Biostatistics (2010), **1**, 1, *pp.* 1–9
doi:10.1093/biostatistics/bio000

Supplementary Materials for "Comparing costs associated with Risk Stratification Rules for t-year Survival"

Tianxi Cai
Department of Biostatistics, Harvard University, Boston, MA, USA
tcai@hsph.harvard.edu

LU TIAN
Department of Health Research & Policy, Stanford University, Stanford, CA, USA
lutian@stanford.edu

DONALD M. LLOYD-JONES
Department of Preventive Medicine, Northwestern University, Chicago, IL, USA

APPENDIX

A. OPTIMAL RISK STRATIFICATION RULE

To show the optimality of $R^{opt}(z)$, we note that an optimal rule assigns subjects with $Z = z$ to the kth risk category if and only if for any $l \neq k$,

$$\mathfrak{C}_z(k) = E\{Yc_{1k} + (1 - Y)c_{0k} \mid Z = z\} \leqslant E\{Yc_{1l} + (1 - Y)c_{0l} \mid Z = z\} = \mathfrak{C}_z(l).$$

This implies that for any $l \neq k$,

$$\mu_0(z)c_{1k} + \{1 - \mu_0(z)\}c_{0k} \leqslant \mu_0(z)c_{1l} + \{1 - \mu_0(z)\}c_{0l}$$

and thus $\mu_0(z)\{(c_{1k} - c_{0k}) - (c_{1l} - c_l)\} \leqslant c_l - c_{0k}$. Coupled with the fact that $c_{1k} - c_{0k} > c_{1l} - c_l, l > k$ and $c_{1k} - c_{0k} < C_{l'}^{(1)} - c_l, l' < k$, we have $P_{kl'} \leqslant \mu_0(z) \leqslant P_{kl}$, for any pair of (l', l) such that $l' < k < l$. It follows that the optimal rule would assign a subject with $Z = z$ to the kth category if and only if $\max_{0 \leqslant l \leqslant k-1} P_{kl} \leqslant \mu_0(z) \leqslant \min_{k+1 \leqslant l \leqslant K+1} P_{kl}$. If $\max_{0 \leqslant l \leqslant k-1} P_{kl} \leqslant \mu_0(z) \leqslant \min_{k+1 \leqslant l \leqslant K+1} P_{kl}$ and $\max_{0 \leqslant l \leqslant k'-1} P_{k'l} \leqslant \mu_0(z) \leqslant \min_{k'+1 \leqslant l \leqslant K+1} P_{k'l}$, then the expected costs associated with assigning subjects with $\mu_0(z)$ to the kth category and to k'th category are equal. For such settings, without loss of generality, one may assign such subjects to category $\min(k, k')$. This concludes that $R^{opt}(z)$ minimizes $\mathfrak{C}_z(R)$.

B. ASYMPTOTIC DERIVATION FOR THE ESTIMATED EXPECTED COST

We assume that the true conditional risk function $\mu_0(z)$ is continuously differentiable with derivative function $d\mu_0(z)/dz$ bounded away from 0 almost everywhere. Throughout, we also assume that the bandwidth $h = O(n^{-\nu})$ with $1/5 \leqslant \nu < 1/2$. Let $\widetilde{\mathfrak{C}}(m,H) = n^{-1}\sum_{i=1}^{n}\mathfrak{C}_i(m,H)$, $\mathfrak{C}(m,H) = E\{\widetilde{\mathfrak{C}}(m,H)\}$, where $V_i(t) = I(X_i \leqslant t)\delta_i + I(X_i > t)$ and $\mathfrak{C}_i(m,H) = \eta\{Y_i, m(Z_i)\}V_i(t)/H(t \wedge X_i)$. Then $\mathfrak{C}(m) = \mathfrak{C}(m,G)$ and $\widetilde{\mathfrak{C}}(m) = \widetilde{\mathfrak{C}}(m,\widehat{G})$. We first write

$$\left|n^{-1}\sum_{i=1}^{n}\mathfrak{C}_i(\widetilde{\mu},\widehat{G}) - \mathfrak{C}(\mu_0)\right| \leqslant \left|n^{-1}\sum_{i=1}^{n}\left\{\mathfrak{C}_i(\widetilde{\mu},\widehat{G}) - \mathfrak{C}_i(\mu_0,G)\right\}\right| + \left|n^{-1}\sum_{i=1}^{n}\mathfrak{C}_i(\mu_0,G) - \mathfrak{C}(\mu_0)\right|.$$

By a law of large numbers, $n^{-1}\sum_{i=1}^{n}\mathfrak{C}_i(\mu_0,G) - \mathfrak{C}(\mu_0) = o_p(1)$. Thus, a sufficient condition for the consistency of $\widetilde{\mathfrak{C}}(\widetilde{\mu})$ is that $n^{-1}\sum_{i=1}^{n}\{\mathfrak{C}_i(\widetilde{\mu},\widehat{G}) - \mathfrak{C}_i(\mu_0,G)\} = o_p(1)$. To show that this condition holds, we note that since $\eta\{Y_i, m(Z_i)\}$ is bounded by a constant η_0, $|n^{-1}\sum_{i=1}^{n}\{\mathfrak{C}_i(\widetilde{\mu},\widehat{G}) - \mathfrak{C}_i(\widetilde{\mu},G)\}| \leqslant n^{-1}\sum_{i=1}^{n}\eta_0|\widehat{G}(t\wedge X_i)^{-1} - G(t\wedge X_i)^{-1}|$. This and the uniform consistency of $\widehat{G}(\cdot)$ (Kalbfleisch and Prentice, 2002) imply that $n^{-1}\sum_{i=1}^{n}\{\mathfrak{C}_i(\widetilde{\mu},\widehat{G}) - \mathfrak{C}_i(\widetilde{\mu},G)\} = o_p(1)$. It remains to show that $n^{-1}\sum_{i=1}^{n}\{\mathfrak{C}_i(\widetilde{\mu},G) - \mathfrak{C}_i(\mu_0,G)\} = o_p(1)$. This convergence holds if for any given non-negative bounded function $\xi(\cdot,\cdot)$,

$$\widetilde{\epsilon}_{\xi k} = n^{-1}\sum_{i=1}^{n}\{I(\widetilde{\mu}(Z_i) > p_k) - I(\mu_0(Z_i) > p_k)\}\xi(X_i,D_i) = o_p(1), \qquad \text{(B.1)}$$

To show (B.1), we let $A_\xi(u) = E[I\{\mu_0(Z_i) > u\}\xi(X_i,D_i)]$, $\widetilde{\varepsilon}_\mu = \sup_z|\widetilde{\mu}(z) - \mu_0(z)|$, $\widetilde{\epsilon}_A = \sup_u|n^{-1}\sum_{i=1}^{n}I\{\mu_0(Z_i) > u\}\xi(X_i,D_i) - A_\xi(u)|$, and note that $|\widetilde{\epsilon}_{\xi k}| \leqslant n^{-1}\sum_{i=1}^{n}I\{p_k + \widetilde{\varepsilon}_\mu \geqslant \mu_0(Z_i) > p_k - \widetilde{\varepsilon}_\mu\}\xi(X_i,D_i) \leqslant 2\widetilde{\epsilon}_A + |A_\xi(p_k+\widetilde{\varepsilon}_\mu) - A_\xi(p_k - \widetilde{\varepsilon}_\mu)|$. By a uniform law of large numbers (Pollard, 1990), $\widetilde{\epsilon}_A = o_p(1)$. On the other hand, using the same arguments as given in Cai et al (2008), we have $\widetilde{\varepsilon}_\mu = o_p(n^{-1/4})$. This, together with the continuity of $A_\xi(u)$, implies that $\widetilde{\epsilon}_{\xi k} = o_p(1)$ and hence the consistency of $\widetilde{\mathfrak{C}}(\widetilde{\mu})$.

To derive the asymptotic distribution for $\widetilde{\mathbb{W}} = n^{\frac{1}{2}}\{\widetilde{\mathfrak{C}}(\widetilde{\mu}) - \mathfrak{C}(\mu_0)\} = n^{\frac{1}{2}}\{\widetilde{\mathfrak{C}}(\widetilde{\mu}, \widehat{G}) - \mathfrak{C}(\mu_0, G)\}$, we write $\widetilde{\mathbb{W}} = \widetilde{\mathbb{W}}_1 + \widetilde{\mathbb{W}}_2 + \widetilde{\mathbb{W}}_3$, where $\widetilde{\mathbb{W}}_1 = n^{\frac{1}{2}}\{\widetilde{\mathfrak{C}}(\widetilde{\mu}, \widehat{G}) - \widetilde{\mathfrak{C}}(\widetilde{\mu}, G)\}$, $\widetilde{\mathbb{W}}_2 = n^{\frac{1}{2}}\{\widetilde{\mathfrak{C}}(\widetilde{\mu}, G) - \widetilde{\mathfrak{C}}(\mu_0, G)\}$, $\widetilde{\mathbb{W}}_3 = n^{\frac{1}{2}}\{\widetilde{\mathfrak{C}}(\mu_0, G) - \mathfrak{C}(\mu_0, G)\}$. For $\widetilde{\mathbb{W}}_1$, we first note that for $s \leqslant t$,

$$n^{\frac{1}{2}}\left\{\frac{G(s)}{\widehat{G}(s)} - 1\right\} \simeq n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^s \frac{dM_i(u)}{\text{pr}(X \geqslant u)} \tag{B.2}$$

converges weakly to a zero-mean Gaussian process indexed by s (Kalbfleisch and Prentice, 2002), where $M_i(u) = I(X_i \leqslant u, \delta_i = 0) - \int_0^u I(X_i \geqslant u)d\Lambda_C(u)$ and $\Lambda_C(\cdot)$ is the cumulative hazard function for the common censoring variable C. It follows that $\widetilde{\mathbb{W}}_1$ is asymptotically equivalent to

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\int_0^t \gamma(u)dM_i(u)\right\}, \tag{B.3}$$

where $\gamma(u) = E[\eta\{Y, \mu_0(Z)\}I(T \geqslant u)]/\text{pr}(X \geqslant u)$.

For $\widetilde{\mathbb{W}}_2$, we write

$$\begin{aligned}\widetilde{\mathbb{W}}_2 &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K-1} d_{0k}w_i\{1 - p_k^{-1}Y_i\}\left[I\{\widetilde{\mu}(Z_i) > p_k\} - I\{\mu_0(Z_i) > p_k\}\right] \\ &= \sum_{k=1}^{K-1}\int \left[I\{\widetilde{\mu}(z) > p_k\} - I\{\mu_0(z) > p_k\}\right] dn^{1/2}\widehat{H}_k(z)\end{aligned}$$

where $w_i = V_i(t)/G(t \wedge X_i)$ and $\widehat{H}_k(z) = n^{-1}\sum_{i=1}^n d_{0k}w_i\{1 - p_k^{-1}Y_i\}I(Z_i \leqslant z)$. By a standard empirical process theory (Pollard, 1990) $n^{\frac{1}{2}}\{\widehat{H}_k(z) - \widetilde{H}_k(z)\}$ converges weakly to a zero-mean Gaussian process, where $\widetilde{H}_k(z) = n^{-1}\sum_{i=1}^n d_{0k}\{1 - p_k^{-1}\mu_0(Z_i)\}I(Z_i \leqslant z)$. This, together with Lemma 1 of Bilias *and others* (1997), implies that $\widetilde{\mathbb{W}}_2$ is asymptotically equivalent to $\widetilde{\varepsilon}_1 + \widetilde{\varepsilon}_2$, where

$$\widetilde{\varepsilon}_1 = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K-1} d_{0k}\{1 - p_k^{-1}\mu_0(Z_i)\} I\{\widetilde{\mu}(Z_i) > p_k, \mu_0(Z_i) \leqslant p_k\}$$
$$\widetilde{\varepsilon}_2 = -n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K-1} d_{0k}\{1 - p_k^{-1}\mu_0(Z_i)\} I\{\widetilde{\mu}(Z_i) \leqslant p_k, \mu_0(Z_i) > p_k\}$$

It follows from $\widetilde{\varepsilon}_\mu = \sup_z |\widetilde{\mu}(z) - \mu_0(z)| = o_p(n^{-1/4})$ that

$$0 \leqslant \widetilde{\varepsilon}_1 \leqslant n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K-1} d_{0k}\{1 - p_k^{-1}\mu_0(Z_i)\} \left[I\{\mu_0(Z_i) > p_k - \widetilde{\varepsilon}_\mu\} - I\{\mu_0(Z_i) > p_k\} \right].$$

Furthermore, the process $\Gamma(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{k=1}^{K-1} [d_{0k}\{1 - p_k^{-1}\mu_0(Z_i)\}[I\{\mu_0(Z_i) > p_k - t\} - I\{\mu_0(Z_i) > p_k\}] - \{\zeta(\mathbf{p}+t) - \zeta(\mathbf{p})\}]$ is stochastic continuous at $t = 0$, where

$$\zeta(\bar{\mathbf{p}}) = E\left[\sum_{k=1}^{K-1} d_{0k}\{1 - p_k^{-1}\mu_0(Z)\} I\{\mu_0(Z) > \bar{p}_k\}\right], \quad \bar{\mathbf{p}} = (\bar{p}_1, \cdots, \bar{p}_K)',$$

and $\mathbf{p} = (p_1, \cdots, p_K)'$. Thus, $\hat{\epsilon}_1$ is bounded above by $\sup_{|t| \leqslant \hat{\epsilon}_m} n^{1/2}\{\zeta(\mathbf{p}+t) - \zeta(\mathbf{p})\}$. Now, since the expected cost function $\zeta(\bar{\mathbf{p}})$ is minimized at $\bar{\mathbf{p}} = \mathbf{p}$, $\partial\zeta(\bar{\mathbf{p}})/\partial\bar{\mathbf{p}} = 0$ when $\bar{\mathbf{p}} = \mathbf{p}$. Therefore, $0 \leqslant \widetilde{\varepsilon}_1 \leqslant O_p(n^{1/2}\widetilde{\varepsilon}_\mu^2)$ and thus $\widetilde{\varepsilon}_1 = o_p(1)$. Similarly, $\widetilde{\varepsilon}_2 = o_p(1)$. It follows that $\widetilde{\widetilde{\mathbb{W}}}_2 = o_p(1)$. This, combined with (B.3), implies that $\widetilde{\widetilde{\mathbb{W}}} \simeq n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{W}_i$, where $\psi_{\mathbb{W}i} = \int_0^t \gamma(u) dM_i(u) + \mathfrak{C}_i(\mu_0, G) - \mathfrak{C}(\mu_0, G)$. By a Central Limit Theorem, $\widetilde{\widetilde{\mathbb{W}}}$ converges to a normal with mean 0 and variance $\sigma^2 = E(\psi_{\mathbb{W}i}^2)$.

C. Asymptotic Properties of $\widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}})$

In this section, we show that $\widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}})$ converges to $\mathbb{C}(\mu_{\boldsymbol{\beta}_0})$ in probability as $n \to \infty$ and derive the asymptotic distribution for $\widehat{\mathbb{W}}$. We assume that $\mathbf{Z}$ is bounded and $\mu_{\boldsymbol{\beta}}(x)$ is continuously differentiable with respect to both $\boldsymbol{\beta}$ and x and $\partial\mu_{\boldsymbol{\beta}_0}(x)/\partial x$ is bounded away from 0 almost everywhere. Furthermore, we assume that $\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z}$ is a continuous random variable with a continuously differentiable density. The consistency of $\widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}})$ follows from the same arguments as given in Appendix B provided that $\widehat{\varepsilon}_m = \sup_{\mathbf{Z}} |\widehat{\mu}_{\widehat{\boldsymbol{\beta}}}(\mathbf{Z}) - \mu_{\boldsymbol{\beta}_0}(\mathbf{Z})| = o_p(n^{-1/4})$. Thus, it remains to establish the convergence rate of $\widehat{\varepsilon}_m$. To this end, we note that

$$\widehat{\varepsilon}_m \leqslant \sup_x |g_0\{\widehat{\theta}_{\widehat{\beta}}(x)\} - g_0\{\theta_{\boldsymbol{\beta}_0}(x)\}| + \sup_{\mathbf{Z}} |g_0\{\theta_{\boldsymbol{\beta}_0}(\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{Z})\} - g_0\{\theta_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z})\}|.$$

where $\theta_{\boldsymbol{\beta}}(x) = g_0^{-1}\{\text{pr}(Y_i = 1|\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z} = x)\}$. From Uno et al (2007) and Cai et al (2008), we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\mathbf{W}_{Bi} + o_p(1) = O_p(1), \qquad \text{(C.1)}$$

$$\text{and} \quad \sup_x |\widehat{\theta}_{\widehat{\boldsymbol{\beta}}}(x) - \theta_{\boldsymbol{\beta}_0}(x)| = O_p\{(nh)^{-\frac{1}{2}}\log(n)^2\}, \qquad \text{(C.2)}$$

$$\text{where} \quad \mathbf{W}_{Bi} = \mathbb{A}^{-1}\left[w_i\mathbf{Z}_i\{Y_i - g(\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z}_i)\} + \int_0^t \boldsymbol{\gamma}_B(u)dM_i(u)\right],$$

$$\mathbb{A} = E\{\dot{g}(\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z}_i)\mathbf{Z}_i\mathbf{Z}_i^{\mathsf{T}}\}, \quad \boldsymbol{\gamma}_B(u) = \frac{E\left[\mathbf{Z}_i\{Y_i - g(\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z}_i)\}I(T_i \geqslant u)\right]}{\text{pr}(X \geqslant u)},$$

and $\dot{g}(x) = dg(x)/dx$. This, together with the boundedness of $\partial\mu_{\boldsymbol{\beta}_0}(x)/\partial x$ and $\mathbf{Z}$, implies that

$$\widehat{\varepsilon}_m \leqslant O_p\{(nh)^{-\frac{1}{2}}\log(n)^2 + n^{-\frac{1}{2}}\} = o_p(n^{-1/4}),$$

since $h^{-1} = o_p(n^{1/2})$. The consistency of $\widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}})$ follows immediately.

To approximate the distribution of $\widehat{\mathbb{W}} = n^{\frac{1}{2}}\{\widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}}) - \mathbb{C}(\mu_{\boldsymbol{\beta}_0})\}$, we write $\widehat{\mathbb{W}} = \widehat{\mathbb{W}}_1 + \widehat{\mathbb{W}}_2 + \widehat{\mathbb{W}}_3$, where $\widehat{\mathbb{W}}_1 = n^{\frac{1}{2}}\{\widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}}, \widehat{G}) - \widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}}, G)\}$, $\widehat{\mathbb{W}}_2 = n^{\frac{1}{2}}\{\widehat{\mathbb{C}}(\widehat{\mu}_{\widehat{\boldsymbol{\beta}}}, G) - \widehat{\mathbb{C}}(\mu_{\boldsymbol{\beta}_0}, G)\}$,

$\widehat{\mathbb{W}}_3 = n^{\frac{1}{2}}\{\widehat{\mathbb{C}}(\mu_{\boldsymbol{\beta}_0}, G) - \mathbb{C}(\mu_{\boldsymbol{\beta}_0})\}$,

$$\widehat{\mathbb{C}}(\mu, G) = n^{-1}\sum_{i=1}^{n}\mathbb{C}_i(\mu, H), \quad \text{and} \quad \mathbb{C}_i(\mu, H) = \frac{V_i(t)\eta\{Y_i, \mu(\mathbf{Z}_i)\}}{H(t \wedge X_i)}.$$

Following similar arguments as given in Appendix B with the uniform consistency of $\widehat{\mu}_{\widehat{\boldsymbol{\beta}}}(x)$ given in (C.2), we have

$$\widehat{\mathbb{W}}_1 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \gamma_{\mathbb{C}}(u)dM_i(u) + o_p(1)$$

where $\gamma_{\mathbb{C}}(u) = E[\eta\{Y_i, \mu_{\boldsymbol{\beta}_0}(\mathbf{Z}_i)\}I(T \geqslant u)]/\text{pr}(X \geqslant u)$. Next, we express $\widehat{\mathbb{W}}_2$ as

$$\begin{aligned}
&\sum_{k=1}^{K-1}\left(\int I[g_0\{\widehat{\theta}_{\widehat{\boldsymbol{\beta}}}(x)\} > p_{k-1}]d\widehat{H}_k(x; \widehat{\boldsymbol{\beta}}) - \int I[g_0\{\theta_{\boldsymbol{\beta}_0}(x)\} > p_{k-1}]d\widehat{H}_k(x; \boldsymbol{\beta}_0)\right) \\
=&\sum_{k=1}^{K-1}\int\left(I[g_0\{\widehat{\theta}_{\widehat{\boldsymbol{\beta}}}(x)\} > p_{k-1}] - I[g_0\{\theta_{\boldsymbol{\beta}_0}(x)\} > p_{k-1}]\right)dn^{\frac{1}{2}}\widehat{H}_k(x; \widehat{\boldsymbol{\beta}}) \\
&+\sum_{k=1}^{K-1}\int I[g_0\{\theta_{\boldsymbol{\beta}_0}(x)\} > p_{k-1}]dn^{\frac{1}{2}}\left\{\widehat{H}_k(x; \widehat{\boldsymbol{\beta}}) - \widehat{H}_k(x; \boldsymbol{\beta}_0)\right\}
\end{aligned}$$

where $\widehat{H}_k(x; \boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} d_{0k}w_i(1 - Y_i p_k^{-1})I(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \leqslant x)$. It follows from the standard empirical process theory (Pollard, 1990) and the convergence of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ given in (C.1) that $n^{\frac{1}{2}}\{\widehat{H}_k(x; \widehat{\boldsymbol{\beta}}) - \widetilde{H}_k(x; \boldsymbol{\beta}_0)\}$ converges weakly to a zero mean Gaussian process, where $\widetilde{H}_k(x; \boldsymbol{\beta}_0) = n^{-1}\sum_{i=1}^{n} d_{0k}w_i\{1 - \mu_{\boldsymbol{\beta}_0}(\mathbf{Z}_i)p_k^{-1}\}I(\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z}_i \leqslant x)$. This, together with same arguments as given in Appendix B, implies that

$$\sum_{k=1}^{K-1}\int\left(I[g_0\{\widehat{\theta}_{\widehat{\boldsymbol{\beta}}}(x)\} > p_{k-1}] - I[g_0\{\theta_{\boldsymbol{\beta}_0}(x)\} > p_{k-1}]\right)dn^{\frac{1}{2}}\widehat{H}_k(x; \widehat{\boldsymbol{\beta}}) = o_p(1)$$

Furthermore, let $\bar{H}_k(x;\boldsymbol{\beta}) = E\{\widehat{H}_k(x;\boldsymbol{\beta})\}$. It follows from the functional central limit theorem (Pollard, 1990) that $n^{\frac{1}{2}}\{\widehat{H}_k(x;\boldsymbol{\beta}) - \bar{H}_k(x;\boldsymbol{\beta})\}$ converges weakly to a zero mean Gaussian process in $(x, \boldsymbol{\beta})$ and thus is equicontinuous in $\boldsymbol{\beta}$. Combining this with the expansion for $\widehat{\boldsymbol{\beta}}$ given in (C.1), we have

$$n^{\frac{1}{2}}\{\widehat{H}_k(x;\widehat{\boldsymbol{\beta}}) - \widehat{H}_k(x;\boldsymbol{\beta}_0)\} = \mathbf{A}_k(x;\boldsymbol{\beta}_0)^{\mathsf{T}} n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbf{W}_{Bi}^{\mathsf{T}} \mathbf{A}_k(x;\boldsymbol{\beta}_0)$$

$\mathbf{A}_k(x;\boldsymbol{\beta}) = \partial \bar{H}_k(x;\boldsymbol{\beta})/\partial\boldsymbol{\beta}$. It follows that

$$\widehat{\mathbb{W}}_2 = n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbf{W}_{Bi}^{\mathsf{T}} \left(\sum_{k=1}^{K-1} \int I[g_0\{\theta_{\boldsymbol{\beta}_0}(x)\} > p_{k-1}] d\mathbf{A}_k(x;\boldsymbol{\beta}_0) \right) + o_p(1).$$

Therefore, $\widehat{\mathbb{W}} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \zeta_{\mathbb{W}i} + o_p(1)$, where

$$\zeta_{\mathbb{W}i} = \int_0^t \gamma_{\mathbb{C}}(u) dM_i(u) + \mathbf{W}_{Bi}^{\mathsf{T}} \sum_{k=1}^{K-1} \int I[g_0\{\theta_{\boldsymbol{\beta}_0}(x)\} > p_{k-1}] d\mathbf{A}_k(x;\boldsymbol{\beta}_0) + \mathbb{C}_i(\mu_{\boldsymbol{\beta}_0}, G) - \mathbb{C}(\mu_{\boldsymbol{\beta}_0}). \tag{C.3}$$

Then it follows from a central limit theorem that $\widehat{\mathbb{W}}$ converges in distribution to a normal with variance $E(\zeta_{\mathbb{W}i}^2)$.

REFERENCES

BILIAS, YANNIS, GU, MINGGAO AND YING, ZHILIANG. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *The Annals of Statistics* **25**(2), 662–682.

KALBFLEISCH, J. D. AND PRENTICE, ROSS L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.*. Hayward, CA: Institute of Mathematical Statistics.