol{Z})\}^2 = o_p(n_{\mathcal{S}}^{-1/2}),$$

$$n_{\mathcal{T}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{T}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i)-\bar{m}(Y_i,\boldsymbol{Z}_i)\}^2 - E_{\mathcal{T}}\{\widehat{m}(Y,\boldsymbol{Z})-\bar{m}(Y,\boldsymbol{Z})\}^2 = o_p(n_{\mathcal{S}}^{-1/2}).$$

Also, by Assumption 1, we have that

$$E_{\mathcal{S}}\{\widehat{\omega}(Y,\boldsymbol{Z})-\bar{\omega}(Y,\boldsymbol{Z})\}^2 = E_{\mathcal{S}}[\bar{\omega}^2(Y,\boldsymbol{Z})\{\widehat{\omega}(Y,\boldsymbol{Z})/\bar{\omega}(Y,\boldsymbol{Z})-1\}^2]$$

$$\leq E_{\mathcal{S}}[\bar{\omega}^2(Y,\boldsymbol{Z})\{2Y^2(\widehat{\eta}_y-\bar{\eta}_y)^2 + Y^4(\widehat{\eta}_y-\bar{\eta}_y)^4 + 2\|\boldsymbol{Z}\|_2^2\|\widehat{\boldsymbol{\eta}}_z-\bar{\boldsymbol{\eta}}_z\|_2^2 + \|\boldsymbol{Z}\|_2^4\|\widehat{\boldsymbol{\eta}}_z-\bar{\boldsymbol{\eta}}_z\|_2^4\}]$$

$$\leq E_{\mathcal{S}}\left[\{2\bar{\omega}^4(Y,\boldsymbol{Z}) + Y^4 + Y^{16} + O_p(n_{\mathcal{S}}^{-1})\}(\widehat{\eta}_y-\bar{\eta}_y)^2\right]$$

$$+ E_{\mathcal{S}}\left[\{2\bar{\omega}^4(Y,\boldsymbol{Z}) + \|\boldsymbol{Z}\|_2^4 + \|\boldsymbol{Z}\|_2^{16} + O_p(n_{\mathcal{S}}^{-1})\}\|\widehat{\boldsymbol{\eta}}_z-\bar{\boldsymbol{\eta}}_z\|_2^2\right]$$

$$= o_p(n_{\mathcal{S}}^{-1/2}).$$

And that for each $\iota \in \{\mathcal{T},\mathcal{S}\}$,

$$E_\iota\{\widehat{m}(Y,\boldsymbol{Z})-\bar{m}(Y,\boldsymbol{Z})\}^2$$

$$\leq E_\iota\{\breve{g}^2(\bar{m}(Y,\boldsymbol{Z}))\left(\|(Y,\boldsymbol{Z}^\top)^\top\|_2^2 + \|\widehat{\boldsymbol{\gamma}}-\bar{\boldsymbol{\gamma}}\|_2^2\right) + C_L^2\left(\|(Y,\boldsymbol{Z}^\top)^\top\|_2^4 + \|\widehat{\boldsymbol{\gamma}}-\bar{\boldsymbol{\gamma}}\|_2^4\right)\}$$

$$= O_p\left(E_\iota\{\breve{g}^2(\bar{m}(Y,\boldsymbol{Z}))\left(\|(Y,\boldsymbol{Z}^\top)^\top\|_2^2 + \|\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}\|_2^2\right)\} + n_{\mathcal{S}}^{-1}\right) = o_p(n_{\mathcal{S}}^{-1/2}),$$

where $\breve{g}(a) = \dot{g}(g^{-1}(a))$. Thus, we have

$$n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i) - \bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2 = o_p(n_{\mathcal{S}}^{-1/2}),$$

$$n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2 = o_p(n_{\mathcal{S}}^{-1/2}),$$

$$n_{\mathcal{T}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{T}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2 = o_p(n_{\mathcal{S}}^{-1/2}).$$

Combining these with Assumption 1 and Eqs (S2.5)-(S2.6), we have

$$\|\Delta_{11}\|_\infty \le n_{\mathcal{S}}^{-1}\max_{i\in\mathcal{I}_{\mathcal{S}}}(Y_i - \beta_x - \boldsymbol{Z}_i^\top\boldsymbol{\beta}_z)\sum_{i\in\mathcal{I}_{\mathcal{S}}}[\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i) - \bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2 + \{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2]$$

$$= o_p(n_{\mathcal{S}}^{-1/2}),$$

$$\|\Delta_{12}\|_\infty \le \max_{i\in\mathcal{I}_{\mathcal{S}}}(Y_i - \beta_x - \boldsymbol{Z}_i^\top\boldsymbol{\beta}_z)\{n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\bar{\omega}^2(Y_i,\boldsymbol{Z}_i)\}^{1/2}[n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2}$$

$$+ \max_{i\in\mathcal{I}_{\mathcal{T}}}(Y_i - \beta_x - \boldsymbol{Z}_i^\top\boldsymbol{\beta}_z)[n_{\mathcal{T}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{T}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2}$$

$$= o_p(n_{\mathcal{S}}^{-1/4}),$$

$$\|\Delta_{13}\|_\infty \le \max_{i\in\mathcal{I}_{\mathcal{S}}}(Y_i - \beta_x - \boldsymbol{Z}_i^\top\boldsymbol{\beta}_z)[n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i) - \bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2}[n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{X_i^2 + \bar{m}^2(y_i,\boldsymbol{Z}_i)\}]^{1/2}$$

$$= o_p(n_{\mathcal{S}}^{-1/4}),$$

$$\|\boldsymbol{\Delta}_{21}\|_\infty \le n_{\mathcal{S}}^{-1}\max_{i\in\mathcal{I}_{\mathcal{S}}}\|\boldsymbol{Z}_i\|_\infty|\beta_x|\sum_{i\in\mathcal{I}_{\mathcal{S}}}[\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i) - \bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2 + \{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2]$$

$$= o_p(n_{\mathcal{S}}^{-1/2}),$$

$$\|\boldsymbol{\Delta}_{22}\|_\infty \le \max_{i\in\mathcal{I}_\mathcal{S}} \|\boldsymbol{Z}_i\|_\infty |\beta_x| \{n_\mathcal{S}^{-1} \sum_{i\in\mathcal{I}_\mathcal{S}} \bar{\omega}^2(Y_i,\boldsymbol{Z}_i)\}^{1/2} [n_\mathcal{S}^{-1} \sum_{i\in\mathcal{I}_\mathcal{S}} \{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2}$$

$$+ \max_{i\in\mathcal{I}_\mathcal{T}} \|\boldsymbol{Z}_i\|_\infty |\beta_x| [n_\mathcal{T}^{-1} \sum_{i\in\mathcal{I}_\mathcal{T}} \{\widehat{m}(Y_i,\boldsymbol{Z}_i) - \bar{m}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2}$$

$$= o_p(n_\mathcal{S}^{-1/4}),$$

$$\|\boldsymbol{\Delta}_{23}\|_\infty \le \max_{i\in\mathcal{I}_\mathcal{S}} \|\boldsymbol{Z}_i\|_\infty |\beta_x| [n_\mathcal{S}^{-1} \sum_{i\in\mathcal{I}_\mathcal{S}} \{\widehat{\omega}(Y_i,\boldsymbol{Z}_i) - \bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2} [n_\mathcal{S}^{-1} \sum_{i\in\mathcal{I}_\mathcal{S}} \{X_i^2 + \bar{m}^2(y_i,\boldsymbol{Z}_i)\}]^{1/2}$$

$$= o_p(n_\mathcal{S}^{-1/4}).$$

Thus, $(\widehat{\beta}_x, \widehat{\boldsymbol{\beta}}_z)$ solve

$$\begin{cases} U(\beta_x,\boldsymbol{\beta}_z) + o_p(n_\mathcal{S}^{-1/4}) = 0, \\ \boldsymbol{V}(\beta_x,\boldsymbol{\beta}_z) + o_p(n_\mathcal{S}^{-1/4}) = \mathbf{0}. \end{cases}$$

Let $(\bar{\beta}_x, \bar{\boldsymbol{\beta}}_z)$ be the solution of

$$\begin{cases} E\{U(\beta_x,\boldsymbol{\beta}_z)\} = 0, \\ E\{\boldsymbol{V}(\beta_x,\boldsymbol{\beta}_z)\} = \mathbf{0}. \end{cases} \tag{S2.7}$$

When $\bar{\omega}(\cdot) = \mathrm{w}(\cdot)$, we have that

$$E\{U(\beta_x,\boldsymbol{\beta}_z)\} = E_\mathcal{S}[\mathrm{w}(Y,\boldsymbol{Z})\{X - \bar{m}(Y,\boldsymbol{Z})\}(Y - \beta_x - \boldsymbol{Z}^\top\boldsymbol{\beta}_z)] + E_\mathcal{T}\{\bar{m}(Y,\boldsymbol{Z})(Y - \beta_x - \boldsymbol{Z}^\top\boldsymbol{\beta}_z)\}$$

$$= E_\mathcal{T}\{X(Y - \beta_x - \boldsymbol{Z}^\top\boldsymbol{\beta}_z)\}.$$

$$E\{\boldsymbol{V}(\beta_x,\boldsymbol{\beta}_z)\} = E_\mathcal{S}[\mathrm{w}(Y,\boldsymbol{Z})\{\bar{m}(Y,\boldsymbol{Z}) - X\}\beta_x] + E_\mathcal{T}[\boldsymbol{Z}\{Y - \bar{m}(Y,\boldsymbol{Z})\beta_x - Z^\top\theta\}]$$

$$= E_\mathcal{T}\{\boldsymbol{Z}(Y - X\beta_x - \boldsymbol{Z}^\top\boldsymbol{\beta}_z)\}.$$

When $\bar{m}(\cdot) = \mathbb{m}(\cdot)$, we have that

$$E\{U(\beta_x, \boldsymbol{\beta}_z)\} = 0 + E_{\mathcal{T}}\{X(Y - \beta_x - \boldsymbol{Z}^\top \boldsymbol{\beta}_z)\},$$

$$E\{\boldsymbol{V}(\beta_x, \boldsymbol{\beta}_z)\} = E_{\mathcal{T}}\{\boldsymbol{Z}(Y - X\beta_x - \boldsymbol{Z}^\top \boldsymbol{\beta}_z)\}.$$

Both cases lead to that $(\beta_{x0}, \boldsymbol{\beta}_{z0})$ be the solution of Eq (S2.7). So under Assumption 2, we have $\bar{\beta}_x = \beta_{x0}$ and $\bar{\boldsymbol{\beta}}_z = \boldsymbol{\beta}_{z0}$. By Assumption 1, both $U(\beta_x, \boldsymbol{\beta}_z)$ and $\boldsymbol{V}(\beta_x, \boldsymbol{\beta}_z)$ are continuous differential on $\beta_x$ and $\boldsymbol{\beta}_z$. Then, using the standard empirical process theory of Van der Vaart (2000)[Theorem 5.21], we have

$$\widehat{\beta}_x - \beta_{x0} = o_p(1), \quad \|\widehat{\boldsymbol{\beta}}_z - \boldsymbol{\beta}_{z0}\|_2 = o_p(1).$$

□

## S3 Proof of Theorem 2

*Proof.* We consider the asymptotic expansion of $\sqrt{n_{\mathcal{S}}}\ \boldsymbol{c}^\top(\widehat{\beta}_x - \beta_{x0}, \widehat{\boldsymbol{\beta}}_z^\top - \boldsymbol{\beta}_{z0}^\top)^\top$. Recall the definition of the information matrix $\boldsymbol{J}_{\beta_x, \boldsymbol{\beta}_z}$ is

$$\boldsymbol{J}_{\beta_x, \boldsymbol{\beta}_z} = -\begin{pmatrix} E\{\frac{\partial U(\beta_x, \boldsymbol{\beta}_z)}{\partial \beta_x}\} & E\{\frac{\partial U(\beta_x, \boldsymbol{\beta}_z)}{\partial \boldsymbol{\beta}_z^\top}\} \\ E\{\frac{\partial \boldsymbol{V}(\beta_x, \boldsymbol{\beta}_z)}{\partial \beta_x}\} & E\{\frac{\partial \boldsymbol{V}(\beta_x, \boldsymbol{\beta}_z)}{\partial \boldsymbol{\beta}_z}\} \end{pmatrix} \in \mathbb{R}^{(1+p)\times(1+p)},$$

where

$$E\left\{\frac{\partial U(\beta_x, \boldsymbol{\beta}_z)}{\partial \beta_x}\right\} = E_{\mathcal{S}}[\bar{\omega}(Y, \boldsymbol{Z})\{\bar{m}(Y, \boldsymbol{Z}) - X\}] - E_{\mathcal{T}}\{\bar{m}(Y, \boldsymbol{Z})\} \in \mathbb{R},$$

$$E\left\{\frac{\partial \boldsymbol{V}(\beta_x,\boldsymbol{\beta}_z)}{\partial \beta_x}\right\} = \left[E\left\{\frac{\partial U(\beta_x,\boldsymbol{\beta}_z)}{\partial \boldsymbol{\beta}_z^\top}\right\}\right]^\top = E_{\mathcal{S}}[\bar{\omega}(Y,\boldsymbol{Z})\boldsymbol{Z}\{\bar{m}(Y,\boldsymbol{Z})-X\}] - E_{\mathcal{T}}\{\boldsymbol{Z}\bar{m}(Y,\boldsymbol{Z})\} \in \mathbb{R}^p,$$

$$E\left\{\frac{\partial \boldsymbol{V}(\beta_x,\boldsymbol{\beta}_z)}{\partial \boldsymbol{\beta}_z}\right\} = -E_{\mathcal{T}}(\boldsymbol{Z}\boldsymbol{Z}^\top) \in \mathbb{R}^{p\times p}.$$

Thus, $\boldsymbol{J}_{\beta_x,\boldsymbol{\beta}_z}$ is independent of $\beta_x, \boldsymbol{\beta}_z$. Let $\boldsymbol{J}_u \in \mathbb{R}^{1+p}, \boldsymbol{J}_v \in \mathbb{R}^{(1+p)\times p}$ satisfying $[\boldsymbol{J}_u, \boldsymbol{J}_v] = \boldsymbol{J}_{\beta_x,\boldsymbol{\beta}_z}^{-1}$. Noting that $(\widehat{\beta}_x, \widehat{\boldsymbol{\beta}}_z^\top)^\top$ is consistent for $(\beta_{x0}, \boldsymbol{\beta}_{z0}^\top)^\top$, by Theorem 5.21 of Van der Vaart (2000), we can expand Eqs (S2.5)-(S2.6) with respect to $\sqrt{n_{\mathcal{S}}}\ \boldsymbol{c}^\top(\widehat{\beta}_x - \beta_{x0}, \widehat{\boldsymbol{\beta}}_z^\top - \boldsymbol{\beta}_{z0}^\top)^\top$ as:

$$\begin{aligned}
\sqrt{n_{\mathcal{S}}}\ \boldsymbol{c}^\top \begin{pmatrix} \widehat{\beta}_x - \beta_{x0} \\ \widehat{\boldsymbol{\beta}}_z^\top - \boldsymbol{\beta}_{z0}^\top \end{pmatrix} &= \sqrt{n_{\mathcal{S}}}\boldsymbol{c}^\top \widehat{\boldsymbol{J}}_{\breve{\beta}_x,\breve{\boldsymbol{\beta}}_z}^{-1} \begin{pmatrix} \widehat{U}_{DR}(\beta_{x0},\boldsymbol{\beta}_{z0}) \\ \widehat{\boldsymbol{V}}_{DR}(\beta_{x0},\boldsymbol{\beta}_{z0}) \end{pmatrix} \\
&= \sqrt{n_{\mathcal{S}}}\boldsymbol{c}^\top(\widehat{\boldsymbol{J}}_u, \widehat{\boldsymbol{J}}_v) \begin{pmatrix} U(\beta_x,\boldsymbol{\beta}_z) + \Delta_{11} + \Delta_{12} + \Delta_{13} \\ \boldsymbol{V}(\beta_x,\boldsymbol{\beta}_z) + \boldsymbol{\Delta}_{21} + \boldsymbol{\Delta}_{22} + \boldsymbol{\Delta}_{23} \end{pmatrix}_{\beta_x=\beta_{x0},\boldsymbol{\beta}_z=\boldsymbol{\beta}_{z0}} \\
&= \sqrt{n_{\mathcal{S}}}\boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u\{U(\beta_x,\boldsymbol{\beta}_z) + \Delta_{11} + \Delta_{12} + \Delta_{13}\}|_{\beta_x=\beta_{x0},\boldsymbol{\beta}_z=\boldsymbol{\beta}_{z0}} \\
&\quad + \sqrt{n_{\mathcal{S}}}\boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v\{\boldsymbol{V}(\beta_x,\boldsymbol{\beta}_z) + \boldsymbol{\Delta}_{21} + \boldsymbol{\Delta}_{22} + \boldsymbol{\Delta}_{23}\}|_{\beta_x=\beta_{x0},\boldsymbol{\beta}_z=\boldsymbol{\beta}_{z0}} \\
&=:\ I + II,
\end{aligned}$$

where $(\breve{\beta}_x, \breve{\boldsymbol{\beta}}_z^\top)^\top$ is some vector lying between $(\beta_{x0}, \boldsymbol{\beta}_{z0}^\top)^\top$ and $(\widehat{\beta}_x, \widehat{\boldsymbol{\beta}}_z^\top)^\top$.

For the first term $I$, we have

$$I = U + T_{11} + T_{12} + T_{13}, \tag{S3.8}$$

where

$$
\begin{aligned}
U &= n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \{X_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}) \\
&\quad + n_{\mathcal{S}}^{1/2} n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \bar{m}(Y_i, \boldsymbol{Z}_i)(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}), \\
T_{11} &= n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \{\bar{m}(Y_i, \boldsymbol{Z}_i) - \widehat{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}), \\
T_{12} &= n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \{\bar{m}(Y_i, \boldsymbol{Z}_i) - \widehat{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}) \\
&\quad + n_{\mathcal{S}}^{1/2} n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}), \\
T_{13} &= n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \{X_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}).
\end{aligned}
$$

We shall show that

$$
\|\widehat{\boldsymbol{J}}_u - \boldsymbol{J}_u\|_\infty = o_p(n_{\mathcal{S}}^{-1/2}), \quad \|\widehat{\boldsymbol{J}}_v - \boldsymbol{J}_v\|_\infty = o_p(n_{\mathcal{S}}^{-1/2}). \tag{S3.9}
$$

Since the dimensionality of $\boldsymbol{Z}$, $p$, is fixed, we have

$$
\begin{aligned}
\|\widehat{\boldsymbol{J}}^{-1}_{\breve{\beta}_x, \breve{\boldsymbol{\beta}}_z} - \boldsymbol{J}^{-1}_{\beta_{x0}, \boldsymbol{\beta}_{z0}}\|_\infty &= \|\widehat{\boldsymbol{J}}^{-1}_{\breve{\beta}_x, \breve{\boldsymbol{\beta}}_z} (\boldsymbol{J}_{\beta_{x0}, \boldsymbol{\beta}_{z0}} - \widehat{\boldsymbol{J}}_{\breve{\beta}_x, \breve{\boldsymbol{\beta}}_z}) \boldsymbol{J}^{-1}_{\beta_{x0}, \boldsymbol{\beta}_{z0}}\|_\infty \\
&\le (1+p)^3 \|\widehat{\boldsymbol{J}}^{-1}_{\breve{\beta}_x, \breve{\boldsymbol{\beta}}_z}\|_\infty \|\boldsymbol{J}_{\beta_{x0}, \boldsymbol{\beta}_{z0}} - \widehat{\boldsymbol{J}}_{\breve{\beta}_x, \breve{\boldsymbol{\beta}}_z}\|_\infty \|\boldsymbol{J}^{-1}_{\beta_{x0}, \boldsymbol{\beta}_{z0}}\|_\infty.
\end{aligned}
$$

By Assumption 1 and the central limit theorem (CLT), we have

$$
\|\boldsymbol{J}_{\beta_{x0}, \boldsymbol{\beta}_{z0}} - \widehat{\boldsymbol{J}}_{\breve{\beta}_x, \breve{\boldsymbol{\beta}}_z}\|_\infty = o_p(n_{\mathcal{S}}^{-1/2}).
$$

Also noting that $\|\widehat{\boldsymbol{J}}^{-1}_{\beta_{x0}, \boldsymbol{\beta}_{z0}}\|_\infty$ and $\|\boldsymbol{J}^{-1}_{\beta_{x0}, \boldsymbol{\beta}_{z0}}\|_\infty$ are bounded by Assumption 1 and Neumann series, we have

$$\|\widehat{\boldsymbol{J}}^{-1}_{\breve{\beta}_x,\breve{\boldsymbol{\beta}}_z} - \boldsymbol{J}^{-1}_{\beta_{x0},\boldsymbol{\beta}_{z0}}\|_\infty = o_p(n_{\mathcal{S}}^{-1/2}),$$

that is, Eq (S3.9) holds. Under Assumption 2, and similar to above deduction ($E\{U(\beta_x, \boldsymbol{\beta}_z)\} = 0$), the below expectation is zero:

$$n_{\mathcal{S}}^{-1/2}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\bar{\omega}(Y_i,\boldsymbol{Z}_i)\{X_i-\bar{m}(Y_i,\boldsymbol{Z}_i)\}(Y_i-\beta_{x0}-\boldsymbol{Z}_i^\top\boldsymbol{\beta}_{z0})+n_{\mathcal{S}}^{1/2}n_{\mathcal{T}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{T}}}\bar{m}(Y_i,\boldsymbol{Z}_i)(Y_i-\beta_{x0}-\boldsymbol{Z}_i^\top\boldsymbol{\beta}_{z0}).$$

So by Assumption 1, Eq (S3.9), CLT and Slutsky's theorem, we have that $U$ weakly converges to $N(0, \sigma_u^2)$, where $\sigma_u^2$ represents the asymptotic variance of $U$ and is order 1. We then consider the remaining terms $(T_{11}, T_{12}, T_{13})$ separately. First, we consider $T_{11}$ in Eq (S3.8). By Assumption 1, the boundness of $|\boldsymbol{c}^\top\widehat{\boldsymbol{J}}_u(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top\boldsymbol{\beta}_{z0})|$, and our derived bounds for $n_{\mathcal{S}}^{-1/2}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i)-\bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2$ and $n_{\mathcal{S}}^{-1/2}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i)-\bar{m}(Y_i,\boldsymbol{Z}_i)\}^2$, we have

$$\begin{aligned}|T_{11}| &= O\Big(n_{\mathcal{S}}^{-1/2}\sum_{i\in\mathcal{I}_{\mathcal{S}}}|\widehat{\omega}(Y_i,\boldsymbol{Z}_i)-\bar{\omega}(Y_i,\boldsymbol{Z}_i)|\cdot|\widehat{m}(Y_i,\boldsymbol{Z}_i)-\bar{m}(Y_i,\boldsymbol{Z}_i)|\Big)\\ &\le \sqrt{n_{\mathcal{S}}}O\Big([n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{\omega}(Y_i,\boldsymbol{Z}_i)-\bar{\omega}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2}[n_{\mathcal{S}}^{-1}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\{\widehat{m}(Y_i,\boldsymbol{Z}_i)-\bar{m}(Y_i,\boldsymbol{Z}_i)\}^2]^{1/2}\Big)\\ &= o_p(1).\end{aligned}$$

For $T_{12}$, we have

$$T_{12} = -n_{\mathcal{S}}^{-1/2}\sum_{i\in\mathcal{I}_{\mathcal{S}}}\bar{\omega}(Y_i,\boldsymbol{Z}_i)\boldsymbol{c}^\top\widehat{\boldsymbol{J}}_u\breve{g}(\bar{m}(Y_i,\boldsymbol{Z}_i))\left[(Y,\boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}}-\bar{\boldsymbol{\gamma}})+O_p(\{(Y,\boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}}-\bar{\boldsymbol{\gamma}})\}^2)\right]$$

$$\times (Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0})$$

$$+ n_{\mathcal{S}}^{1/2} n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_\mathcal{T}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \breve{g}(\bar{m}(Y_i, \boldsymbol{Z}_i))[(Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}) + O_p(\{(Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}})\}^2)](Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}),$$

Again using Eq (S3.9) and Assumption 1, we have

$$n_{\mathcal{S}}^{-1} \sum_{i \in \mathcal{I}_\mathcal{S}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \breve{g}(\bar{m}(Y_i, \boldsymbol{Z}_i))(Y_i, \boldsymbol{Z}_i^\top)^\top (Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0})$$

$$+ n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_\mathcal{T}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \breve{g}(\bar{m}(Y_i, \boldsymbol{Z}_i))(Y_i, \boldsymbol{Z}_i^\top)^\top (Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}) \xrightarrow{p} \boldsymbol{\xi}_\gamma^u,$$

where

$$\boldsymbol{\xi}_\gamma^u = -E_\mathcal{S}\{\bar{\omega}(Y, \boldsymbol{Z}) \boldsymbol{c}^\top \boldsymbol{J}_u \ \breve{g}(\bar{m}(Y, \boldsymbol{Z}))(Y, \boldsymbol{Z}^\top)^\top (Y - \beta_{x0} - \boldsymbol{Z}^\top \boldsymbol{\beta}_{z0})\}$$

$$+ E_\mathcal{T}\{\boldsymbol{c}^\top \boldsymbol{J}_u \ \breve{g}(\bar{m}(Y, \boldsymbol{Z}))(Y, \boldsymbol{Z}^\top)^\top (Y - \beta_{x0} - \boldsymbol{Z}^\top \boldsymbol{\beta}_{z0})\}.$$

By standard analysis, it can be shown that

$$\|\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}\|_1 = O_p(n_{\mathcal{S}}^{-1/2}),$$

and $\sqrt{n_\mathcal{S}}(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}})$ converges in distribution to multivariate normal distribution with mean $\mathbf{0}$ and covariance of order 1. Combining these with Assumption 1 and using Slutsky's theorem, we have that $T_{12}$ is asymptotically equivalent with $\sqrt{n_\mathcal{S}}(\boldsymbol{\xi}_\gamma^u)^\top(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}})$, which weakly converges to normal distribution with mean 0 and variance of order 1.

Similarly, we write the term $T_{13}$ as

$$T_{13} = n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_\mathcal{S}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \{X_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\}(Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0})$$

$$\times \left[(Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}}) + O_p(\{(Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}})\}^2)\right] (Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}).$$

Using Eq (S3.9) and Assumption 1, we have

$$n_{\mathcal{S}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_u \{X_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\} (Y_i, \boldsymbol{Z}_i^\top)^\top (Y_i - \beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}) \xrightarrow{p} \boldsymbol{\xi}_\eta^u,$$

where

$$\boldsymbol{\xi}_\eta^u = E_{\mathcal{S}}\{\bar{\omega}(Y, \boldsymbol{Z}) \boldsymbol{c}^\top \boldsymbol{J}_u (X - \bar{m}(Y, \boldsymbol{Z}))(Y, \boldsymbol{Z}^\top)^\top (Y - \beta_{x0} - \boldsymbol{Z}^\top \boldsymbol{\beta}_{z0})\}.$$

Also, by standard analysis, it can be shown that

$$\|\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}}\|_1 = O_p(n_{\mathcal{S}}^{-1/2}),$$

and $\sqrt{n_{\mathcal{S}}}(\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}})$ converges in distribution to multivariate normal distribution with mean $\mathbf{0}$ and covariance of order 1. Combining these with Assumption 1 and using Slutsky's theorem, we have that $T_{13}$ is asymptotically equivalent with $\sqrt{n_{\mathcal{S}}}(\boldsymbol{\xi}_\eta^u)^\top (\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}})$, which weakly converges to normal distribution with mean 0 and variance of order 1.

For the second term $II$, we have

$$II = V + T_{21} + T_{22} + T_{23}, \tag{S3.10}$$

where

$$V = n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \{\bar{m}(Y_i, \boldsymbol{Z}_i) - X_i\} \beta_{x0}$$

$$+ n_{\mathcal{S}}^{1/2} n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \{Y_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}\},$$

$$T_{21} = n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}\beta_{x0},$$

$$T_{22} = n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}\beta_{x0}$$

$$- n_{\mathcal{S}}^{1/2} n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}\beta_{x0},$$

$$T_{23} = n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \{\bar{m}(Y_i, \boldsymbol{Z}_i) - X_i\}\beta_{x0}.$$

Under Assumption 2, and similar to above deduction ($E\{\boldsymbol{V}(\beta_x, \boldsymbol{\beta}_z)\} = 0$), the expectation of the following expression equals zero:

$$n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{Z}_i \{\bar{m}(Y_i, \boldsymbol{Z}_i) - X_i\}\beta_{x0} + n_{\mathcal{S}}^{1/2} n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{Z}_i \{Y_i - \bar{m}(Y_i, \boldsymbol{Z}_i)\beta_{x0} - \boldsymbol{Z}_i^\top \boldsymbol{\beta}_{z0}\}.$$

So by Assumption 1, Eq (S3.9), CLT and Slutsky's theorem, we have that $V$ weakly converges to $N(0, \sigma_v^2)$, where $\sigma_v^2$ represents the asymptotic variance of $V$ and is order 1. We then consider the remaining terms ($T_{21}, T_{22}, T_{23}$) separately in Eq (S3.10). First, we consider $T_{21}$. By Assumption 1, the boundedness of $|\boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \beta_{x0}|$, and our derived bounds for $n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\}^2$ and $n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}^2$, we have

$$|T_{21}| = O\Big(n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} |\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)| \cdot |\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)|\Big)$$

$$\leq \sqrt{n_{\mathcal{S}}} O\Big([n_{\mathcal{S}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{\omega}(Y_i, \boldsymbol{Z}_i) - \bar{\omega}(Y_i, \boldsymbol{Z}_i)\}^2]^{1/2} [n_{\mathcal{S}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \{\widehat{m}(Y_i, \boldsymbol{Z}_i) - \bar{m}(Y_i, \boldsymbol{Z}_i)\}^2]^{1/2}\Big)$$

$$= o_p(1).$$

For $T_{22}$, we have

$$T_{22} = n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \breve{g}(\bar{m}(Y_i, \boldsymbol{Z}_i)) \left[ (Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}) + O_p(\{(Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}})\}^2) \right] \beta_{x0}$$
$$- n_{\mathcal{S}}^{1/2} n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \left( \breve{g}(\bar{m}(Y_i, \boldsymbol{Z}_i)) \left[ (Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}}) + O_p(\{(Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}})\}^2) \right] \beta_{x0} \right).$$

Again using Eq (S3.9) and Assumption 1, we have

$$n_{\mathcal{S}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \breve{g}(\bar{m}(Y_i, \boldsymbol{Z}_i))(Y, \boldsymbol{Z}^\top) \beta_{x0} - n_{\mathcal{T}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{T}}} \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \breve{g}(\bar{m}(Y_i, \boldsymbol{Z}_i))(Y_i, \boldsymbol{Z}_i^\top) \beta_{x0} \xrightarrow{p} \boldsymbol{\xi}_\gamma^v,$$

where

$$\boldsymbol{\xi}_\gamma^v = E_{\mathcal{S}}\{\bar{\omega}(Y, \boldsymbol{Z}) \boldsymbol{c}^\top \boldsymbol{J}_v \boldsymbol{Z} \breve{g}(\bar{m}(Y, \boldsymbol{Z}))(Y, \boldsymbol{Z}^\top)^\top \beta_{x0}\} - E_{\mathcal{T}}\{\boldsymbol{c}^\top \boldsymbol{J}_v \boldsymbol{Z} \breve{g}(\bar{m}(Y, \boldsymbol{Z}))(Y, \boldsymbol{Z}^\top)^\top \beta_{x0}\}.$$

Recall that $\sqrt{n_{\mathcal{S}}}(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}})$ converges in distribution to a multivariate normal distribution with mean $\mathbf{0}$ and covariance of order 1. Combining these with Assumption 1 and using Slutsky's theorem, we have that $T_{22}$ is asymptotically equivalent with $\sqrt{n_{\mathcal{S}}}(\boldsymbol{\xi}_\gamma^v)^\top(\widehat{\boldsymbol{\gamma}} - \bar{\boldsymbol{\gamma}})$, which weakly converges to normal distribution with mean 0 and variance of order 1.

Similarly, we write the term $T_{23}$ as

$$T_{23} = n_{\mathcal{S}}^{-1/2} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \left[ (Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}}) + O_p(\{(Y, \boldsymbol{Z}^\top)(\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}})\}^2) \right] \{\bar{m}(Y_i, \boldsymbol{Z}_i) - X_i\} \beta_{x0}.$$

Using Eq (S3.9) and Assumption 1, we have

$$n_{\mathcal{S}}^{-1} \sum_{i \in \mathcal{I}_{\mathcal{S}}} \bar{\omega}(Y_i, \boldsymbol{Z}_i) \boldsymbol{c}^\top \widehat{\boldsymbol{J}}_v \boldsymbol{Z}_i \{\bar{m}(Y_i, \boldsymbol{Z}_i) - X_i\} \beta_{x0} \xrightarrow{p} \boldsymbol{\xi}_\eta^v,$$

where

$$\boldsymbol{\xi}_\eta^v = E_{\mathcal{S}}[\bar{\omega}(Y, \boldsymbol{Z})\boldsymbol{c}^\top \boldsymbol{J}_v \boldsymbol{Z}(Y, \boldsymbol{Z}^\top)^\top \{\bar{m}(Y, \boldsymbol{Z}) - X\}\beta_{x0}].$$

Also, $\sqrt{n_{\mathcal{S}}}(\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}})$ converges in distribution to multivariate normal distribution with mean $\mathbf{0}$ and covariance of order 1. Combining these with Assumption 1 and using Slutsky's theorem, we have that $T_{23}$ is asymptotically equivalent with $\sqrt{n_{\mathcal{S}}}(\boldsymbol{\xi}_\eta^v)^\top(\widehat{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}})$, which weakly converges to normal distribution with mean 0 and variance of order 1.

Combining with the asymptotic properties derived for $U, V, T_{11}, T_{12}, T_{13}, T_{21}, T_{22}, T_{23}$ and the expansions Eqs (S3.8)-(S3.10), we finish the proof for the asymptotic expansion and distribution of $\sqrt{n_{\mathcal{S}}}\ \boldsymbol{c}^\top(\widehat{\beta}_x - \beta_{x0}, \widehat{\boldsymbol{\beta}}_z^\top - \boldsymbol{\beta}_{z0}^\top)^\top$. $\square$

## S4 Additional simulation results

In this section, we present additional simulation results under three cases that deviate from the ideal conditions: (i) unbalanced labels of the binary variable $X$, (ii) reduced overlap between the two populations, and (iii) smaller sample size ratios between the target and source populations. These results provide further insights into the robustness of the proposed method and give potential limitations that warrant further investigation.

### S4.1 Simulation with unbalanced two labels of binary $X$

In this section, we conduct simulations with unbalanced two labels of binary $X$. The difference from the simulation setting in the main text is that here we consider two new models to generate $X_i$:

$$\text{M*}_{\textbf{cor}} : \text{logit}\{P(X_i = 1 \mid Y_i, \boldsymbol{Z}_i)\} = 4 + 0.8Y_i + 3.2Z_{1,i} - 3.2Z_{2,i},$$

$$\text{M*}_{\textbf{mis}} : \text{logit}\{P(X_i = 1 \mid Y_i, \boldsymbol{Z}_i)\} = 4 + 0.8Y_i + 3.2Z_{1,i} - 3.2Z_{2,i} + 3Y_iZ_{1,i}.$$

This modification increases the proportion of 1's of binary $X$ to the range $(0.75, 0.90)$, compared with the earlier range of $(0.40, 0.50)$. Similarly, the imputation model $m(y, \boldsymbol{z}) = (y, 1, z_1, z_2)^\top \boldsymbol{\gamma}$ is correctly specified under $\text{M*}_{\textbf{cor}}$ but misspecified under $\text{M*}_{\textbf{mis}}$, as $\text{M*}_{\textbf{mis}}$ includes the interaction term. We then consider the following three configurations: (I) $\text{M*}_{\textbf{cor}}$ and $\text{W}_{\textbf{cor}}$, (II) $\text{M*}_{\textbf{mis}}$ and $\text{W}_{\textbf{cor}}$, and (III) $\text{M*}_{\textbf{cor}}$ and $\text{W}_{\textbf{mis}}$.

We compare our DRTL-comb estimator with the estimator (denoted as "Naive") obtained by directly regressing $Y$ on $\boldsymbol{Z}$ while ignoring completely missing $X$ in the target population. We also present the preliminary IW and IM estimators as two benchmark estimators. For the variance estimator of the Naive method, we use standard error of linear regression. For the variance estimator of the IW, IM, and DRTL-comb methods, we use bootstrap in practice.

For each configuration, 500 bootstrap samples for variance estimation and 500 simulation replications are generated to summarize the average performance measures. For the given estimators $\widehat{\beta}_0, \widehat{\beta}_x, \widehat{\beta}_{z_1}, \widehat{\beta}_{z_2}$ that correspond to the coefficients of the intercept and $X, Z_1, Z_2$ respectively, we report the empirical average bias, root mean square error (RMSE), standard error, and coverage rate of the nominal 95% confidence interval. We present the statistical inference results for $\beta_x$ and $\boldsymbol{\beta}_z$ in Tables S4.1-S4.2. Here, there are no inference results for $\beta_x$ using the Naive method, as it excludes $X$ from the regression models.

As shown in Tables S4.1–S4.2, the Naive method performs poorly across all configurations because it ignores the information in the binary variable $X$, which is related to $Y$ and $\boldsymbol{Z}$. When both nuisance models are correctly specified (configuration (I)), the IM and DRTL-comb methods exhibit similar performance, with slightly smaller bias and RMSE than the IW method. When the imputation model is misspecified (configuration (II)), IM shows larger bias than IW and DRTL-comb, whereas under a misspecified density ratio model (configuration (III)), IW yields greater bias and RMSE than IM and DRTL-comb. In contrast, DRTL-comb produces nearly unbiased point estimates for $\beta_x$ and $\boldsymbol{\beta}_z$ in configurations (I) and (III), demonstrating its double robustness.

Table S4.1: Point estimator results for $\beta_x$ and $\boldsymbol{\beta}_z$ with unbalanced two labels of binary $X$.

| | **Bias** | | | | **RMSE** | | | |
|---|---|---|---|---|---|---|---|---|
| True | Naive | IW | IM | DRTL-comb | Naive | IW | IM | DRTL-comb |
| Configuration (I): M*$_{\mathbf{cor}}$ and W$_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = 0.834$ | 0.843 | -0.028 | 0.000 | -0.008 | 0.845 | 0.216 | 0.157 | 0.177 |
| $\beta_x = 0.929$ | / | 0.028 | 0.003 | 0.012 | / | 0.234 | 0.166 | 0.188 |
| $\beta_{z_1} = -0.212$ | 0.088 | -0.007 | -0.005 | -0.005 | 0.102 | 0.104 | 0.053 | 0.054 |
| $\beta_{z_2} = 0.163$ | -0.089 | 0.009 | 0.004 | 0.005 | 0.102 | 0.105 | 0.051 | 0.051 |
| Configuration (II): M*$_{\mathbf{mis}}$ and W$_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = 1.929$ | -0.255 | -0.034 | -0.062 | -0.036 | 0.260 | 0.192 | 0.129 | 0.153 |
| $\beta_x = -0.336$ | / | 0.031 | 0.079 | 0.045 | / | 0.233 | 0.156 | 0.190 |
| $\beta_{z_1} = -0.036$ | -0.088 | -0.016 | -0.037 | -0.019 | 0.101 | 0.116 | 0.069 | 0.075 |
| $\beta_{z_2} = 0.033$ | 0.038 | 0.012 | -0.002 | 0.001 | 0.065 | 0.098 | 0.057 | 0.055 |
| Configuration (III): M*$_{\mathbf{cor}}$ and W$_{\mathbf{mis}}$ | | | | | | | | |
| $\beta_0 = 0.269$ | 0.821 | 0.356 | 0.003 | 0.005 | 0.822 | 0.376 | 0.129 | 0.142 |
| $\beta_x = 0.920$ | / | -0.397 | -0.001 | -0.002 | / | 0.420 | 0.134 | 0.149 |
| $\beta_{z_1} = -0.880$ | 0.129 | 1.592 | 0.002 | 0.002 | 0.137 | 1.593 | 0.048 | 0.050 |
| $\beta_{z_2} = 0.178$ | 0.148 | -0.063 | 0.000 | 0.000 | 0.156 | 0.085 | 0.053 | 0.055 |

Table S4.2: Variance estimator results for $\beta_x$ and $\boldsymbol{\beta}_z$ with unbalanced two labels of binary $X$.

| | **Standard Error** | | | | **Coverage Rate** | | | |
|---|---|---|---|---|---|---|---|---|
| True | Naive | IW | IM | DRTL-comb | Naive | IW | IM | DRTL-comb |
| Configuration (I): M*$_{\mathbf{cor}}$ and W$_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = 0.834$ | 0.049 | 0.198 | 0.157 | 0.173 | 0.000 | 0.912 | 0.934 | 0.928 |
| $\beta_x = 0.929$ | / | 0.215 | 0.163 | 0.182 | / | 0.912 | 0.940 | 0.934 |
| $\beta_{z_1} = -0.212$ | 0.051 | 0.093 | 0.053 | 0.053 | 0.600 | 0.910 | 0.944 | 0.938 |
| $\beta_{z_2} = 0.163$ | 0.051 | 0.090 | 0.052 | 0.053 | 0.574 | 0.924 | 0.958 | 0.956 |
| Configuration (II): M*$_{\mathbf{mis}}$ and W$_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = 1.929$ | 0.049 | 0.176 | 0.113 | 0.145 | 0.000 | 0.922 | 0.918 | 0.936 |
| $\beta_x = -0.336$ | / | 0.215 | 0.136 | 0.180 | / | 0.930 | 0.922 | 0.932 |
| $\beta_{z_1} = -0.036$ | 0.051 | 0.101 | 0.059 | 0.072 | 0.622 | 0.904 | 0.902 | 0.934 |
| $\beta_{z_2} = 0.033$ | 0.051 | 0.086 | 0.055 | 0.053 | 0.864 | 0.912 | 0.944 | 0.936 |
| Configuration (III): M*$_{\mathbf{cor}}$ and W$_{\mathbf{mis}}$ | | | | | | | | |
| $\beta_0 = 0.269$ | 0.048 | 0.131 | 0.128 | 0.142 | 0.000 | 0.204 | 0.960 | 0.958 |
| $\beta_x = 0.920$ | / | 0.137 | 0.131 | 0.149 | / | 0.152 | 0.956 | 0.960 |
| $\beta_{z_1} = -0.880$ | 0.048 | 0.058 | 0.049 | 0.051 | 0.238 | 0.000 | 0.964 | 0.958 |
| $\beta_{z_2} = 0.178$ | 0.047 | 0.056 | 0.052 | 0.053 | 0.128 | 0.798 | 0.936 | 0.930 |

For variance estimation, DRTL-comb generally lies between the IW and IM methods, indicating that it does not inflate the standard error. A new phenomenon is that the IW method produces larger standard errors in most cases compared with the results in the main text, which is potentially due to the unbalanced labels of the binary variable $X$. Regarding coverage, the IW method shows poor performance in configuration (III), with coverage far below the nominal 95% level (all less than 80%), and the IM method exhibits unsatisfactory coverage in configuration (II) (e.g., 90.2% for $\beta_{z_1}$). In contrast, DRTL-comb maintains coverage rates close to the nominal level in most cases.

## S4.2 Simulation with reduced overlap between the two populations

In this case, we conduct simulations with reduced overlap of the distribution $Y$ across the two populations. Specifically, under the same data generation presented in simulation studies in the main text, we artificially delete the source samples with negative $Y$ values. Then, the total sample size $n_{\mathcal{S}} + n_{\mathcal{T}}$ ranges from 2000 to less than 1700, and the sample size ratio $(n_{\mathcal{T}}/n_{\mathcal{S}})$ ranges from $(0.7, 0.9)$ to $(0.95, 2.5)$. We also consider three same configurations in the main text: (I) $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}_{\mathbf{cor}}$, (II) $\mathrm{M}_{\mathbf{mis}}$ and $\mathrm{W}_{\mathbf{cor}}$, and (III) $\mathrm{M}_{\mathbf{cor}}$ and

$\mathrm{W}_{\mathbf{mis}}$.

We also compare the DRTL-comb estimator with the Naive, IW, and IM estimators as three benchmarks. For each configuration, 500 bootstrap samples for variance estimation of the IW, IM, and DRTL-comb methods and 500 simulation replications are generated to summarize the average performance measures. We present the statistical inference results for $\beta_x$ and $\boldsymbol{\beta}_z$ in Tables S4.3-S4.4. Here, there are no inference results for $\beta_x$ using the Naive method, as it excludes $X$ from the regression models.

As shown in Tables S4.3–S4.4, the Naive method performs poorly across all configurations because it ignores the information in the binary variable $X$, which is related to $(Y, \boldsymbol{Z})$. The IW method exhibits substantial bias and unsatisfactory coverage in every configuration. This is because removing source samples with negative $Y$ reduces population overlap, leading to inaccurate estimation of the density ratio for the distribution of $(Y, \boldsymbol{Z})$ across the two populations.

Both the IM and DRTL-comb methods produce nearly unbiased point estimates and achieve nominal coverage rates in configurations (I) and (III). In configuration (II), where the imputation model is misspecified, IM exhibits smaller but still non-negligible bias for $\beta_x$ compared with IW. Due to the combined effects of inaccurate density ratio estimation and misspecified

Table S4.3: Point estimator results for $\beta_x$ and $\boldsymbol{\beta}_z$ with reduced population overlap.

| | **Bias** | | | | **RMSE** | | | |
|---|---|---|---|---|---|---|---|---|
| True | Naive | IW | IM | DRTL-comb | Naive | IW | IM | DRTL-comb |
| Configuration (I): $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = 1.105$ | 0.572 | 0.141 | 0.003 | 0.002 | 0.574 | 0.158 | 0.102 | 0.102 |
| $\beta_x = 1.103$ | / | -0.484 | 0.001 | 0.002 | / | 0.501 | 0.173 | 0.173 |
| $\beta_{z_1} = -0.437$ | 0.313 | 0.183 | -0.005 | -0.005 | 0.317 | 0.196 | 0.072 | 0.071 |
| $\beta_{z_2} = 0.392$ | -0.319 | -0.165 | 0.003 | 0.003 | 0.322 | 0.180 | 0.071 | 0.071 |
| Configuration (II): $\mathrm{M}_{\mathbf{mis}}$ and $\mathrm{W}_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = 1.263$ | 0.413 | 0.103 | 0.040 | 0.050 | 0.416 | 0.129 | 0.115 | 0.124 |
| $\beta_x = 0.854$ | / | -0.458 | -0.101 | -0.120 | / | 0.478 | 0.222 | 0.236 |
| $\beta_{z_1} = -0.427$ | 0.304 | 0.201 | 0.018 | 0.022 | 0.308 | 0.218 | 0.095 | 0.099 |
| $\beta_{z_2} = 0.219$ | -0.147 | -0.101 | -0.005 | -0.016 | 0.156 | 0.118 | 0.064 | 0.065 |
| Configuration (III): $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}_{\mathbf{mis}}$ | | | | | | | | |
| $\beta_0 = 0.618$ | 0.471 | 0.289 | 0.003 | -0.003 | 0.474 | 0.293 | 0.081 | 0.119 |
| $\beta_x = 0.999$ | / | -0.616 | 0.002 | 0.016 | / | 0.623 | 0.151 | 0.244 |
| $\beta_{z_1} = -1.004$ | 0.252 | 1.290 | 0.002 | -0.006 | 0.256 | 1.291 | 0.056 | 0.085 |
| $\beta_{z_2} = 0.339$ | -0.310 | -0.223 | 0.003 | 0.004 | 0.314 | 0.227 | 0.069 | 0.085 |

Table S4.4: Variance estimator results for $\beta_x$ and $\boldsymbol{\beta}_z$ with reduced population overlap.

| | **Standard Error** | | | | **Coverage Rate** | | | |
|---|---|---|---|---|---|---|---|---|
| True | Naive | IW | IM | DRTL-comb | Naive | IW | IM | DRTL-comb |
| Configuration (I): $\text{M}_{\textbf{cor}}$ and $\text{W}_{\textbf{cor}}$ | | | | | | | | |
| $\beta_0 = 1.105$ | 0.049 | 0.068 | 0.100 | 0.101 | 0.000 | 0.440 | 0.942 | 0.938 |
| $\beta_x = 1.103$ | / | 0.122 | 0.164 | 0.167 | / | 0.046 | 0.926 | 0.928 |
| $\beta_{z_1} = -0.437$ | 0.051 | 0.069 | 0.068 | 0.069 | 0.000 | 0.222 | 0.936 | 0.938 |
| $\beta_{z_2} = 0.392$ | 0.051 | 0.067 | 0.069 | 0.069 | 0.000 | 0.314 | 0.940 | 0.946 |
| Configuration (II): $\text{M}_{\textbf{mis}}$ and $\text{W}_{\textbf{cor}}$ | | | | | | | | |
| $\beta_0 = 1.263$ | 0.049 | 0.076 | 0.109 | 0.113 | 0.000 | 0.736 | 0.940 | 0.936 |
| $\beta_x = 0.854$ | / | 0.136 | 0.195 | 0.201 | / | 0.088 | 0.942 | 0.926 |
| $\beta_{z_1} = -0.427$ | 0.051 | 0.080 | 0.091 | 0.093 | 0.000 | 0.274 | 0.944 | 0.932 |
| $\beta_{z_2} = 0.219$ | 0.051 | 0.060 | 0.062 | 0.061 | 0.188 | 0.642 | 0.948 | 0.942 |
| Configuration (III): $\text{M}_{\textbf{cor}}$ and $\text{W}_{\textbf{mis}}$ | | | | | | | | |
| $\beta_0 = 0.618$ | 0.048 | 0.047 | 0.086 | 0.126 | 0.000 | 0.000 | 0.964 | 0.964 |
| $\beta_x = 0.999$ | / | 0.086 | 0.158 | 0.253 | 0.000 | 0.000 | 0.956 | 0.964 |
| $\beta_{z_1} = -1.004$ | 0.048 | 0.044 | 0.058 | 0.085 | 0.000 | 0.000 | 0.964 | 0.968 |
| $\beta_{z_2} = 0.339$ | 0.047 | 0.040 | 0.070 | 0.089 | 0.000 | 0.000 | 0.952 | 0.954 |

imputation in configuration (II), DRTL-comb does not achieve an almost unbiased estimate for $\beta_x$; however, it still yields acceptable bias and maintains a nominal coverage rate, demonstrating its robustness.

## S4.3 Simulation with smaller sample size ratios

In this section, we conduct simulations with a smaller sample size ratio range of two populations ($n_{\mathcal{T}}/n_{\mathcal{S}} \in (0.3, 0.6)$). The difference from the simulation setting in the main text is that here we consider two new models to generate a membership variable $S_i$ to assign the $i$th observation to the source population when $S_i = 1$ and to the target data when $S_i = 0$:

$$\mathrm{W}^*_{\mathbf{cor}}: \ \mathrm{logit}\{P(S_i = 1 \mid Y_i, Z_i)\} = 0.8 + 0.3Y_i - 0.5Z_{1,i} + 0.3Z_{2,i},$$

$$\mathrm{W}^*_{\mathbf{mis}}: \ \mathrm{logit}\{P(S_i = 1 \mid Y_i, Z_i)\} = 0.8 + 0.3Y_i - 0.5Z_{1,i} + 0.3Z_{2,i} + 2Y_iZ_{1,i}.$$

The density ratio model $\widehat{\omega}(y, \boldsymbol{z}) = (1, y, z_1, z_2)^\top \widehat{\boldsymbol{\eta}}$ is correctly specified under $\mathrm{W}^*_{\mathbf{cor}}$ but misspecified under $\mathrm{W}^*_{\mathbf{mis}}$. We then consider the following three configurations: (I) $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}^*_{\mathbf{cor}}$, (II) $\mathrm{M}_{\mathbf{mis}}$ and $\mathrm{W}^*_{\mathbf{cor}}$, and (III) $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}^*_{\mathbf{mis}}$.

We also compare the DRTL-comb estimator with the Naive, IW, and IM estimators as three benchmarks. For each configuration, 500 bootstrap samples for variance estimation of the IW, IM, and DRTL-comb methods and 500 simulation replications are generated to summarize the average performance measures. We present the statistical inference results for $\beta_x$ and $\boldsymbol{\beta}_z$ in Tables S4.5-S4.6. Here, there are no inference results for $\beta_x$ using the Naive method, as it excludes $X$ from the regression models.

As shown in Tables S4.5-S4.6, the Naive method, due to the neglect of the information of the binary $X$ which is related to $Y, \boldsymbol{Z}$, has inferior performance in all configurations. When both nuisance models are correct (configuration (I)), the two preliminary methods (IW and IM) and the DRTL-comb method demonstrate similar performance in terms of bias and RMSE. When the imputation model is misspecified (configuration (II)), IM exhibits a larger bias and RMSE than IW and DRTL-comb, whereas with a misspecified density ratio model (configuration (III)), IW shows a greater bias and RMSE than IM and DRTL-comb. However, DRTL-comb achieves almost unbiased point estimators for $\beta_x$ and $\boldsymbol{\beta}_z$ in three configurations, showing its double robustness. For the variance estimator, DRTL-comb typically falls between the IW and IM methods, which indicates that the proposed DRTL-comb method will not introduce a large standard error. Regarding the coverage rate, the IW method has poor coverage rates below the nominal level of 95% in configuration (III), and the IM method has unsatisfactory coverages in configuration (II). However, DRTL-comb maintains a nominal coverage rate in most cases. The lowest coverages are 62.0%

Table S4.5: Point estimator results for $\beta_x$ and $\boldsymbol{\beta}_z$ with a smaller $n_{\mathcal{T}}/n_{\mathcal{S}}$ range.

| | **Bias** | | | | **RMSE** | | | |
|---|---|---|---|---|---|---|---|---|
| True | Naive | IW | IM | DRTL-comb | Naive | IW | IM | DRTL-comb |
| | | | | Configuration (I): $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}^*_{\mathbf{cor}}$ | | | | |
| $\beta_0 = -0.004$ | 0.528 | 0.003 | 0.006 | 0.006 | 0.532 | 0.088 | 0.081 | 0.082 |
| $\beta_x = 1.224$ | / | 0.002 | -0.006 | -0.006 | / | 0.139 | 0.133 | 0.137 |
| $\beta_{z_1} = -0.300$ | 0.364 | -0.002 | -0.002 | -0.002 | 0.370 | 0.069 | 0.079 | 0.080 |
| $\beta_{z_2} = 0.324$ | -0.364 | 0.001 | 0.000 | 0.000 | 0.370 | 0.070 | 0.079 | 0.080 |
| | | | | Configuration (II): $\mathrm{M}_{\mathbf{mis}}$ and $\mathrm{W}^*_{\mathbf{cor}}$ | | | | |
| $\beta_0 = 0.079$ | 0.449 | 0.001 | 0.160 | 0.098 | 0.454 | 0.087 | 0.182 | 0.131 |
| $\beta_x = 1.040$ | / | 0.010 | -0.369 | -0.223 | / | 0.153 | 0.394 | 0.263 |
| $\beta_{z_1} = -0.213$ | 0.279 | -0.006 | 0.053 | 0.024 | 0.286 | 0.081 | 0.093 | 0.077 |
| $\beta_{z_2} = 0.195$ | -0.232 | 0.005 | -0.097 | -0.059 | 0.241 | 0.064 | 0.120 | 0.094 |
| | | | | Configuration (III): $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}^*_{\mathbf{mis}}$ | | | | |
| $\beta_0 = 0.212$ | 0.436 | 0.174 | 0.011 | 0.011 | 0.439 | 0.188 | 0.068 | 0.070 |
| $\beta_x = 1.005$ | / | 0.020 | -0.019 | -0.019 | / | 0.110 | 0.119 | 0.124 |
| $\beta_{z_1} = -1.158$ | 0.237 | 1.551 | 0.004 | 0.004 | 0.242 | 1.552 | 0.052 | 0.053 |
| $\beta_{z_2} = 0.321$ | -0.309 | -0.036 | -0.002 | -0.002 | 0.314 | 0.065 | 0.068 | 0.068 |

Table S4.6: Variance estimator results for $\beta_x$ and $\boldsymbol{\beta}_z$ with a smaller $n_{\mathcal{T}}/n_{\mathcal{S}}$ range.

| | **Standard Error** | | | | **Coverage Rate** | | | |
|---|---|---|---|---|---|---|---|---|
| True | Naive | IW | IM | DRTL-comb | Naive | IW | IM | DRTL-comb |
| Configuration (I): $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}^*_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = -0.004$ | 0.067 | 0.083 | 0.081 | 0.082 | 0.000 | 0.930 | 0.958 | 0.956 |
| $\beta_x = 1.224$ | / | 0.132 | 0.129 | 0.131 | / | 0.936 | 0.932 | 0.932 |
| $\beta_{z_1} = -0.300$ | 0.068 | 0.073 | 0.079 | 0.079 | 0.000 | 0.964 | 0.950 | 0.936 |
| $\beta_{z_2} = 0.324$ | 0.067 | 0.069 | 0.078 | 0.078 | 0.002 | 0.940 | 0.948 | 0.948 |
| Configuration (II): $\mathrm{M}_{\mathbf{mis}}$ and $\mathrm{W}^*_{\mathbf{cor}}$ | | | | | | | | |
| $\beta_0 = 0.079$ | 0.067 | 0.088 | 0.085 | 0.086 | 0.000 | 0.942 | 0.542 | 0.798 |
| $\beta_x = 1.040$ | / | 0.149 | 0.135 | 0.137 | / | 0.938 | 0.242 | 0.620 |
| $\beta_{z_1} = -0.213$ | 0.068 | 0.079 | 0.083 | 0.078 | 0.010 | 0.936 | 0.920 | 0.948 |
| $\beta_{z_2} = 0.195$ | 0.067 | 0.062 | 0.072 | 0.072 | 0.068 | 0.938 | 0.734 | 0.858 |
| Configuration (III): $\mathrm{M}_{\mathbf{cor}}$ and $\mathrm{W}^*_{\mathbf{mis}}$ | | | | | | | | |
| $\beta_0 = 0.212$ | 0.051 | 0.070 | 0.068 | 0.070 | 0.000 | 0.308 | 0.954 | 0.954 |
| $\beta_x = 1.005$ | / | 0.111 | 0.119 | 0.126 | / | 0.950 | 0.950 | 0.956 |
| $\beta_{z_1} = -1.158$ | 0.049 | 0.058 | 0.053 | 0.054 | 0.000 | 0.000 | 0.960 | 0.958 |
| $\beta_{z_2} = 0.321$ | 0.051 | 0.053 | 0.064 | 0.065 | 0.000 | 0.886 | 0.942 | 0.942 |

for $\beta_x$ and 79.8% for $\beta_0$ in configuration (II); although below nominal, it still represents a substantial improvement over IM (24.2% for $\beta_x$ and 54.2% for $\beta_0$).

## S5 Additional data analysis results

In this section, we first present the detailed linear regression results for the two populations separately from an oracle perspective, assuming that binary $X$, smoking status, is known in the target data (Table S5.7). Here, the source population has 4713 subjects with the negative polygenic risk score for BMI (prs.BMI), while the target population includes 3196 subjects with the positive prs.BMI.

As shown in Table S5.7, there are clear differences between the two populations in both cases regarding the estimates of covariate coefficients (Estimate), standard error (SE), 95% confidence intervals (95%CI), and $p$-values. The coefficient of energy for the target population is larger than that of the source population, suggesting an increased effect of energy on BMI among the target individuals. The effects of $X$, sex, and age on BMI also differ between the two populations (sex = 1 for males and sex = 0 for females). There are larger standard errors for covariates in the target data compared to those in the source data, resulting in wider 95%

Table S5.7: Detailed results for two populations without missing $X$.

| Population | Covariate | Estimate | SE | 95%CI | $p$-value |
|---|---|---|---|---|---|
| Source | $X$ | 0.752 | 0.119 | (0.520,0.985) | $< 5e-4$ |
| | energy | 0.074 | 0.058 | (-0.040,0.188) | 0.201 |
| | sex | 0.888 | 0.119 | (0.656,1.120) | $< 5e-4$ |
| | age | 0.053 | 0.058 | (-0.060,0.166) | 0.358 |
| Target | $X$ | 0.774 | 0.180 | (0.422,1.126) | $< 5e-4$ |
| | energy | 0.242 | 0.091 | (0.065,0.420) | 0.008 |
| | sex | 1.004 | 0.181 | (0.650,1.359) | $< 5e-4$ |
| | age | 0.008 | 0.088 | (-0.165,0.181) | 0.927 |

confidence intervals in the target data. These differences demonstrate the heterogeneity between the source and target populations.

Table S5.8: Coefficients of the logistic regression of $X$ on $Y$ and $\boldsymbol{Z}$ for two populations.

| Population | $Y$ | energy | sex | age |
|---|---|---|---|---|
| Source | 0.049 | 0.000 | 0.446 | 0.188 |
| Target | 0.032 | -0.015 | 0.355 | 0.243 |

# Bibliography

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, C1–C68.

Liu, M., Y. Zhang, K. P. Liao, and T. Cai (2023). Augmented transfer regression learning with semi-non-parametric nuisance models. *Journal of Machine Learning Research 24*(293), 1–50.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Yan, H. and S. X. Chen (2024). Transfer learning with general estimating equations. *arXiv preprint arXiv:2410.04398*.