Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Robust Estimation of the Proportion of Treatment Effect Explained by Surrogate Marker Information: Supplementary Materials

Appendix A:. Equivalence of the surrogacy of S and $W = \mu_T(\mathbf{S})$

We aim to show that the surrogacy of $\mathbf{S}$, as defined by $R_{\mathbf{S}}$, is equivalent to the surrogacy of $W = \mu_T(\mathbf{S}) = E(Y^{(T)}|\mathbf{S}^{(T)} = \mathbf{S})$, as defined by R_W, where

$$R_W = 1 - \frac{\Delta_W}{\Delta} \qquad \text{and} \qquad \Delta_W = \int E(Y^{(T)}|W^{(T)} = w)dF_{W^{(C)}}(w) - E(Y^{(C)}).$$

That is, we must show that $\Delta_{\mathbf{S}} = \Delta_W$ when $W = \mu_T(\mathbf{S})$, i.e. W is obtained using the true model. Note that

$$\begin{aligned}
\Delta_{\mathbf{S}} &= \int E(Y^{(T)}|\mathbf{S}^{(T)} = \mathrm{s})dF_C(\mathbf{s}) - E(Y^{(C)}) \\
&= \int \mu_T(\mathbf{s})dF_C(\mathbf{s}) - E(Y^{(C)}) \\
&= \int E\{Y^{(T)}|\mu_T(\mathbf{S}^{(T)}) = \mu_T(\mathbf{s})\}dF_C(\mathbf{s}) - E(Y^{(C)}) \\
&= \int E\{Y^{(T)}|W^{(T)} = \mu_T(\mathbf{s})\}dF_C(\mathbf{s}) - E(Y^{(C)}) \\
&= E\left[E\{Y^{(T)}|W^{(T)} = \mu_T(\mathbf{S}^{(C)})\} \mid \mathbf{S}^{(C)}\right] - E(Y^{(C)}) \\
&= E\left[E\{Y^{(T)}|W^{(T)} = W^{(C)})\} \mid W^{(C)}\right] - E(Y^{(C)}) \\
&= \int E(Y^{(T)}|W^{(T)} = w)dF_{W^{(C)}}(w) - E(Y^{(C)}) = \Delta_W,
\end{aligned}$$

where $W^{(T)} = \mu_T(\mathbf{S}^{(T)})$, $W^{(C)} = \mu_T(\mathbf{S}^{(C)})$, $F_{W^{(C)}}(\cdot)$ is the cumulative distribution function of $W^{(C)}$, and we have used the fact that

$$E\{Y^{(T)}|\mu_T(\mathbf{S}^{(T)}) = w\} = \int E\{Y^{(T)}|\mathbf{S}^{(T)} = \mathrm{s}\}dF_w(\mathbf{s})d\mathbf{s} = \int \mu_T(\mathbf{s})dF_w(\mathbf{s}) = w,$$

where $f_w(\mathbf{s})$ is the cumulative distribution of $\mathbf{S}^{(T)}$ conditional on $\mu_T(\mathbf{S}^{(T)}) = w$.

Appendix B:. Asymptotic Properties of $\widehat{\Delta}_S$ and $\widehat{R}_S$

We assume the following regularity conditions:

1. $n = n_T + n_C$, $n_T/n = \pi_T > 0$ and $n_C/n = \pi_C > 0$
2. Both $S^{(T)}$ and $S^{(C)}$ are continuous random variables and $S^{(C)}$ has a finite support contained by the interval $[a, b]$, which is the support of $S^{(T)}$.

3. The functions $\mu_T(s)r(s)$ and $r(s)$ have continuous second derivatives over the interval $s \in [a, b]$, where $\mu_T(s) = E(Y^{(T)}|S^{(T)} = s)$, $r(s) = f_C(s)/f_T(s)$, $f_C(s)$ is the density function of $S^{(C)}$ and $f_T(s)$ is the density function of $S^{(T)}$.
4. The kernel function $K(s)$ is a smooth function with finite support, symmetric at zero and $\int K(s)ds = 1$. Denote $K_h(s) = K(s/h)/h$, where h is the smoothing bandwidth. Here we assume that $h = O_p(n^{-\delta}), \delta \in (1/4, 1/2)$.

Recall that

$$\widehat{\Delta}_S = \int \hat{\mu}_T(s)d\hat{F}_C(s) - n_C^{-1}\sum_{i=1}^{n_C} Y_{Ci} = n_C^{-1}\sum_{i=1}^{n_C}\widehat{\mu}_T(S_{Ci}) - n_C^{-1}\sum_{i=1}^{n_C} Y_{Ci}$$

where $\hat{F}_C(s)$ is the empirical estimate of $F_C(s)$ and

$$\hat{\mu}_T(s) = \frac{n_T^{-1}\sum_{j=1}^{n_T} Y_{Tj}K_h(S_{Tj}-s)}{\hat{f}_T(s)}, \text{ and } \hat{f}_T(s) = n_T^{-1}\sum_{j=1}^{n_T} K_h(S_{Tj}-s)$$

are the nonparametric estimators for $\mu_T(s)$ and $f_T(s)$, respectively. We first note the fact that

$$\sup_s \left\{ |n_T^{-1}\sum_{j=1}^{n_T} Y_{Tj}K_h(S_{Tj}-s) - \mu_T(s)f_T(s)| + |\hat{f}_T(s) - f_T(s)| \right\} = O_p(h^2 + \log(n)(nh)^{-\frac{1}{2}}) = O_p(\log(n)n^{\frac{\delta-1}{2}}),$$

which implies that

$$\sup_s |\hat{\mu}_T(s) - \mu_T(s)| = O_p(h^2 + \log(n)(nh)^{-\frac{1}{2}}) = O_p(\log(n)n^{\frac{\delta-1}{2}}). \quad (1)$$

To show the consistency of $\widehat{\Delta}_S$, it is sufficient to demonstrate that

$$\left|\int \hat{\mu}_T(s)d\hat{F}_C(s) - \int \mu_T(s)dF_C(s)\right| = o_p(1),$$

where $\hat{F}_C(s)$ is the empirical cumulative distribution functions based on $\{S_{Cj}, j = 1, \cdots, n_C\}$ and $F_C(s)$ are the true cumulative distribution function of $S^{(C)}$, since $n_C^{-1}\sum_{i=1}^{n_C} Y_{Ci} \to E(Y^{(C)})$ in probability. It follows that

$$\begin{aligned} & \left|\int \hat{\mu}_T(s)d\hat{F}_C(s) - \int \mu_T(s)dF_C(s)\right| \\ \leq & \left|\int \{\hat{\mu}_T(s) - \mu_T(s)\}dF_C(s)\right| + \left|\int \mu_T(s)d\{\hat{F}_C(s) - F_C(s)\}\right| + \left|\int \{\hat{\mu}_T(s) - \mu_T(s)\}d\{\hat{F}_C(s) - F_C(s)\}\right| \\ = & \; o_p(1) \end{aligned}$$

where the first two terms are bounded by the uniform consistency of $\hat{\mu}_T(s)$ and $\hat{F}_C(s)$ while the variance of the last term is $o_p(n^{-1})$ when $nh \to \infty$. The consistency of $\widehat{R}_S$ then follows since $\widehat{\Delta}$ is a consistent estimator of Δ.

In order to derive the asymptotic distribution of $n^{\frac{1}{2}}\left\{\widehat{\Delta}_S - \Delta_S\right\}$ and $n^{\frac{1}{2}}\left\{\widehat{R}_S - R_S\right\}$, we first derive the asymptotic distribution of

$$n^{\frac{1}{2}}\left\{\int \hat{\mu}_T(s)d\hat{F}_C(s) - \int \mu_T(s)dF_C(s)\right\}.$$

Similar to the above, we have the decomposition:

$$\begin{aligned} & \int \hat{\mu}_T(s)d\hat{F}_C(s) - \int \mu_T(s)dF_C(s) \\ = & \int \{\hat{\mu}_T(s) - \mu_T(s)\}dF_C(s) + \int \mu_T(s)d\{\hat{F}_C(s) - F_C(s)\} + o_p(n^{-\frac{1}{2}}). \end{aligned}$$

Therefore,

$$\begin{aligned} & n^{\frac{1}{2}}\left\{\int \hat{\mu}_T(s)d\hat{F}_C(s) - \int \mu_T(s)dF_C(s)\right\} \\ = & n^{\frac{1}{2}}\int \{\hat{\mu}_T(s) - \mu_T(s)\}dF_C(s) + n^{\frac{1}{2}}\int \mu_T(s)d\{\hat{F}_C(s) - F_C(s)\} + o_p(1) \\ = & I_1 + I_2 + o_p(1). \end{aligned}$$

Now consider the first term I_1 :

$$\begin{aligned}
I_1 =& n^{\frac{1}{2}} \int \{\hat{\mu}_T(s) - \mu_T(s)\} dF_C(s) \\
=& n^{\frac{1}{2}} \int \left\{ \frac{n_T^{-1} \sum_{j=1}^{n_T} Y_{Tj} K_h(S_{Tj} - s)}{n_T^{-1} \sum_{j=1}^{n_T} K_h(S_{Tj} - s)} - \mu_T(s) \right\} dF_C(s) \\
=& n^{\frac{1}{2}} \int \left[-\hat{\mu}_T(s) n_T^{-1} \sum_{j=1}^{n_T} \left\{ \frac{K_h(S_{Tj} - s)}{f_T(s)} - 1 \right\} + n_T^{-1} \sum_{j=1}^{n_T} \left\{ \frac{Y_{Tj} K_h(S_{Tj} - s)}{f_T(s)} - \mu_T(s) \right\} \right] dF_C(s) \\
=& n^{\frac{1}{2}} \int \left[-\mu_T(s) n_T^{-1} \sum_{j=1}^{n_T} \left\{ \frac{K_h(S_{Tj} - s)}{f_T(s)} - 1 \right\} + n_T^{-1} \sum_{j=1}^{n_T} \left\{ \frac{Y_{Tj} K_h(S_{Tj} - s)}{f_T(s)} - \mu_T(s) \right\} \right] dF_C(s) + O_p(\log(n)^2 n^{\delta - \frac{1}{2}}) \\
=& (\pi_T n_T)^{-\frac{1}{2}} \sum_{j=1}^{n_T} \int \left[-\frac{\mu_T(s)}{f_T(s)} \{K_h(S_{Tj} - s) - f_T(s)\} + \left\{ \frac{Y_{Tj} K_h(S_{Tj} - s)}{f_T(s)} - \mu_T(s) \right\} \right] dF_C(s) + o_p(1).
\end{aligned}$$

In the derivations, we used the uniform bounds the differences $|\hat{\mu}_T(s) - \mu_T(s)|$ and $|\widehat{f}_T(s) - f_T(s)|$. Since

$$\begin{aligned}
& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \int \frac{\mu_T(s)}{f_T(s)} \{K_h(S_{Tj} - s) - f_T(s)\} dF_C(s) \\
=& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \left\{ \int \mu_T(S_{Tj} + hs) r(S_{Tj} + hs) K(s) ds - \int \mu_T(s) dF_C(s) \right\} \\
=& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \left[\mu_T(S_{Tj}) r(S_{Tj}) - \int \mu_T(s) dF_C(s) \right] + O_p(n^{\frac{1}{2} - 2\delta})
\end{aligned}$$

and

$$\begin{aligned}
& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \int \left\{ \frac{Y_{Tj} K_h(S_{Tj} - s)}{f_T(s)} - \mu_T(s) \right\} dF_C(s) \\
=& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \left\{ Y_{Tj} \int r(S_{Tj} + hs) K(s) ds - \int \mu_T(s) dF_C(s) \right\} \\
=& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \left[Y_{Tj} r(S_{Tj}) - \int \mu_T(s) dF_C(s) \right] + O_p(n^{\frac{1}{2} - 2\delta}).
\end{aligned}$$

Using a change of variables and Taylor series expansion, it follows that

$$I_1 = (\pi_T n_T)^{-\frac{1}{2}} \sum_{j=1}^{n_T} \{Y_{Tj} - \mu_T(S_{Tj})\} r(S_{Tj}) + o_p(1).$$

Coupled with the fact that

$$I_2 = (\pi_C n_C)^{-\frac{1}{2}} \sum_{j=1}^{n_C} \left\{ \mu_T(S_{Cj}) - \int \mu_T(s) F_C(s) \right\}$$

it suggests that

$$\begin{aligned}
& n^{\frac{1}{2}} \left\{ \int \hat{\mu}_T(s) d\hat{F}_C(s) - \int \mu_T(s) dF_C(s) \right\} \\
=& (\pi_T n_T)^{-\frac{1}{2}} \sum_{j=1}^{n_T} \{Y_{Tj} - \mu_T(S_{Tj})\} r(S_{Tj}) + (\pi_C n_C)^{-\frac{1}{2}} \sum_{j=1}^{n_C} \left\{ \mu_T(S_{Cj}) - \int \mu_T(s) F_C(s) \right\} + o_p(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
n^{\frac{1}{2}} \begin{pmatrix} \widehat{\Delta}_S - \Delta_S \\ \widehat{\Delta} - \Delta \end{pmatrix} =& (\pi_T n_T)^{-\frac{1}{2}} \sum_{j=1}^{n_T} \begin{pmatrix} \{Y_{Tj} - \mu_T(S_{Tj})\} r(S_{Tj}) \\ Y_{Tj} - E(Y^{(T)}) \end{pmatrix} \\
& - (\pi_C n_C)^{-\frac{1}{2}} \sum_{j=1}^{n_C} \begin{pmatrix} Y_{jC} - E(Y^{(C)}) - \{\mu_T(S_{Cj}) - \int \mu_T(s) F_C(s)\} \\ Y_{jC} - E(Y^{(C)}) \end{pmatrix} + o_p(1)
\end{aligned}$$

By the central limit theorem $n^{\frac{1}{2}}(\widehat{\Delta}_S - \Delta_S, \widehat{\Delta} - \Delta)'$ converges to a bivariate normal with a variance-covariance matrix of

$$\Sigma_\Delta = \pi_T^{-1} E \left(\begin{array}{c} \{Y_{Tj} - \mu_T(S_{Tj})\}\, r(S_{Tj}) \\ Y_{Tj} - E(Y^{(T)}) \end{array} \right)^{\otimes 2} + \pi_C^{-1} E \left(\begin{array}{c} Y_{jC} - E(Y^{(C)}) - \{\mu_T(S_{Cj}) - \int \mu_T(s) F_C(s)\} \\ Y_{jC} - E(Y^{(C)}) \end{array} \right)^{\otimes 2},$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}'$ for a vector $\boldsymbol{a}$. By the multivariate delta method, $n^{\frac{1}{2}}(\widehat{R}_S - R_S)$ also converges to a mean zero Gaussian distribution weakly as the sample size $n \to \infty$ as long as $\Delta \neq 0$.

Appendix C:. Asymptotic Properties of $\widehat{\Delta}_{\mathbf{S}}$ and $R_{\mathbf{S}}$.

In this section, we additionally assume the following regularity conditions:

1. $\widehat{\boldsymbol{\beta}}$ is a regular estimator in that $\widehat{\boldsymbol{\beta}}$ converges to a deterministic vector $\boldsymbol{\beta}_0$ in probability and has the expansion,

$$n_T^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \tau_{Tj} + o_p(1)$$

   for i.i.d. τ_{Tj}, which is a mean zero random vector with a finite second moment. One consequence is that $|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| = O_p(n^{-\frac{1}{2}})$.
2. Both $\mathbf{S}^{(T)}$ and $\mathbf{S}^{(C)}$ are bounded. The random variables $\boldsymbol{\beta}_0'\mathbf{S}^{(T)}$ and $\boldsymbol{\beta}_0'\mathbf{S}^{(C)}$ have continuous density function and $\boldsymbol{\beta}_0'\mathbf{S}^{(C)}$ has a finite support contained by the interval $[a, b]$, which is the support of $\boldsymbol{\beta}_0'\mathbf{S}^{(T)}$.
3. The function $\mu_{\boldsymbol{\beta}_0 T}(s) r_{\boldsymbol{\beta}_0}(s)$, has a continuous second derivative over the interval $s \in [a, b]$, where $\mu_{\boldsymbol{\beta}T}(s) = E(Y^{(T)} | \boldsymbol{\beta}'\mathbf{S}^{(T)} = s)$, $r_{\boldsymbol{\beta}}(s) = f_{\boldsymbol{\beta}C}(s)/f_{\boldsymbol{\beta}T}(s)$, $f_{\boldsymbol{\beta}C}(s)$ is the density function of $\boldsymbol{\beta}'\mathbf{S}^{(C)}$ and $f_{\boldsymbol{\beta}T}(s)$ is the density function of $\boldsymbol{\beta}'\mathbf{S}^{(T)}$.
4. For sufficiently small $\delta > 0$

$$|r_{\boldsymbol{\beta}_2}(s) - r_{\boldsymbol{\beta}_1}(s)| + |\mu_{\boldsymbol{\beta}_2 T}(s) r_{\boldsymbol{\beta}_2}(s) - \mu_{\boldsymbol{\beta}_1 T}(s) r_{\boldsymbol{\beta}_1}(s)| \leq C_0 |\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|$$

   for $|\boldsymbol{\beta}_i - \boldsymbol{\beta}_0| < \delta, i = 1, 2$.

Recall that

$$\widehat{\Delta}_{\mathbf{S}} = \int \hat{\mu}_{\hat{\boldsymbol{\beta}}T}(s) d\hat{F}_{\hat{\boldsymbol{\beta}}C}(s) - n_C^{-1} \sum_{i=1}^{n_C} Y_{Ci} = n_C^{-1} \sum_{i=1}^{n_C} \widehat{\mu}_T(\widehat{Q}_{Ci}) - n_C^{-1} \sum_{i=1}^{n_C} Y_{Ci}$$

where $\hat{F}_{\boldsymbol{\beta}C}(s)$ is the empirical estimate of $F_{\boldsymbol{\beta}C}(s)$, $\widehat{Q}_{Ci} = \widehat{\boldsymbol{\beta}}'\mathbf{S}_{Ci}$, and

$$\hat{\mu}_{\boldsymbol{\beta}T}(s) = \frac{n_T^{-1} \sum_{j=1}^{n_T} Y_{Tj} K_h(\boldsymbol{\beta}'\mathbf{S}_{Tj} - s)}{\hat{f}_{\boldsymbol{\beta}T}(s)} \quad \text{and} \quad \hat{f}_{\boldsymbol{\beta}T}(s) = n_T^{-1} \sum_{j=1}^{n_T} K_h(\boldsymbol{\beta}'\mathbf{S}_{Tj} - s)$$

are the nonparametric estimators for $\mu_{\boldsymbol{\beta}T}(s) = E(Y^{(T)} | \boldsymbol{\beta}'\mathbf{S}^{(T)} = s)$ and $f_{\boldsymbol{\beta}T}(s)$, the density function of $\boldsymbol{\beta}'\mathbf{S}^{(T)}$, based on $\{(Y_{Tj}, \mathbf{S}_{Tj}), j = 1, \cdots, n_T\}$, respectively. We first note that

$$n_T^{\frac{1}{2}} \{\hat{f}_{\hat{\boldsymbol{\beta}}T}(s) - \hat{f}_{\boldsymbol{\beta}_0 T}(s)\} = O_p\left(n^{\delta - \frac{1}{4}} \log(n)^2 + 1\right) \tag{2}$$

since the difference

$$\begin{aligned}
& n_T^{\frac{1}{2}} \{\hat{f}_{\hat{\boldsymbol{\beta}}T}(s) - \hat{f}_{\boldsymbol{\beta}_0 T}(s)\} \\
=& n_T^{\frac{1}{2}} \int K_h(u - s) d\{\hat{F}_{\hat{\boldsymbol{\beta}}T}(u) - \hat{F}_{\boldsymbol{\beta}_0 T}(u)\} \\
=& n_T^{\frac{1}{2}} \int K_h(u - s) d\{\hat{F}_{\hat{\boldsymbol{\beta}}T}(u) - \hat{F}_{\boldsymbol{\beta}_0 T}(u) - F_{\hat{\boldsymbol{\beta}}T}(u) + F_{\boldsymbol{\beta}_0 T}(u)\} + n_T^{\frac{1}{2}} \int K_h(u - s) d\{F_{\hat{\boldsymbol{\beta}}T}(u) - F_{\boldsymbol{\beta}_0 T}(u)\} \\
\leq& \sup_u |n_T^{\frac{1}{2}} \left\{\hat{F}_{\hat{\boldsymbol{\beta}}T}(u) - \hat{F}_{\boldsymbol{\beta}_0 T}(u) - F_{\hat{\boldsymbol{\beta}}T}(u) + F_{\boldsymbol{\beta}_0 T}(u)\right\}| \int h^{-1} |\dot{K}(u)| du + n_T^{\frac{1}{2}} |f_{\hat{\boldsymbol{\beta}}T}(u) - f_{\boldsymbol{\beta}_0 T}(u)| \\
=& O_p\left(n^{\delta - \frac{1}{4}} \log(n)^2 + 1\right).
\end{aligned}$$

The last approximation holds since the class of function $\mathcal{F} = \{I(\boldsymbol{\beta}'\mathbf{S} \leq s) - I(\boldsymbol{\beta}_0'\mathbf{S} \leq s) | s \in R, |\boldsymbol{\beta} - \boldsymbol{\beta}_0| < \delta\}$ is Donsker with an envelop function 1. Specifically, we can calculate the bracketing entropy $\log\{N_{[]}(\epsilon, \mathcal{F}, \| \cdot \|)\} \leq C_1 |\log(\epsilon)|$ for a positive constant C_1, which implies that the bracketing entropy integral

$$J_{[]}(\delta, \mathcal{F}, \| \cdot \|) = \int_0^\delta \left[1 + \log\{N_{[]}(\epsilon, \mathcal{F}, \| \cdot \|)\}\right]^{\frac{1}{2}} d\epsilon \leq C_2 \delta \log(\delta)^2 < \infty,$$

where the norm $\|f\| = [E\{f^2(\mathbf{S}^{(T)})\}]^{\frac{1}{2}}$ for $f \in \mathcal{F}$ and C_2 is a constant. Furthermore since $E\{f^2(\mathbf{S}^{(C)})\} \leq C_3 \delta$ for $f \in \mathcal{F}$, it follows from the maximum inequality that

$$\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0| \leq \delta, u} |n_T^{\frac{1}{2}} \left\{ \hat{F}_{\boldsymbol{\beta}T}(u) - \hat{F}_{\boldsymbol{\beta}_0 T}(u) - F_{\boldsymbol{\beta}T}(u) + F_{\boldsymbol{\beta}_0 T}(u) \right\} | \leq C_4 \delta^{\frac{1}{2}} \log(\delta)^2,$$

where $C_j, j = 3, 4$ are positive constants. Considering the assumption that $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| = O_p(n^{-\frac{1}{2}})$, thus we have

$$\sup_u |n_T^{\frac{1}{2}} \left\{ \hat{F}_{\hat{\boldsymbol{\beta}}T}(u) - \hat{F}_{\boldsymbol{\beta}_0 T}(u) - F_{\hat{\boldsymbol{\beta}}T}(u) + F_{\boldsymbol{\beta}_0 T}(u) \right\} | = O_p(n^{-\frac{1}{4}}).$$

Therefore

$$\sup_s |\hat{f}_{\hat{\boldsymbol{\beta}}T}(s) - f_{\boldsymbol{\beta}_0 T}(s)| \leq \sup_s |\hat{f}_{\hat{\boldsymbol{\beta}}T}(s) - \hat{f}_{\boldsymbol{\beta}_0 T}(s)| + \sup_s |\hat{f}_{\boldsymbol{\beta}_0 T}(s) - f_{\boldsymbol{\beta}_0 T}(s)| = O_p(\log(n) n^{\frac{\delta - 1}{2}})$$

Similarly, it can be shown that

$$\sup_s |\hat{\mu}_{\hat{\boldsymbol{\beta}}T}(s) - \mu_{\boldsymbol{\beta}_0 T}(s)| = O_p(\log(n) n^{\frac{\delta - 1}{2}}). \tag{3}$$

Therefore, the consistency of $\widehat{\Delta}_{\mathbf{S}}$ follows using similar arguments as in the preceding single surrogate marker setting given (2) and (3). In addition, it follows from the same arguments used above that the process

$$\begin{aligned} & n^{\frac{1}{2}} \left\{ \int \hat{\mu}_{\hat{\boldsymbol{\beta}}T}(s) d\hat{F}_{\hat{\boldsymbol{\beta}}C}(s) - \int \mu_{\boldsymbol{\beta}_0 T}(s) dF_{\boldsymbol{\beta}_0 C}(s) \right\} \\ =& n^{\frac{1}{2}} \int \{\hat{\mu}_{\hat{\boldsymbol{\beta}}T}(s) - \mu_{\boldsymbol{\beta}_0 T}(s)\} dF_{\boldsymbol{\beta}C}(s) + n^{\frac{1}{2}} \int \mu_{\boldsymbol{\beta}_0 T}(s) d\{\hat{F}_{\hat{\boldsymbol{\beta}}C}(s) - F_{\boldsymbol{\beta}_0 C}(s)\} + o_p(1) \\ =& \tilde{I}_1 + \tilde{I}_2 + o_p(1). \end{aligned}$$

Here

$$\begin{aligned} \tilde{I}_1 =& n^{\frac{1}{2}} \int \{\hat{\mu}_{\hat{\boldsymbol{\beta}}T}(s) - \mu_{\boldsymbol{\beta}_0 T}(s)\} dF_{\boldsymbol{\beta}_0 C}(s) \\ =& (\pi_T n_T)^{-\frac{1}{2}} \sum_{j=1}^{n_T} \int \left[-\frac{\mu_{\hat{\boldsymbol{\beta}}T}(s)}{f_{\hat{\boldsymbol{\beta}}T}(s)} \left\{ K_h(\hat{\boldsymbol{\beta}}'\mathbf{S}_{Tj} - s) - f_{\hat{\boldsymbol{\beta}}T}(s) \right\} + \left\{ \frac{Y_{Tj} K_h(\hat{\boldsymbol{\beta}}' S_{Tj} - s)}{f_{\hat{\boldsymbol{\beta}}T}(s)} - \mu_{\boldsymbol{\beta}_0 T}(s) \right\} \right] dF_{\boldsymbol{\beta}_0 C}(s) + o_p(1). \end{aligned}$$

Consider the processes

$$\begin{aligned} Q_{1n}(\boldsymbol{\beta}) =& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \left[\int \frac{\mu_{\boldsymbol{\beta}T}(s)}{f_{\boldsymbol{\beta}T}(s)} K_h(\boldsymbol{\beta}'\mathbf{S}_{Tj} - s) dF_{\boldsymbol{\beta}C}(s) - \int \mu_{\boldsymbol{\beta}T}(s) dF_{\boldsymbol{\beta}T}(s) \right] \\ =& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \left[\frac{\mu_{\boldsymbol{\beta}T}(\boldsymbol{\beta}'\mathbf{S}_{Tj}) f_{\boldsymbol{\beta}C}(\boldsymbol{\beta}'\mathbf{S}_{Tj})}{f_{\boldsymbol{\beta}T}(\boldsymbol{\beta}'\mathbf{S}_{Tj})} - \int \mu_{\boldsymbol{\beta}T}(s) dF_{\boldsymbol{\beta}T}(s) \right] + o_p(1) \end{aligned}$$

and

$$\begin{aligned} Q_{2n}(\boldsymbol{\beta}) =& n_T^{-\frac{1}{2}} \sum_{j=1}^{n_T} \left[\int \frac{Y_{Tj}}{f_{\boldsymbol{\beta}T}(s)} K_h(\boldsymbol{\beta}' S_{Tj} - s) dF_{\boldsymbol{\beta}C}(s) - \int \mu_{\boldsymbol{\beta}T}(s) dF_{\boldsymbol{\beta}T}(s) \right] \\ =& n_T^{-1/2} \sum_{j=1}^{n_T} \left[Y_{Tj} \frac{f_C(\boldsymbol{\beta}'\mathbf{S}_{Tj})}{f_{\boldsymbol{\beta}T}(\boldsymbol{\beta}'\mathbf{S}_{Tj})} - \int \mu_{\boldsymbol{\beta}T}(s) dF_{\boldsymbol{\beta}T}(s) \right] + o_p(1), \end{aligned}$$

where $|\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\delta$. Therefore

$$Q_{2n}(\boldsymbol{\beta})-Q_{1n}(\boldsymbol{\beta})=n_T^{-1/2}\sum_{j=1}^{n_T}r_{\boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{S}_{Tj})\left\{Y_{Tj}-\mu_{\boldsymbol{\beta}T}(\boldsymbol{\beta}'\mathbf{S}_{Tj})\right\}.$$

The class of function

$$\left\{r_{\boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{S})\{y-\mu_{\boldsymbol{\beta}T}(\boldsymbol{\beta}'\mathbf{S})\}\ \middle|\ |\boldsymbol{\beta}-\boldsymbol{\beta}_0|\le\delta\right\}$$

is Donsker under the regularity conditions and therefore

$$Q_{2n}(\hat{\boldsymbol{\beta}})-Q_{1n}(\hat{\boldsymbol{\beta}})=Q_{2n}(\boldsymbol{\beta}_0)-Q_{1n}(\boldsymbol{\beta}_0)+o_p(1)$$

which implies that

$$\tilde{I}_1=(\pi_Tn_T)^{-\frac{1}{2}}\sum_{j=1}^{n_T}\left\{Y_{Tj}-\mu_T(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})\right\}r_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})+n^{\frac{1}{2}}\int\{\mu_{\hat{\boldsymbol{\beta}}T}(s)-\mu_{\boldsymbol{\beta}_0T}(s)\}dF_{\hat{\boldsymbol{\beta}}C}(s)+o_p(1).$$

Next

$$\begin{aligned}\tilde{I}_2&=\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{\hat{F}_{\hat{\boldsymbol{\beta}}C}(s)-F_{\boldsymbol{\beta}_0C}(s)\}\\&=\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{\hat{F}_{\hat{\boldsymbol{\beta}}C}(s)-F_{\hat{\boldsymbol{\beta}}C}(s)+F_{\hat{\boldsymbol{\beta}}C}(s)-F_{\boldsymbol{\beta}_0C}(s)\}\\&=\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{\hat{F}_{\hat{\boldsymbol{\beta}}C}(s)-F_{\hat{\boldsymbol{\beta}}C}(s)\}+\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{F_{\hat{\boldsymbol{\beta}}C}(s)-F_{\boldsymbol{\beta}_0C}(s)\}\\&=\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{\hat{F}_{\boldsymbol{\beta}_0C}(s)-F_{\boldsymbol{\beta}_0C}(s)\}+\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{F_{\hat{\boldsymbol{\beta}}C}(s)-F_{\boldsymbol{\beta}_0C}(s)\}\\&\quad+\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{\hat{F}_{\hat{\boldsymbol{\beta}}C}(s)-F_{\hat{\boldsymbol{\beta}}C}(s)-\hat{F}_C(s)+F_{\boldsymbol{\beta}_0C}(s)\}\\&=\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{\hat{F}_{\boldsymbol{\beta}_0C}(s)-F_{\boldsymbol{\beta}_0C}(s)\}+\int\mu_{\boldsymbol{\beta}_0T}(s)dn^{\frac{1}{2}}\{F_{\hat{\boldsymbol{\beta}}C}(s)-F_{\boldsymbol{\beta}_0C}(s)\}+o_p(1)\end{aligned}$$

Therefore

$$\begin{aligned}&n^{\frac{1}{2}}\left\{\int\hat{\mu}_{\hat{\boldsymbol{\beta}}T}(s)d\hat{F}_{\hat{\boldsymbol{\beta}}C}(s)-\int\mu_{\boldsymbol{\beta}_0T}(s)dF_{\boldsymbol{\beta}_0C}(s)\right\}\\=&(\pi_Tn_T)^{-\frac{1}{2}}\sum_{j=1}^{n_T}\left\{Y_{Tj}-\mu_T(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})\right\}r_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})+(\pi_Cn_C)^{-\frac{1}{2}}\sum_{i=1}^{n_C}\left\{\mu_{\boldsymbol{\beta}_0T}(\boldsymbol{\beta}_0'S_{Cj})-\int\mu_{\boldsymbol{\beta}_0T}(s)F_{\boldsymbol{\beta}_0C}(s)\right\}\\&+n^{\frac{1}{2}}\left\{\int\mu_{\hat{\boldsymbol{\beta}}T}(s)dF_{\hat{\boldsymbol{\beta}}C}(s)-\int\mu_{\boldsymbol{\beta}_0T}(s)dF_{\boldsymbol{\beta}_0C}(s)\right\}+o_p(1)\\=&(\pi_Tn_T)^{-\frac{1}{2}}\sum_{j=1}^{n_T}\left\{Y_{Tj}-\mu_T(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})\right\}r_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})+(\pi_Cn_C)^{-\frac{1}{2}}\sum_{i=1}^{n_C}\left\{\mu_{\boldsymbol{\beta}_0T}(\boldsymbol{\beta}_0'S_{Cj})-\int\mu_{\boldsymbol{\beta}_0T}(s)F_{\boldsymbol{\beta}_0C}(s)\right\}\\&+n^{\frac{1}{2}}\mathbf{a}_0'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(1),\end{aligned}$$

where $\mathbf{a}_0=\partial\int\mu_{\boldsymbol{\beta}_0T}(s)F_{\boldsymbol{\beta}_0C}(s)/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0}$. Therefore

$$\begin{aligned}n^{\frac{1}{2}}\begin{pmatrix}\widehat{\Delta}_{\mathbf{S}}-\Delta_Q\\\widehat{\Delta}-\Delta\end{pmatrix}=&(\pi_Tn_T)^{-\frac{1}{2}}\sum_{j=1}^{n_T}\begin{pmatrix}\left\{Y_{Tj}-\mu_T(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})\right\}r(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})+\mathbf{a}_0'\tau_{Tj}\\Y_{Tj}-E(Y^{(T)})\end{pmatrix}\\&-(\pi_Cn_C)^{-\frac{1}{2}}\sum_{j=1}^{n_C}\begin{pmatrix}Y_{jC}-E(Y^{(C)})-\left\{\mu_T(\boldsymbol{\beta}_0'\mathbf{S}_{Cj})-\int\mu_T(s)F_C(s)\right\}\\Y_{jC}-E(Y^{(C)})\end{pmatrix}+o_p(1)\end{aligned}$$

By the central limit theorem $n^{\frac{1}{2}}(\widehat{\Delta}_{\mathbf{S}}-\Delta_Q,\widehat{\Delta}-\Delta)'$ converges to a mean-zero bivariate normal with a variance-covariance matrix of

$$\pi_T^{-1}E\begin{pmatrix}\left\{Y_{Tj}-\mu_T(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})\right\}r(\boldsymbol{\beta}_0'\mathbf{S}_{Tj})+\mathbf{a}_0'\tau_{Tj}\\Y_{Tj}-E(Y^{(T)})\end{pmatrix}^{\otimes2}+\pi_C^{-1}E\begin{pmatrix}Y_{jC}-E(Y^{(C)})-\left\{\mu_T(\boldsymbol{\beta}_0'\mathbf{S}_{Cj})-\int\mu_T(s)F_C(s)\right\}\\Y_{jC}-E(Y^{(C)})\end{pmatrix}^{\otimes2}.$$

It follows from the multivariate delta method that $n^{\frac{1}{2}}(\widehat{R}_{\mathbf{S}}-R_Q)$ also converges to a mean zero Gaussian distribution as long as $\Delta\neq0$.

Appendix D:. Justification of the Resampling Procedure

To justify the resampling procedure in the single surrogate marker setting, we first derive the asymptotic approximation for the perturbed estimators. Since the perturbation weights are independent (in contrast to the correlated implicit weights from the regular bootstrap method), the derivation of the asymptotical linear expansion of the perturbed estimator is almost the same as that for the original estimator in Appendix B under weak moment conditions for the perturbation weights, e.g., the existence of a finite third moment. Similar derivations have been used in other settings [1, 2, 3]. Specifically, we have

$$
\begin{aligned}
n^{\frac{1}{2}}\begin{pmatrix} \widehat{\Delta}_S^{(b)}-\Delta_S \\ \widehat{\Delta}^{(b)}-\Delta \end{pmatrix} =&(\pi_T n_T)^{-\frac{1}{2}}\sum_{j=1}^{n_T}\begin{pmatrix} \{Y_{Tj}-\mu_T(S_{Tj})\}\, r(S_{Tj}) \\ Y_{Tj}-E(Y^{(T)}) \end{pmatrix} V_{Tj}^{(b)} \\
&-(\pi_C n_C)^{-\frac{1}{2}}\sum_{j=1}^{n_C}\begin{pmatrix} Y_{jC}-E(Y^{(C)})-\left\{\mu_T(S_{Cj})-\int \mu_T(s)F_C(s)\right\} \\ Y_{jC}-E(Y^{(C)}) \end{pmatrix} V_{Cj}^{(b)}+o_p(1),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
n^{\frac{1}{2}}\begin{pmatrix} \widehat{\Delta}_S^{(b)}-\widehat{\Delta}_S \\ \widehat{\Delta}^{(b)}-\widehat{\Delta} \end{pmatrix} =&(\pi_T n_T)^{-\frac{1}{2}}\sum_{j=1}^{n_T}\begin{pmatrix} \{Y_{Tj}-\mu_T(S_{Tj})\}\, r(S_{Tj}) \\ Y_{Tj}-E(Y^{(T)}) \end{pmatrix} (V_{Tj}^{(b)}-1) \\
&-(\pi_C n_C)^{-\frac{1}{2}}\sum_{j=1}^{n_C}\begin{pmatrix} Y_{jC}-E(Y^{(C)})-\left\{\mu_T(S_{Cj})-\int \mu_T(s)F_C(s)\right\} \\ Y_{jC}-E(Y^{(C)}) \end{pmatrix} (V_{Cj}^{(b)}-1)+o_p(1).
\end{aligned}
$$

It follows from the central limit theorem that conditional on the data, $n^{\frac{1}{2}}\left(\widehat{\Delta}_S^{(b)}-\widehat{\Delta}_S, \widehat{\Delta}^{(b)}-\widehat{\Delta}\right)$ converges weakly to a mean zero Gaussian distribution with the variance covariance matrix of

$$
\begin{aligned}
&\lim_{n_T\to\infty}\frac{1}{\pi_T n_T}\sum_{j=1}^{n_T}\begin{pmatrix} \{Y_{Tj}-\mu_T(S_{Tj})\}\, r(S_{Tj}) \\ Y_{Tj}-E(Y^{(T)}) \end{pmatrix}^{\otimes 2} \\
+&\lim_{n_C\to\infty}\frac{1}{\pi_C n_C}\sum_{j=1}^{n_C}\begin{pmatrix} Y_{jC}-E(Y^{(C)})-\left\{\mu_T(S_{Cj})-\int \mu_T(s)F_C(s)\right\} \\ Y_{jC}-E(Y^{(C)}) \end{pmatrix}^{\otimes 2},
\end{aligned}
$$

which converges to Σ_Δ in probability. This implies that

$$
\sup_{\mathbf{d}\in R^2}\left|\text{pr}\left\{ n^{\frac{1}{2}}\begin{pmatrix} \widehat{\Delta}_S^{(b)}-\widehat{\Delta}_S \\ \widehat{\Delta}^{(b)}-\widehat{\Delta} \end{pmatrix} \le \mathbf{d} \;\middle|\; (Y_{Ti}, S_{Ti}, Y_{Cj}, S_{Cj}), 1\le i\le n_T, 1\le j\le n_C \right\} - \text{pr}\left(\mathbf{Z}_\Delta \le \mathbf{d}\right)\right| = o_p(1),
$$

where $\mathbf{Z}_\Delta$ is a bivariate Gaussian with mean zero and variance-covariance matrix of Σ_Δ. Therefore, the conditional distributions of $n^{\frac{1}{2}}(\widehat{\Delta}_S^{(b)}-\widehat{\Delta}_S)$, $n^{\frac{1}{2}}(\widehat{\Delta}^{(b)}-\widehat{\Delta})$, and $n^{\frac{1}{2}}(\widehat{R}_S^{(b)}-\widehat{R}_S)$ can be used to approximate the distributions of $n^{\frac{1}{2}}(\widehat{\Delta}_S-\Delta_S)$, $n^{\frac{1}{2}}(\widehat{\Delta}-\Delta)$, and $n^{\frac{1}{2}}(\widehat{R}_S-R_S)$ for large n.

The justification of the resampling procedure in the multivariate marker setting is similar under the additional assumption that

$$
n_T^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}^{(b)}-\boldsymbol{\beta}_0) = n_T^{-\frac{1}{2}}\sum_{j=1}^{n_T}\tau_{Tj}V_{Tj}^{(b)}+o_p(1)
$$

which holds in general when $\widehat{\boldsymbol{\beta}}$ is the root of an estimating equation. As mentioned in the text, this type of resampling approach is similar to the wild bootstrap [4, 5, 6].

References

1. Park Y, Wei LJ. Estimating subjectspecific survival functions under the accelerated failure time model. *Biometrika* 2003; **90**(3):717–723.
2. Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association* 2007; **102**(478):527–537.
3. Cai T, Tian L, Uno H, Solomon S, Wei LJ. Calibrating parametric subject-specific risk estimation. *Journal of the American Statistical Association* 2010; **97**(2):389–404.
4. Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* 1986; :1261–1295.
5. Hardle W. *Applied Nonparametric Regression*, vol. 27. Cambridge Univ Press, 1990.
6. Mammen E. Bootstrap, wild bootstrap, and asymptotic normality. *Probability Theory and Related Fields* 1992; **93**(4):439–455, doi: 10.1007/BF01192716.