JBS 000, 000–000 DOI: 000
000 0000

Supplementary Material for Bias-adjusted Kaplan-Meier Survival Curves for Marginal Treatment Effect in Observational Studies

Xiaofei Wang[1*], Fangfang Bai[2], Herbert Pang[3], and Stephen L George[1]

[1]Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, U.S.A.

[2]School of Statistics, University of International Business and Economics, Beijing, 100029, China

[3]Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China

[*]*email:* xiaofei.wang@duke.edu

This paper has been submitted for consideration for publication in *JBS*

Theoretical Results

Here, we give the detailed and complete proof of the main results in Section 3 in the paper.

Let

$$
\begin{aligned}
N_{1i}(t) &= I(Y_{1i} \leqslant t, \delta_{1i} = 1) & N_{1i}^{EL}(t) &= p_i D_i I(Y_{1i} \leqslant t, \delta_{1i} = 1) \\
Y_{1i}(t) &= I(Y_{1i} \geqslant t) & Y_{1i}^{EL}(t) &= p_i D_i I(Y_{1i} \geqslant t) \\
M_{1i}(t) &= N_{1i}(t) - \int_0^t Y_{1i}(s) d\Lambda_1(s) & M_{1i}^{EL}(t) &= N_{1i}^{EL}(t) - \int_0^t Y_{1i}^{EL}(s) d\Lambda_1(s) \\
\overline{N}_1(t) &= \sum_{i=1}^{n_1} I(Y_{1i} \leqslant t, \delta_{1i} = 1) & \overline{Y}_1(t) &= \sum_{i=1}^{n_1} I(Y_{1i} \geqslant t),\ \overline{Y}_1^{EL}(t) = \sum_{i=1}^{n} Y_{1i}^{EL}(t) \\
\overline{M}_1(t) &= \sum_{i=1}^{n_1} M_{1i}(t),\ \pi_1(t) = P(Y_1 \geqslant t) & \overline{M}_1^{EL}(t) &= \sum_{i=1}^{n} M_{1i}^{EL}(t) = \sum_{i=1}^{n} p_i D_i M_{1i}(t).
\end{aligned}
$$

Similarly, if the subscript 1 is replaced with subscript 0 in all the aforementioned equations, it will denote the corresponding quantities of $S_0(t)$. For simplicity, we omit it here.

Let $\xi = (\xi_1, \xi_2)^T$, $\xi_1 = \frac{n_1 - \gamma^T a}{n}$, $\xi_2 = \frac{\gamma^T}{n}$ and $\psi(Z_i) = (1, \phi^T(Z_i))^T$. Then $p_i D_i = \frac{1}{n} \frac{D_i}{\xi^T \psi(Z_i)}$ and $p_i(1 - D_i) = \frac{1}{n} \frac{1 - D_i}{1 - \xi^T \psi(Z_i)}$. Let $\widehat{\xi}$ be the solution of equations (2.1) and (2.2). Then we have

$$
\frac{1}{n} \sum_{i=1}^{n} \Big[\frac{D_i}{\widehat{\xi}^T \psi(Z_i)} - \frac{1 - D_i}{1 - \widehat{\xi}^T \psi(Z_i)} \Big] \psi(Z_i) = 0. \tag{0.1}
$$

Lemma 1: *$\widehat{\xi}$ is the solution of equation (0.1), and ξ_0 is the value satisfies $\xi_0^T \psi(Z) = e(Z)$, which is the true value of the propensity score $P(D = 1|Z)$. Then under the conditions in Theorem 1, we have*

$$
\sqrt{n}(\widehat{\xi} - \xi_0) \quad = \quad A^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(D_i - e(Z_i)) \psi(Z_i)}{e(Z_i)(1 - e(Z_i))} + o_p(1),
$$

where $A = E\Big[\frac{\psi(Z_i) \psi^T(Z_i)}{e(Z_i)(1 - e(Z_i))} \Big]$.

Proof. Since $\widehat{\xi}$ is the solution of (0.1), using a Taylor series expansion at ξ_0, we have

$$
\frac{1}{n} \sum_{i=1}^{n} \Big[\frac{D_i}{e(Z_i)} - \frac{1 - D_i}{1 - e(Z_i)} \Big] \psi(Z_i) - E\Big[\frac{\psi(Z_i) \psi^T(Z_i)}{e(Z_i)(1 - e(Z_i))} \Big] (\widehat{\xi} - \xi_0) + o_p(n^{-\frac{1}{2}}) = 0
$$

Then the results can be obtained immediately.

LEMMA 2: *When $S_1(t) > 0$, we have*

$$\frac{\widehat{S}_1(t)}{S_1(t)} = 1 - \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} \Big\{ \frac{d\overline{N}_1^{EL}(s)}{\overline{Y}_1^{EL}(s)} - d\Lambda_1(s) \Big\}, \tag{0.2}$$

where $\Lambda_1(\cdot)$ is the cumulative hazard function of T_1.

Lemma 2 can be easily proved by the formulas for integration by parts and the differential of a reciprocal, similar to the proof of Theorem 3.2.3 in Flemming and Harrington (1991).

Proof of Theorem 1

Proof. Based on (0.2) in Lemma 2, we can obtain

$$\begin{aligned}
\frac{S_1(t) - \widehat{S}_1(t)}{S_1(t)} &= \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} \frac{I(\overline{Y}_1^{EL}(s) > 0)}{\overline{Y}_1^{EL}(s)} dM_1^{EL}(t) + \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} I(\overline{Y}_1^{EL}(s) = 0) d\Lambda_1(s) \\
&= \sum_{i=1}^n \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} \frac{I(\overline{Y}_1^{EL}(s) > 0)}{\overline{Y}_1^{EL}(s)} p_i D_i dM_{1i}(t) + \mathcal{B}_1(t),
\end{aligned}$$

where $\mathcal{B}_1(t) = \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} I(\overline{Y}_1^{EL}(s) = 0) d\Lambda_1(s)$.

Let $\tau = \inf\{s : \overline{Y}_1^{EL}(s) = 0\}$ and $\mathcal{B}(t) = \mathcal{B}_1(t) S_1(t)$. So $\forall\, t > \tau$, $\overline{Y}_1^{EL}(t) = \sum_{i=1}^n p_i D_i I(Y_{1i} \geqslant t) = 0$.

And we can obtain

$$\begin{aligned}
\mathcal{B}(t) &= S_1(t) \int_\tau^t \frac{\widehat{S}_1(s-)}{S_1(s)} d\Lambda_1(s) I(\tau < t) \\
&= S_1(t) \widehat{S}_1(\tau) \frac{S_1(\tau) - S_1(t)}{S_1(\tau) S_1(t)} I(\tau < t) \\
&= \widehat{S}_1(\tau) \Big(1 - \frac{S_1(t))}{S_1(\tau)}\Big) I(\tau < t),
\end{aligned}$$

therefore, as $n \to \infty$,

$$\begin{aligned}
E[\mathcal{B}(t)] &= E\Big\{\widehat{S}_1(\tau) \Big(1 - \frac{S_1(t))}{S_1(\tau)}\Big) I(\tau < t)\Big\} \\
&\leqslant E\Big\{I(\tau \leqslant t)(1 - S(t))\Big\} = (1 - S(t))(1 - \pi_1(t))^{n_1} \longrightarrow 0,
\end{aligned}$$

which implies that as $n \to \infty$, $\mathcal{B}(t) \stackrel{P}{\longrightarrow} 0$.

The derivatives of the consistency theorem can be divided into two steps. First, we show that for any fixed $u \in (0, \infty)$ such that as $n_1 \to \infty$, $\overline{Y}_1(u) \stackrel{P}{\longrightarrow} \infty$, then $\sup_{0 \leqslant s \leqslant u} |\widehat{S}_1(s) -$

$S_1(s)| \xrightarrow{P} 0, \quad as \quad n \to \infty$. Second, we show that if $t \in (0, \infty]$ is such that for any $u < t$,$\overline{Y}_1(u) \xrightarrow{P} \infty$, as $n \to \infty$. Then $\sup_{0 \leqslant s \leqslant t} |\widehat{S}_1(s) - S_1(s)| \xrightarrow{P} 0, \quad as \quad n \to \infty.$

Under the conditions in this theorem, for any $t \leqslant u$, $S_1(t) > 0$. Followed by Lemma 2, we have as $n \to \infty$,

$$P\Big\{\frac{S_1(t) - \widehat{S}_1(t)}{S_1(t)} = U(t) \quad \text{on} \quad [0, u]\Big\},$$

where

$$U(t) = \sum_{i=1}^{n} \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} \frac{I(\overline{Y}_1^{EL}(s) > 0)}{\overline{Y}_1^{EL}(s)} p_i D_i dM_{1i}(t). \tag{0.3}$$

Note that $\xi_0 \psi(Z_i) = e(Z_i)$, based on (2.3), we have

$$\begin{aligned} q_i = p_i D_i &= \frac{1}{n} \frac{D_i}{\widehat{\xi} \psi(Z_i)} \\ &= \frac{1}{n} \frac{D_i}{e(Z_i)} - \frac{1}{n} \frac{D_i \psi^T(Z_i)}{e^2(Z_i)} (\widehat{\xi} - \xi_0) + o_p(n^{-\frac{1}{2}}), \end{aligned} \tag{0.4}$$

hence (0.3) becomes

$$\begin{aligned} U(t) &= \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} \frac{I(\overline{Y}_1^{EL}(s) > 0)}{\overline{Y}_1^{EL}(s)} \frac{D_i}{e(Z_i)} dM_{1i}(s) - \\ & \quad \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} \frac{I(\overline{Y}_1^{EL}(s) > 0)}{\overline{Y}_1^{EL}(s)} \frac{D_i \psi^T(Z_i)}{e^2(Z_i)} dM_{1i}(s)(\widehat{\xi} - \xi_0) + o_p(n^{-\frac{1}{2}}) \\ &=: U_1(t) - U_2(t)(\widehat{\xi} - \xi_0) + o_p(n^{-\frac{1}{2}}), \end{aligned}$$

where "=:" means "defined as" throughout our paper.

Combining equation (0.4) and Lemma 1, as $n \to \infty$, we can obtain

$$\frac{\widehat{S}_1(s-)}{S_1(s)} \frac{I(\overline{Y}_1^{EL}(s) > 0)}{n\overline{Y}_1^{EL}(s)} \frac{D_i}{e(Z_i)} = \frac{1}{n} \quad \frac{D_i}{e(Z_i)} \frac{\widehat{S}_1^{(1)}(s-)}{S_1(s)} \frac{I(\overline{Y}_1(s) > 0)}{\pi_1(s)},$$

where $\widehat{S}_1^{(1)}(s-) = \prod_{u<s} \Big\{1 - \frac{\sum \frac{D_i}{e(Z_i)} \Delta N_{1i}(u)}{\sum \frac{D_i}{e(Z_i)} Y_{1i}(u)}\Big\}$.

Let $U_1^*(t) = \frac{1}{n_1} \sum_{i=1}^{n_1} \int_0^t \frac{\widehat{S}_1(s-)}{S_1(s)} \frac{I(\overline{Y}_1(s)>0)}{\pi_1(s)} dM_{1i}(s)$. Followed the assumption about D and (Y_1, C), we have $\sup_{0 \leqslant s \leqslant t} |U_1(t) - U_1^*(t)| \xrightarrow{P} 0$.

The process $M_{1i}(t)$ is a martingale and the process $\frac{\widehat{S}_1^{(1)}(s-)}{S_1(s)} \frac{I(\overline{Y}_1(s)>0)}{\pi_1(s)}$ is a predictable and bounded process, by Corollary 3.4.1 in Flemming and Harrington (1991), $\forall \epsilon, \eta > 0$, for any

$0 \leqslant t < u$, we can obtain

$$\begin{aligned}
P\Big\{ \sup_{0\leqslant s\leqslant t} |U_1^*(s)|^2 \geqslant \epsilon \Big\} &\leqslant \frac{\eta}{\epsilon} + P\Big\{ \int_0^t \frac{\widehat{S}_1^{(1)}(s-)^2}{S_1^2(s)} \frac{\sum_{i=1}^{n_1} I(Y_{1i} \geqslant s)}{n_1^2 \pi_1^2(s)} I(\overline{Y}_1(s) > 0) d\Lambda_1(s) \geqslant \eta \Big\} \\
&\leqslant \frac{\eta}{\epsilon} + P\Big\{ \frac{\Lambda_1(t)}{n S_1^2(t)\pi_1(t)} \geqslant \eta \Big\}.
\end{aligned}$$

Therefore, as $n \to \infty$, $\sup_{0\leqslant s<u} |U_1(s)| \xrightarrow{P} 0$. Under the conditions in this theorem, as $n \to \infty$, $U_2(s) = O_p(1)$. On the other side, we have showed that $||\widehat{\xi} - \xi_0|| = O_p(n^{-\frac{1}{2}})$, therefore as $n \to \infty$, $\sup_{0\leqslant s<u} |U_2(s)(\widehat{\xi} - \xi_0)| \xrightarrow{P} 0$. The first step of consistency theory is finished.

Next, we are going to show that if $u = \sup\{t : \pi_1(t) > 0\}$. Then $\sup_{0\leqslant s\leqslant u} |\widehat{S}_1(s) - S_1(s)| \xrightarrow{P} 0, \quad as \quad n \to \infty$.

So it implies to find a t_0 such that $as \quad n \to \infty$, $\sup_{t_0\leqslant s\leqslant u} |\widehat{S}_1(s) - S_1(s)| \xrightarrow{P} 0$.

For $t_0 \leqslant s \leqslant u$,

$$\begin{aligned}
S_1(u) \leqslant & \ S_1(s) \leqslant \ S(t_0) \\
\widehat{S}_1(u) \leqslant & \ \widehat{S}_1(s) \leqslant \ \widehat{S}_1(t_0),
\end{aligned}$$

hence, after simple algorithms, we have

$$\sup_{t_0\leqslant s\leqslant u} \Big|\widehat{S}_1(s) - S_1(s)\Big| \leqslant \Big|\widehat{S}_1(t_0) - S_1(t_0)\Big| + 2\Big|S_1(t_0) - S_1(u)\Big| + \Big|\widehat{S}_1(u) - S_1(u)\Big|.$$

We have shown that $\Big|\widehat{S}_1(t_0) - S_1(t_0)\Big| \xrightarrow{P} 0, \quad as \quad n \to \infty$ in the first step, and the second term in last equation tends to 0 by the continuity of $S_1(\cdot)$ if we choose some t_0 close to u. Therefore, it suffices to show that as $n \to \infty$,

$$\Big|\widehat{S}_1(u) - S_1(u)\Big| \xrightarrow{P} 0.$$

We discuss the two cases: $S_1(u) = 0$ and $S_1(u) > 0$, separately.

When $S_1(u) = 0$, there exists a t_0 such that $\Big|S_1(t_0) - S_1(u)\Big| \xrightarrow{P} 0$, which implies

$\Big|S_1(t_0)\Big| \xrightarrow{P} 0$, note that $u > t_0$, then

$$\begin{aligned} \Big|\widehat{S}_1(u) - S_1(u)\Big| &\leqslant \Big|\widehat{S}_1(u)\Big| + \Big|S_1(u)\Big| \\ &\leqslant \Big|\widehat{S}_1(t_0)\Big| + \Big|S_1(t_0)\Big| \\ &\leqslant \Big|\widehat{S}_1(t_0) - S_1(t_0)\Big| + 2\Big|S_1(t_0)\Big|. \end{aligned}$$

Together with the results in the first step, the first term in last equation converges in probability to 0 as $n \to \infty$.

When $S_1(u) > 0$, since

$$\Big|\widehat{S}_1(u) - S_1(u)\Big| \leqslant \Big|\widehat{S}_1(u) - \widehat{S}_1(t_0)\Big| + \Big|\widehat{S}_1(t_0) - S_1(t_0)\Big| + \Big|S_1(t_0) - S_1(u)\Big|,$$

so it suffices to show that as $n \to \infty$, $\Big|\widehat{S}_1(u) - \widehat{S}_1(t_0)\Big| \xrightarrow{P} 0$.

For any $\epsilon > 0$,

$$\begin{aligned} P\Big\{\Big|\widehat{S}_1(t) - \widehat{S}_1(t_0)\Big| > \epsilon\Big\} &= P\Big\{\Big|\int_{t_0}^{t} I\{\overline{Y}^{EL}(t) > 0\}\widehat{S}_1(s-)\frac{d\overline{N}_1^{EL}(s)}{\overline{Y}_1^{EL}(s)}\Big| > \epsilon\Big\} \\ &\leqslant P\Big\{\Big|\int_{t_0}^{t} I\{\overline{Y}(t) > 0\}\Big\{\frac{d\overline{N}_1^{EL}(s)}{\overline{Y}_1^{EL}(s)} - d\Lambda_1(s)\Big\}\Big| > \epsilon/2\Big\} + \\ &\quad P\Big\{\Big|\int_{t_0}^{t} d\Lambda_1(s)\Big| > \epsilon/2\Big\}. \end{aligned} \tag{0.5}$$

By the continuity of $S_1(t)$, we can find a t_0 such that

$$\Big|\int_{t_0}^{t} d\Lambda_1(s)\Big| = \Big| - \int_{t_0}^{t} \frac{dS_1(t)}{S_1(t)}\Big| \leqslant \Big|S_1(t) - S_1(t_0)\Big| < \epsilon/4.$$

Similar to the proof of uniformly consistency of the process $U_1(t)$ in the first step, we can easily show that as $n \to \infty$

$$P\Big\{\Big|\int_{t_0}^{t} I\{\overline{Y}(t) > 0\}\Big\{\frac{d\overline{N}_1^{EL}(s)}{\overline{Y}_1^{EL}(s)} - d\Lambda_1(s)\Big\}\Big| > \epsilon/2\Big\} < \epsilon/2.$$

Here we omit the detailed proof. Hence as $n \to \infty$, $\Big|\widehat{S}_1(t) - \widehat{S}_1(t_0)\Big| \xrightarrow{P} 0$ when $S_1(t) > 0$.

We have finished the proof of the consistency theorem.

Proof of Theorem 2

Proof. By the derivations of the consistency theorem and the consistency of $\widehat{S}_1(t)$, we can obtain:

$$\begin{aligned}\frac{S_1(t)-\widehat{S}_1(t)}{S_1(t)} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{1}{\pi_1(s)}\frac{D_i}{e(Z_i)}dM_{1i}(s)\\ &- B_1\cdot(\widehat{\xi}-\xi_0)+o_p(n^{-\frac{1}{2}}),\end{aligned}$$

where $B_1 = E\Big[\int_0^t \frac{1}{\pi_1(s)}\frac{D_i\psi^T(Z_i)}{e^2(Z_i)}dM_{1i}(s)\Big]$.

Together with Lemma 1, we have

$$\begin{aligned}\frac{S_1(t)-\widehat{S}_1(t)}{S_1(t)} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{1}{\pi_1(s)}\frac{D_i}{e(Z_i)}dM_{1i}(s)\\ &- B_1A^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{(D_i-e(Z_i))\psi(Z_i)}{e(Z_i)(1-e(Z_i))}+o_p(n^{-\frac{1}{2}}).\end{aligned} \tag{0.6}$$

By the central limit theory of independent identically random variables, we can obtain for any $t\in\mathcal{I}_1$, $\sqrt{n}S_1(t)^{-1}(\widehat{S}_1(t)-S_1(t))$ converges to a normal distribution $N(0,V_1)$,where

$$V_1 = E\Big[\frac{D}{e(Z)}\int_0^t\frac{1}{\pi_1(s)}dM_1(s)\Big]^2 - B_1A^{-1}B_1^T.$$

Note that the two terms in equation (0.6) are correlated. We have finished the proof of this theorem.

Proof of Theorem 4

Proof. Followed by the proof of Theorem 2, for a fix time point t, we have

$$\begin{aligned}\widehat{\Delta}(t)-\Delta(t) &= \widehat{S}_1(t)-\widehat{S}_0(t)-(S_1(t)-S_0(t))\\ &= S_1(t)\frac{1}{n}\sum_{i=1}^{n}\frac{D_i}{e(Z_i)}\int_0^t\frac{1}{\pi_1(s)}dM_{1i}(s) - S_1(t)B_1(\widehat{\xi}-\xi_0) -\\ &\quad S_0(t)\frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{1-e(Z_i)}\int_0^t\frac{1}{\pi_0(s)}dM_{0i}(s)+S_0(t)B_0(\widehat{\xi}-\xi_0)+o_p(n^{-\frac{1}{2}}).\end{aligned}$$

Then the large sample properties can be easily obtained.

Additional Simulation Results

The simulation setups are the same as simulation study 2 except that the sample size is set to $n = 100$ and $n = 300$.

Results of simulation study 2 with $n = 100$

method	para	true	est.hat	bias	se	sd	RMSE
KM	$S_1(t)$	0.504	0.572	0.135	0.078	0.074	0.156
KM	$S_0(t)$	0.265	0.221	0.165	0.065	0.062	0.177
KM	$\Delta(t)$	0.239	0.351	0.467	0.103	0.097	0.478
IPW incorrect	$S_1(t)$	0.504	0.529	0.05	0.076	0.075	0.091
IPW incorrect	$S_0(t)$	0.265	0.254	0.041	0.073	0.07	0.084
IPW incorrect	$\Delta(t)$	0.239	0.275	0.149	0.103	0.103	0.181
IPW correct	$S_1(t)$	0.504	0.514	0.02	0.073	0.074	0.075
IPW correct	$S_0(t)$	0.265	0.265	0.002	0.074	0.072	0.074
IPW correct	$\Delta(t)$	0.239	0.248	0.039	0.099	0.104	0.107
EL	$S_1(t)$	0.504	0.508	0.008	0.071	0.068	0.071
EL	$S_0(t)$	0.265	0.272	0.027	0.076	0.067	0.081
EL	$\Delta(t)$	0.239	0.236	0.012	0.097	0.091	0.098

Results of simulation study 2 with $n = 300$

method	para	true	est.hat	bias	se	sd	RMSE	
KM	$S_1(t)$	0.504	0.569	0.129	0.044	0.044	0.136	67.5
KM	$S_0(t)$	0.265	0.224	0.155	0.039	0.037	0.16	76.7
KM	$\Delta(t)$	0.239	0.345	0.443	0.059	0.057	0.447	53.7
IPW incorrect	$S_1(t)$	0.504	0.524	0.04	0.043	0.044	0.059	92.2
IPW incorrect	$S_0(t)$	0.265	0.256	0.033	0.043	0.042	0.054	92.8
IPW incorrect	$\Delta(t)$	0.239	0.268	0.121	0.058	0.061	0.134	92.5
IPW correct	$S_1(t)$	0.504	0.509	0.01	0.042	0.043	0.043	95.1
IPW correct	$S_0(t)$	0.265	0.267	0.007	0.043	0.043	0.044	94.1
IPW correct	$\Delta(t)$	0.239	0.242	0.013	0.056	0.061	0.057	96.6
EL	$S_1(t)$	0.504	0.506	0.004	0.041	0.041	0.041	94.5
EL	$S_0(t)$	0.265	0.269	0.016	0.043	0.041	0.046	93.1
EL	$\Delta(t)$	0.239	0.237	0.01	0.054	0.055	0.055	95.2

References

Flemming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis.* New York: Wiley.