Web-based Supplementary Material for

"Nonparametric Maximum Likelihood Estimators of Time-Dependent Accuracy Measures for Survival Outcome Under Two-Stage Sampling Designs"

by Dandan Liu, Tianxi Cai, Anna Lok and Yingye Zheng

Appendix A Regularity Conditions

We impose the following regularity conditions:

(a) C has a finite support $\mathcal{B}_C = [0, \tau]$ and $\Pr(C \geq \tau | Z) = \Pr(C = \tau | Z)$ is bounded away from zero.

(b) The function $\Lambda_0(t)$ is twice continuously differentiable in $\mathcal{B}_C$ and $\lambda(t) > 0$ for $t \in \mathcal{B}_C$ and $G_{0k}(c)$ is twice continuously differentiable in $\mathcal{B}_Y$ for $k = 1, \ldots, K$.

(c) We consider estimating the accuracy measures for $t \in [\underline{\tau}, \overline{\tau}] \subset \mathcal{B}_C$ and $c \in [\underline{c}, \overline{c}] \subset \mathcal{B}_Y$, such that $\Pr(T > \overline{\tau}) \Pr(T < \underline{\tau}) \Pr(Y > \overline{c}) \Pr(T < \underline{c}) > 0$;

(d) There exists a function $p(v; x, \delta, z)$ $(v = 0, 1)$ such that $\sum_{v=0}^{1} p(v; x, \delta, z) = 1$, $p(v = 1; x, \delta, z) > c_1$ almost surely for some positive constant c_1, and under the probability measure associated with $\prod_{i=1}^{n} p(V_i; X_i, \delta_i, Z_i)$ and conditional on $\mathcal{O}$,

$$\left\{ \log \frac{\Pr(V_1 = v_1, \ldots, V_n = r_n | \mathcal{O})}{\prod_{i=1}^{n} p(V_i; X_i, \delta_i, Z_i)} + w(\mathcal{O})/2 \right\} / f(\mathbb{O})^{1/2}$$

is asymptotically standard normal, where w is a positive and measurable function.

Conditions (a) and (b) are standard boundness and smoothness conditions for censored data. Condition (c) defines the estimable region of accuracy measures. Condition (d) was imposed for the application of Le Cam's third lemma to establish the asymptotic equivalence of the two-phase sampling mechanism to random sampling such that $\mathcal{O}_i$, $(i = 1, \ldots, n)$ can be treated as i.i.d. observations (Zeng and Lin, 2014).

Appendix B Asymptotic Results of $\widehat{\boldsymbol{\theta}}$

Theorem B.1 below summarized the asymptotic results of $\widehat{\beta}$, $\widehat{\Lambda}$ and $\widehat{G}$ and are similar to those provided in Zeng and Lin (2014).

Theorem B.1. *Under conditions (a)-(d), with probability one,*

$$|\widehat{\beta} - \beta_0| + \sup_{t\in[0,\tau]} |\widehat{\Lambda}(t) - \Lambda_0(t)| + \sum_{k=1}^{K} \sup_{c\in[c_l,c_u]} |\widehat{G}_k(c) - G_{0k}(c)| \to 0$$

almost surely. In addition, $\sqrt{n}(\widehat{\beta} - \beta_0, \widehat{\Lambda} - \Lambda_0, \widehat{G} - G_0)$ *weakly converge to a zero-mean Gaussian process in* $R \times \ell^\infty[0,\tau] \times \ell^\infty[c_l, c_u]^K$*, where* $\ell^\infty[0,\tau]$ *and* $\ell^\infty[c_l, c_u]$ *are normed spaces consisting of all bounded functions and the norm is defined as the supremum norm on* $\mathcal{B}_C$ *and* $\mathcal{B}_y$ *respectively.*

The proof of the consistency of $\widehat{\beta}$, $\widehat{\Lambda}$ and $\widehat{G}$ follows from the proof of Theorem 1 in Zeng and Lin (2014) when Z is discrete and is thus omitted here. The weak convergence could be proved using arguments in the proof of Theorem 2 in Zeng and Lin (2014). The proof consists of four major steps and are briefly summarized below.

- Step 1: Proving the invertibility of the information operator for $\boldsymbol{\theta}_0$.
- Step 2: Deriving the score equation for $\boldsymbol{\theta}_0$.
- Step 3: Obtaining the asymptotic linear expansion of the score function for $\boldsymbol{\theta}_0$.
- Step 4: Proving the weak convergence of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$.

Steps 1-3 were proved in Zeng and Lin (2014). Let $\|f(\cdot)\|_{V[a,b]}$ denote total variation of $f(\cdot)$ in $[a,b]$. Let BV$[a,b]$ denote the space of functions with bounded total variation in $[a,b]$. Consider the set

$$\mathcal{H} = \{(v, q, \mathbf{h}) : \mathbf{h} = (h_1, \ldots, h_K), |v| \le 1, \|q(t)\|_{V[0,\tau]} \le 1, \|h_k(c)\|_{V[c_l,c_u]} \le 1\}$$

where $q(\cdot) \in BV[0,\tau]$ and $h_k \in BV[c_l, c_u]$, $k = 1, \ldots, K$. We identify $(\widehat{\beta} - \beta_0, \widehat{\Lambda} - \Lambda_0, \widehat{G} - G_0)$ as a random element in $\ell^\infty(\mathcal{H})$ through definition $v(\widehat{\beta} - \beta_0) + \int_0^\tau q(t) d(\widehat{\Lambda} - \Lambda_0)(t) + \sum_{k=1}^K \int_{c_l}^{c_u} [h_k(c) - \mathrm{E}_k\{h_k(Y)\}] d(\widehat{G}_k - G_{0k})(c)$ where E_k denote conditional expectation of Y given $Z = z_k$. Let $\ell^F(\boldsymbol{\theta})$ denote the full log-likelihood function of one observation assuming Y is observed, where

$$\ell^F(\boldsymbol{\theta}) = [\delta \log\{\lambda(X)\} + \delta\beta Y - \Lambda(X)\exp(\beta Y) + \log\{dG(Y|Z)\}].$$

Let $\ell(\boldsymbol{\theta}) = \mathrm{E}\{\ell^F(\boldsymbol{\theta})|\mathcal{O}\}$ denote the corresponding observed log-likelihood function, $\dot{\ell}_\beta(\boldsymbol{\theta})$ be the derivative of $\ell(\boldsymbol{\theta})$ with respect to β, $\dot{\ell}_\Lambda(\boldsymbol{\theta})[q]$ be the path-wise derivative along the path $\Lambda + \epsilon \int q d\Lambda$, and $\dot{\ell}_k(\boldsymbol{\theta})[h_k]$ be the derivative along the path $dG_k + \epsilon\{h_k - \mathrm{E}_k(h_k)\}dG_k$ for $k = 1, \ldots, K$. To be specific

$$\begin{aligned}
\dot{\ell}_\beta(\boldsymbol{\theta}) =& \mathrm{E}\left[\delta Y - Y\Lambda(X)\exp(\beta Y)|\mathcal{O}\right], \\
\dot{\ell}_\Lambda(\boldsymbol{\theta})[q] =& \mathrm{E}\left[\delta q(X) - \exp(\beta Y)\int_0^X q(t)d\Lambda(t)|\mathcal{O}\right], \\
\dot{\ell}_k(\boldsymbol{\theta})[h_k] =& \mathrm{E}\left[h_k(Y) - \mathrm{E}_k\{h_k(Y)\}|\mathcal{O}\right],
\end{aligned} \tag{B.1}$$

where the expression of the expectation given the observed data is given in section 3.1 of the paper. Let $\dot{\ell}(\boldsymbol{\theta})[v, q, \mathbf{h}]$ denote the score operator of $\boldsymbol{\theta}$, mapping a neighborhood of $\boldsymbol{\theta}_0$ into $\ell^\infty(\mathcal{H})$. Then $\dot{\ell}(\boldsymbol{\theta})[v, q, \mathbf{h}] = \dot{\ell}_\beta(\boldsymbol{\theta}) + \dot{\ell}_\Lambda(\boldsymbol{\theta})[q] + \sum_{k=1}^K \dot{\ell}_k(\boldsymbol{\theta})[h_k]$

Let $\mathcal{P}_n$ and $\mathcal{P}$ denote empirical measure and true measure, respectively. Since $\widehat{\boldsymbol{\theta}}$ maximizes $\ell(\boldsymbol{\theta})$, $\dot{\ell}_\beta(\widehat{\boldsymbol{\theta}}) = 0$, $\dot{\ell}_\Lambda(\widehat{\boldsymbol{\theta}})[q] = 0$ and $\dot{\ell}_k(\widehat{\boldsymbol{\theta}})[h_k] = 0$, $k = 1, \ldots, K$. Therefore, $\widehat{\boldsymbol{\theta}}$

is the solution of the functional score equation $\mathcal{P}_n\{\dot{\ell}(\boldsymbol{\theta})\} = 0$. Using Theorem 2.11.22 of van der Vaart and Wellner (1996), we could show that

$$\sqrt{n}\mathcal{P}\{\dot{\ell}(\widehat{\boldsymbol{\theta}})\} = -\sqrt{n}(\mathcal{P}_n - \mathcal{P})\{\dot{\ell}(\boldsymbol{\theta}_0)\} + o_p(1). \tag{B.2}$$

Let $\mathcal{B}$ denote the information operator mapping $\boldsymbol{\theta} - \boldsymbol{\theta}_0$ to $\ell^\infty(\mathcal{H})$, where $\boldsymbol{\theta}$ is in the neighborhood of $\boldsymbol{\theta}_0$. On the left-hand side of (B.2), $\mathcal{P}\{\dot{\ell}(\widehat{\boldsymbol{\theta}})\}$ can be linearized around β_0, Λ_0, G_0 and is asymptotically equivalent to

$$\begin{aligned}
&\mathcal{B}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] \\
=&\mathcal{B}_\beta[v, q, \mathbf{h}](\widehat{\beta} - \beta_0) + \mathcal{B}_\Lambda(\widehat{\Lambda} - \Lambda_0)[v, q, \mathbf{h}] + \mathcal{B}_G(\widehat{G} - G_0)[v, q, \mathbf{h}],
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{B}_\beta[v, q, \mathbf{h}] &= v\left\{\ddot{\ell}_{\beta\beta} + \ddot{\ell}_{\Lambda\beta}[q] + \sum_{k=1}^K \ddot{\ell}_{k\beta}[h_k]\right\} \\
\mathcal{B}_\Lambda(\Lambda - \Lambda_0)[v, q, \mathbf{h}] &= \ddot{\ell}_{\beta\Lambda}[\Lambda - \Lambda_0] + \ddot{\ell}_{\Lambda\Lambda}[q, \Lambda - \Lambda_0] + \sum_{k=1}^K \ddot{\ell}_{k\Lambda}[h_k, \Lambda - \Lambda_0] \\
\mathcal{B}_G(G - G_0)[v, q, \mathbf{h}] &= \sum_{k=1}^K \left\{v\ddot{\ell}_{\beta k}[G_k - G_{0k}] + \ddot{\ell}_{\Lambda k}[q, G_k - G_{0k}] + \sum_{l=1}^K \ddot{\ell}_{kl}[h_k, G_l - G_{0l}]\right\},
\end{aligned}$$

$\ddot{\ell}_{\beta\beta}$, $\ddot{\ell}_{\Lambda\beta}[q]$ and $\ddot{\ell}_{k\beta}[h_k]$ are the derivatives of $\dot{\ell}_\beta$, $\dot{\ell}_\Lambda[q]$, and $\dot{\ell}_k[h_k]$ with respect to β evaluated at $\boldsymbol{\theta}_0$; $\ddot{\ell}_{\beta\Lambda}[q^*]$, $\ddot{\ell}_{\Lambda\Lambda}[q, q^*]$ and $\ddot{\ell}_{k\Lambda}[h_k, q^*]$ are the path-wise derivative of $\dot{\ell}_\beta$, $\dot{\ell}_\Lambda[q]$, $\dot{\ell}_k[h_k]$ with respect to Λ along q^* evaluated at $\boldsymbol{\theta}_0$; and $\ddot{\ell}_{\beta k}[h_k]$, $\ddot{\ell}_{\Lambda k}[q, h_k]$ and $\ddot{\ell}_{lk}[h_l, h_k]$ are the path-wise derivative of $\dot{\ell}_\beta$, $\dot{\ell}_\Lambda[q]$, $\dot{\ell}_l[h_l]$ with respect to G_k along h_k evaluated at $\boldsymbol{\theta}_0$. It was proved by Zeng and Lin (2014) that B is invertible. Therefore,

$$\begin{aligned}
&\sqrt{n}\left[v(\widehat{\beta} - \beta_0) + \int_0^\tau q(t)d\{\widehat{\Lambda}(t) - \Lambda_0(t)\} + \sum_{k=1}^K \int_{c_u}^{c_l} h_k(c)d\{\widehat{G}_k(c) - G_{0k}(c)\}\right] \\
=& -\sqrt{n}(\mathcal{P}_n - \mathcal{P})\left\{\dot{\ell} \circ \mathcal{B}^{-1}(\boldsymbol{\theta}_0)[v, q, \mathbf{h}]\right\} + o_p(1)
\end{aligned}$$

The weak convergence of $\widehat{\boldsymbol{\theta}}$ is proved. Therefore, for any constant $t^* \in (0, \tau)$ and $c^* \in (c_l, c_u)$

$$\begin{aligned}
\sqrt{n}(\widehat{\beta} - \beta_0) =& \sqrt{n}\mathcal{P}_n(\mathcal{W}_\beta), & \mathcal{W}_\beta =& \dot{\ell} \circ \mathcal{B}^{-1}(\boldsymbol{\theta}_0)(1, 0, \mathbf{0}); \\
\sqrt{n}\{\widehat{\Lambda}(t^*) - \Lambda_0(t^*)\} =& \sqrt{n}\mathcal{P}_n(\mathcal{W}_\Lambda), & \mathcal{W}_\Lambda =& \dot{\ell} \circ \mathcal{B}^{-1}(\boldsymbol{\theta}_0)(0, q^*, \mathbf{0}); \\
\sqrt{n}\{\widehat{G}_k(c^*) - G_{0k}(c^*)\} =& \sqrt{n}\mathcal{P}_n(\mathcal{W}_{G_k}), & \mathcal{W}_{G_k} =& \dot{\ell} \circ \mathcal{B}^{-1}(\boldsymbol{\theta}_0)(0, 0, \mathbf{h}_k^*);
\end{aligned}$$

where $q^*(t) = I(t < t^*)$ and $\mathbf{h}_k^*(c)$ is a K-vector of functions with the k-th element equal to $I(c < c^*)$ and 0 otherwise.

Appendix C Asymptotic property of one-step estimator $\widehat{\boldsymbol{\theta}}^{(1)}$

C.1 Consistency

In the E-step of the one-step estimation, the observed log-likelihood function given the IPW estimators $\widetilde{\boldsymbol{\theta}}$ is $\ell(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \mathrm{E}\{\ell^F(\boldsymbol{\theta})|\mathcal{O}, \widetilde{\boldsymbol{\theta}}\}$. In the M-step of the one-step estimation, we aim to maximize $\mathcal{P}_n\{\ell(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})\}$ with respect to $\boldsymbol{\theta}$. Using similar notations in Appendix B, let $\dot{\ell}_\beta(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ be the derivative of $\ell(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ with respect to β, $\dot{\ell}_\Lambda(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[q]$ be the path-wise derivative along the path $\Lambda + \epsilon \int q d\Lambda$, and $\dot{\ell}_k(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[h_k]$ be the derivative along the path $dG_k + \epsilon\{h_k - \mathrm{E}_k(h_k)\}dG_k$ for $k = 1, \ldots, K$. The corresponding score function is $\dot{\ell}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[v, q, \mathbf{h}] = \dot{\ell}_\beta(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) + \dot{\ell}_\Lambda(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[q] + \sum_{k=1}^K \dot{\ell}_k(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[h_k]$, where

$$\begin{aligned}
\dot{\ell}_\beta(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) =& \mathrm{E}\left[\delta Y - Y\Lambda(X)\exp(\beta Y)|\mathcal{O}, \widetilde{\boldsymbol{\theta}}\right], \\
\dot{\ell}_\Lambda(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[q] =& \mathrm{E}\left[\delta q(X) - \exp(\beta Y)\int_0^X q(t)d\Lambda(t)|\mathcal{O}, \widetilde{\boldsymbol{\theta}}\right], \\
\dot{\ell}_k(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[h_k] =& \mathrm{E}\left[h_k(Y) - \mathrm{E}_k\{h_k(Y)\}|\mathcal{O}, \widetilde{\boldsymbol{\theta}}\right].
\end{aligned}$$

Obviously, $\dot{\ell}_\beta(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$, $\dot{\ell}_\Lambda(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[q]$ and $\dot{\ell}_k(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[h_k]$ have the same format as their counterparts in (B.1) for NPMLE, but with the expectation conditioning on both the observed data $\mathcal{O}$ and the initial parameter $\widetilde{\boldsymbol{\theta}}$. The calculation of the conditional expectation is given in Section 3.2 of the paper. In addition, $\dot{\ell}(\boldsymbol{\theta}, \boldsymbol{\theta}) = \dot{\ell}(\boldsymbol{\theta})$ and $\mathcal{P}\{\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)\} = \mathcal{P}\{\dot{\ell}(\boldsymbol{\theta}_0)\} = 0$. With the consistency of $\widetilde{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$ (Liu et al., 2012), $\mathcal{P}\{\dot{\ell}(\boldsymbol{\theta}_0, \widetilde{\boldsymbol{\theta}})\} \to \mathcal{P}\{\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)\} = 0$. Therefore, $\widehat{\boldsymbol{\theta}}^{(1)}$, the solution of $\mathcal{P}_n\{\dot{\ell}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})\} = 0$ is also a consistent estimator of $\boldsymbol{\theta}_0$. The consistency of one-step estimator $\widehat{\boldsymbol{\theta}}^{(1)}$ could be proved iteratively.

C.2 Weak Convergence

Since $\dot{\ell}(\widehat{\boldsymbol{\theta}}^{(1)}, \widetilde{\boldsymbol{\theta}})$ could be taken as a function indexed by $\boldsymbol{\theta}^{(1)}$ and $\widetilde{\boldsymbol{\theta}}$, using Theorem 2.11.22 of van der Vaart and Wellner (1996), we could show that

$$\sqrt{n}\mathcal{P}\{\dot{\ell}(\widehat{\boldsymbol{\theta}}^{(1)}, \widetilde{\boldsymbol{\theta}})\} = -\sqrt{n}(\mathcal{P}_n - \mathcal{P})\{\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)\} + o_p(1). \tag{C.1}$$

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be linear operators mapping $\boldsymbol{\theta} - \boldsymbol{\theta}_0$ to $\ell^\infty(\mathcal{H})$, where $\boldsymbol{\theta}$ is in the neighborhood of $\boldsymbol{\theta}_0$. On the left-hand side of (C.1), $\mathcal{P}\{\dot{\ell}(\widehat{\boldsymbol{\theta}}^{(1)}, \widetilde{\boldsymbol{\theta}})\}$ can be linearized around β_0, Λ_0, G_0 and is asymptotically equivalent to

$$\begin{aligned}
&\mathcal{B}_1(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] + \mathcal{B}_2(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] \\
=&\mathcal{B}_{1\beta}(\widehat{\beta}^{(1)} - \beta_0)[v, q, \mathbf{h}] + \mathcal{B}_{1\Lambda}(\widehat{\Lambda}^{(1)} - \Lambda_0)[v, q, \mathbf{h}] + \mathcal{B}_{1G}(v, q, \mathbf{h})(\widehat{G}^{(1)} - G_0)[v, q, \mathbf{h}] \\
+&\mathcal{B}_{2\beta}(\widetilde{\beta} - \beta_0)[v, q, \mathbf{h}] + \mathcal{B}_{2\Lambda}(\widetilde{\Lambda} - \Lambda_0)[v, q, \mathbf{h}] + \mathcal{B}_{2G}(\widetilde{G} - G_0)[v, q, \mathbf{h}],
\end{aligned}$$

where $\mathcal{B}_{1\beta}$, $\mathcal{B}_{1\Lambda}$ and $\mathcal{B}_{1G}$ are the derivative of $\dot{\ell}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[v, q, \mathbf{h}]$ with respect to $\boldsymbol{\theta}$ and $\mathcal{B}_{2\beta}$, $\mathcal{B}_{2\Lambda}$ and $\mathcal{B}_{2G}$ are the derivative of $\dot{\ell}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[v, q, \mathbf{h}]$ with respect to $\widetilde{\boldsymbol{\theta}}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ along

$$\left[\beta - \beta_0, \int_0^\tau q(t) d\{\Lambda(t) - \Lambda_0(t)\}, \sum_{k=1}^K \int_{c_u}^{c_l} h_k(c) d\{G_k(c) - G_{0k}(c)\}\right]$$

respectively. Since $\mathcal{B}_1$ is the information operator given fixed $\widetilde{\boldsymbol{\theta}}$, the invertibility of $\mathcal{B}_1$ could be proved similarly along the lines of the proof of the invertibility of $\mathcal{B}$ in Zeng and Lin (2014) using the fact that $\mathbf{h}$ is orthogonal to $\dot{\ell}_\beta(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ and $\dot{\ell}_\Lambda(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. Specifically, by Theorem 4.7 of Rudin (1973), $\mathcal{B}_1$ is invertible if it is one to one, or equivalently, is non-zero along any non-trivial submodel. Suppose B_1 is zero along some submodel

$$\left[\beta + \epsilon\overline{v}, \Lambda + \epsilon \int \overline{q} d\Lambda, dG + \epsilon\{\overline{\mathbf{h}} - \mathrm{E}(\overline{\mathbf{h}})\}dG\right],$$

where $\mathrm{E}\{\overline{\mathbf{h}}(Y)\} = 0$. Then the score function given fixed $\widetilde{\boldsymbol{\theta}}$ along this submodel is zero, i.e. $\dot{\ell}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})[\overline{v}, \overline{q}, \overline{\mathbf{h}}] = 0$. When $V = 1$, the score function is

$$\{\delta - \Lambda(X)\exp(\beta Y)\}Y\overline{v} + \delta\overline{q}(X) - \exp(\beta Y)\int_0^X \overline{q}(t)d\Lambda(t) + \sum_{k=1}^K \overline{h}_k(Y) = 0.$$

Multiplying both sides by $\overline{h}_k(Y)$ and take the expectation, it follows that $\overline{\mathbf{h}} = \mathbf{0}$. Furthermore, setting $\delta = 1$ and $X = 0$ in the above equation, it follows that $Y\overline{v} + \overline{q}(0) = 0$ and hence $\overline{v} = 0$. With $\overline{\mathbf{h}} = \mathbf{0}$ and $\overline{v} = 0$, setting $\delta = 1$ in the above equation, it follows that $\overline{q}(X) - \int_0^X \overline{q}(t)d\Lambda(t) = 0$ which is a homogeneous equation for $\overline{q}$ with only zero solution, implying that $\overline{q} = 0$. Therefore B_1 is invertible.

It should also be noted that $\mathcal{B} = \mathcal{B}_1 + \mathcal{B}_2$. Together with (C.1), it follows that

$$\mathcal{B}_1(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] + \mathcal{B}_2(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] = -\sqrt{n}(\mathcal{P}_n - \mathcal{P})\{\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)\} + o_p(1),$$

and thus

$$\begin{aligned}
&\sqrt{n}\left[v(\widehat{\beta}^{(1)} - \beta_0) + \int_0^\tau q(t)d\{\widehat{\Lambda}^{(1)}(t) - \Lambda_0(t)\} + \sum_{k=1}^K \int_{c_u}^{c_l} h_k(c)d\{\widehat{G}_k^{(1)}(c) - G_{0k}(c)\}\right] \\
&= -\sqrt{n}\mathcal{P}_n\left\{\dot{\ell} \circ \mathcal{B}_1^{-1}(\boldsymbol{\theta}_0)[v, q, \mathbf{h}] + \mathcal{B}_2 \circ \mathcal{B}_1^{-1}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}]\right\} + o_p(1).
\end{aligned} \tag{C.2}$$

Therefore

$$\begin{aligned}
\sqrt{n}(\widehat{\beta}^{(1)} - \beta_0) =& \sqrt{n}\mathcal{P}_n(\mathcal{W}_\beta^{(1)}), \quad \mathcal{W}_\beta^{(1)} = \dot{\ell} \circ \mathcal{B}_1^{-1}(\boldsymbol{\theta}_0)(1, 0, \mathbf{0}) + \mathcal{B}_2 \circ \mathcal{B}_1^{-1}(\mathbb{W}^{\text{IPW}})(1, 0, \mathbf{0}); \\
\sqrt{n}\{\widehat{\Lambda}^{(1)}(t^*) - \Lambda_0(t^*)\} =& \sqrt{n}\mathcal{P}_n(\mathcal{W}_\Lambda^{(1)}), \quad \mathcal{W}_\Lambda = \dot{\ell} \circ \mathcal{B}_1^{-1}(\boldsymbol{\theta}_0)(0, q^*, \mathbf{0}) + \mathcal{B}_2 \circ \mathcal{B}_1^{-1}(\mathbb{W}^{\text{IPW}})(0, q^*, \mathbf{0}); \\
\sqrt{n}\{\widehat{G}_k^{(1)}(c^*) - G_{0k}(c^*)\} =& \sqrt{n}\mathcal{P}_n(\mathcal{W}_{G_k}^{(1)}), \quad \mathcal{W}_{G_k}^{(1)} = \dot{\ell} \circ \mathcal{B}_1^{-1}(\boldsymbol{\theta}_0)(0, 0, \mathbf{h}_k^*) + \mathcal{B}_2 \circ \mathcal{B}_1^{-1}(\mathbb{W}^{\text{IPW}})(0, 0, \mathbf{h}_k^*),
\end{aligned}$$

where $\mathbb{W}^{\text{IPW}}$ is the expression of the weighted asymptotic linear expansion of $\widetilde{\boldsymbol{\theta}}$ which was given in Liu et al. (2012). Let $\mathcal{W}_G^{(1)} = (\mathcal{W}_{G_1}^{(1)}, \ldots, \mathcal{W}_{G_K}^{(1)})$ and $\mathbb{W}^{(1)} = (\mathcal{W}_\beta^{(1)}, \mathcal{W}_\Lambda^{(1)}, \mathcal{W}_G^{(1)})$.

Appendix D Asymptotic Results of Parameter-wise Best Linear Unbiased Estimators

In the following, we will show NPMLE $\widehat{\boldsymbol{\theta}}$ is asymptotically equivalent to an estimator $\widehat{\boldsymbol{\theta}}^{pBLUE}$ which linearly combines $\widehat{\boldsymbol{\theta}}^{(1)}$ and $\widetilde{\boldsymbol{\theta}}$. Since NPMLE is the most efficient estimators, $\widehat{\boldsymbol{\theta}}^{pBLUE}$ is the parameter-wise best linear unbiased estimator with minimum variance, i.e. the same variance as $\widehat{\boldsymbol{\theta}}$.

Since $\dot{\ell}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) = \dot{\ell}(\boldsymbol{\theta}_0)$, the left-hand sides of B.2 and C.1 are asymptotically equivalent, i.e.

$$\sqrt{n}\mathcal{B}_1(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] + \sqrt{n}\mathcal{B}_2(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] = \sqrt{n}\mathcal{B}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] + o_p(1),$$

which suggests a linear combined estimator $\widehat{\boldsymbol{\theta}}^{pBLUE}$ such that

$$\begin{aligned}
&\sqrt{n}\left[v(\widehat{\beta}^{pBLUE} - \beta_0) + \int_0^\tau q(t) d\{\widehat{\Lambda}^{pBLUE}(t) - \Lambda_0(t)\} + \sum_{k=1}^K \int_{c_u}^{c_l} h_k(c) d\{\widehat{G}_k^{pBLUE}(c) - G_{0k}(c)\}\right] \\
=& \sqrt{n}\mathcal{B}_1 \circ \mathcal{B}^{-1}(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}] + \sqrt{n}\mathcal{B}_2 \circ \mathcal{B}^{-1}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)[v, q, \mathbf{h}]
\end{aligned}$$

Obviously, $\widehat{\boldsymbol{\theta}}$ is asymptotically equivalent to $\widehat{\boldsymbol{\theta}}^{pBLUE}$ and $\widehat{\boldsymbol{\theta}}^{pBLUE}$ is the best linear unbiased estimators with optimal weight operators $\mathcal{B}_1 \circ \mathcal{B}^{-1}$ and $\mathcal{B}_2 \circ \mathcal{B}^{-1}$ for $\widehat{\boldsymbol{\theta}}^{(1)}$ and $\widetilde{\boldsymbol{\theta}}$ respectively.

Appendix E Asymptotic Linear Expansions

Using the results in Appendix B.3 of Liu et al. (2012), with the consistency of $\widehat{F}(c)$ and $\widehat{S}(t,c)$, we could show that

$$\sqrt{n}\{\widehat{\mathrm{TPR}}_t(c) - \mathrm{TPR}_t(c)\} = \sqrt{n}\mathcal{P}_n\{\mathcal{W}_{\mathrm{TPR}_t}(c)\}, \quad \sqrt{n}\{\widehat{\mathrm{FPR}}_t(c) - \mathrm{FPR}_t(c)\} = \sqrt{n}\mathcal{P}_n\{\mathcal{W}_{\mathrm{FPR}_t}(c)\},$$
$$\sqrt{n}\{\widehat{\mathrm{PPV}}_t(c) - \mathrm{PPV}_t(c)\} = \sqrt{n}\mathcal{P}_n\{\mathcal{W}_{\mathrm{PPV}_t}(c)\}, \quad \sqrt{n}\{\widehat{\mathrm{NPV}}_t(c) - \mathrm{NPV}_t(c)\} = \sqrt{n}\mathcal{P}_n\{\mathcal{W}_{\mathrm{NPV}_t}(c)\},$$
$$\sqrt{n}\{\widehat{\mathrm{ROC}}_t(c) - \mathrm{ROC}_t(c)\} = \sqrt{n}\mathcal{P}_n\{\mathcal{W}_{\mathrm{ROC}_t}(c)\}, \quad \sqrt{n}\{\widehat{\mathrm{AUC}}_t(c) - \mathrm{AUC}_t(c)\} = \sqrt{n}\mathcal{P}_n\{\mathcal{W}_{\mathrm{AUC}_t}(c)\},$$
$$\sqrt{n}\{\widehat{\mathrm{DMR}}_t(c) - \mathrm{DMR}_t(c)\} = \sqrt{n}\mathcal{P}_n\{\mathcal{W}_{\mathrm{DMR}_t}(c)\},$$

where

$$\begin{aligned}
\mathcal{W}_{\mathrm{TPR}_t}(c) &= \frac{\mathrm{TPR}_t(c)\mathcal{W}_S(t,c_u) - \mathcal{W}_F(c) - \mathcal{W}_S(t,c)}{1 - S(t,c_u)}, \\
\mathcal{W}_{\mathrm{FPR}_t}(c) &= \frac{\mathcal{W}_S(t,c) - \mathrm{FPR}_t(c)\mathcal{W}_S(t,c_u)}{S(t,c_u)}, \\
\mathcal{W}_{\mathrm{PPV}_t}(c) &= \frac{\{\mathrm{PPV}_t(c) - 1\}\mathcal{W}_F(c) - \mathcal{W}_S(t,c)}{1 - F(c)}, \\
\mathcal{W}_{\mathrm{NPV}_t}(c) &= \frac{\mathcal{W}_S(t,c_u) - \mathcal{W}_S(t,c) - \mathrm{NPV}_t(c)\mathcal{W}_F(c)}{F(c)}, \\
\mathcal{W}_{\mathrm{ROC}_t}(c) &= \mathcal{W}_{\mathrm{TPR}_t}\{\mathrm{FPR}^{-1}(c)\} - \dot{\mathrm{ROC}}_t(c)\mathcal{W}_{\mathrm{FPR}_t}\{\mathrm{FPR}^{-1}(c)\}, \\
\mathcal{W}_{\mathrm{AUC}_t} &= \int_0^1 \mathcal{W}_{\mathrm{ROC}_t}(c)dc, \\
\mathcal{W}_{\mathrm{DMR}_t} &= \int_{c_l}^{c_u} \{\mathcal{W}_{\mathrm{TPR}_t}(c) - \mathcal{W}_{\mathrm{FPR}_t}(c)\}dF(c) + \int_{c_l}^{c_u} \{\mathrm{TPR}_t(c) - \mathrm{FPR}_t(c)\}d\mathcal{W}_F(c),
\end{aligned}$$

where $\mathcal{W}_F(c)$ and $\mathcal{W}_S(t,c)$ were given in Section 4.1 of the main paper.

The asymptotic properties of $\sqrt{n}\{\widehat{A}_t^{(1)}(c) - A_t(c)\}$ could be proved following the same strategy as the weak convergece of $\sqrt{n}\{\widehat{A}_t(c) - A_t(c)\}$.

References

Liu, D., Cai, T., and Zheng, Y. (2012). Evaluating the predictive value of biomarkers with stratified case-cohort design. *Biometrics*, 68(4):1219–1227.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Process: With Applications to Statistics*. springer.

Zeng, D. and Lin, D. (2014). Efficient estimation of semiparametric transformation models for two-phase cohort studies. *Journal of the American Statistical Association*, 0(ja):null.