Supporting Information for "Combining primary cohort data with external aggregate information without assuming comparability" by Ziqi Chen, Jing Ning, Yu Shen and Jing Qin

1. Regularity Conditions and Asymptotic Proofs

We summarize the regularity conditions as follows:

C.1 The covariate vector $\mathbf{X}_1$ is in a bounded compact set $\mathcal{X}$ in $\mathbf{R}^p$.

C.2 The true regression parameter vector $\boldsymbol{\beta}_0$ lies in a compact set and $\alpha_0 = \Lambda_0(t^*) < \infty$.

C.3 Both T and C are absolutely continuous.

C.4 $E\{(\boldsymbol{\Psi}(\mathbf{X}_1;\boldsymbol{\beta}_0,\alpha_0) - \mathbf{I}_1 \circ \boldsymbol{\tau}_0)(\boldsymbol{\Psi}(\mathbf{X}_1;\boldsymbol{\beta}_0,\alpha_0) - \mathbf{I}_1 \circ \boldsymbol{\tau}_0)^T\}$ is positive definite.

C.5 The tuning parameter $\gamma \to 0$ and $\gamma n^{1/3} \to \infty$ as n goes to the infinity.

1.1 *Proof of Theorem 1*

Similar to Lemma 1 in Qin and Lawless (1994), we take the first derivative of the following function with respect to $(\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \boldsymbol{\tau})$,

$$\sum_{i=1}^{n} \Delta_i[\mathbf{X}_i^T\boldsymbol{\beta} - \log\{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta}) + \nu I(Y_i \leqslant t^*)\}] + n\nu\alpha$$
$$-\sum_{i=1}^{n} \log\{1 + \boldsymbol{\xi}^T(\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha) - \mathbf{I}_i \circ \boldsymbol{\tau})\} - n\sum_{k=1}^{K} p_\gamma(|\tau_k|),$$

and have five sets of estimation equations,

$$\begin{aligned}
\mathbf{S}_1(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha,\boldsymbol{\tau}) &= \sum_{i=1}^{n}\Delta_i\left\{\mathbf{X}_i-\frac{\mathbf{S}^{(1)}(Y_i,\boldsymbol{\beta})}{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu I(Y_i\leqslant t^*)}\right\} \\
&\quad -\sum_{i=1}^{n}\frac{\boldsymbol{\xi}^T\partial\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)/\partial\boldsymbol{\beta}}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}, \\
\mathbf{S}_2(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha,\boldsymbol{\tau}) &= \sum_{i=1}^{n}\frac{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}, \\
\mathbf{S}_3(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha,\boldsymbol{\tau}) &= \sum_{i=1}^{n}\left\{\frac{\Delta_i I(Y_i\leqslant t^*)}{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu I(Y_i\leqslant t^*)}-\alpha\right\}, \\
\mathbf{S}_4(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha,\boldsymbol{\tau}) &= \sum_{i=1}^{n}\left\{\frac{\boldsymbol{\xi}^T\partial\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)/\partial\alpha}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}-\nu\right\}, \\
\mathbf{S}_5(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha,\boldsymbol{\tau}) &= -\sum_{i=1}^{n}\left\{\frac{\mathbf{I}_i\circ\boldsymbol{\xi}}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}+b(\boldsymbol{\tau})\right\},
\end{aligned}$$

where $b(\boldsymbol{\tau}) = (p_\gamma^{'}(|\tau_1|)\text{sign}(\tau_1),\cdots,p_\gamma^{'}(|\tau_K|)\text{sign}(\tau_K))^T$ and $\mathbf{S}^{(1)}(t,\boldsymbol{\beta}) = n^{-1}\sum_{j=1}^{n} I(Y_j \geqslant t)\exp(\mathbf{X}_j^T\boldsymbol{\beta})\mathbf{X}_j$. By using Lemma 1 in Qin and Lawless (1994) and arguments in Huang et al. (2016), we can show that the maximizer of the penalized profile log-likelihood function in (11), denoted as $(\widehat{\boldsymbol{\beta}},\widehat{\alpha},\widehat{\boldsymbol{\tau}},\widehat{\boldsymbol{\xi}},\widehat{\nu})$, satisfies $\mathbf{S}_j(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha,\boldsymbol{\tau})=0$ for $j=1,\cdots,5$.

Let the penalized profile log-likelihood function (11) after removing the penalty term $n\sum_{k=1}^{K}p_\gamma(|\tau_k|)$ be l_3. Assume $(\tilde{\boldsymbol{\beta}},\tilde{\alpha},\tilde{\boldsymbol{\tau}})$ maximize l_3. By the argument of Lemma 1 in Qin and Lawless (1994), $(\tilde{\boldsymbol{\beta}},\tilde{\alpha},\tilde{\boldsymbol{\tau}})$ are in the interior of the ball $B_n = \{(\boldsymbol{\beta},\alpha,\boldsymbol{\tau}) : ||(\boldsymbol{\beta},\alpha,\boldsymbol{\tau}) - (\boldsymbol{\beta}_0,\alpha_0,\boldsymbol{\tau}_0)|| \leqslant n^{-1/3}\}$ with probability 1. Note that

$$p_\gamma(t) = \gamma|t|I(|t|\leqslant\gamma) - \frac{t^2-2a\gamma|t|+\gamma^2}{2(a-1)}I(\gamma<|t|\leqslant a\gamma) + \frac{(a+1)\gamma^2}{2}I(|t|>a\gamma).$$

By regularity condition C.5, $\gamma n^{1/3}\to\infty$, thus $\max_k\{p_\gamma(|\tau_k|) : |\tau_k-\tau_{k0}|<n^{-1/3}\}$ is smaller than $\min_k\{p_\gamma(|\tau_k|) : |\tau_k-\tau_{k0}|>n^{-1/3}\}$. This implies that $(\widehat{\boldsymbol{\beta}},\widehat{\alpha},\widehat{\boldsymbol{\tau}})$ are in the interior of the ball $B_n^* = \{(\boldsymbol{\beta},\alpha,\boldsymbol{\tau}) : ||(\boldsymbol{\beta},\alpha,\boldsymbol{\tau}) - (\boldsymbol{\beta}_0,\alpha_0,\boldsymbol{\tau}_0)|| \leqslant C_1 n^{-1/3}\}$ with probability 1, where C_1 is a constant. Given $||(\boldsymbol{\beta},\alpha,\boldsymbol{\tau}) - (\boldsymbol{\beta}_0,\alpha_0,\boldsymbol{\tau}_0)|| \leqslant C_1 n^{-1/3}$, similar to the arguments of Owen

(1990) and Qin and Lawless (1994), we have

$$\begin{aligned}\boldsymbol{\xi}(\boldsymbol{\beta},\alpha,\boldsymbol{\tau}) &= \left[\frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}^T\right]^{-1} \\ &\quad \times\left[\frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}\right]+o(n^{-1/3}) \quad (a.s.) \\ &= O(n^{-1/3}) \quad (a.s.)\end{aligned}$$

It follows that $||(\boldsymbol{\beta},\boldsymbol{\xi},\alpha,\boldsymbol{\tau})-(\boldsymbol{\beta}_0,\mathbf{0},\alpha_0,\boldsymbol{\tau}_0)|| \leqslant C_2 n^{-1/3}$ holds with C_2 being a constant. For each component of $\boldsymbol{\tau}$,

$$n^{-1}\mathbf{S}_{5k}=\frac{1}{n}\sum_{i=1}^{n}\frac{I(\mathbf{X}_i\in\Omega_k)\xi_k}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}-p_\gamma'(|\tau_k|)\text{sign}(\tau_k)=A_k+B_k.$$

Since $\max_k |A_k| \leqslant \max_k |\xi_k|\ |n^{-1}\sum_{i=1}^n\{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha)-\mathbf{I}_i\circ\boldsymbol{\tau}\}^{-1}| = O_p(n^{-1/3})$ $\times\, O_p(1) = o_p(\gamma)$, $P(\max_k |A_k| > \gamma/2) \to 0$. When $\boldsymbol{\tau}_0=\mathbf{0}$, $|\tau_k| \leqslant cn^{-1/3}$, thus we have $p_\gamma'(|\tau_k|)=\gamma$, and $B_k=-p_\gamma'(|\tau_k|)\text{sign}(\tau_k)=-\gamma\text{sign}(\tau_k)$. This implies that $\text{sign}(\tau_k)$ dominates $n^{-1}\mathbf{S}_{5k}$ asymptotically. As $n\to\infty$, with probability tending to one, $n^{-1}\mathbf{S}_{5k}<0$ for $\tau_k \in (0, cn^{-1/3})$ and $n^{-1}\mathbf{S}_{5k}>0$ for $\tau_k\in(-cn^{-1/3},0)$. So, $\widehat{\tau}_k=0$ $(k=1,\cdots,K)$ with probability tending to one when $\boldsymbol{\tau}_0=\mathbf{0}$.

Define $\mathbf{S}_{0k}=\mathbf{S}_k(\boldsymbol{\beta}_0,\mathbf{0},0,\alpha_0,\mathbf{0})$, for $k=1,2,3,4,5$. Let $\boldsymbol{\theta}=(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha)$. By using the same arguments as those on Page 790 of Huang et al. (2016), we can show that $\widehat{\boldsymbol{\theta}}$ satisfies $\mathbf{S}_j(\boldsymbol{\beta},\boldsymbol{\xi},\nu,\alpha,\mathbf{0})=0$ for $j=1,\cdots,4$. Therefore, $n^{1/2}\{\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\}$ converges to a multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Gamma}^{-1}=(\boldsymbol{\Sigma}+\tilde{\mathbf{B}}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{B}}^T)^{-1}$, where

$$\begin{aligned}\boldsymbol{\Sigma} &:= \text{Var}(n^{-1/2}\mathbf{S}_{01}),\ \mathbf{J}:=\text{Var}(n^{-1/2}\mathbf{S}_{02}),\ \mathbf{H}:=-n^{-1}E\{\partial\boldsymbol{\Psi}(\mathbf{X}_1;\boldsymbol{\beta}_0,\alpha_0)/\partial\alpha\}, \\ \tilde{\mathbf{B}} &:= \left[E\left\{\frac{\partial\boldsymbol{\Psi}(\mathbf{X}_1;\boldsymbol{\beta}_0,\alpha_0)^T}{\partial\boldsymbol{\beta}}\right\},-\int_0^{t^*}\frac{E\{\mathbf{X}_1\exp(\mathbf{X}_1^T\boldsymbol{\beta}_0)I(Y_1\geqslant u)\}}{E\{\exp(\mathbf{X}_1^T\boldsymbol{\beta}_0)I(Y_1\geqslant u)\}}d\Lambda_0(u),\mathbf{0}_{p\times 1}\right], \\ L &:= \text{Var}(n^{-1/2}\mathbf{S}_{03})=\int_0^{t^*}\left[E\{\exp(\mathbf{X}_1^T\boldsymbol{\beta}_0)I(Y\geqslant u)\}\right]^{-1}d\Lambda_0(u), \\ \tilde{\mathbf{Q}} &:= \begin{pmatrix}\mathbf{J} & \mathbf{0} & \mathbf{H}\\ \mathbf{0}^T & L & 1\\ \mathbf{H}^T & 1 & 0\end{pmatrix}.\end{aligned}$$

Similarly, the same arguments for the weak convergence of $\widehat{\Lambda}_C(t)$ in Huang et al. (2016) can be applied to derive the weak convergence of $\widehat{\Lambda}(t)$.

1.2 *Proof of Theorem 2*

Using the same arguments for Theorem 1, we can show $\widehat{\boldsymbol{\tau}}_1 \to^p \boldsymbol{\tau}_{01}$ and $P(\widehat{\boldsymbol{\tau}}_{02} = \mathbf{0}) \to 1$, as $n \to \infty$. We assume $\boldsymbol{\tau}_{01}$ is $l-$dimensional and denote $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \mathbf{0}, 0, \alpha_0, \boldsymbol{\tau}_0)$ and $\mathbf{S}^*_{0k} = \mathbf{S}_k(\boldsymbol{\beta}_0, \mathbf{0}, 0, \alpha_0, \boldsymbol{\tau}_0)$, for $k = 1, 2, 3, 4, 5$. Then $\mathbf{S}^*_{0k}$ can be rewritten as

$$\begin{aligned}
\mathbf{S}^*_{01} &= \sum_{i=1}^n \int_0^\infty \left\{ \mathbf{X}_i - \frac{\mathbf{S}^{(1)}(u, \boldsymbol{\beta}_0)}{\mathbf{S}^{(0)}(u, \boldsymbol{\beta}_0)} \right\} dM_i(u), \\
\mathbf{S}^*_{02} &= \sum_{i=1}^n \{ \boldsymbol{\Psi}(\mathbf{X}_i; \boldsymbol{\beta}_0, \alpha_0) - \mathbf{I}_i \circ \boldsymbol{\tau}_0 \}, \\
\mathbf{S}^*_{03} &= \sum_{i=1}^n \int_0^\infty \frac{I(u \leqslant t^*)}{\mathbf{S}^{(0)}(u, \boldsymbol{\beta}_0)} dM_i(u), \\
\mathbf{S}^*_{04} &= 0, \\
\mathbf{S}^*_{05} &= \mathbf{0}_{l \times 1},
\end{aligned}$$

where $M_i(t) = N_i(t) - \int_0^t \exp(\mathbf{X}_i^T \boldsymbol{\beta}_0) I(Y_i \geqslant t) d\Lambda_0(t)$ and $N_i(t) = \Delta_i I(Y_i \leqslant t)$. Clearly, $E\mathbf{S}^*_{01} = \mathbf{0}$, $E\mathbf{S}^*_{02} = \mathbf{0}$ and $E\mathbf{S}^*_{03} = 0$. Simple algebra shows that

$$\boldsymbol{\Sigma} := \mathrm{Var}(n^{-1/2}\mathbf{S}^*_{01}) = \int_0^\infty \mathbf{m}(u) d\Lambda_0(u),$$

where $\mathbf{m}(u) = E\left[\{\mathbf{X}_1 - \boldsymbol{\mu}(u)\}\{\mathbf{X}_1 - \boldsymbol{\mu}(u)\}^T \exp(\mathbf{X}_1^T \boldsymbol{\beta}_0) I(Y_1 \geqslant u) \right]$ with

$$\boldsymbol{\mu}(u) = \frac{E\{\mathbf{X}_1 \exp(\mathbf{X}_1^T \boldsymbol{\beta}_0) I(Y_1 \geqslant u)\}}{E\{\exp(\mathbf{X}_1^T \boldsymbol{\beta}_0) I(Y_1 \geqslant u)\}}.$$

By the double expectation formula, $\mathrm{Cov}(\mathbf{S}^*_{01}, \mathbf{S}^*_{02}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{S}^*_{01}, \mathbf{S}^*_{03}) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{S}^*_{02}, \mathbf{S}^*_{03}) = \mathbf{0}$. Then, by the martingale central limit theorem and classic central limit theorem, $(n^{-1/2}\mathbf{S}^{*T}_{01}, n^{-1/2}\mathbf{S}^{*T}_{02}, n^{-1/2}\mathbf{S}^*_{03}, \mathbf{S}^*_{04}, \mathbf{S}^{*T}_{05})^T$ converges to a zero mean multivariate normal distribution with covariance matrix $\mathbf{\Omega}$ defined by

$$\begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & L & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $L := \mathrm{Var}(n^{-1/2}\mathbf{S}_{03}^*) = \int_0^{t^*} \left[E\{\exp(\mathbf{X}_1^T\boldsymbol{\beta}_0)I(Y_1 \geqslant u)\}\right]^{-1} d\Lambda_0(u)$ and $\mathbf{J} := \mathrm{Var}(n^{-1/2}\mathbf{S}_{02}^*)$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \boldsymbol{\tau}_1)$ and $\mathbf{S}(\boldsymbol{\theta}) = (\mathbf{S}_1(\boldsymbol{\theta})^T, \mathbf{S}_2(\boldsymbol{\theta})^T, \mathbf{S}_3(\boldsymbol{\theta}), \mathbf{S}_4(\boldsymbol{\theta}), \mathbf{S}_5(\boldsymbol{\theta})^T)^T$. Note that

$$-n^{-1}E\left(\frac{\partial \mathbf{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\Big|_{\boldsymbol{\theta}_0} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{Q} \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{B} &= -n^{-1}\left[E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\xi}}\right\}\Big|_{\boldsymbol{\theta}_0}, E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \nu}\right\}\Big|_{\boldsymbol{\theta}_0}, E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \alpha}\right\}\Big|_{\boldsymbol{\theta}_0}, E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}_1}\right\}\Big|_{\boldsymbol{\theta}_0}\right] \\ &= \left[E\left\{\frac{\partial \mathbf{\Psi}(\mathbf{X}_1; \boldsymbol{\beta}_0, \alpha_0)^T}{\partial \boldsymbol{\beta}}\right\}, -\int_0^{t^*} \frac{E\{\mathbf{X}_1 \exp(\mathbf{X}_1^T\boldsymbol{\beta}_0)I(Y_1 \geqslant u)\}}{E\{\exp(\mathbf{X}_1^T\boldsymbol{\beta}_0)I(Y_1 \geqslant u)\}} d\Lambda_0(u), \mathbf{0}_{p\times 1}, \mathbf{0}_{p\times l}\right], \end{aligned}$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{J} & \mathbf{0} & \mathbf{H} & \mathbf{E} \\ \mathbf{0} & L & 1 & \mathbf{0} \\ \mathbf{H}^T & 1 & 0 & \mathbf{0} \\ \mathbf{E}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} \mathrm{diag}(\mathbf{I}^{(l)}) \\ \mathbf{0}_{(K-l)\times l} \end{pmatrix},$$

and $\mathbf{H} = -n^{-1}E\{\partial \mathbf{\Psi}(\mathbf{X}_1; \boldsymbol{\beta}_0, \alpha_0)/\partial \alpha\}$, where $\mathbf{I}^{(l)} = (E\{I(\mathbf{X}_1 \in \Omega_1)\}, \cdots, E\{I(\mathbf{X}_1 \in \Omega_l)\})^T$. Therefore,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{Q} \end{pmatrix}^{-1} \begin{pmatrix} n^{-1/2}\mathbf{S}_{01}^* \\ n^{-1/2}\dot{\mathbf{S}}^* \end{pmatrix} + o_p(1), \tag{1}$$

where $\dot{\mathbf{S}}^* = (\mathbf{S}_{02}^{*T}, \mathbf{S}_{03}^*, 0, \mathbf{0}^T)^T$. Since

$$\begin{pmatrix} \mathbf{\Sigma} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{Q} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1} & -(\mathbf{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ \mathbf{Q}^{-1}\mathbf{B}^T(\mathbf{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1} & (\mathbf{Q} + \mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B})^{-1} \end{pmatrix},$$

$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = (\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}(n^{-1/2}\mathbf{S}^*_{01}) - (\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}^{-1}(n^{-1/2}\dot{\mathbf{S}}^*) + o_p(1)$. It implies that the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ is

$$
\begin{aligned}
&(\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1} \\
&\quad + (\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}^{-1}\begin{pmatrix} \mathbf{J} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & L & 0 & \mathbf{0} \\ \mathbf{0} & 0 & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{Q}^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1} \\
&= (\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1} \\
&\quad - (\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}^{-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{H} & \mathbf{E} \\ \mathbf{0} & 0 & 1 & \mathbf{0} \\ \mathbf{H}^T & 1 & 0 & \mathbf{0} \\ \mathbf{E}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{Q}^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}.
\end{aligned}
$$

Through some algebra, we can show that

$$
\mathbf{B}\mathbf{Q}^{-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{H} & \mathbf{E} \\ \mathbf{0} & 0 & 1 & \mathbf{0} \\ \mathbf{H}^T & 1 & 0 & \mathbf{0} \\ \mathbf{E}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{Q}^{-1}\mathbf{B}^T = \mathbf{0},
$$

which follows that the asymptotic covariance matrix of $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is $(\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}$. Similarly, the same arguments for the weak convergence of $\widehat{\Lambda}_C(t)$ in Huang et al. (2016) can be applied to derive the weak convergence of $\widehat{\Lambda}(t)$.

2. Variance Estimation

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \boldsymbol{\tau}_1)$ and $\mathbf{S}(\boldsymbol{\theta}) = (\mathbf{S}_1(\boldsymbol{\theta})^T, \mathbf{S}_2(\boldsymbol{\theta})^T, \mathbf{S}_3(\boldsymbol{\theta}), \mathbf{S}_4(\boldsymbol{\theta}), \mathbf{S}_5(\boldsymbol{\theta})^T)^T$. The information matrix of the proposed penalized likelihood is

$$-n^{-1}E\left(\frac{\partial \mathbf{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\Big|_{\boldsymbol{\theta}_0} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{Q} \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{B} &= -n^{-1}\left[E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\xi}}\right\}\Big|_{\boldsymbol{\theta}_0}, E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \nu}\right\}\Big|_{\boldsymbol{\theta}_0}, E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \alpha}\right\}\Big|_{\boldsymbol{\theta}_0}, E\left\{\frac{\partial \mathbf{S}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}_1}\right\}\Big|_{\boldsymbol{\theta}_0}\right] \\ &= \left[E\left\{\frac{\partial \boldsymbol{\Psi}(\mathbf{X}_1; \boldsymbol{\beta}_0, \alpha_0)^T}{\partial \boldsymbol{\beta}}\right\}, -\int_0^{t^*} \frac{E\{\mathbf{X}_1 \exp(\mathbf{X}_1^T\boldsymbol{\beta}_0) I(Y_1 \geqslant u)\}}{E\{\exp(\mathbf{X}_1^T\boldsymbol{\beta}_0) I(Y_1 \geqslant u)\}} d\Lambda_0(u), \mathbf{0}_{p\times 1}, \mathbf{0}_{p\times l}\right], \end{aligned}$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{J} & \mathbf{0} & \mathbf{H} & \mathbf{E} \\ \mathbf{0} & L & 1 & \mathbf{0} \\ \mathbf{H}^T & 1 & 0 & \mathbf{0} \\ \mathbf{E}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} \mathrm{diag}(\mathbf{I}^{(l)}) \\ \mathbf{0}_{(K-l)\times l} \end{pmatrix},$$

$\mathbf{H} = -n^{-1}E\{\partial \boldsymbol{\Psi}(\mathbf{X}_1; \boldsymbol{\beta}_0, \alpha_0)/\partial \alpha\}$ and $\mathbf{I}^{(l)} = (E\{I(\mathbf{X}_1 \in \Omega_1)\}, \cdots, E\{I(\mathbf{X}_1 \in \Omega_l)\})^T$. On the other hand, we have derived that

$$\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{Q} \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1} & -(\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ \mathbf{Q}^{-1}\mathbf{B}^T(\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1} & (\mathbf{Q} + \mathbf{B}^T\boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1} \end{pmatrix},$$

and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a zero mean multivariate normal distribution with covariance matrix $(\boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^T)^{-1}$. Therefore, the asymptotic covariance of $\widehat{\boldsymbol{\beta}}$ can be estimated by the sub-matrix consisting of the first p rows and the first p columns of $-\left\{E\left(\frac{\partial \mathbf{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\Big|_{\boldsymbol{\theta}_0}\right\}^{-1}$. The estimators of the asymptotic covariances of $\widehat{\boldsymbol{\beta}}_{DEL}$ and $\widehat{\boldsymbol{\beta}}_{DEL-E}$ can be obtained similarly. For the estimated cumulative baseline hazard function, by the Taylor's expansion, we have

$$\begin{aligned} \widehat{\Lambda}(t) &= \widehat{\Lambda}(t; \widehat{\boldsymbol{\beta}}, \widehat{\nu}) := \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i I(Y_i \leqslant t)}{\mathbf{S}^{(0)}(Y_i, \widehat{\boldsymbol{\beta}}) + \widehat{\nu} I(Y_i \leqslant t^*)} \\ &= \widehat{\Lambda}(t; \boldsymbol{\beta}_0, 0) + \left\{\frac{\partial \widehat{\Lambda}(t; \boldsymbol{\beta}, \nu)}{\partial \boldsymbol{\beta}}\Big|_{(\boldsymbol{\beta}^*, 0)}\right\}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \left\{\frac{\partial \widehat{\Lambda}(t; \boldsymbol{\beta}, \nu)}{\partial \nu}\Big|_{(\boldsymbol{\beta}_0, \nu^*)}\right\}(\widehat{\nu} - 0), \quad (2) \end{aligned}$$

where $(\boldsymbol{\beta}^*, \nu^*)$ lies between $(\widehat{\boldsymbol{\beta}}, \widehat{\nu})$ and $(\boldsymbol{\beta}_0, 0)$. Then, for each time point t, the standard error (SE) of $\widehat{\Lambda}(t)$ (denoted as $SE(t)$) can be obtained based on (2).

In the simulation studies, to report the overall performance of the $\widehat{\Lambda}(t)$, we defined a set of 100 equidistant time points within the interquartile range, denoted as $(t_1, \cdots, t_{100})$. We then averaged estimation biases, standard errors, and converge probabilities over these time points, and then used these aggregate statistics to compare the performance of the four methods. We further defined the empirical mean integrated squared error (EMISE) to summarize the estimation accuracy and efficiency of the estimated $\Lambda(\cdot)$ as follows:

$$\text{EMISE}(\widehat{\Lambda}) = \sum_{q=1}^{500}\sum_{l=1}^{100}\{\widehat{\Lambda}^{(q)}(t_l) - \Lambda(t_l)\}^2/(50000),$$

where $\widehat{\Lambda}^{(q)}(\cdot)$ is the estimator of $\Lambda(\cdot)$ using the qth data set, for $q = 1, \ldots, 500$.

3. Extension to Incorporate Aggregate Survival Information at Multiple Time Points

Without loss of generality, we consider external survival information available at two time points,

$$Pr(T > t_1^*|\mathbf{X} \in \Omega_k) = \phi_{1k}, \quad k = 1, 2, \cdots, K \tag{3}$$

and

$$Pr(T > t_2^*|\mathbf{X} \in \Omega_k) = \phi_{2k}, \quad k = 1, 2, \cdots, K. \tag{4}$$

We introduce two additional parameters, defined as $\alpha_1 = \Lambda(t_1^*)$ and $\alpha_2 = \Lambda(t_2^*)$, whose sample analogues are

$$\sum_{i=1}^{n} \lambda_i I(Y_i \leqslant t_1^*) - \alpha_1 = 0 \tag{5}$$

and

$$\sum_{i=1}^{n} \lambda_i I(Y_i \leqslant t_2^*) - \alpha_2 = 0. \tag{6}$$

We then maximize the conditional log-likelihood l_1 with constraints (5) and (6). Using the Lagrange multipliers method, the objective function to be maximized is

$$\sum_{i=1}^{n}\Delta_i\{\mathbf{X}_i^T\boldsymbol{\beta}+\log(\lambda_i)\}-n\sum_{i=1}^{n}\lambda_i\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})-n\nu_1\left\{\sum_{i=1}^{n}\lambda_iI(Y_i\leqslant t_1^*)-\alpha_1\right\}$$
$$-n\nu_2\left\{\sum_{i=1}^{n}\lambda_iI(Y_i\leqslant t_2^*)-\alpha_2\right\}. \quad (7)$$

Taking a derivative of (7) with respect to λ_i, and letting the derivative be 0, we have

$$\lambda_i=\frac{\Delta_i}{n\{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu_1I(Y_i\leqslant t_1^*)+\nu_2I(Y_i\leqslant t_2^*)\}}, \quad (8)$$

where ν_1 and ν_2 are determined by

$$\sum_{i=1}^{n}\left\{\frac{\Delta_iI(Y_i\leqslant t_1^*)}{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu_1I(Y_i\leqslant t_1^*)+\nu_2I(Y_i\leqslant t_2^*)}-\alpha_1\right\}=0,$$

and

$$\sum_{i=1}^{n}\left\{\frac{\Delta_iI(Y_i\leqslant t_2^*)}{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu_1I(Y_i\leqslant t_1^*)+\nu_2I(Y_i\leqslant t_2^*)}-\alpha_2\right\}=0.$$

After plugging Equation (8) into (7), we have the double empirical log-likelihood,

$$l=\sum_{i=1}^{n}\Delta_i[\mathbf{X}_i^T\boldsymbol{\beta}-\log\{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu_1I(Y_i\leqslant t_1^*)+\nu_2I(Y_i\leqslant t_2^*)\}]+n\nu_1\alpha_1+n\nu_2\alpha_2+\sum_{i=1}^{n}\log(p_i). \quad (9)$$

Given the external aggregate information, the estimators of $\boldsymbol{\beta}$, α_1, α_2, ν_1, and ν_2 can be derived by maximizing the likelihood function (9) with the following constraints:

$$p_i\geqslant 0,\ \ \sum_{i=1}^{n}p_i=1,\ \ \sum_{i=1}^{n}p_i\Psi_k(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\phi_{1k})=0,\ \ \sum_{i=1}^{n}p_i\Psi_k(\mathbf{X}_i;\boldsymbol{\beta},\alpha_2,\phi_{2k})=0, k=1,\cdots,K, \quad (10)$$

where $\Psi_k(\mathbf{X}_i;\boldsymbol{\beta},\alpha_l,\phi_{lk})=I(\mathbf{X}_i\in\Omega_k)\left[\exp\{-\alpha_l\exp(\mathbf{X}_i^T\boldsymbol{\beta})\}-\phi_{lk}\right]$ for $l=1,2$. We denote the corresponding estimators of $\boldsymbol{\beta}$, α_1, α_2, ν_1, and ν_2 here as $\widehat{\boldsymbol{\beta}}_C$, $\widehat{\alpha}_{1C}$, $\widehat{\alpha}_{2C}$, $\widehat{\nu}_{1C}$, and $\widehat{\nu}_{2C}$, respectively. Using the combined information, the baseline cumulative hazard function $\Lambda(t)$ can be estimated by:

$$\widehat{\Lambda}_C(t)=\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_iI(Y_i\leqslant t)}{\mathbf{S}^{(0)}(Y_i,\widehat{\boldsymbol{\beta}}_C)+\widehat{\nu}_{1C}I(Y_i\leqslant t_1^*)+\widehat{\nu}_{2C}I(Y_i\leqslant t_2^*)}.$$

Let $\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)=(\Psi_1(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\phi_{11}),\cdots,\Psi_K(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\phi_{1K}),\Psi_1(\mathbf{X}_i;\boldsymbol{\beta},\alpha_2,\phi_{21}),\cdots,$

$\Psi_K(\mathbf{X}_i;\boldsymbol{\beta},\alpha_2,\phi_{2K}))^T$. The primary cohort may not be comparable to the aggregate survival information, we introduce additional parameters $\boldsymbol{\tau}=(\tau_{11},\cdots,\tau_{1K},\tau_{21},\cdots,\tau_{2K})^T$ to incorporate such information. Following similar arguments in Section 3.2, we use the following estimating equations:

$$
\begin{aligned}
\mathbf{S}_1(\boldsymbol{\beta},\boldsymbol{\xi},\nu_1,\nu_2,\alpha_1,\alpha_2,\boldsymbol{\tau}) &= \sum_{i=1}^n \Delta_i\left\{\mathbf{X}_i - \frac{\mathbf{S}^{(1)}(Y_i,\boldsymbol{\beta})}{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu_1 I(Y_i\leqslant t_1^*)+\nu_2 I(Y_i\leqslant t_2^*)}\right\}\\
&\quad -\sum_{i=1}^n \frac{\boldsymbol{\xi}^T\partial\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)/\partial\boldsymbol{\beta}}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)-\mathbf{I}_i\circ\boldsymbol{\tau}\}},\\
\mathbf{S}_2(\boldsymbol{\beta},\boldsymbol{\xi},\nu_1,\nu_2,\alpha_1,\alpha_2,\boldsymbol{\tau}) &= \sum_{i=1}^n \frac{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)-\mathbf{I}_i\circ\boldsymbol{\tau}}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)-\mathbf{I}_i\circ\boldsymbol{\tau}\}},\\
\mathbf{S}_3(\boldsymbol{\beta},\boldsymbol{\xi},\nu_1,\nu_2,\alpha_1,\alpha_2,\boldsymbol{\tau}) &= \sum_{i=1}^n\left\{\frac{\Delta_i I(Y_i\leqslant t_1^*)}{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu_1 I(Y_i\leqslant t_1^*)+\nu_2 I(Y_i\leqslant t_2^*)}-\alpha_1\right\},\\
\mathbf{S}_4(\boldsymbol{\beta},\boldsymbol{\xi},\nu_1,\nu_2,\alpha_1,\alpha_2,\boldsymbol{\tau}) &= \sum_{i=1}^n\left\{\frac{\Delta_i I(Y_i\leqslant t_2^*)}{\mathbf{S}^{(0)}(Y_i,\boldsymbol{\beta})+\nu_1 I(Y_i\leqslant t_1^*)+\nu_2 I(Y_i\leqslant t_2^*)}-\alpha_2\right\}\\
\mathbf{S}_5(\boldsymbol{\beta},\boldsymbol{\xi},\nu_1,\nu_2,\alpha_1,\alpha_2,\boldsymbol{\tau}) &= \sum_{i=1}^n\left\{\frac{\boldsymbol{\xi}^T\partial\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)/\partial\alpha_1}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}-\nu_1\right\},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{S}_6(\boldsymbol{\beta},\boldsymbol{\xi},\nu_1,\nu_2,\alpha_1,\alpha_2,\boldsymbol{\tau}) &= \sum_{i=1}^n\left\{\frac{\boldsymbol{\xi}^T\partial\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)/\partial\alpha_2}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}-\nu_2\right\},\\
\mathbf{S}_7(\boldsymbol{\beta},\boldsymbol{\xi},\nu_1,\nu_2,\alpha_1,\alpha_2,\boldsymbol{\tau}) &= -\sum_{i=1}^n\left\{\frac{\mathbf{I}_i\circ\boldsymbol{\xi}}{1+\boldsymbol{\xi}^T\{\boldsymbol{\Psi}(\mathbf{X}_i;\boldsymbol{\beta},\alpha_1,\alpha_2)-\mathbf{I}_i\circ\boldsymbol{\tau}\}}+\mathbf{b}(\boldsymbol{\tau})\right\},
\end{aligned}
$$

where $\mathbf{I}_i=(I(\mathbf{X}_i\in\Omega_1),\cdots,I(\mathbf{X}_i\in\Omega_K),I(\mathbf{X}_i\in\Omega_1),\cdots,I(\mathbf{X}_i\in\Omega_K))^T$ and $\mathbf{b}(\boldsymbol{\tau})=(p'_\gamma(|\tau_{11}|)\text{sign}(\tau_{11}),\cdots,p'_\gamma(|\tau_{1K}|)\text{sign}(\tau_{1K}),p'_\gamma(|\tau_{21}|)\text{sign}(\tau_{21}),\cdots,p'_\gamma(|\tau_{2K}|)\text{sign}(\tau_{2K}))^T$. We get estimators $\widehat{\boldsymbol{\beta}}$, $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\nu}_1$, and $\widehat{\nu}_2$ by solving these estimating equations. The cumulative baseline hazard function $\Lambda(t)$ can be estimated by

$$
\widehat{\Lambda}(t)=\frac{1}{n}\sum_{i=1}^n\frac{\Delta_i I(Y_i\leqslant t)}{\mathbf{S}^{(0)}(Y_i,\widehat{\boldsymbol{\beta}})+\widehat{\nu}_1 I(Y_i\leqslant t_1^*)+\widehat{\nu}_2 I(Y_i\leqslant t_2^*)}.
$$

In practice, we recommend not using the external survival information from two approximate time points. It would not increase the statistical efficiency much, but it could cause singularity in the covariance matrix of the aforementioned estimating equations.

4. Additional Simulation Studies

We conducted simulation studies to evaluate the finite sample performance of the proposed method when the external aggregate survival rate $P(T > t^*|X_2 = 0)$ is comparable with the primary cohort, but $P(T > t^*|X_2 = 1)$ is not, i.e., $(\tau_1, \tau_2) = (0, 0.06)$. The data generation mechanism of the primary cohort was the same as that of Section 4.1. Table S3 summarizes the simulation results of the estimated regression coefficients, including the empirical biases, empirical standard deviation (SD), and estimated standard errors (SE); square root of mean squared errors (RMSE); and coverage probabilities (CP) of 95% Wald-type confidence intervals.

All the four methods produced similar results for the regression coefficient of X_1 (e.g., β_1), since the external aggregate information was given in the subgroups determined by X_2. For β_2, our proposed method was more efficient than the standard Cox regression with the partial likelihood (PL). On the other hand, the double empirical likelihood method (DEL) and its extension (DEL-E) by Huang et al. (2016) had large estimation biases and inaccurate inference conclusions by directly borrowing from incomparable external information.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

We have conducted additional simulation studies to compare estimators by using two different penalties (SCAD and LASSO) under four settings. The simulation results (see Table S4) suggest that two different penalties produced almost identical results under our settings, confirming the robustness of our method in terms of the penalty functions.

[Table 4 about here.]

References

Huang, C. Y., Qin, J., and Tsai, H. T. (2016). Efficient estimation of the cox model with auxiliary subgroup survival information. *Journal of the American Statistical Association* **111,** 787–799.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22,** 300–325.

Table S1: Simulation results (all the entries are multiplied by 100) of the estimated cumulative baseline hazard functions under Settings 1 and 2. PL, the standard Cox regression with the partial likelihood; DEL and DEL-E, the double empirical likelihood method and its extension by Huang et al. (2016); ADEL, the proposed adaptive double empirical likelihood method.

		Setting 1					Setting 2				
PC	Method	Bias	SD	SE	EMISE	CP	Bias	SD	SE	EMISE	CP
Sample Size=100											
0%	PL	−0.59	9.19	8.94	1.04	92.8	−0.59	9.19	8.94	1.04	92.8
	DEL	0.36	6.04	6.95	0.52	96.9	−3.79	6.05	6.78	0.69	86.4
	DEL-E	−0.51	8.90	8.83	0.97	93.5	−4.36	8.70	8.21	1.16	84.0
	ADEL	−0.29	7.03	7.72	0.66	96.2	−0.63	7.14	8.55	0.67	96.1
15%	PL	0.19	10.47	9.67	1.39	91.5	0.19	10.47	9.67	1.39	91.5
	DEL	0.63	6.73	7.54	0.68	96.4	−3.41	6.72	7.37	0.81	87.5
	DEL-E	0.43	10.36	9.57	1.36	91.9	−3.47	10.07	8.92	1.45	84.0
	ADEL	0.16	8.16	8.51	0.90	95.3	−0.55	8.17	9.28	0.91	94.7
30%	PL	0.05	11.32	10.43	1.62	91.0	0.05	11.32	10.43	1.62	91.0
	DEL	0.68	7.21	8.14	0.79	96.3	−3.32	7.28	7.96	0.93	87.9
	DEL-E	0.35	11.17	10.31	1.59	91.5	−3.61	10.95	9.61	1.71	84.2
	ADEL	0.03	8.55	9.36	1.01	96.0	−0.58	8.70	10.00	1.04	93.9
Sample Size=200											
0%	PL	−0.37	6.39	6.33	0.50	93.3	−0.37	6.39	6.33	0.50	93.3
	DEL	0.18	4.33	4.92	0.27	96.8	−4.06	4.31	4.80	0.46	79.1
	DEL-E	−0.32	6.19	6.26	0.46	93.9	−4.20	6.06	5.83	0.66	82.5
	ADEL	−0.15	5.21	5.43	0.35	95.5	−0.82	5.30	6.19	0.37	95.5
15%	PL	0.33	7.00	6.81	0.61	94.3	0.33	7.00	6.81	0.61	94.3
	DEL	0.19	4.49	5.29	0.29	96.8	−3.92	4.47	5.16	0.47	81.6
	DEL-E	0.52	6.85	6.76	0.59	95.1	−3.36	6.72	6.29	0.71	85.1
	ADEL	0.02	5.62	5.89	0.42	95.6	−0.50	5.74	6.65	0.43	95.9
30%	PL	0.36	7.79	7.37	0.78	93.8	0.36	7.79	7.37	0.78	93.8
	DEL	0.26	5.00	5.74	0.37	96.8	−3.83	4.97	5.59	0.54	82.3
	DEL-E	0.56	7.60	7.30	0.74	94.5	−3.42	7.45	6.79	0.86	84.4
	ADEL	0.12	6.33	6.50	0.54	95.9	−0.40	6.36	7.26	0.55	96.1

Table S2: Simulation results (all the entries are multiplied by 100) of the estimated cumulative baseline hazard functions under Settings 3 and 4. PL, the standard Cox regression with the partial likelihood; DEL and DEL-E, the double empirical likelihood method and its extension by Huang et al. (2016); ADEL, the proposed adaptive double empirical likelihood method.

		Setting 3					Setting 4				
PC	Method	Bias	SD	SE	EMISE	CP	Bias	SD	SE	EMISE	CP
					Sample Size=100						
0%	PL	−0.59	9.19	8.94	1.04	92.8	−0.59	9.19	8.94	1.04	92.8
	DEL	5.17	6.30	7.17	0.86	84.6	9.97	6.56	7.37	1.70	65.5
	DEL-E	0.12	8.93	8.92	0.97	94.3	0.59	8.96	9.00	0.98	94.8
	ADEL	0.82	7.26	8.29	0.69	94.8	2.02	8.02	8.74	0.86	93.6
15%	PL	0.19	10.47	9.67	1.39	91.5	0.19	10.47	9.67	1.39	91.5
	DEL	5.48	7.00	7.78	1.06	85.6	10.35	7.28	7.99	1.97	68.2
	DEL-E	1.04	10.40	9.67	1.38	92.8	1.48	10.42	9.74	1.39	93.3
	ADEL	1.18	8.41	9.13	0.97	94.6	2.17	8.84	9.59	1.08	94.3
30%	PL	0.05	11.32	10.43	1.62	91.0	0.05	11.32	10.43	1.62	91.0
	DEL	5.61	7.51	8.41	1.20	86.4	10.56	7.80	8.65	2.15	70.5
	DEL-E	0.95	11.19	10.42	1.59	92.2	1.35	11.13	10.49	1.58	92.7
	ADEL	0.91	8.86	9.95	1.08	95.1	1.96	9.28	10.44	1.19	94.9
					Sample Size=200						
0%	PL	−0.37	6.39	6.33	0.50	93.3	−0.37	6.39	6.33	0.50	93.3
	DEL	5.01	4.52	5.08	0.57	76.1	9.83	4.72	5.22	1.38	47.6
	DEL-E	0.30	6.20	6.33	0.46	94.9	0.77	6.21	6.39	0.47	95.3
	ADEL	0.82	5.36	6.11	0.37	95.0	2.04	5.70	6.37	0.45	95.1
. 15%	PL	0.33	7.00	6.81	0.61	94.3	0.33	7.00	6.81	0.61	94.3
	DEL	5.04	4.70	5.46	0.60	78.4	9.89	4.90	5.61	1.43	52.5
	DEL-E	1.13	6.88	6.83	0.60	95.5	1.57	6.89	6.88	0.62	95.7
	ADEL	1.27	5.86	6.43	0.46	93.2	2.39	6.20	6.86	0.56	94.5
30%	PL	0.36	7.79	7.37	0.78	93.8	0.36	7.79	7.37	0.78	93.8
	DEL	5.19	5.23	5.93	0.70	79.2	10.10	5.46	6.10	1.57	56.0
	DEL-E	1.17	7.62	7.37	0.76	95.2	1.61	7.64	7.43	0.77	95.5
	ADEL	1.19	6.50	7.05	0.58	93.8	2.29	6.89	7.50	0.68	95.3

Table S3: Simulation results (all the entries are multiplied by 100) of estimated regression coefficients when the external aggregate survival rate $P(T > t^*|X_2 = 0)$ is comparable with the primary cohort and $P(T > t^*|X_2 = 1)$ is not, i.e., $(\tau_1, \tau_2) = (0, 0.06)$.

	β_1					β_2				
Method	Bias	SD	SE	RMSE	CP	Bias	SD	SE	RMSE	CP
					Sample Size=100					
PL	−1.27	11.09	11.56	11.15	95.4	1.93	21.69	21.25	21.76	95.4
DEL	−1.44	10.85	11.34	10.93	94.8	16.84	9.36	9.53	19.26	53.0
DEL-E	−1.94	11.03	11.50	11.19	94.8	17.40	9.38	9.61	19.76	50.6
ADEL	−1.22	11.04	11.54	11.10	95.4	1.79	18.88	19.88	18.93	92.6
					Sample Size=200					
PL	−0.18	8.15	8.00	8.15	95.4	0.44	15.17	14.78	15.16	95.2
DEL	−0.34	8.05	7.86	8.05	95.2	16.99	6.84	06.54	18.31	24.2
DEL-E	−0.79	8.16	7.96	8.19	94.4	17.49	6.89	06.59	18.80	23.0
ADEL	−0.11	8.14	7.99	8.13	95.6	0.58	14.45	14.86	14.45	93.0

Table S4: Simulation results (all the entries are multiplied by 100) of estimated regression coefficients by using SCAD and LASSO penalties.

	β_1				β_2			
Method	Bias	SD	SE	RMSE	Bias	SD	SE	RMSE
Setting 1: $(\tau_1, \tau_2) = (0.01, 0.01)$								
ADEL-SCAD	−1.32	10.20	9.87	10.27	0.24	19.09	19.67	19.07
ADEL-LASSO	−1.32	10.19	9.86	10.26	0.22	19.07	19.67	19.05
Setting 2: $(\tau_1, \tau_2) = (0.04, -0.04)$								
ADEL-SCAD	−2.80	10.92	11.04	11.26	1.17	19.29	20.21	19.30
ADEL-LASSO	−2.80	10.92	11.03	11.26	1.16	19.28	20.20	19.29
Setting 3: $(\tau_1, \tau_2) = (0.06, 0.06)$								
ADEL-SCAD	−1.68	10.79	10.82	10.91	−3.27	19.48	20.43	19.73
ADEL-LASSO	−1.69	10.80	10.82	10.92	−3.31	19.49	20.43	19.74
Setting 4: $(\tau_1, \tau_2) = (0.08, 0.08)$								
ADEL-SCAD	−2.18	10.92	11.49	11.13	−3.93	19.29	20.99	19.66
ADEL-LASSO	−2.19	10.93	11.49	11.14	−3.94	19.30	20.99	19.67