Supplementary Materials

Appendix A Influence functions of $\widehat{\Delta}, \widehat{\Delta}_g, \widehat{\sigma}$ and $\widehat{\sigma}_g^2$

We derive asymptotic distributions for $\widehat{\Delta}$ and $\widehat{\Delta}_g$. To this end, we first write

$$\widehat{\mu}_a - \mu_a = \frac{\sum_{i=1}^n I(A_i = a)Y_i}{\sum_{i=1}^n I(A_i = a)} - \mu_a = \frac{\sum_{i=1}^n I(A_i = a)(Y_i - \mu_a)}{\sum_{i=1}^n I(A_i = a)} := n^{-1}\sum_{i=1}^n \psi_{a,i} + o_p(n^{-1/2}),$$

where $\psi_{a,i} = 2I(A_i = a)(Y_i - \mu_a)$. It follows that

$$\sqrt{n}(\widehat{\Delta} - \Delta) = n^{-1/2}\sum_{i=1}^n(\psi_{1,i} - \psi_{0,i}) + o_p(1) := n^{-1/2}\sum_{i=1}^n \psi_i + o_p(1).$$

By the central limit theorem we have that $\sqrt{n}(\widehat{\Delta} - \Delta)$ converges in distribution to a normal distribution $N(0, \sigma^2)$ with $\sigma^2 = E[\psi_i^2]$.

Similarly, we have

$$\widehat{\mu}_{g,a} - \mu_{g,a} = \frac{\sum_{A_i=a} g(S_i)}{\sum_{A_i=a} 1} - \mu_{g,a} = \frac{\sum_{A_i=a}\{g(S_i) - \mu_{g,a}\}}{\sum_{A_i=a} 1} := n^{-1}\sum_{i=1}^n \psi_{g,a,i} + o_p(n^{-1/2}).$$

It follows that

$$\sqrt{n}(\widehat{\Delta}_g - \Delta_g) = n^{-1/2}\sum_{i=1}^n(\psi_{g,1,i} - \psi_{g,0,i}) + o_p(1) := n^{-1/2}\sum_{i=1}^n \psi_{g,i} + o_p(1).$$

By the central limit theorem we have that $\sqrt{n}(\widehat{\Delta}_g - \Delta_g)$ converges in distribution to a normal distribution $N(0, \sigma_g^2)$ with $\sigma_g^2 = E[\psi_{g,i}^2]$.

We next derive estimators for the asymptotic variances σ^2 and σ_g^2. To this end, we first note that the variance of $\psi_{a,i}$ is

$$E\psi_{a,i}^2 = E[4I(A_i = a)(Y_i - \mu_a)^2],$$

which can be estimated by

$$\hat{\Sigma}_a = n^{-1}\sum_{i=1}^n 4I(A_i = a)(Y_i - \hat{\mu}_a)^2.$$

Therefore, the asymptotic variance of $\sqrt{n}(\widehat{\Delta} - \Delta)$, σ^2, can be estimated by $\widehat{\sigma}^2 := \hat{\Sigma} := \hat{\Sigma}_1 + \hat{\Sigma}_0$.

It follows from the above formulas that $n^{1/2}(\widehat{\sigma}^2 - \sigma^2)$ can be written as the form

$$n^{1/2}(\widehat{\sigma}^2 - \sigma^2) = n^{-1/2} \sum_{i=1}^{n} \psi_{\sigma^2,i} + o_p(1).$$

Similarly, we can get the estimate for the asymptotic variance of $\sqrt{n}(\widehat{\Delta}_g - \Delta_g)$, $\widehat{\sigma}_g^2$, with a given g, and

$$n^{1/2}(\widehat{\sigma}_g^2 - \sigma_g^2) = n^{-1} \sum_{i=1}^{n} \psi_{\sigma_g^2,i} + o_p(1).$$

With the above influence functions for $\widehat{\sigma}^2$ and $\widehat{\sigma}_g^2$, the variance estimates for $\widehat{\sigma}^2$ and $\widehat{\sigma}_g^2$ can be obtained by perturbation resampling method.

Appendix B Derivations of the optimal g

In this section, we derive the specific form for the optimal transformation function of the surrogate information, $g_{\text{opt}}(\cdot)$. We aim to solve the following problem for g:

$$\min_g L(g) = E\{Y^{(1)} - g(S^{(1)})\}^2, \quad \text{given} \quad E\{Y^{(0)} - g(S^{(0)})\} = 0$$

with $g(s) = m_0(s) + c, s \in D_0$ and $g(s)$ is continuous.

Without loss of generality, we assume that S is continuous with conditional densities given $A = a$, $\dot{F}_a(s) := f_a(s)$, with respect to the Lebesgue measure. Similar arguments as given below can be used to derive g_{opt} when S is discrete. It can be shown that

$$E\{Y^{(1)} - g(S^{(1)})\}^2 \propto E[g^2(S^{(1)})] - 2E[Y^{(1)}g(S^{(1)})] = E[g^2(S^{(1)})] - 2E[m_1(S^{(1)})g(S^{(1)})].$$

And thus the problem is equivalent to finding a function $g_{\text{opt}}(\cdot)$ such that

$$\min_g \frac{1}{2}E[g^2(S^{(1)})] - E[m_1(S^{(1)})g(S^{(1)})] \quad \text{given} \quad E[g(S)|A = 0] = \mu_0.$$

Our optimization problem is thus,

$$\min_g \int \frac{1}{2}g^2(s)f_1(s)ds - \int m_1(s)g(s)f_1(s)ds \quad \text{given that} \int g(s)f_0(s)ds = \mu_0,$$

which is equivalent to

$$\min_g \mathcal{L}(g), \quad \text{given that} \quad \mathbb{G}(g) = \mu_0,$$

where we used the functional notation

$$\mathcal{L}(g) = \int \frac{1}{2}g^2(s)f_1(s)ds - \int m_1(s)g(s)f_1(s)ds, \quad \text{and} \quad \mathbb{G}(g) = \int g(s)f_0(s)ds.$$

Taking the Frechet derivatives of the functionals, we have that for all measurable h such that $\int h^2(s) f_1(s) ds < \infty$,

$$\frac{d}{dg}\Big[\mathcal{L}(g) - \lambda \mathbb{G}(g)\Big](h) = \int g_{\text{opt}}(s) h(s) f_1(s) ds - \int g_{\text{opt}}(s) h(s) m_1(s) f_1(s) ds - \lambda \int h(s) f_0(s) ds = 0.$$

Setting $h = \delta(s)$, this implies that

$$g_{\text{opt}}(s) = m_1(s) + \lambda f_0(s)/f_1(s) = m_1(s) + \lambda\, r(s), s \in D_c \cup D_1.$$

By the constraint $\int_{D_c} \{m_1(s) + \lambda\, r(s)\} f_0(s) ds + \int_{D_0} \{m_0(s) + c\} f_0(s) ds = \mu_0 = \int m_0(s) f_0(s) ds$ and $g_{\text{opt}}(s)$ is continuous at s^*, or $m_1(s^*) + \lambda\, r(s^*) = m_0(s^*) + c$, we have

$$\begin{aligned} \lambda &= \{K_2 + K_1 r(s^*)\}^{-1} \int_{D_c} \Delta_{01}(s) f_0(s) ds + K_1 \{K_2 + K_1 r(s^*)\}^{-1} \Delta_{01}(s^*), \\ c &= \{K_2 + r(s^*) K_1\}^{-1} \left[-K_2 \Delta_{01}(s^*) + r(s^*) \int_{D_c} \Delta_{01}(s) f_0(s) ds \right] \end{aligned}$$

with $\Delta_{01}(s) = m_0(s) - m_1(s)$, $K_1 = \int_{D_0} f_0(s) ds$, $K_2 = \int_{D_c} f_0^2(s)/f_1(s) ds = \int_{D_c} r(s) f_0(s) ds$.

Finally, the optimal function $g_{\text{opt}}(\cdot)$ can be expressed as

$$g_{\text{opt}}(s) = \begin{cases} m_1(s) + \lambda\, r(s), & s \in D_c \cup D_1 \\ m_0(s) + c, & s \in D_0. \end{cases}$$

Appendix C Relationship between PTE and PTE$_L$

In this section, we show the relationship between our proposed PTE and the PTE of Parast et al. (2016). To this end, let Δ_{rte} denote the "residual treatment effect" defined in Parast et al. (2016) as:

$$\Delta_{\text{rte}} = \int E(Y^{(1)} - Y^{(0)} | S^{(1)} = S^{(0)} = s) d\mathscr{F}(s) = \int \{m_1(s) - m_0(s)\} d\mathscr{F}(s),$$

where $\mathscr{F}(\cdot)$ is a reference distribution function. It follows that

$$\Delta_L = \Delta - \Delta_{\text{rte}} = \int m_1(s)\{dF_1(s) - d\mathscr{F}(s)\} - \int m_0(s)\{dF_0(s) - d\mathscr{F}(s)\}. \tag{1}$$

and $\text{PTE}_L = \Delta_L / \Delta$.

To relate Δ_L to $\Delta_{g_{\text{opt}}(S)}$, recall that

$$\begin{aligned}
g_{\text{opt}}(s) &= m_1(s)+\lambda f_0(s)/f_1(s), s\in D_c\cup D_1,\\
g_{\text{opt}}(s) &= m_0(s)+c, s\in D_0,\\
\lambda &= \left\{K_2+K_1\frac{f_0(s^*)}{f_1(s^*)}\right\}^{-1}\int_{D_c}\{m_0(s)-m_1(s)\}f_0(s)ds+K_1\left\{K_2+K_1\frac{f_0(s^*)}{f_1(s^*)}\right\}^{-1}\{m_0(s^*)-m_1(s^*)\},\\
c &= \left\{1+\frac{f_0(s^*)}{f_1(s^*)}\frac{K_1}{K_2}\right\}^{-1}\left[m_1(s^*)-m_0(s^*)+\frac{f_0(s^*)}{f_1(s^*)}\frac{1}{K_2}\int_{D_c}\{m_0(s)-m_1(s)\}f_0(s)ds\right] \text{ and}\\
\Delta_{g_{\text{opt}}(S)} &= E\{g_{\text{opt}}(S^{(1)})-g_{\text{opt}}(S^{(0)})\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Delta_{g_{\text{opt}}(S)} &= \int_{D_1}m_1(s)f_1(s)ds+\int_{D_c}\{m_1(s)+\lambda f_0(s)/f_1(s)\}f_1(s)ds-\mu_0\\
&= \int_{D_1\cup D_c}m_1(s)f_1(s)ds-\int_{D_0\cup D_c}m_0(s)f_0(s)ds+\int_{D_c}f_0(s)ds\\
&\times\left[\left\{K_2+K_1\frac{f_0(s^*)}{f_1(s^*)}\right\}^{-1}\int_{D_c}\{m_0(s)-m_1(s)\}f_0(s)ds+K_1\left\{K_2+K_1\frac{f_0(s^*)}{f_1(s^*)}\right\}^{-1}\{m_0(s^*)-m_1(s^*)\}\right]\\
&= \int_{D_1}m_1(s)f_1(s)ds+\int_{D_c}m_1(s)\left[f_1(s)-f_0(s)\frac{\int_{D_c}f_0(s)d(s)}{K_2+K_1f_0(s^*)/f_1(s^*)}\right]ds\\
&-\int_{D_0}m_0(s)f_0(s)ds-\int_{D_c}m_0(s)\left[f_0(s)-f_0(s)\frac{\int_{D_c}f_0(s)d(s)}{K_2+K_1f_0(s^*)/f_1(s^*)}\right]ds\\
&+\int_{D_c}f_0(s)ds\ K_1\left\{K_2+K_1\frac{f_0(s^*)}{f_1(s^*)}\right\}^{-1}\{m_0(s^*)-m_1(s^*)\}\\
&:= \int m_1(s)\{dF_1(s)-d\mathscr{F}_{\text{new}}(s)\}-\int m_0(s)\{dF_0(s)-d\mathscr{F}_{\text{new}}(s)\}\\
&+\int_{D_c}f_0(s)ds\ K_1\left\{K_2+K_1\frac{f_0(s^*)}{f_1(s^*)}\right\}^{-1}\{m_0(s^*)-m_1(s^*)\}
\end{aligned}$$

where

$$\mathscr{F}_{\text{new}}(s)=\int_{D_c}I(v\le s)f_0(v)dv\frac{\int_{D_c}f_0(s)ds}{K_2+K_1f_0(s^*)/f_1(s^*)}. \tag{2}$$

If $\mathscr{F}(s)$ in (1) is replaced by $\mathscr{F}_{\text{new}}(s)$, then $\Delta_{g_{\text{opt}}(S)}=\Delta_L+\int_{D_c}f_0(s)ds\ K_1\left\{K_2+K_1\frac{f_0(s^*)}{f_1(s^*)}\right\}^{-1}\{m_0(s^*)-m_1(s^*)\}$; and thus, when D_0 is empty (K_1=0), or $m_0(s^*)=m_1(s^*)$ we have PTE $\equiv$ PTE$_L$.

We next show that only assumptions (C1) and (C2) are required to ensure that the proposed PTE is between 0 and 1.

(C1) $\mathbb{S}_1(u) \geq \mathbb{S}_0(u)$ for all u,

(C2) $\mathbb{M}_1(u) \geq \mathbb{M}_0(u)$ for all u in the common support of $g_{\text{opt}}(S^{(1)})$ and $g_{\text{opt}}(S^{(0)})$,

where $\mathbb{S}_a(u) = P\{g_{\text{opt}}(S^{(a)}) > u \mid A = a\}$, $\mathbb{M}_a(u) = E\{Y^{(a)} \mid g_{\text{opt}}(S^{(a)}) = u\}$, $a = 0, 1$, which are assumed to be continuous functions. Following arguments given in Appendices A and B, we have

$$\Delta = E\{Y^{(1)}\} - E\{Y^{(0)}\} = \int \mathbb{M}_1(u) d\mathbb{F}_1(u) - \int \mathbb{M}_0(u) d\mathbb{F}_0(u),$$

$$\Delta_{g_{\text{opt}}} = \int \mathbb{M}_1(u)\{d\mathbb{F}_1(u) - d\mathbb{F}_{\text{new}}(u)\} - \int \mathbb{M}_0(u)\{d\mathbb{F}_0(u) - d\mathbb{F}_{\text{new}}(u)\} + H(u^*)\{\mathbb{M}_0(u^*) - \mathbb{M}_1(u^*)\},$$

$$\Delta - \Delta_{g_{\text{opt}}} = \int_{\mathbb{D}_c} \{\mathbb{M}_1(u) - \mathbb{M}_0(u)\}\dot{\mathbb{F}}_{\text{new}}(u) du + H(u^*)\{\mathbb{M}_1(u^*) - \mathbb{M}_0(u^*)\}, \quad (3)$$

where $H(u^*)$ is a non-negative function of u^*, $\mathbb{F}_a(u) = 1 - \mathbb{S}_a(u)$, $\dot{\mathbb{F}}_a(u) = d\mathbb{F}_a(u)/du$, and $\dot{\mathbb{F}}_{\text{new}}(u)$ is similarly defined as (2) but for $g_{\text{opt}}(S)$ instead of S and $\mathbb{D}_c$ is the common support of $g_{\text{opt}}(S^{(1)})$ and $g_{\text{opt}}(S^{(0)})$. It is also straightforward to see from an integration by parts that

$$\Delta_{g_{\text{opt}}(S)} = \int u \, d\mathbb{F}_1(u) - \int u \, d\mathbb{F}_0(u) = \int \{\mathbb{S}_1(u) - \mathbb{S}_0(u)\} du.$$

Thus, from condition (C1), we have $\Delta_{g_{\text{opt}}(S)} \geq 0$. On the other hand, since $\dot{\mathbb{F}}_{\text{new}}(u) \geq 0$, we see from (3) that $\Delta - \Delta_{g_{\text{opt}}(S)} \geq 0$ under condition (C2). It follows that PTE $\in [0, 1]$ under conditions (C1) and (C2). Furthermore, $\Delta_{g_{\text{opt}}(S)} = 0$ when $\Delta = 0$.

Appendix D Asymptotic properties for $\widehat{g}(\cdot)$

Throughout, we assume that $m_a(s), a = 0, 1$ is continuously differentiable. In addition, we assume that $f_a(s), a = 0, 1$ is continuously differentiable with finite support. For inference, we require under-smoothing with $h = o_p(n^{-1/5})$ for interval estimation of g_{opt} and $h = o_p(n^{-1/4})$ for the interval estimation of RE and PTE. Since $\widehat{m}_a(s)$ and $\widehat{f}_a(s), a = 0, 1$ are standard kernel estimators, we have that they are consistent w.r.t their true values with rate $(\log n)^{\frac{1}{2}}(nh)^{-\frac{1}{2}} + h^2$. It follows immediately that $|\widehat{g}(s) - g_{\text{opt}}(s)| = O_p\{(\log n)^{\frac{1}{2}}(nh)^{-\frac{1}{2}} + h^2\}$.

We firstly derive the influence functions for each estimator in Section 2.3. The influence functions can be derived following exactly the derivations of $\widehat{\mu}_a - \mu_a$ and $\widehat{\mu}_{g,a} - \widehat{\mu}_{g,a}$. Direct calculations show that

$$
\begin{aligned}
\widehat{f}_a(s) - f_a(s) &= \frac{n^{-1}\sum_{A_i=a}\{K_h(S_i - s) - f_a(s)\}}{n^{-1}\sum_{A_i=a} 1} \\
&:= (nh)^{-1}\sum_{i=1}^{n}\phi_{a,i}(s) + o_p\{(nh)^{-1/2}\}, \\
\widehat{m}_a(s) - m_a(s) &= \frac{\sum_{i=1}^{n} I(A_i = a)K_h(S_i - s)\{Y_i - m_a(s)\}}{\sum_{i=1}^{n} I(A_i = a)K_h(S_i - s)} \\
&:= (nh)^{-1}\sum_{i=1}^{n}\phi_{m,i}(s) + o_p\{(nh)^{-1/2}\}, \\
\hat{K}_1 - K_1 &= \int_{D_0}\{\widehat{f}_0(s) - f_0(s)\}ds = n^{-1}\sum_{i=1}^{n} 2I(A_i = 0)I(S_i \in D_0) - \int_{D_0} f_0(s)ds \\
&:= n^{-1}\sum_{i=1}^{n}\phi_{K_1,i} + o_p\{n^{-1/2}\}, \\
\hat{K}_2 - K_2 &= \int_{D_c} 2\frac{f_0(s)}{f_1(s)}\{\widehat{f}_0(s) - f_0(s)\}ds - \int_{D_c}\frac{f_0^2(s)}{f_1^2(s)}\{\widehat{f}_1(s) - f_1(s)\}ds \\
&= n^{-1}\sum_{i=1}^{n} 2I(A_i = 0)I(S_i \in D_c)2\frac{f_0(S_i)}{f_1(S_i)} - \int_{D_c} 2\frac{f_0^2(s)}{f_1(s)}ds \\
&\quad - \left\{n^{-1}\sum_{i=1}^{n} 2I(A_i = 1)I(S_i \in D_c)\frac{f_0^2(S_i)}{f_1^2(S_i)} - \int_{D_c}\frac{f_0^2(s)}{f_1(s)}ds\right\} \\
&:= n^{-1}\sum_{i=1}^{n}\phi_{K_2,i} + o_p\{n^{-1/2}\}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\widehat{\lambda}-\lambda &= \frac{\int_{D_c}\{\widehat{m}_0(s)-\widehat{m}_1(s)\}d\widehat{F}_0(s)}{\hat{K}_2+\hat{K}_1\widehat{f}_0(s^*)/\widehat{f}_1(s^*)}+\hat{K}_1\frac{\widehat{m}_0(s^*)-\widehat{m}_1(s^*)}{\hat{K}_2+\hat{K}_1\widehat{f}_0(s^*)/\widehat{f}_1(s^*)} \\
&\quad -\frac{\int_{D_c}\{m_0(s)-m_1(s)\}dF_0(s)}{K_2+K_1f_0(s^*)/f_1(s^*)}-K_1\frac{m_0(s^*)-m_1(s^*)}{K_2+K_1f_0(s^*)/f_1(s^*)} \\
&= \frac{\left[\int_{D_c}\{\widehat{m}_0(s)-m_0(s)-\widehat{m}_1(s)+m_1(s)\}f_0(s)ds+\int_{D_c}\{m_0(s)-m_1(s)\}\{\widehat{f}_0(s)-f_0(s)\}ds\right]}{K_2+K_1f_0(s^*)/f_1(s^*)} \\
&\quad +K_1\frac{\widehat{m}_0(s^*)-m_0(s^*)-\widehat{m}_1(s^*)+m_1(s^*)}{K_2+K_1f_0(s^*)/f_1(s^*)}+\frac{m_0(s^*)-m_1(s^*)}{K_2+K_1f_0(s^*)/f_1(s^*)}n^{-1}\sum_{i=1}^n\phi_{K_1,i} \\
&\quad -\frac{\int_{D_c}\{m_0(s)-m_1(s)\}dF_0(s)+K_1\{m_0(s^*)-m_1(s^*)\}}{\{K_2+K_1f_0(s^*)/f_1(s^*)\}^2} \\
&\quad \times\left[\frac{1}{n}\sum_{i=1}^n\phi_{K_2,i}+\frac{1}{n}\sum_{i=1}^n\phi_{K_1,i}\frac{f_0(s^*)}{f_1(s^*)}+(nh)^{-1}\left\{\frac{\sum_{i=1}^n\phi_{0,i}(s^*)}{f_1(s^*)}-\frac{f_0(s^*)\sum_{i=1}^n\phi_{1,i}(s^*)}{f_1^2(s^*)}\right\}\right] \\
&\quad +o_p\{(nh)^{-1/2}\} \\
&= \frac{1}{K_2+K_1f_0(s^*)/f_1(s^*)}\left[n^{-1}\sum_{i=1}^n 2I(A_i=0)Y_iI(S_i\in D_c)-\int_{D_c}m_0(s)f_0(s)ds\right. \\
&\quad -n^{-1}\sum_{i=1}^n 2I(A_i=1)Y_iI(S_i\in D_c)f_0(S_i)/f_1(S_i)+\int_{D_c}m_1(s)f_0(s)ds \\
&\quad \left.+n^{-1}\sum_{i=1}^n 2I(A_i=0)I(S_i\in D_c)\{m_0(S_i)-m_1(S_i)\}-\int_{D_c}\{m_0(s)-m_1(s)\}f_0(s)ds\right] \\
&\quad +\frac{K_1}{K_2+K_1f_0(s^*)/f_1(s^*)}(nh)^{-1}\sum_{i=1}^n\{\phi_{m,0,i}(s^*)-\phi_{m,1,i}(s^*)\} \\
&\quad +\frac{m_0(s^*)-m_1(s^*)}{K_2+K_1f_0(s^*)/f_1(s^*)}n^{-1}\sum_{i=1}^n\phi_{K_1,i}+o_p\{(nh)^{-1/2}\} \\
&\quad -\frac{\int_{D_c}\{m_0(s)-m_1(s)\}dF_0(s)+K_1\{m_0(s^*)-m_1(s^*)\}}{\{K_2+K_1f_0(s^*)/f_1(s^*)\}^2} \\
&\quad \times\left[\frac{1}{n}\sum_{i=1}^n\phi_{K_2,i}+\frac{1}{n}\sum_{i=1}^n\phi_{K_1,i}\frac{f_0(s^*)}{f_1(s^*)}+(nh)^{-1}\left\{\frac{\sum_{i=1}^n\phi_{0,i}(s^*)}{f_1(s^*)}-\frac{f_0(s^*)\sum_{i=1}^n\phi_{1,i}(s^*)}{f_1^2(s^*)}\right\}\right] \\
&:= (nh)^{-1}\sum_{i=1}^n\phi_{\lambda,i}+o_p\{(nh)^{-1/2}\}.
\end{aligned}
$$

Similarly, we can get $\hat{c}-c=(nh)^{-1}\sum_{i=1}^n\phi_{c,i}+o_p\{(nh)^{-1/2}\}$.

Using above results we can obtain the influence functions for the optimal transformation function estimators by coupling delta method with the fact that

$$\begin{aligned} g_{\text{opt}}(s) &= \widetilde{G}\left(m_0(s), m_1(s), f_0(s), f_1(s), \lambda, c\right) \\ \text{and } \hat{g}(s) &= \widetilde{G}\left(\widehat{m}_0(s), \widehat{m}_1(s), \widehat{f}_0(s), \widehat{f}_1(s), \widehat{\lambda}, \hat{c}\right). \end{aligned}$$

Specifically, we can show that

$$\hat{g}(s) - g_{\text{opt}}(s) = (nh)^{-1} \sum_{i=1}^{n} \phi_{g,i}(s) + o_p\{(nh)^{-1/2}\},$$

where $E(\phi_{g,i}^2(s)) < \infty$.

Appendix E Perturbation resampling

For resampling, we may generate $\mathbf{V} = (V_1, ..., V_n)$ from independent and identically distributed non-negative random variables with mean 1 and variance 1 such as the unit exponential distribution. For each set of $\mathbf{V}$, we let $\widehat{\mu}_a^* = \{\sum_{i:A_i=a} Y_i V_i\}/\{\sum_{i:A_i=a} V_i\}$, $\widehat{\mu}_{a,g(S)}^* = \{\sum_{i:A_i=a} g(S_i) V_i\}/\{\sum_{i:A_i=a} V_i\}$,

$$\widehat{f}_a^*(s) = \frac{\sum_{i:A_i=a} K_h(S_i - s) V_i}{\sum_{i:A_i=a} V_i}, \ \widehat{m}_a^*(s) = \frac{\sum_{i:A_i=a} K_h(S_i - s) Y_i V_i}{\sum_{i:A_i=a} K_h(S_i - s) V_i}, \ \widehat{\Delta}_{01}^*(s) = \widehat{m}_0^*(s) - \widehat{m}_1^*(s)$$

$$\widehat{\lambda}^* = \left\{\hat{K}_2^* + \hat{K}_1^* \hat{r}(s^*)\right\}^{-1} \left\{\int_{D_c} \widehat{\Delta}_{01}^*(s) \widehat{f}_0^*(s) ds + \hat{K}_1^* \widehat{\Delta}_{01}^*(s^*)\right\},$$

$$\hat{c}^* = \left\{\hat{K}_2^* + \hat{K}_1^* \hat{r}^*(s^*)\right\}^{-1} \left\{\hat{r}^*(s^*) \int_{D_c} \widehat{\Delta}_{01}^*(s) f_0^*(s) ds - \hat{K}_2^* \widehat{\Delta}_{01}^*(s^*)\right\},$$

where $\widehat{r}^*(s) = \widehat{f}_0^*(s)/\widehat{f}_1^*(s)$, $\hat{K}_1^* = \int_{D_0} \widehat{f}_0^*(s) ds$, $\hat{K}_2^* = \int_{D_c} \widehat{r}^*(s) \widehat{f}_0^*(s) ds$. Then we may obtain the perturbed counterparts of $\widehat{g}(s)$, $\widehat{\text{PTE}}_{\widehat{g}}$, and $\widehat{\text{RP}}_{\widehat{g}}$ as

$$\widehat{g}^*(s) = \begin{cases} \widehat{m}_1^*(s) + \widehat{\lambda}^* \widehat{r}_0^*(s), \ s \in D_c \cup D_1 \\ \widehat{m}_0^*(s) + \hat{c}^*, \ s \in D_0. \end{cases}$$

$$\widehat{\text{PTE}}_{\widehat{g}^*}^* = \widehat{\Delta}_{\widehat{g}^*}^* / \widehat{\Delta}^*,$$

$$\widehat{\text{RP}}_{\widehat{g}^*}^*(\bar{n}) := \widehat{\text{RP}}_{\widehat{g}^*}^*(\bar{n}, \bar{n}) \text{ where } \widehat{\text{RP}}_g^*(n_1, n_2) = \frac{\mathcal{P}(\widehat{\Delta}_g^*/\widehat{\sigma}_g^*, n_1)}{\mathcal{P}(\widehat{\Delta}^*/\widehat{\sigma}^*, n_2)}.$$

where $\widehat{\Delta}^* = \widehat{\mu}_1^* - \widehat{\mu}_0^*$, $\widehat{\Delta}_{\widehat{g}^*}^* = \widehat{\mu}_{1,\widehat{g}^*(S)}^* - \widehat{\mu}_{0,\widehat{g}^*(S)}^*$, $\widehat{\sigma}^{*2} = n^{-1} \sum_{i=1}^n V_i \widehat{\psi}_i^2$ and $\widehat{\sigma}_g^{*2} = n^{-1} \sum_{i=1}^n V_i \widehat{\psi}_{g,i}^2$. In practice, we may generate a large number, say B, realizations for $\mathbf{V}$, and then obtain B realizations of $\widehat{g}^*(s)$, $\widehat{\text{PTE}}_{\widehat{g}^*}^*$ and $\widehat{\text{RP}}_{\widehat{g}^*}^*(\bar{n})$. The variance estimation and the confidence interval (CI) can be constructed based on the empirical variances and quantiles of these realizations.

Appendix F More simulation results

	Proposed				PTE_{W2020}		PTE_L		PTE_W		PTE_F	
	True	Est	ESE_{ASE}	CP	Est	ESE	Est	ESE	Est	ESE	Est	ESE
$(1)_{new}$	.587	.567	$.069_{.086}$	.972	.520	.055	.264	.068	.106	.043	.103	.041
$(2)_{new}$	.096	.099	$.050_{.056}$	.972	.063	.049	-.266	.061	.036	.016	.027	.012
$(3)_{new}$	.198	.192	$.028_{.030}$	.952	.157	.020	-.014	.014	-.024	.009	-.019	.007
$(4)_{new}$	.588	.598	$.061_{.073}$	.982	.575	.048	.319	.087	.341	.081	.315	.069

Estimates (Est) of PTE (using our proposed g_{opt}), PTE_{W2020}, PTE_L, PTE_W, and PTE_F along with their empirical standard errors (ESE) under settings $(1)_{new}$-$(4)_{new}$; for PTE estimates using our proposed g_{opt}, we also present the averages of the estimate standard errors (ASE, shown in subscript) along with the empirical coverage probabilities (CP) of the 95% confidence intervals.

References

Parast, L., McDermott, M. M., and Tian, L. (2016). Robust estimation of the proportion of treatment effect explained by surrogate marker information. *Statistics in medicine*, 35(10):1637–1653.