Statistics and Its Interface Volume 0 (2022) 1–4

Optimal subsampling proportional subdistribution hazards regression with rare events in big data*

Li Erqian, Tang Man-Lai Tian Maozai and Yu Keming†

The proportional subdistribution hazards (PSH) model has been widely employed for analyzing competing risks data which have mutually exclusive events with multiple causes and commonly occur in clinical research. With the rapid development of healthcare industry, massively sized survival data sets are becoming increasingly prevalent and classical PSH models are computationally intensive with large data sets. In this article, we propose the optimal subsampling estimators and two-step algorithm for the Fine-Gray model. Asymptotic properties of the proposed estimators are established and an extensive simulation study is conducted to demonstrate the efficiency of the estimators. Our proposed methodology is then illustrated with the large dataset from the SEER (Surveillance, Epidemiology, and End Results) database.

AMS 2000 subject classifications: Primary 62N01; secondary 62P10.
Keywords and phrases: Big data, Competing risks data, Optimal subsampling.

1. INTRODUCTION

Massively sized survival data sets are becoming increasingly prevailing with the rapid development of the healthcare industry [1]. For example, the world has recently experienced a record-breaking pressure placed on healthcare systems since the outbreak of COVID-19 in Wuhan at the end of 2019 [2]. As a result of the fast-growing demand for medical care in hospitals, with limited space and number of intensive care units, estimation of the length of stay of patients with COVID-19 in hospitals can provide insightful information to decision makers for efficient allocation of equipment and managing hospital overload in different countries. Competing risks models extend the classical survival setting to consider a collection of mutually exclusive potential event types, which is common in clinical trails [3]. The Fine-Gray subdistribution hazard model [4] has garnered its popularity to analyze data in the presence of competing risks [5, 6, 7]. Some popular computational approaches for estimating the Fine-Gray model include the weighted estimating equations [4] and pseudo-observations [8, 9]. However, there is evidence that both approaches could be computationally intensive and impractical even for data sets with moderate sample sizes [10, 11]. As a result, subsampling has been a popular technique to extract useful information from massive data.

Subsampling-based methods have recently been developed for various areas. For linear regression models, subsampling methods include subsampling the covariance matrix [12], leveraging methods [13] and information-based optimal subdata selection [14]. [15] and [16] extended subsampling methods to logistic regression analysis. [17] investigated the optimal subsampling method under the A- and L-optimality criteria for generalized linear models. [18, 19] developed efficient subsampling method for quantile regression. [20] relaxed the assumption of distribution and derived optimal Poisson subsampling probabilities in the context of quasi-likelihood estimation.

Recently, subsampling methods have been successfully extended for handling massive data in survival analysis, such as illness-death model under semi-competing risks settings [21, 22], additive hazards model [23] and Cox regression [1, 24]. To the best of our knowledge, subsampling methods for competing risks data have not been fully considered. The occurrence of massive data sets as well as relatively rare events of failures motivates us to propose subsampling methods based on Fine-Gray model. This paper extends the subsampling methods of [1] to the Fine-Gray model and derives explicit L- and A-optimal probabilities. Although our work is motivated by [1], we make the following contributions:

- We propose an explicit and efficient subsampling approach for rare-event competing risks data. The consistency and asymptotic normality are established for

arXiv: 0000.0000
*This research was funded by National Natural Science Funds of China (Grant No. 12101015), Scientific Research Foundation of North China University of Technology (No. 110051360002), the Fundamental Research Funds for Beijing Universities, NCUT (No.110052971921/007), National Natural Science Foundation of China (No. 11861042), and the China Statistical Research Project (No. 2020LZ25). M. L. TANG's research was partially supported by the Research Matching Grant (project: 700006 Applications of SAS Viya in Big Data Analytics) and the FDS research grant (UGC/FDS14/P05/20) from the Research Grants Council of Hong Kong Special Administrative Region. and by a grant from the National Natural Science Foundation of China (Grant 11871124). Also, this project is partially supported by the Big Data Intelligence Centre of the Hang Seng University of Hong Kong.
†Corresponding author.

regression coefficients and cumulative baseline hazard function based on the Fine-Gray model.

- According to our simulation study with large sample sizes (e.g., 10^6), we found that the existing R package `cmprsk` would encounter severe computational challenges. Our subsampling method however requires substantially less computing time with comparable root mean squared error (RMSE). Moreover, the higher the censoring rate, the less running time, indicating high efficiency of our proposed estimators.

The rest of this paper is organized as follows. Section 2 introduces the model setup and the subsampling methodology under Fine-Gray model. Section 3 presents the asymptotic properties of the coefficient estimates, the L- and A-optimal probabilities as well as a two-step algorithm. Section 4 illustrates the performance of our methodology through numerical simulation and presents a real application based on the publicly available data set from SEER. The results indicate that our subsampling methods are not only theoretically sound but also computationally highly efficient and fast.

2. METHODOLOGY

Assume that there exist K causes of failure, denoted by $\epsilon \in \{1, ..., K\}$. Without loss of generality, set $K = 2$. Let T and C denote the failure and censoring time, respectively. Denote $X = \min(T, C)$, and censoring indicator $\delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Denote $\nu \times 1$-dimensional bounded time-independent covariate vector as $\boldsymbol{Z}$. The observed data are then given as $\{X_i, \delta_i, \delta_i\epsilon_i, \boldsymbol{Z}_i; i = 1, ..., n\}$. Denote the number of censored observation as n_c. $n_e = n - n_c = \sum_{i=1}^n \delta_i$ is then the number of observed failures.

The cumulative incidence function (CIF) for failure from Cause 1 conditional on the covariate is $F_1(t|\boldsymbol{Z}) = P(T \leq t, \epsilon = 1|\boldsymbol{Z})$. The subdistribution hazard [25] is then defined as

$$\begin{aligned}\lambda_1(t|\boldsymbol{Z}) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\{t < T \leq t + \Delta t, \epsilon = 1 \\ &\quad \big|(T \geq t) \cup (T \leq t \cap \epsilon \neq 1), \boldsymbol{Z}\} \\ &= \frac{dF_1(t|\boldsymbol{Z})/dt}{1 - F_1(t|\boldsymbol{Z})}.\end{aligned}$$

Under the proportional hazards specification, assume $\lambda_1(t|\boldsymbol{Z}) = \lambda_{10}(t)\exp(\boldsymbol{Z}^\top\boldsymbol{\beta}^0)$, where $\lambda_{10}(\cdot)$ is an unspecified nonnegative function and $\Lambda_{10}(t)$ is the cumulative baseline hazard function $\int_0^t \lambda_{10}(u)du$. Let $\boldsymbol{\beta}^0$, λ_{10}^0 and Λ_{10}^0 be the true values of $\boldsymbol{\beta}$, λ_{10}, and Λ_{10}, respectively.

With inverse probability of censoring weighting (IPCW [26]) for right censored data, the log partial likelihood for PSH model is given by

$$\begin{aligned}\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty &\left[\boldsymbol{Z}_i^\top\boldsymbol{\beta} - \log\left\{\sum_j w_j(u)Y_j(u)\exp\{\boldsymbol{Z}_j^\top\boldsymbol{\beta}\}\right\}\right] \\ &\times w_i(u)dN_i(u),\end{aligned}$$

where $N_i(t) = I(T_i \leq t, \epsilon_i = 1), Y_i(t) = 1 - N_i(t-)$ [28], and $w_i(t) = I(C_i \geq T_i \wedge t)\hat{G}(t)/\hat{G}(X_i \wedge t)$ is the time-dependent weight with $G(t) = P(C \geq t)$ being the survival function of censoring variable C and $\hat{G}(t)$ being the Kaplan-Meier estimator of $G(t)$ [27]. For given t, if individual i failed due to right censoring or event of interest, $w_i(t)Y_i(t) = 0$; if individual i failed because of competing risks, $w_i(t)Y_i(t)$ lies between 0 and 1 and decreases with time; otherwise, $w_i(t)Y_i(t) = 1$.

Remark 2.1. *By using IPCW, we assume censoring variable C is independent of covariate $\boldsymbol{Z}$ for simplicity. As discussed in Section 4 of [4], IPCW can also be generalized for the case when C and $\boldsymbol{Z}$ are dependent.*

In what follows, we adopt some conventional notations to simplify our presentation for survival analysis. Let

$$\hat{S}_2^{(k)}(\boldsymbol{\beta}, t) = \frac{1}{n}\sum_i^n w_i(t)Y_i(t)\boldsymbol{Z}_i^{\otimes k}\exp\{\boldsymbol{Z}_i^\top\boldsymbol{\beta}\}, \tag{1}$$

where $\boldsymbol{Z}_i^{\otimes 0} = 1, \boldsymbol{Z}_i^{\otimes 1} = \boldsymbol{Z}_i, \boldsymbol{Z}_i^{\otimes 2} = \boldsymbol{Z}_i\boldsymbol{Z}_i^\top$. It is noted that $\int_0^\infty w_i(t)dN_i(t) = w_i(X_i)N_i(X_i) = I(T_i \leq X_i, \epsilon = 1)\frac{I(C_i \geq X_i)\hat{G}(X_i)}{\hat{G}(X_i)} = I(\delta_i\epsilon_i = 1)$, which leads to another useful expression of the log partial likelihood

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n I(\delta_i\epsilon_i = 1)\left[\boldsymbol{Z}_i^\top\boldsymbol{\beta} - \log\left\{\sum_{j \in R_i} w_j(X_i)\exp\{\boldsymbol{Z}_j^\top\boldsymbol{\beta}\}\right\}\right],$$

where $R_i = \{j : (X_j \geq X_i) \cup ((X_j \leq X_i) \cap (\delta_j = 1) \cap (\epsilon_j \neq 1))\}$. The score function is then given as

$$U(\boldsymbol{\beta}) = \frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^n I(\delta_i\epsilon_i = 1)\left[\boldsymbol{Z}_i^\top - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, X_i)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, X_i)}\right].$$

Let $\hat{\boldsymbol{\beta}}_{\text{full}}$ denote the estimator based on full sample and τ be the maximal follow-up time.

To conduct subsampling, denote $\boldsymbol{p} = (p_1, \cdots, p_{n_c})^\top$ as the vector of sampling probabilities for the censored observations, where $\sum_{i=1}^{n_c} p_i = 1$, and set

$$\pi_i = \begin{cases} \frac{1}{p_i q}, & \text{if } \delta_i = 0, p_i > 0 \\ 0, & \text{if } \delta_i = 0, p_i = 0 \\ 1, & \text{if } \delta_i = 1 \end{cases}$$

For the subsample, let

$$\hat{S}_{\pi 2}^{(k)}(\boldsymbol{\beta}, t) \quad = \quad \frac{1}{n}\sum_{i \in \mathcal{Q}} \pi_i w_i(t)Y_i(t)\boldsymbol{Z}_i^{\otimes k}\exp\{\boldsymbol{Z}_i^\top\boldsymbol{\beta}\},$$

and $\tilde{\boldsymbol{\beta}}$ be the estimator based on $\mathcal{Q}$, which is the set containing all observed failures and a subsample of size q drawn from censored observations. Thus, $\tilde{\boldsymbol{\beta}}$ solves the following estimating equation:

$$0 = U^*(\boldsymbol{\beta}) = \frac{\partial \ell^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i \in \mathcal{Q}} I(\delta_i \epsilon_i = 1) \left[\boldsymbol{Z}_i^\top - \frac{\hat{S}_{\pi 2}^{(1)}(\boldsymbol{\beta}, X_i)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, X_i)} \right]$$

where

$$\ell^*(\boldsymbol{\beta}) = \sum_{i \in \mathcal{Q}} I(\delta_i \epsilon_i = 1) \left[\boldsymbol{Z}_i^\top \boldsymbol{\beta} - \log\{n \hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, X_i)\} \right].$$

Finally, let h_i be the random variable which counts the number of times observation i drawn into the subsample $\mathcal{Q}$, $\boldsymbol{h} = (h_1, \cdots, h_n)^\top$. Conditional on the $\sigma-$algebra $\mathcal{D}_n = \mathfrak{F}\{X_i, \delta_i, \delta_i \epsilon_i, \boldsymbol{Z}_i; i = 1, \cdots, n\}$, we have

$$\begin{aligned} h_i &= 1, \text{if } \delta_i = 1, \\ \boldsymbol{h}_c | \mathcal{D}_n &\sim Multinomial(q, \boldsymbol{p}), \end{aligned}$$

where $\boldsymbol{h}_c$ is the $n_c \times 1$ subvector of $\boldsymbol{h}$, corresponding to the censored observations. Hence, we have $\sum_{i \in \mathcal{Q}} \pi_i w_i(t) Y_i(t) \boldsymbol{Z}_i^{\otimes k} \exp\{\boldsymbol{Z}_i^\top \boldsymbol{\beta}\} = \sum_{i=1}^n h_i \pi_i w_i(t) Y_i(t) \boldsymbol{Z}_i^{\otimes k} \exp\{\boldsymbol{Z}_i^\top \boldsymbol{\beta}\}$, and

$$\hat{\Lambda}_{10}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{1}{\hat{S}_2^{(0)}(\tilde{\boldsymbol{\beta}}, u)} w_i(u) dN_i(u)$$

is a modified version of Breslow's estimator [4], which converges uniformly in probability to the true baseline hazard of the subdistribution on the interval $[0,\tau)$ with τ being chosen chosen such that $P(X \geq \tau) > 0$.

3. ASYMPTOTIC PROPERTIES

In this section, we derive the consistency and asymptotic normality of the estimator from the subsampling algorithm to the full-data partial-likelihood estimator. The establishment of the asymptotic properties follows from [1].

We assume that $n_e = o(n)$ as $n \to \infty$, the number of sampled subsets from censored observations is q with $q/n \to 0$ as $n \to \infty$. Let $\mathcal{C}$ be the set of all censored observations in the full data, $\mathcal{E}$ the set of all observed failure observations, and $\mathcal{Q}$ the set consisting of $\mathcal{E}$ and all censored observations in the subsample. Hence, we have $|\mathcal{C}| = n_c, |\mathcal{E}| = n_e, |\mathcal{Q}| = n_e + q$.

Let

$$\begin{aligned} S_1^{(k)}(\boldsymbol{\beta}, t) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) \boldsymbol{Z}_i^{\otimes k} \exp\{\boldsymbol{Z}_i^\top \boldsymbol{\beta}\}, \\ \alpha^{(k)}(\boldsymbol{\beta}, t) &= \mathrm{E}[S_1^{(k)}(\boldsymbol{\beta}, t)]. \end{aligned}$$

To characterize asymptotic properties of subsample estimators, we require the following regularity assumptions from [1].

A.1 $\Lambda_{10}(\tau) = \int_0^\tau \lambda_{10}(t) dt < \infty$.

A.2 There exist the compact neighbourhood $\mathcal{B}$ of $\boldsymbol{\beta}^0$ and $\alpha^{(k)}$ defined above over $\mathcal{B} \times [0, \tau]$, satisfying that for $k = 0, 1, 2$

$$\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}} \|S_1^{(k)}(\boldsymbol{\beta}, t) - \alpha^{(k)}(\boldsymbol{\beta}, t)\|_2 \xrightarrow{P} 0.$$

A.3 There exist $C_1, C_2 > 0$, such that $\max_{ij} |z_{ij}| < C_1$ and $\max_i |\boldsymbol{Z}_i^\top \boldsymbol{\beta}^0| < C_2$.

A.4 For $\mathcal{B}, \alpha^{(0)}, \alpha^{(1)}, \alpha^{(2)}$ defined in (A.2) , let $e = \alpha^{(1)}/\alpha^{(0)}$ and $\mathrm{v} = \alpha^{(2)}/\alpha^{(0)} - e^{\otimes 2}$. thus, for all $\boldsymbol{\beta} \in \mathcal{B}, t \in [0, \tau]$

$$a^{(1)}(\boldsymbol{\beta}, t) = \frac{\partial \alpha^{(0)}(\boldsymbol{\beta}, t)}{\partial \boldsymbol{\beta}}, \quad \alpha^{(2)}(\boldsymbol{\beta}, t) = \frac{\partial^2 \alpha^{(0)}(\boldsymbol{\beta}, t)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top},$$

$\alpha^{(0)}(\cdot, t), \alpha^{(1)}(\cdot, t)$ and $\alpha^{(2)}(\cdot, t)$, $\boldsymbol{\beta} \in \mathcal{B}$ are uniformly continuous function on $t \in [0, \tau]$, $\alpha^{(0)}, \alpha^{(1)} and \alpha^{(2)}$ are bounded over $\mathcal{B} \times [0, \tau]$; and $\alpha^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$.

A.5 The matrix

$$\mathbf{A} = \int_0^\tau \mathrm{v}(\boldsymbol{\beta}^0, t) \alpha^{(0)}(\boldsymbol{\beta}^0, t) \lambda_{10}(t) dt$$

is positive definite.

A.6 (T, ϵ) is independent of C given $\boldsymbol{Z}$.

A.7 $\mathrm{E}\{w_i(\tau) Y_i(\tau)\} > 0$, for all $i = 1, \cdots, n$.

A.8 As $n \to \infty$, $p_i n$ is bounded away from 0 for all $i \in \mathcal{C}$.

A.9 The Hessian matrix $\partial^2 \ell(\boldsymbol{\beta})/(\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top)$ and $\partial^2 \ell^*(\boldsymbol{\beta})/(\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top)$ are non-singular with probability going to 1, as $n \to \infty$ and $q \to \infty$.

A.10 $n_e = o(n)$ as $n \to \infty$, and the number of sampled censored observations is substantially smaller than the full sample, i.e., $q = o(n)$.

A.1-A.7 are the standard assumptions for the consistency and asymptotic normality of competing risks regression in the unconditional space. A.8-A.10 are regularity conditions for subsampling procedure. A.8 guarantees that the sampling probabilities do not approach 0 too fast as the sample size increases. A.9 ensures that the subsample-based and full-sample-based information matrices are invertible with increasing sample size. A.10 assures that on the basis of rare events, the number of censored observations is downsampled to be substantially smaller than n_c.

Remark 3.1. *We assume A.10 to assure $q \ll n_c$ and the efficiency of subsampling methods. However, this assumption can be relaxed to $n_e = O(n)$. The convergence rate $O(q^{-1})$ and $O(q^{-1/2})$ in the asymptotic results will then be replaced by $O(n^{-1})$ and $O(n^{-1/2})$, respectively.*

Theorem 3.1. *Assume that A.1,A.3 and A.7-A.9 hold. Condition on $\mathcal{D}_n$, we have*

$$\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\|_2 = O_{P|\mathcal{D}_n}(q^{-1/2}) \tag{2}$$

and for each $t \in [0, \tau]$,

$$\hat{\Lambda}_{10}(t, \tilde{\boldsymbol{\beta}}) - \hat{\Lambda}_{10}(t, \hat{\boldsymbol{\beta}}_{\text{full}}) = O_{P|\mathcal{D}_n}(q^{-1/2}) \tag{3}$$

where $O_{P|\mathcal{D}_n}$ stands for "in probability" big-O notation in the conditional probability measure given $\mathcal{D}_n$, with $n \to \infty$ and $q \to \infty$.

The proof of Theorem 3.1 requires the following three lemmas.

Lemma 3.1. *Based on A.1,A.3 and A.8, and for a fixed vector of coefficients $\boldsymbol{\beta}$, it holds that*

$$\begin{aligned}(4)\qquad &\sup_{t\in[0,\tau]} |\hat{S}^{(k)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(k)}_{2}(\boldsymbol{\beta}, t)| \\ &= O_{P|\mathcal{D}_n}(q^{-1/2})\end{aligned}$$

for k = 0, 1, 2, where τ is the maximal follow-up time.

Proof of Lemma 3.1 Without loss of generality, it suffices to prove for the case when $k = 1$ since the proof is similar for $k = 0$ and 2. When $k = 2$, it can be done by treating a general element in the matrix. We can firstly rewrite Eq.(4) as follows

$$\begin{aligned}(5)\qquad &\hat{S}^{(1)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(1)}_{2}(\boldsymbol{\beta}, t) \\ &= \frac{1}{n}\sum_{i=1}^{n} w_i(t)Y_i(t)\exp\{\boldsymbol{Z}_i^\top \boldsymbol{\beta}\}\boldsymbol{Z}_i(\pi_i h_i - 1)\end{aligned}$$

The conditional expectation of the above expression is 0 since $\mathrm{E}(h_i|\mathcal{D}_n) = \pi_i^{-1}$ for all observations with a non-zero sampling probability, and $Y_i(t) = 0$ for all observations with a zero sampling probability [1]. Examining the conditional variance, with the characteristics of Multinomial distribution, we have

$$\begin{aligned}
&\mathrm{Var}\left\{\hat{S}^{(1)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(1)}_{2}(\boldsymbol{\beta}, t)\big|\mathcal{D}_n\right\} \\
&= \frac{1}{n^2}\Bigg\{\sum_{i\in\mathcal{C}} \exp\{2\boldsymbol{Z}_i^\top\boldsymbol{\beta}\}w_i(t)Y_i(t)\boldsymbol{Z}_i^{\otimes 2}\mathrm{Var}(\pi_i h_i|\mathcal{D}_n) \\
&\quad + \sum_{i,j\in\mathcal{C}, i\neq j} \exp\{(\boldsymbol{Z}_i + \boldsymbol{Z}_j)^\top\boldsymbol{\beta}\}w_i(t)w_j(t)Y_i(t)Y_j(t) \\
&\quad \times \boldsymbol{Z}_i^\top\boldsymbol{Z}_j \mathrm{Cov}(\pi_i h_i, \pi_j h_j|\mathcal{D}_n)\Bigg\} \\
&= \frac{1}{n^2}\Bigg\{\sum_{i\in\mathcal{C}} \exp\{2\boldsymbol{Z}_i^\top\boldsymbol{\beta}\}w_i(t)Y_i(t)\boldsymbol{Z}_i^{\otimes 2}\frac{qp_i(1-p_i)}{p_i^2q_i^2} \\
&\quad + \sum_{i,j\in\mathcal{C}, i\neq j} \exp\{(\boldsymbol{Z}_i + \boldsymbol{Z}_j)^\top\boldsymbol{\beta}\}w_i(t)w_j(t)Y_i(t)Y_j(t) \\
&\quad \times \boldsymbol{Z}_i^\top\boldsymbol{Z}_j \frac{-qp_ip_j}{q^2p_ip_j}\Bigg\} \\
&= \frac{1}{n^2}\Bigg\{\sum_{i\in\mathcal{C}} \frac{1}{p_i q}\exp\{2\boldsymbol{Z}_i^\top\boldsymbol{\beta}\}w_i(t)Y_i(t)\boldsymbol{Z}_i^{\otimes 2} \\
&\quad - \sum_{i,j\in\mathcal{C}} \frac{1}{q}\exp\{(\boldsymbol{Z}_i + \boldsymbol{Z}_j)^\top\boldsymbol{\beta}\}w_i(t)w_j(t)Y_i(t)Y_j(t)\boldsymbol{Z}_i^\top\boldsymbol{Z}_j\Bigg\} \\
&= O_{|\mathcal{D}_n}(q^{-1}),
\end{aligned}$$

where $O_{|\mathcal{D}_n}(q^{-1})$ denotes the standard big-O notation in the conditional space. The last equality can be derived from A.3 and A.8.

Since $\mathrm{Var}\left\{q^{\frac{1}{2}}\left(\hat{S}^{(1)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(1)}_{2}(\boldsymbol{\beta}, t)\right)|\mathcal{D}_n\right\} = O_{|\mathcal{D}_n}(1)$, by the definition of the order relation, there exist $M, q_0 > 0$ such that

$$\mathrm{Var}\left\{q^{\frac{1}{2}}\left(\hat{S}^{(1)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(1)}_{2}(\boldsymbol{\beta}, t)\right)|\mathcal{D}_n\right\} \leq M, \text{for all } q \geq q_0.$$

For any $\varepsilon > 0$, there exist $M_\varepsilon = (\frac{M}{\varepsilon})^{1/2}$ and q_0, by Chebyshev's inequality, such that for all $q > q_0$

$$\begin{aligned}
&P\left(\left|q^{\frac{1}{2}}\left(\hat{S}^{(1)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(1)}_{2}(\boldsymbol{\beta}, t)\right)\right| \geq M_\varepsilon|\mathcal{D}_n\right) \\
&\leq \frac{\mathrm{Var}\left\{q^{\frac{1}{2}}\left(\hat{S}^{(1)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(1)}_{2}(\boldsymbol{\beta}, t)\right)|\mathcal{D}_n\right\}}{\frac{M}{\epsilon}} \\
&\leq \varepsilon.
\end{aligned}$$

Thus, $\hat{S}^{(1)}_{\pi 2}(\boldsymbol{\beta}, t) - \hat{S}^{(1)}_{2}(\boldsymbol{\beta}, t) = O_{P|\mathcal{D}_n}(q^{-\frac{1}{2}})$. Finally, note that time t affects only the deterministic part in Eq.(5) and it is bounded, the result holds also for the supremum over t. Therefore, the proof of Lemma 3.1 is complete.

Lemma 3.2. *If A.1,A.3 and A.7-A.8 are satisfied, then conditioning on $\mathcal{D}_n$ we have*

$$\frac{1}{n}\frac{\partial \ell^*(\hat{\boldsymbol{\beta}}_{\text{full}})}{\partial \boldsymbol{\beta}} = O_{P|\mathcal{D}_n}(q^{-1/2}),$$

in an element-wise sense.

Lemma 3.2 shows that as q and n go to infinity, the subsample-based pseudo-score function approaches to 0 at the value $\hat{\boldsymbol{\beta}}_{\text{full}}$.

Proof of Lemma 3.2. Firstly, we have

(6)
$$\frac{1}{n}\frac{\partial \ell^*(\hat{\boldsymbol{\beta}}_{\text{full}})}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{\boldsymbol{Z}_i - \frac{\hat{S}^{(1)}_{\pi 2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)}{\hat{S}^{(0)}_{\pi 2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)}\right\} w_i(t)dN_i(t).$$

In the conditional space, $w_i(t)N_i(t)$ is deterministic for all t, and so is $\hat{\boldsymbol{\beta}}_{\text{full}}$. For the integrand of Eq.(6), it is a function of $\boldsymbol{x}_0 + \boldsymbol{\eta} = \left(\hat{S}^{(0)}_{\pi 2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t), \hat{S}^{(1)}_{\pi 2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)\right)^\top$ [29]. The first order Taylor expansion about $\boldsymbol{x}_0 = \left(\hat{S}^{(0)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t), \hat{S}^{(1)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)\right)^\top$ yields

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Bigg[\boldsymbol{Z}_i - \frac{\hat{S}^{(1)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)}{\hat{S}^{(0)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)} \\
&\quad + \frac{1}{\hat{S}^{(0)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)}\left(\hat{S}^{(1)}_{\pi 2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t) - \hat{S}^{(1)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)\right) \\
&\quad - \frac{\hat{S}^{(1)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)}{\hat{S}^{(0)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)^2}\left(\hat{S}^{(0)}_{\pi 2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t) - \hat{S}^{(0)}_{2}(\hat{\boldsymbol{\beta}}_{\text{full}}, t)\right) \\
&\quad + \boldsymbol{\xi}_t\Bigg] w_i(t)dN_i(t)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\frac{\partial \ell(\hat{\boldsymbol{\beta}}_{\text{full}})}{\partial \boldsymbol{\beta}} \\
&+ \frac{1}{n}\int_0^\tau \left[\frac{1}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)}\left(\hat{S}_{\pi 2}^{(1)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)-\hat{S}_2^{(1)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)\right)\right. \\
&\quad -\frac{\hat{S}_2^{(1)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)^2}\left(\hat{S}_{\pi 2}^{(0)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)-\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)\right) \\
&\quad \left. +\boldsymbol{\xi}_t\right] w_i(t)dN_i(t)
\end{aligned}
$$

where

$$
\xi_t = \boldsymbol{\eta}^\top \int_0^1\int_0^1 v\ddot{L}dudv\boldsymbol{\eta},
$$

$\ddot{L}$ is the second order derivative of the integrand of Eq.(6) with respect to $\boldsymbol{x}_0+\boldsymbol{\eta}$ taking value at $\boldsymbol{x}_0+uv\boldsymbol{\eta}$. By Lemma 3.1, $\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)$ and $\hat{S}_2^{(1)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)$ are conditionally bounded in probability due to the continuous mapping theorem, A.3, A.7 and $\boldsymbol{\xi}_t = O_{P|\mathcal{D}_n}(q^{-1})$. Therefore, based on Lemma 3.1, for $k=0,1$, we have

$$
\begin{aligned}
\frac{1}{n}\frac{\partial \ell^*(\hat{\boldsymbol{\beta}}_{\text{full}})}{\partial \boldsymbol{\beta}} &= O_{|\mathcal{D}_n}\left(\hat{S}_{\pi 2}^{(k)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)-\hat{S}_2^{(k)}(\hat{\boldsymbol{\beta}}_{\text{full}},t)\right) \\
&= O_{P|\mathcal{D}_n}(q^{-1/2}),
\end{aligned}
$$

Hence, Lemma 3.2 is proved.

For Lemma 3.3, we denote

$$
\begin{aligned}
\mathcal{I}(\boldsymbol{\beta}) &= \frac{1}{n}\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} \\
&= -\frac{1}{n}\sum_{i=1}^n\int_0^\tau \left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)} - \left(\frac{\hat{S}_2^{(1)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}\right\} w_i(t)dN_i(t), \\
\tilde{\mathcal{I}}(\boldsymbol{\beta}) &= \frac{1}{n}\frac{\partial^2 \ell^*(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} \\
&= -\frac{1}{n}\sum_{i=1}^n\int_0^\tau \left\{\frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)} - \left(\frac{\hat{S}_{\pi 2}^{(1)}(\boldsymbol{\beta},t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}\right\} w_i(t)dN_i(t),
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{S}_{2,k}^{(1)}(\boldsymbol{\beta},t) &= \frac{1}{n}\sum_{i=1}^n \exp(\boldsymbol{Z}_i^\top\boldsymbol{\beta})w_i(t)Y_i(t)Z_{ik}, \\
\hat{S}_{\pi 2,k}^{(1)}(\boldsymbol{\beta},t) &= \frac{1}{n}\sum_{i=1}^n \exp(\boldsymbol{Z}_i^\top\boldsymbol{\beta})w_i(t)Y_i(t)Z_{ik}\pi_i h_i,
\end{aligned}
$$

where Z_{ik} stands for the $k-$th element of the vector $\boldsymbol{Z}_i$.

Lemma 3.3. *If A.1,A.3 and A.7-A.8 are satisfied, then condition on $\mathcal{D}_n$, for a vector of fixed coefficients $\boldsymbol{\beta}$, we have*

$$
\tilde{\mathcal{I}}(\boldsymbol{\beta}) - \mathcal{I}(\boldsymbol{\beta}) = O_{P|\mathcal{D}_n}(q^{-1/2}), \tag{7}
$$

in the sense that it holds for each element in the matrix.

Lemma 3.3 shows that the subsample-based observed information matrix for $\boldsymbol{\beta}$ converges to the corresponding observed information matrix based on the full sample.

Proof of Lemma 3.3.

$$
\begin{aligned}
&\tilde{\mathcal{I}}(\boldsymbol{\beta}) - \mathcal{I}(\boldsymbol{\beta}) \\
&= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)} - \frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)}\right\} w_i(t)dN_i(t) \\
&- \frac{1}{n}\sum_{i=1}^n\int_0^\tau \left[\left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2} - \left\{\frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2}\right] \\
&\times w_i(t)dN_i(t).
\end{aligned}
$$

According to to Lemma 3.1, A.3 and A.7, the first addend can be shown to be $O_{P|\mathcal{D}_n}(q^{-1/2})$ by rewriting it as

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\int_0^\tau \left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)} - \frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)}\right\} w_i(t)dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)} - \frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right. \\
&\quad \left. +\frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)} - \frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)}\right\} w_i(t)dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)-\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right\} w_i(t)dN_i(t) \\
&\quad + \frac{1}{n}\sum_{i=1}^n\int_0^\tau \frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)\left\{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)-\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)\right\}}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)} \\
&\quad \times w_i(t)dN_i(t) \\
&= O_{P|\mathcal{D}_n}(q^{-1/2}).
\end{aligned}
$$

For the second addend, we can use similar arguments given in the proof of Lemma 3.2 for the general element of the matrix in the $m-$th row and l-th column with the Taylor expansion for $\hat{S}_{\pi 2,m}^{(1)}(\boldsymbol{\beta},t)\hat{S}_{\pi 2,l}^{(1)}(\boldsymbol{\beta},t)/\{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)\}^2$ with respect to $\boldsymbol{x_0}+\boldsymbol{\eta} = \left(\hat{S}_{\pi 2,m}^{(1)}(\boldsymbol{\beta},t), \hat{S}_{\pi 2,l}^{(1)}(\boldsymbol{\beta},t), \hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)\right)^\top$ at $\boldsymbol{x}_0 = \left(\hat{S}_{2,m}^{(1)}(\boldsymbol{\beta},t), \hat{S}_{2,l}^{(1)}(\boldsymbol{\beta},t), \hat{S}_2^{(0)}(\boldsymbol{\beta},t)\right)^\top$. As a result,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\int_0^\tau \left[\left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2}_{ml} - \left\{\frac{\hat{S}_{\pi 2}^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2}_{ml}\right] w_i(t)dN_i(t) \\
&= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \left[\left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2}_{ml} - \left\{\frac{\hat{S}_2^{(2)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right\}^{\otimes 2}_{ml}\right. \\
&\quad + \frac{\hat{S}_{2,l}^{(1)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)^2}\left(\hat{S}_{\pi 2,m}^{(1)}(\boldsymbol{\beta},t) - \hat{S}_{2,m}^{(1)}(\boldsymbol{\beta})\right) \\
&\quad + \frac{\hat{S}_{2,m}^{(1)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)^2}\left(\hat{S}_{\pi 2,l}^{(1)}(\boldsymbol{\beta},t) - \hat{S}_{2,l}^{(1)}(\boldsymbol{\beta},t)\right)
\end{aligned}
$$

$$-\frac{2\hat{S}^{(1)}_{2,l}(\boldsymbol{\beta},t)\hat{S}^{(1)}_{2,m}(\boldsymbol{\beta},t)}{\hat{S}^{(0)}_{2}(\boldsymbol{\beta},t)^3}\left(\hat{S}^{(0)}_{\pi 2}(\boldsymbol{\beta},t)-\hat{S}^{(0)}_{2}(\boldsymbol{\beta},t)\right)+\boldsymbol{\xi}_t\Bigg]$$
$$\times\, w_i(t)dN_i(t),$$

where

$$\xi_t=\boldsymbol{\eta}^\top\int_0^1\int_0^1 v\ddot{S}dudv\boldsymbol{\eta},$$

$\ddot{S}$ is the second order derivative of $\hat{S}^{(1)}_{\pi 2,m}(\boldsymbol{\beta},t)\hat{S}^{(1)}_{\pi 2,l}(\boldsymbol{\beta},t)/\{\hat{S}^{(0)}_{\pi 2}(\boldsymbol{\beta},t)\}^2$ with respect to $\boldsymbol{x}_0+\boldsymbol{\eta}$ taking value at $\boldsymbol{x}_0+uv\boldsymbol{\eta}$. Based on the continuous mapping theorem, A.3, A.7 and Lemma 3.1, we have $\boldsymbol{\xi}_t=O_{P|\mathcal{D}_n}(q^{-1})$, and Eq.(7) follows.

Proof of Theorem 3.1, Eq.(2). Firstly, we introduce the following notations

$$S^{(1)}_{\pi 2,kl}(\boldsymbol{\beta},t)=\frac{1}{n}\sum_{i=1}^n\pi_i\exp(\boldsymbol{Z}_i^\top\boldsymbol{\beta})w_i(t)Y_i(t)h_iZ_{ik}Z_{il},$$
$$S^{(1)}_{\pi 2,klm}(\boldsymbol{\beta},t)=\frac{1}{n}\sum_{i=1}^n\pi_i\exp(\boldsymbol{Z}_i^\top\boldsymbol{\beta})w_i(t)Y_i(t)h_iZ_{ik}Z_{il}Z_{im}.$$

Let $\ell_k^{*'}(\boldsymbol{\beta})$ be the derivative of $\ell^*(\boldsymbol{\beta})$ with respect to β_k and $\tilde{\mathcal{I}}_k$ be the $k-$th row of the matrix $\tilde{\mathcal{I}}$. The first order Taylor expansion for $n^{-1}\ell_k^{*'}(\tilde{\boldsymbol{\beta}})$ about $\hat{\boldsymbol{\beta}}_{\text{full}}$ is given as

(8)
$$0=\frac{1}{n}\ell_k^{*'}(\tilde{\boldsymbol{\beta}})=\frac{1}{n}\ell_k^{*'}(\hat{\boldsymbol{\beta}}_{\text{full}})+\tilde{\mathcal{I}}_k^\top(\hat{\boldsymbol{\beta}}_{\text{full}})(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}})+\frac{1}{n}\text{Res}_k,$$

where

$$\frac{1}{n}\text{Res}_k=(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}})^\top\int_0^1\int_0^1 v\frac{\partial\tilde{\mathcal{I}}_k}{\partial\boldsymbol{\beta}}\bigg|_{\hat{\boldsymbol{\beta}}_{\text{full}}+uv(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}})}dudv$$
$$\cdot(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}).$$

For any general element, say (l,m), of $\partial\tilde{\mathcal{I}}_k(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$, we get

$$\begin{aligned}
&\frac{1}{n}\frac{\partial^2\ell_k^{*'}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_l\boldsymbol{\partial}\beta_m}\\
&=\frac{1}{n}\sum_{i=1}^n\int_0^\tau\frac{1}{S^{(0)3}_{\pi 2}(\boldsymbol{\beta},t)}\Big[-S^{(1)}_{\pi 2,klm}(\boldsymbol{\beta},t)S^{(0)2}_{\pi 2}(\boldsymbol{\beta},t)\\
&\quad+\Big(S^{(1)}_{\pi 2,kl}(\boldsymbol{\beta},t)S^{(1)}_{\pi 2,m}(\boldsymbol{\beta},t)+S^{(1)}_{\pi 2,km}(\boldsymbol{\beta},t)S^{(1)}_{\pi 2,l}(\boldsymbol{\beta},t)\\
&\quad+S^{(1)}_{\pi 2,lm}(\boldsymbol{\beta},t)S^{(1)}_{\pi 2,k}(\boldsymbol{\beta},t)\Big)S^{(0)}_{\pi 2}(\boldsymbol{\beta},t)\\
&\quad-2S^{(1)}_{\pi 2,k}(\boldsymbol{\beta},t)S^{(1)}_{\pi 2,l}(\boldsymbol{\beta},t)S^{(1)}_{\pi 2,m}(\boldsymbol{\beta},t)\Big]w_i(t)dN_i(t).
\end{aligned}$$

According to Lemma 3.1, the continuous mapping theorem and A.7, the subsample-based pseudo-score function $n^{-1}\ell^*(\boldsymbol{\beta})/\boldsymbol{\beta}$ converges in conditional probability uniformly to the full sample score function $n^{-1}\partial\ell(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$. Since $\mathcal{B}$ is compact, by the Appendix A of [4],

$$\hat{\boldsymbol{\beta}}_{\text{full}}-\boldsymbol{\beta}^0=o_p(1),$$

and the function $\partial\ell^*(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ converges in probability to a continuous and deterministic function of $\boldsymbol{\beta}$, denoted as $U_0(\boldsymbol{\beta})$, which has a unique 0 at $\boldsymbol{\beta}^0$ uniformly for $\boldsymbol{\beta}\in\mathcal{B}$. By the uniform convergence in conditional probability of $n^{-1}\ell^*(\boldsymbol{\beta})/\boldsymbol{\beta}$ to $U_0(\boldsymbol{\beta})$, and Theorem 5.9 of [30], we have

$$\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^0=o_{P|\mathcal{D}_n}(1).$$

Consequently,

(9)
$$\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}=o_{P|\mathcal{D}_n}(1).$$

Therefore, $\tilde{\boldsymbol{\beta}}$ converges consistently to $\hat{\boldsymbol{\beta}}_{\text{full}}$ in the conditional space.

For the rate of convergence, due to continuity, and $\boldsymbol{\eta}=\hat{\boldsymbol{\beta}}_{\text{full}}+uv(\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}})$ is on the line segment between $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\text{full}}$ for $u,v\in[0,1]$, it follows that $\boldsymbol{\eta}-\hat{\boldsymbol{\beta}}_{\text{full}}=o_{P|\mathcal{D}_n}(1)$. Hence, $\boldsymbol{\eta}=O_{P|\mathcal{D}_n}(1)$. Combining A.3 and A.8, we can verify that $\hat{S}^{(k)}_{\pi 2}(\boldsymbol{\eta},t), k=0,1,2,t\in[0,\tau]$ are all bounded in conditional probability. In particular, we can write

$$\hat{S}^{(0)}_{\pi 2}(\boldsymbol{\eta},t)=\frac{1}{q}\sum_{i=1}^q\frac{\exp\{\boldsymbol{Z}_i^{*\top}\boldsymbol{\eta}\}w_i^*(t)Y_i^*(t)}{np_i^*}$$
$$+\frac{1}{n}\sum_{i\in\mathcal{E}}\exp\{\boldsymbol{Z}_i^\top\boldsymbol{\eta}\}w_i(t)Y_i(t),$$

which is conditionally bounded away from 0, where "*" denotes sampling with replacement. Similar results hold for $\hat{S}^{(k)}_{\pi 2}(\boldsymbol{\eta},t), k=1,2, \hat{S}^{(1)}_{\pi 2,kl}(\boldsymbol{\eta},t)$ and $\hat{S}^{(1)}_{\pi 2,klm}(\boldsymbol{\eta},t)$, for all $t\in[0,\tau]$. Therefore, based on A.7, we know that $\partial\tilde{\mathcal{I}}_k(\boldsymbol{\eta})/\partial\boldsymbol{\beta}=O_{P|\mathcal{D}_n}(1)$ for all k,

(10)
$$\frac{1}{n}\text{Res}_k=O_{P|\mathcal{D}_n}\left(\|\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}\|_2^2\right).$$

From Eq.(8)-(10) and A.9, we have

(11)
$$\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}=-\tilde{\mathcal{I}}^{-1}(\hat{\boldsymbol{\beta}}_{\text{full}})\left\{\frac{1}{n}\frac{\partial\ell^*(\hat{\boldsymbol{\beta}}_{\text{full}})}{\partial\boldsymbol{\beta}}+O_{P|\mathcal{D}_n}\left(\|\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}\|_2^2\right)\right\}.$$

Since matrix inversion is a continuous operation, according to Lemma 3.3 and the continuous mapping theorem, $\tilde{\mathcal{I}}^{-1}(\hat{\boldsymbol{\beta}}_{\text{full}})-\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}_{\text{full}})=o_{P|\mathcal{D}_n}(1)$, yielding $\tilde{\mathcal{I}}^{-1}(\hat{\boldsymbol{\beta}}_{\text{full}})=O_{P|\mathcal{D}_n}(1)$. Therefore, combining Lemma 3.1-3.2 with Eq.(9)-(11), we have

$$\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}=O_{P|\mathcal{D}_n}(q^{-1/2})+o_{P|\mathcal{D}_n}\left(\|\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}\|_2\right).$$

Hence,

(12)
$$\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{\text{full}}=O_{P|\mathcal{D}_n}(q^{-1/2}).$$

The proof of Eq.(2) in Theorem 3.1 is completed.

Proof of Theorem 3.1 Eq.(3). Write

$$\hat{\Lambda}_{10}(t,\tilde{\boldsymbol{\beta}})-\hat{\Lambda}_{10}(t,\hat{\boldsymbol{\beta}}_{\text{full}})$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \left\{\frac{1}{\hat{S}_2^{(0)}(\tilde{\boldsymbol{\beta}},u)} - \frac{1}{\hat{S}_2^{(0)}(\hat{\boldsymbol{\beta}}_{\text{full}},u)}\right\} w_i(u)dN_i(u).$$

A Taylor expansion about $\hat{\boldsymbol{\beta}}_{\text{full}}$ yields

$$\begin{aligned}(13)\qquad &\hat{\Lambda}_{10}(t,\tilde{\boldsymbol{\beta}}) - \hat{\Lambda}_{10}(t,\hat{\boldsymbol{\beta}}_{\text{full}})\\ &= (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}})\frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\hat{S}_2^{(1)}(\boldsymbol{\xi},u)}{\hat{S}_2^{(0)2}(\boldsymbol{\xi},u)} w_i dN_i(u),\end{aligned}$$

where $\boldsymbol{\xi}$ is on the line segment between $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{\text{full}}$. Based on A.3, A.7 and Eq.(12), Eq.(13) is a product of a term of $O_{P|\mathcal{D}_n}(q^{-1/2})$ and a term bounded in conditional probability, thus proving Eq.(3) and completing Theorem 3.1.

Theorem 3.2 below is about the asymptotic properties of the subsample-based estimators in the unconditional space.

Theorem 3.2. *Given that A.1-A.10 hold, as $q \to \infty$ and $n \to \infty$,*

$$(14)\qquad \sqrt{q}\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\boldsymbol{p},\boldsymbol{\beta}^0)(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{D} N(0,\mathbf{I}),$$

and for all $t \in [0,\tau]$,

$$(15)\qquad \sqrt{q}\mathbb{V}_{\hat{\Lambda}_{10}(t,\tilde{\boldsymbol{\beta}})}^{-1/2}(\boldsymbol{p},\boldsymbol{\beta}^0,t)(\hat{\Lambda}_{10}(t,\tilde{\boldsymbol{\beta}}) - \Lambda_{10}^0(t)) \xrightarrow{D} N(0,1),$$

where

$$\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{p},\boldsymbol{\beta}) = qn^{-1}\mathcal{I}^{-1}(\boldsymbol{\beta})\boldsymbol{\Sigma}\mathcal{I}^{-1}(\boldsymbol{\beta}) + \mathcal{I}^{-1}(\boldsymbol{\beta})\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta})\mathcal{I}^{-1}(\boldsymbol{\beta}),$$

$$\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}) = \frac{1}{n^2}\left\{\sum_{i\in\mathcal{C}} \frac{\mathbf{a}_i(\boldsymbol{\beta})\mathbf{a}_i(\boldsymbol{\beta})^\top}{p_i} - \sum_{i,j\in\mathcal{C}} \mathbf{a}_i(\boldsymbol{\beta})\mathbf{a}_j(\boldsymbol{\beta})^\top\right\},$$

$$\begin{aligned}\mathbf{a}_i(\boldsymbol{\beta}) = \sum_{j=1}^{n}\int_0^\tau &\left\{\boldsymbol{Z}_i - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta},t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\right\} \frac{w_i(t)Y_i(t)\exp(\boldsymbol{Z}_i^\top\boldsymbol{\beta})}{\hat{S}_2^{(0)}(\boldsymbol{\beta},t)}\\ &\times w_j(t)dN_j(t)\end{aligned}$$

$$\begin{aligned}\mathbb{V}_{\hat{\Lambda}_{10}(t,\tilde{\boldsymbol{\beta}})}(\boldsymbol{p},\boldsymbol{\beta},t) &= \frac{q}{n}\sum_{i=1}^{n}\int_0^t \frac{w_i(u)dN_i(u)}{n^{-1}\hat{S}_2^{(0)2}(\boldsymbol{\beta},u)}\\ &+ \mathbf{H}^\top(\boldsymbol{\beta},t)\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{p},\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta},t)\end{aligned}$$

$$\mathbf{H}(\boldsymbol{\beta},t) = \sum_{i=1}^{n}\int_0^t \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta},u)}{\hat{S}_2^{(0)2}(\boldsymbol{\beta},u)} w_i(u)dN_i(u).$$

Here, $\mathbf{I}$ is the $\nu \times \nu$ identity matrix and $\boldsymbol{\Sigma}$ is the variance-covariance matrix of $n^{-1/2}U(\boldsymbol{\beta}^0)$ defined as in [4].

Remark 3.2. *The rate of convergence of subsampling-based estimator is theoretically slightly slower than that of the full-data parametric version. However, in our simulation studies with large sample sizes, we find that the RMSEs of subsampling-based estimators are very close to those of the full-data estimator. Besides, if we relax the assumption $n_e = o(n)$, $\sqrt{q}$ would become $\sqrt{n}$.*

Lemma 3.4 establishes the consistency of $\tilde{\boldsymbol{\beta}}$.

Lemma 3.4. *If A.3-A.9 hold, then as $q \to \infty$ and $n \to \infty$, $\forall\varepsilon > 0$, we have*

$$(16)\qquad \lim_{n,q\to\infty} \Pr(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2 > \varepsilon) = 0.$$

Proof of Lemma 3.4. In Theorem 3.1, it is established that for all $\varepsilon > 0$,

$$\lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\| \geq \varepsilon|\mathcal{D}_n\right) = 0.$$

In the unconditional probability space, $\Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\| \geq \varepsilon|\mathcal{D}_n\right)$ is itself a random variable, denoted as $\pi_{n,q}$. It then follows that

$$\Pr\left(\lim_{n,q\to\infty} \pi_{n,q} = 0\right) = 1,$$

meaning that $\pi_{n,q} \xrightarrow[n,q\to\infty]{a.s.} 0$. For all $\varepsilon > 0$,

$$\begin{aligned}(17)\qquad &\lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\|_2 > \varepsilon\right)\\ &= \lim_{n,q\to\infty} E(\pi_{n,q}) = E(\lim_{n,q\to\infty} \pi_{n,q}) = 0,\end{aligned}$$

where the interchange of expectation and limit is allowed due to the dominated convergence theorem (as $\pi_{n,q}$ is trivially bounded by 1). Hence,

$$\begin{aligned}&\Pr\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2 \geq \varepsilon\right)\\ &= \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}} + \hat{\boldsymbol{\beta}}_{\text{full}} - \boldsymbol{\beta}^0\|_2 \geq \varepsilon\right)\\ &\leq \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\|_2 + \|\hat{\boldsymbol{\beta}}_{\text{full}} - \boldsymbol{\beta}^0\|_2 \geq \varepsilon\right)\\ &\leq \Pr\left(\{\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\|_2 \geq \varepsilon/2\} \cup \{\|\hat{\boldsymbol{\beta}}_{\text{full}} - \boldsymbol{\beta}^0\|_2 \geq \varepsilon/2\}\right)\\ &\leq \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\|_2 \geq \varepsilon/2\right) + \Pr\left(\|\hat{\boldsymbol{\beta}}_{\text{full}} - \boldsymbol{\beta}^0\|_2 \geq \varepsilon/2\right).\end{aligned}$$

Taking limits on both sides, we have

$$\begin{aligned}&\lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2 \geq \varepsilon\right)\\ &\quad\leq \lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\text{full}}\|_2 \geq \varepsilon/2\right)\\ &\quad\quad+ \lim_{n,q\to\infty} \Pr\left(\|\hat{\boldsymbol{\beta}}_{\text{full}} - \boldsymbol{\beta}^0\|_2 \geq \varepsilon/2\right) = 0.\end{aligned}$$

Here, the first addend on the right hand side tends to 0 due to Eq.(17) while the second addend becomes 0 based on the results for competing risks in [4] and A.3-A.7. As a result, Eq.(16) is proved.

Proof of Theorem 3.2. Similar to Eq.(11), and based on Lemma 3.4, a Taylor expansion for the subsample-based pseudo-score function evaluated at $\tilde{\boldsymbol{\beta}}$ about $\boldsymbol{\beta}^0$ yields

$$(18)\quad \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = -\tilde{\mathcal{I}}^{-1}(\boldsymbol{\beta}^0)\left\{\frac{1}{n}\frac{\partial\ell^*(\boldsymbol{\beta}^0)}{\partial\boldsymbol{\beta}} + o_P\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2\right)\right\}.$$

Similar to [31], the subsample-based pseudo-score function can be decomposed into two separate components, i.e.,

$$(19)\qquad \frac{1}{n}\frac{\partial\ell^*(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \frac{1}{n}\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} + \frac{1}{n}\sum_{i\in\mathcal{C}}(1 - \pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}),$$

where

$$\tilde{\mathbf{a}}_i(\boldsymbol{\beta}) = \sum_{j=1}^{n} \int_0^{\tau} \left\{ \boldsymbol{Z}_i - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)} \right\} \times \frac{\exp(\boldsymbol{Z}_i^\top \boldsymbol{\beta}) Y_i(t) w_i(t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} w_j(t) dN_j(t).$$

To verify Eq. (19), we notice that

$$\begin{aligned}
&\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}) = \sum_{i=1}^{n}(1-\pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}) \\
&= \sum_{i=1}^{n}(1-\pi_i h_i)\sum_{j=1}^{n}\int_0^{\tau}\left\{ \boldsymbol{Z}_i - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)} \right\} \\
&\quad \times \frac{\exp(\boldsymbol{Z}_i^\top \boldsymbol{\beta}) Y_i(t) w_i(t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} w_j(t) dN_j(t). \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^{\tau}\left\{ \boldsymbol{Z}_i - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)} \right\} \\
&\quad \times \frac{\exp(\boldsymbol{Z}_i^\top \boldsymbol{\beta}) Y_i(t) w_i(t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} w_j(t) dN_j(t) \\
&- \sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^{\tau}\left\{ \boldsymbol{Z}_i - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)} \right\} \\
&\quad \times \frac{\exp(\boldsymbol{Z}_i^\top \boldsymbol{\beta}) Y_i(t) w_i(t)\pi_i h_i}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} w_j(t) dN_j(t) \\
&= \sum_{j=1}^{n}\int_0^{\tau}\left\{ \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)}\frac{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} \right\} w_j(t) dN_j(t) \\
&- \sum_{j=1}^{n}\int_0^{\tau}\left\{ \frac{\hat{S}_{\pi 2}^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} - \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)}\frac{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} \right\} w_j(t) dN_j(t) \\
&= \sum_{j=1}^{n}\int_0^{\tau}\left\{ \frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)} - \frac{\hat{S}_{\pi 2}^{(1)}(\boldsymbol{\beta}, t)}{\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} \right\} w_j(t) dN_j(t) \\
&= \frac{\partial \ell^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}.
\end{aligned}$$

Based on Eqs.(18)-(19), we can write

$$\begin{aligned}
(20)\quad &\sqrt{q}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \\
&= -\tilde{\mathcal{I}}^{-1}(\boldsymbol{\beta}^0)\left\{ \frac{\sqrt{q}}{n}\frac{\partial \ell(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} + \frac{\sqrt{q}}{n}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}^0) \right. \\
&\quad \left. + o_P(\sqrt{q}\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_2) \right\}.
\end{aligned}$$

Based on the asymptotic results in [4] and A.3-A.7, we have

$$n^{-1/2}\frac{\partial \ell(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} \xrightarrow{D} N(0, \boldsymbol{\Sigma}).$$

where $\boldsymbol{\Sigma}$ is defined as in [4]. By continuous mapping theorem and multivariate Slutsky theorem [32], $\mathcal{I}(\boldsymbol{\beta}^0)^{-1} \to \boldsymbol{\Omega}^{-1}$, and

$$\left\{-\mathcal{I}(\boldsymbol{\beta}^0)\right\}^{-1/2} n^{-1/2}\frac{\partial \ell(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} \xrightarrow{D} N(0, \mathcal{I}^{-1/2}(\boldsymbol{\beta}^0)\boldsymbol{\Sigma}\mathcal{I}^{-1/2}(\boldsymbol{\beta}^0)).$$

Based on Lemma 3.3 and the dominated convergence theorem, we can show that $\tilde{\mathcal{I}}(\boldsymbol{\beta}^0)$ is consistent to $\mathcal{I}(\boldsymbol{\beta}^0)$. Based on A.9 and similar to the proof of Lemma 3.4, we have

$$(21)\quad \left\{-\tilde{\mathcal{I}}(\boldsymbol{\beta}^0)\right\}^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial \ell(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}} \xrightarrow{D} N(0, \mathcal{I}^{-1/2}(\boldsymbol{\beta}^0)\boldsymbol{\Sigma}\mathcal{I}^{-1/2}(\boldsymbol{\beta}^0)).$$

If this term is multiplied by a factor of $q^{1/2}n^{-1/2}\{-\tilde{\mathcal{I}}(\boldsymbol{\beta}^0)\}^{-1/2}$, as in Eq. (20), then according to Slutsky's theorem and theorem 5.13 of [32], the first addend has asymptotic variance $q/n\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\boldsymbol{\Sigma}\mathcal{I}^{-1}(\boldsymbol{\beta}^0)$ which will vanish to 0.

Next, we will show that the second addend is asymptotically normally distributed. Firstly, we have

$$\begin{aligned}
(22)\quad &\frac{\sqrt{q}}{n}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}) \\
&= \frac{\sqrt{q}}{n}\sum_{i=1}^{n}\int_0^{\tau}\frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t) - \hat{S}_{\pi 2}^{(1)}(\boldsymbol{\beta}, t)\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} \\
&\quad \times w_i(t) dN_i(t),
\end{aligned}$$

and

$$\begin{aligned}
(23)\quad &\frac{\sqrt{q}}{n}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\mathbf{a}_i(\boldsymbol{\beta}) \\
&= \frac{\sqrt{q}}{n}\sum_{i=1}^{n}\int_0^{\tau}\frac{\hat{S}_2^{(1)}(\boldsymbol{\beta}, t)\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t) - \hat{S}_{\pi 2}^{(1)}(\boldsymbol{\beta}, t)\hat{S}_2^{(0)}(\boldsymbol{\beta}, t)}{\hat{S}_2^{(0)2}(\boldsymbol{\beta}, t)} \\
&\quad \times w_i(t) dN_i(t),
\end{aligned}$$

By Lemma 3.1 and continuous mapping theorem, Eq.(22) and Eq.(23) share the same asymptotic distribution since we have

$$\begin{aligned}
&\frac{\sqrt{q}}{n}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}) - \frac{\sqrt{q}}{n}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\mathbf{a}_i(\boldsymbol{\beta}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\frac{\sqrt{q}(\hat{S}_2^{(0)}(\boldsymbol{\beta}, t) - \hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t))}{\hat{S}_2^{(0)2}(\boldsymbol{\beta}, t)\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t)} \\
&\quad \times \left\{ \hat{S}_2^{(1)}(\boldsymbol{\beta}, t)\left[\hat{S}_{\pi 2}^{(0)}(\boldsymbol{\beta}, t) - \hat{S}_2^{(0)}(\boldsymbol{\beta}, t)\right] \right. \\
&\quad \left. + \hat{S}_2^{(0)}(\boldsymbol{\beta}, t)\left[\hat{S}_2^{(1)}(\boldsymbol{\beta}, t) - \hat{S}_{\pi 2}^{(1)}(\boldsymbol{\beta}, t)\right] \right\} w_i(t) dN_i(t) \\
&= O_{P|\mathcal{D}_n}(q^{-1/2}),
\end{aligned}$$

by assumption A.3 and A.7 and Lemma 3.1.

We then establish the asymptotic normality of Eq.(23) and identify its covariance matrix since Eq.(22) and Eq.(23)

are asymptotically equivalent. It should be noticed $\mathbf{a}_i(\boldsymbol{\beta}^0)$ is constant conditioning on $\mathcal{D}_n$. Hence, $\frac{\sqrt{q}}{n}\sum_{i\in\mathcal{C}}\pi_i h_i \mathbf{a}_i(\boldsymbol{\beta}^0)$ can be alternatively expressed as a sum of q iid observations in the conditional space as follows

$$\frac{\sqrt{q}}{n}\sum_{i=1}^{q}\pi_i^*\mathbf{a}_i^*(\boldsymbol{\beta}^0) = \frac{1}{\sqrt{q}}\sum_{i=1}^{q}\frac{\mathbf{a}_i^*(\boldsymbol{\beta}^0)}{np_i^*} \equiv \frac{1}{\sqrt{q}}\sum_{i=1}^{q}\boldsymbol{\gamma}_i(\boldsymbol{p},\boldsymbol{\beta}^0).$$

Since the distribution of $\boldsymbol{\gamma}_i(\boldsymbol{p},\boldsymbol{\beta}^0)$ changes as n and q increase, the Lindeberg-Feller condition is established ([30], proposition 2.27) as it covers the settings of triangular arrays. Denote $\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}) \equiv \mathrm{Var}(\boldsymbol{\gamma}(\boldsymbol{p},\boldsymbol{\beta})|\mathcal{D}_n)$. We have

$$\begin{aligned}\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}^0) &= E\left(\boldsymbol{\gamma}(\boldsymbol{p},\boldsymbol{\beta}^0)\boldsymbol{\gamma}^\top(\boldsymbol{p},\boldsymbol{\beta}^0)|\mathcal{D}_n\right) \\ &\quad - \mathrm{E}\left(\boldsymbol{\gamma}(\boldsymbol{p},\boldsymbol{\beta}^0)|\mathcal{D}_n\right)\mathrm{E}\left(\boldsymbol{\gamma}(\boldsymbol{p},\boldsymbol{\beta})|\mathcal{D}_n\right)^\top \\ &= \frac{1}{n^2}\left\{\sum_{i\in\mathcal{C}}\frac{\mathbf{a}_i(\boldsymbol{\beta}^0)\mathbf{a}_i(\boldsymbol{\beta}^0)^\top}{p_i}\right. \\ &\quad \left. - \sum_{i,j\in\mathcal{C}}\mathbf{a}_i(\boldsymbol{\beta}^0)\mathbf{a}_j(\boldsymbol{\beta}^0)^\top\right\} = O_{|\mathcal{D}_n}(1).\end{aligned}$$

For every $\varepsilon > 0$ and for some $\delta > 0$, by Markov inequality [32], we have

$$\begin{aligned}&\sum_{i=1}^{q}\mathrm{E}\{\|q^{-1/2}\boldsymbol{\gamma}_i(\boldsymbol{p},\boldsymbol{\beta}^0)\|_2^2 I(\|q^{-1/2}\boldsymbol{\gamma}_i(\boldsymbol{p},\boldsymbol{\beta}^0)\|_2 > \varepsilon)|\mathcal{D}_n\} \\ &\le \frac{1}{q^{1+\delta/2}\varepsilon^\delta}\sum_{i=1}^{q}\mathrm{E}\left\{\left\|\frac{\mathbf{a}_i^*(\boldsymbol{\beta}^0)}{np_i^*}\right\|_2^{2+\delta}|\mathcal{D}_n\right\} \\ &= \frac{1}{q^{\delta/2}\varepsilon^\delta n^{2+\delta}}\sum_{i\in\mathcal{C}}\frac{\|\mathbf{a}_i(\boldsymbol{\beta}^0)\|_2^{2+\delta}}{p_i^{1+\delta}} \\ &= O_{|\mathcal{D}_n}(q^{-\delta/2}) = o_{|\mathcal{D}_n}(1).\end{aligned}$$

Hence, $\sum_{i=1}^{n}(q^{-1/2}\boldsymbol{\gamma}_i(\boldsymbol{p},\boldsymbol{\beta}^0) - \mathrm{E}[q^{-1/2}\gamma_i(\boldsymbol{p},\boldsymbol{\beta}^0)|\mathcal{D}_n])$ converges in distribution to $N(0,\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}^0))$. By the reconstruction of $\boldsymbol{\gamma}_i$ and $\mathrm{E}(1-\pi_i h_i|\mathcal{D}_n) = 0$ for any observation $i, i = 1,\cdots,n$ with $\mathbf{a}_i(\boldsymbol{\beta}^0) \neq 0$, we have $n^{-1}\sqrt{q}\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}^0)^{-1/2}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\mathbf{a}_i(\boldsymbol{\beta}^0)$ converges conditionally on $\mathcal{D}_n$ to a standard multivariate normal distribution. Replacing $\mathbf{a}_i(\boldsymbol{\beta}^0)$ by $\tilde{\mathbf{a}}_i(\boldsymbol{\beta}^0)$, for all $\boldsymbol{u} \in \mathcal{R}^\nu$, where ν is the dimension of covariates, we have

(24)

$$\Pr\left\{n^{-1}\sqrt{q}\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}^0)^{-1/2}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}^0) \le \boldsymbol{u}\,|\mathcal{D}_n\right\} \xrightarrow{P} \Phi(\boldsymbol{u}),$$

where Φ is the standard multivariate normal cumulative distribution function. Since the conditional probability is a bounded random variable in the unconditional space (which converges to a constant), by Eq.(24) and the dominated convergence theorem, we have

(25)

$$\Pr\left\{n^{-1}\sqrt{q}\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}^0)^{-1/2}\sum_{i\in\mathcal{C}}(1-\pi_i h_i)\tilde{\mathbf{a}}_i(\boldsymbol{\beta}^0) \le \boldsymbol{u}\right\} \xrightarrow{P} \Phi(\boldsymbol{u}).$$

It should also be observed that since the first addend in Eq. (20) goes to 0 as $q \to \infty$ and $n \to \infty$, based on A.10, the two addends are asymptotically independent. Combining Eq. (20)- (21), with Eq.(25), we arrive at Eq.(14). Here, we replace $\tilde{\mathcal{I}}$ with $\mathcal{I}$ by Lemma 3.3 and Theorem 5.14 of [32]. It should be noted that the first expression within $\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{p},\boldsymbol{\beta}^0)$, $qn^{-1}\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\boldsymbol{\Sigma}\mathcal{I}^{-1}(\boldsymbol{\beta}^0)$ goes to 0 as $q \to \infty$ and $n \to \infty$. Here, we prefer to retain it in order to obtain a more accurate representation for finite samples.

3.1 Optimal Sampling Probabilities

Here, we use the Average-optimal design criterion proposed by [33] due to its analytical convenience. We aim to derive the sampling probability vector which minimizes $Tr(\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{p},\boldsymbol{\beta}^0))$, where Tr is the trace operator.

Theorem 3.3. *The Average-optimal sampling probabilities vector $\boldsymbol{p}^A$ is of the form*

$$(26)\qquad p_m^A = \frac{\|\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\mathbf{a}_m(\boldsymbol{\beta}^0)\|_2}{\sum_{i\in\mathcal{C}}\|\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\mathbf{a}_i(\boldsymbol{\beta}^0)\|_2} \quad \textit{for all } m \in \mathcal{C}.$$

Proof of Theorem 3.3.

$$\begin{aligned}&Tr(\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{p},\boldsymbol{\beta}^0)) \\ &= Tr\{\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\boldsymbol{\phi}(\boldsymbol{p},\boldsymbol{\beta}^0)\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\} + d \\ &= Tr\left[\frac{1}{n^2}\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\left\{\sum_{i\in\mathcal{C}}\frac{1}{p_i}\mathbf{a}_i(\boldsymbol{\beta}^0)\mathbf{a}_i^\top(\boldsymbol{\beta}^0)\right.\right. \\ &\quad \left.\left. - \sum_{i,j\in\mathcal{C}}\mathbf{a}_i(\boldsymbol{\beta}^0)\mathbf{a}_j^\top(\boldsymbol{\beta}^0)\right\}\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\right] + d,\end{aligned}$$

where d is a constant. Omitting the part involving $\boldsymbol{p}$ yields

$$\begin{aligned}&Tr\left\{\frac{1}{n^2}\sum_{i\in\mathcal{C}}\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\frac{1}{p_i}\mathbf{a}_i(\boldsymbol{\beta}^0)\mathbf{a}_i^\top(\boldsymbol{\beta}^0)\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\right\} \\ &= \frac{1}{n^2}\sum_{i\in\mathcal{C}}\frac{1}{p_i}Tr\{\mathbf{a}_i^\top(\boldsymbol{\beta}^0)\mathcal{I}^{-2}(\boldsymbol{\beta}^0)\mathbf{a}_i(\boldsymbol{\beta}^0)\} \\ &= \frac{1}{n^2}\sum_{i\in\mathcal{C}}\frac{1}{p_i}\|\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\mathbf{a}_i(\boldsymbol{\beta}^0)\|_2^2.\end{aligned}$$

Removing the factor of n^{-2} does not alter the optimization solution. Next, we define the Lagrangian function with multiplier γ as

$$g(\boldsymbol{p}) = \sum_{i\in\mathcal{C}}\frac{1}{p_i}\|\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\mathbf{a}_i(\boldsymbol{\beta}^0)\|_2^2 + \gamma(1-\sum_{i\in\mathcal{C}}p_i).$$

Differentiating it with respect to p_m with $m \in \mathcal{C}$ and setting the derivative equal to 0, we get

$$\frac{\partial g(\boldsymbol{p})}{\partial p_m} = -\frac{\|\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\mathbf{a}_m(\boldsymbol{\beta}^0)\|_2^2}{p_m^2} - \gamma \equiv 0,$$

yielding

$$p_m = \frac{\|\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\mathbf{a}_m(\boldsymbol{\beta}^0)\|_2}{\sqrt{-\gamma}}.$$

Since all probabilities sum up to 1, we have

$$\sqrt{-\gamma} = \sum_{i \in \mathcal{C}} \|\mathcal{I}^{-1}(\boldsymbol{\beta}^0)\mathbf{a}_i(\boldsymbol{\beta}^0)\|_2,$$

which yields Eq. (26). Since $\boldsymbol{\beta}^0$ is unknown, we use a two-step estimator of [15].

Another strategy is the L-optimal ('Linear'-optimal) criterion which requires to minimize the trace of covariance matrix of $\mathcal{I}(\boldsymbol{\beta}^0)\tilde{\boldsymbol{\beta}}$.

Theorem 3.4. *The Linear-optimal sampling probabilities vector $\mathbf{p}^L$ is of the form*

$$p_m^L = \frac{\|\mathbf{a}_m(\boldsymbol{\beta}^0)\|_2}{\sum_{i \in \mathcal{C}} \|\mathbf{a}_i(\boldsymbol{\beta}^0)\|_2} \quad \textit{for all } m \in \mathcal{C}. \tag{27}$$

The proof of Theorem 3.4 is straightforward and its arguments are similar to those presented in the proof of Theorem 3.3, and is thus omitted.

3.2 Two-Step Procedure

Since the results obtained in Eq. (26) and Eq.(27) contain the unknown $\boldsymbol{\beta}^0$, we propose the following two-step algorithm:

Step 1 Sample q observations uniformly from $\mathcal{C}$ and combine them with $\mathcal{E}$ to form $\mathcal{Q}_{pilot}$. Obtain a crude estimator $\tilde{\boldsymbol{\beta}}_U$ by a weighted PSH regression on $\mathcal{Q}_{pilot}$ and use it to derive approximated optimal sampling probabilities using Eq. (26) or Eq.(27).

Step 2 Sample another q observations from $\mathcal{C}$ using the probabilities computed at Step 1. Combine these observations with $\mathcal{E}$ to form $\mathcal{Q}$ and conduct weighted PSH regression on $\mathcal{Q}$ to obtain the two step estimators $\tilde{\boldsymbol{\beta}}_{TS}$ of $\boldsymbol{\beta}^0$.

Similar to Theorem 3.2, Theorem 3.5 establishes the asymptotic properties of $\tilde{\boldsymbol{\beta}}_{TS}$ and $\hat{\Lambda}_{10}(0, \tilde{\boldsymbol{\beta}}_{TS})$.

Theorem 3.5. *Under Assumptions A1-A10, the following asymptotic properties hold*

$$\sqrt{q}\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\boldsymbol{p}^{opt}, \boldsymbol{\beta}^0)(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^0) \xrightarrow{D} N(0, \mathbf{I}), \tag{28}$$

$$\sqrt{q}\mathbb{V}_{\hat{\Lambda}_{10}, \tilde{\boldsymbol{\beta}}}^{-1/2}(\boldsymbol{p}^{opt}, \boldsymbol{\beta}^0, t)\{\hat{\Lambda}_{10}(t, \tilde{\boldsymbol{\beta}}_{TS}) - \Lambda_{10}(t)\} \xrightarrow{D} N(0, 1), \tag{29}$$

where $\boldsymbol{p}^{opt}$ is either $\mathbf{p}^L$ or $\mathbf{p}^A$, depending on the chosen optimality criterion.

Proof of Theorem 3.5.

If $\tilde{\boldsymbol{\beta}}_U$ is also conditioned upon the conditional space, then the sampling probabilities become deterministic, and we return to the settings of Theorem 3.2. The consistency and normality results derived for $\tilde{\boldsymbol{\beta}}$ can be applied to any vector of deterministic sampling probabilities that satisfy A.8. For example, with each component of $\boldsymbol{p}$ equal to each other, we can get the asymptotic consistency and normality of $\tilde{\boldsymbol{\beta}}_U$ based on Theorem 3.2.

Thus, for all $\boldsymbol{u} \in \mathbb{R}^\nu$,

$$\begin{aligned}
&\Pr\left\{\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\boldsymbol{p}^{opt}, \boldsymbol{\beta}^0)\left(\sqrt{q}(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^0)\right) \le \boldsymbol{u}\right\} \\
&= E\left[P\left\{\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\boldsymbol{p}^{opt}, \boldsymbol{\beta}^0)\left(\sqrt{q}(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^0)\right) \le \boldsymbol{u} | \mathcal{D}_n, \tilde{\boldsymbol{\beta}}_U\right\}\right].
\end{aligned}$$

By Theorem 3.2 and asymptotic properties for $\tilde{\boldsymbol{\beta}}_U$, we have

$$\Pr\left\{\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\boldsymbol{p}^{opt}, \boldsymbol{\beta}^0)\left(\sqrt{q}(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^0)\right) \le u | \mathcal{D}_n, \tilde{\boldsymbol{\beta}}_U\right\} \to \Phi(\boldsymbol{u}).$$

Since the conditional probability is a bounded random variable in the unconditional space, which converges to a constant, Eq.(28) is implied by the dominated convergence theorem with the consistency and normality results derived for $\tilde{\boldsymbol{\beta}}$. The same arguments hold for proving Eq.(29). Thus, the proof is omitted.

Remark 3.3. *We can compute the two-step optimal subsampling probabilities directly by replacing the $\boldsymbol{\beta}^0$ with $\tilde{\boldsymbol{\beta}}_U$ in the expression of optimal sampling probability and the computation covariance matrix is similar. Thus, we omit the cumbersome and replicated expressions and notations in Theorem 3.2 here for simplicity.*

3.3 Variance Estimation

Based on Eq.(28), a natural estimator for the covariance matrix of $\tilde{\boldsymbol{\beta}}_{TS}$ is

$$\begin{aligned}
&q^{-1}\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS}) \\
&= n^{-1}\mathcal{I}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})\boldsymbol{\Sigma}\mathcal{I}^{-1}(\tilde{\boldsymbol{\beta}}_{TS}) \\
&\quad + q^{-1}\mathcal{I}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})\boldsymbol{\phi}(\boldsymbol{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS})\mathcal{I}^{-1}(\tilde{\boldsymbol{\beta}}_{TS}).
\end{aligned}$$

However, calculation of $\mathcal{I}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})$, $\boldsymbol{\Sigma}$ or $\boldsymbol{\phi}(\boldsymbol{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS})$ involves the full data, and may be avoided by replacing these matrices with their subsampling-based counterparts, i.e., $\tilde{\mathcal{I}}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})$, $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\phi}}(\boldsymbol{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS})$ where

$$\tilde{\boldsymbol{\phi}}(\boldsymbol{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS}) = \frac{1}{n^2 q}\left\{\sum_{i=1}^{q} \frac{\mathbf{a}_i^*(\tilde{\boldsymbol{\beta}}_{TS})\mathbf{a}_i^*(\tilde{\boldsymbol{\beta}}_{TS})^\top}{p_i^{*2}}\right.$$

$$-\frac{1}{q^2}\sum_{i=1}^{q}\frac{\mathbf{a}_i^*(\tilde{\boldsymbol{\beta}}_{TS})}{p_i^*}\left(\sum_{i=1}^{q}\frac{\mathbf{a}_i^*(\tilde{\boldsymbol{\beta}}_{TS})}{p_i^*}\right)^\top\Bigg\}.$$

The variance estimator for $\hat{\Lambda}_{10}(t,\tilde{\boldsymbol{\beta}}_{TS})$ is simply $q^{-1}\mathbb{V}_{\hat{\Lambda}_{10}(t,\tilde{\boldsymbol{\beta}})}(\boldsymbol{p}^{opt},\tilde{\boldsymbol{\beta}}_{TS},t)$.

4. NUMERICAL RESUTLS

4.1 Simulation Study

In this section, we compare the performance of our proposed methods with the full-data partial-likelihood estimator. The data are generated based on [4], where

$$P(T_i \le t, \epsilon_i = 1|\boldsymbol{Z}_i) = 1 - [1-\theta\{1-\exp(-t)\}]^{\exp(\boldsymbol{Z}^\top\boldsymbol{\beta}^0)}$$

and the sub-distribution for type 2 failure is obtained by taking $P(\epsilon_i = 2|\boldsymbol{Z}_i) = 1 - P(\epsilon_i = 1|\boldsymbol{Z}_i)$ and using an exponential distribution with rate $\exp(-\boldsymbol{Z}^\top\boldsymbol{\beta}^0)$ for $P(T_i \le t|\epsilon_i = 2, \boldsymbol{Z}_i)$. We set $\theta = 0.3$ and the true vector of coefficient $\boldsymbol{\beta}^0 = (0.3, -0.5, 0.1, -0.1, 0.1, -0.3)^\top$.

The covariates $\boldsymbol{Z}_i, i = 1, 2, \cdots, 6$ are generated from $N(0, \boldsymbol{\Sigma}^0)$, with different covariance matrics $\boldsymbol{\Sigma}^0 = (\rho_{ij})$ as follows

A $\rho_{ii} = 1$ and $\rho_{ij} = 0$ for $i \ne j$, $i, j = 1, 2, \cdots, 6$. In this case, Z_is are independently distributed with the same variance.

B $(\rho_i)_{i=1,2,\cdots,6} = (1, 1.5, 2, 2, 1.5, 1)$ respectively. In this case, Z_is are independently distributed with different variance.

C $\rho_{ii} = 1$ and $\rho_{ij} = 0.5$ for $i \ne j$, $i, j = 1, 2, \cdots, 6$. In this case, Z_is are mildly correlated.

We also simulate two other cases as follows

D $Z_i \sim \text{Uniform}(-2, 2.5), i = 1, 2, \cdots, 6$.

E $Z_i \sim \text{Exp}(0.5), i = 1, 2, \cdots, 6$.

We use Full, Unif, L-opt and A-opt to represent the full-data partial likelihood estimator while Uniform, L-optimal and A-optimal to represent the subsampling estimators, respectively. Our simulation studies compare the root mean squared errors (RMSEs) of different methods, based on the average of the 100 samples, with respect to the real vector of coefficient $\boldsymbol{\beta}^0$ and with respect to $\hat{\boldsymbol{\beta}}_{\text{full}}$, serving as the "golden standard" .

Table 1 reports the simulation results for 100 replications with $n = 15000$, and the censoring time is generated from Exp(0.02) and the resulting averaged censoring rate is about 97%. q is set equal to the number of failures. We also conduct the classical full-data partial-likelihood estimation based on R package `cmprsk` for comparison. Simulation results show that our subsampling programs cost much less time than R package. The L-opt and A-opt methods require much less time than the Full estimator with similar RMSEs. Unif method uses the least computing time but yields the largest RMSEs. Besides, our subsampling methods still well behave even when $\boldsymbol{Z}$ follows from uniform distribution or exponential distribution.

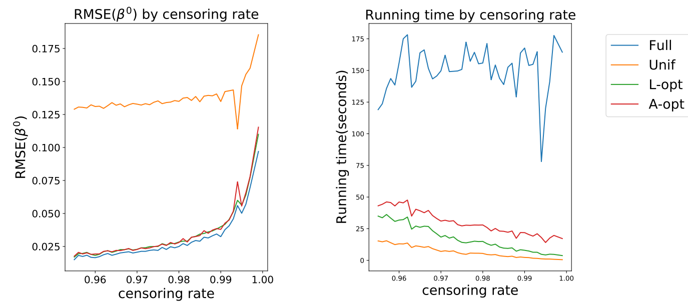

Figure 1. RMSE(β^0) and running time variations via censoring rates.n=1500000.

Table 2 displays the simulation results from much bigger sample size $n = 1500000$, with a censoring rate being 99.9%. Regarding RMSEs, our L-opt and A-opt methods perform similarly to the Full estimator, especially under settings "C","D" and"E", and have much smaller RMSEs than Unif method. Unif requires the least running time, followed by L-opt and A-opt. Full estimator takes roughly 20 times longer than L-opt and 120 times longer than Unif when q equals to the number of the failures.

Figure 1 illustrates the performance of our subsampling methods varying with censoring rates based on case "A". The censoring rate ranges from 95.5% to 99.9% and q equals to the number of failures. As censoring rate increases, RMSE($\boldsymbol{\beta}^0$) of each estimator increases and the L-opt and A-opt estimators yield similar RMSEs to those of the full-data estimator, but cost much less computing time. The computing time of the L-opt estimator approaches to that of the Unif estimator for large censoring rates.

Remark 4.1. *Since n_e/n approaches to 0 while the censoring rate approaches to 1, Figure 1 represents the variations of RMSE and running time via censoring rates and thus also portrays the performance under assumption A.10.*

Figure 2 displays the simulation results for case "A" at censoring rate 99.9% with q/n ranging from 0.0040 to 0.0145 (q/n_e ranging from 1 to 11). When q/n increases, the RMSE($\boldsymbol{\beta}^0$) of Unif decreases and converges to the RMSE($\boldsymbol{\beta}^0$) of the Full estimator. The convergence occurs after q/n exceeds $0.0145(q = 11n_e)$. The RMSE($\boldsymbol{\beta}^0$)s of the optimal sampling methods show a slight decrease and converge to the RMSE($\boldsymbol{\beta}^0$) of the Full estimator for very small q/n. The computing time of each subsampling estimator displays small increase as q/n increases and their efficiencies are comparable to that of the Full estimator. In conclusion, the approximation of Unif estimator to full-data estimator is predictable when q/n approaches to 1. The Unif estimator demonstrates very limited improvement of RMSE

Table 1. Simulation result of right-censored data: q was set equal to the number of failures. The mean $\bar{q}$, and standard deviation of q, SD(q), are reported for $n = 15000$

Setting	$\bar{q}$	SD(q)	Method	RMSE$\boldsymbol{\beta}^0$	RMSE$\hat{\boldsymbol{\beta}}_{full}$	Run Time (sec.)
A(97.4%)	393.0	21.0	Full	0.224	0.0	0.189
			cmprsk	0.229	-	10.916
			L-opt	0.253	0.122	0.048
			A-opt	0.261	0.119	0.061
			Unif	0.354	0.307	0.019
B(97.2%)	424.0	20.0	Full	0.185	0.0	0.177
			cmprsk	0.181	-	11.707
			L-opt	0.21	0.1	0.048
			A-opt	0.216	0.099	0.063
			Unif	0.254	0.215	0.019
C(97.6%)	356.0	18.0	Full	0.295	0.0	0.165
			cmprsk	0.295	-	9.806
			L-opt	0.351	0.182	0.045
			A-opt	0.337	0.167	0.058
			Unif	0.451	0.342	0.016
D(96.8%)	485.0	22.0	Full	0.173	0.0	0.174
			cmprsk	0.181	-	12.698
			L-opt	0.194	0.083	0.051
			A-opt	0.193	0.087	0.066
			Unif	0.263	0.219	0.02
E(97.6%)	365.0	22.0	Full	0.552	0.0	0.156
			cmprsk	0.604	-	7.796
			L-opt	0.601	0.249	0.042
			A-opt	0.62	0.244	0.057
			Unif	0.74	0.53	0.015

Table 2. Simulation results of right-censored data: q was set equal to the number of failures. the mean $\bar{q}$, and standard deviation of q, SD(q), are reported for $n = 1500000$

Setting	$\bar{q}$	SD(q)	Method	RMSE $\boldsymbol{\beta}^0$	RMSE $\hat{\boldsymbol{\beta}}_{full}$	Run Time (sec.)
A(99.9%)	2004.0	44.031	Full	0.104	0.0	90.59
			L-opt	0.113	0.057	4.585
			A-opt	0.113	0.055	11.514
			Unif	0.182	0.174	0.643
B(99.9%)	2203.4	43.916	Full	0.08	0.0	72.547
			L-opt	0.09	0.042	4.622
			A-opt	0.09	0.042	11.422
			Unif	0.129	0.12	0.668
C(99.9%)	1827.2	41.497	Full	0.131	0.0	83.502
			L-opt	0.15	0.075	4.567
			A-opt	0.158	0.077	11.503
			Unif	0.256	0.243	0.568
D(99.8%)	2563.5	43.893	Full	0.076	0.0	86.892
			L-opt	0.086	0.039	4.758
			A-opt	0.086	0.038	11.702
			Unif	0.152	0.143	0.72
E(99.9%)	1877.4	44.561	Full	0.237	0.0	65.288
			L-opt	0.273	0.116	4.011
			A-opt	0.262	0.106	10.029
			Unif	0.47	0.424	0.547

Table 3. Empirical Bias (Bias),Empirical Variance Estimates (Var) , 95% Confidence Interval (CI), Empirical Coverage Probability(CP) for case "D", $n = 150000$, with q equal to the number of failures

Methods		$\beta_1^0 = 0.2$	$\beta_2^0 = -0.5$	$\beta_3^0 = 0.1$	$\beta_4^0 = -0.1$	$\beta_5^0 = 0.1$	$\beta_6^0 = -0.3$
Full	Bias(Var)	0.0(0.002)	-0.004(0.003)	0.002(0.002)	0.007(0.011)	-0.003(0.002)	-0.0(0.002)
	CI	[0.205,0.396]	[-0.604,-0.404]	[0.012,0.192]	[-0.298,0.112]	[0.007,0.188]	[-0.398,-0.203]
	CP	0.946	0.94	0.946	0.998	0.95	0.952
L-opt	Bias(Var)	-0.001(0.003)	-0.003(0.003)	0.0(0.003)	0.002(0.003)	-0.002(0.003)	0.0(0.003)
	CI	[0.194,0.403]	[-0.615,-0.391]	[0.002,0.199]	[-0.201,0.004]	[-0.002,0.198]	[-0.407,-0.193]
	CP	0.956	0.942	0.95	0.944	0.95	0.952
A-opt	Bias(Var)	0.001(0.003)	-0.004(0.003)	0.003(0.003)	0.001(0.003)	-0.003(0.003)	0.001(0.003)
	CI	[0.195,0.407]	[-0.615,-0.393]	[0.001,0.204]	[-0.198,0.001]	[-0.004,0.197]	[-0.405,-0.193]
	CP	0.954	0.952	0.94	0.954	0.95	0.954
Uniform	Bias(Var)	0.04(0.003)	-0.106(0.003)	0.009(0.003)	-0.008(0.003)	0.009(0.003)	-0.04(0.003)
	CI	[0.229,0.452]	[-0.711,-0.5]	[0.003,0.214]	[-0.213,-0.002]	[-0.002,0.22]	[-0.452,-0.228]
	CP	0.892	0.491	0.942	0.946	0.946	0.892

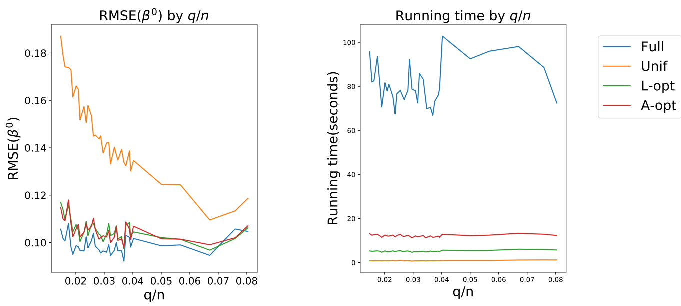

Figure 2. RMSE(β^0) and running time variation via q/n.n=1500000.

at moderate q/n. On the other hand, the optimal subsampling methods perform well with respect to RMSE at very small q/n.

Finally, Tables 3-5 show the results for empirical bias (EmpBias), empirical variance (EmpVar), Empirical 95% confidence interval and coverage probability for each estimated coefficient under case "D". The observations are similar to those presented in previous simulations. Generally, regardless of computing time, Full estimator behaves better than L-opt and A-opt in terms of smaller bias and narrower 95% confidence interval. Again, Uniform estimator performs the worst. It is because the estimator formula and variance formula are actually both unreliable when using the uniform method and the confidence interval is then unreliable. Thus, we recommend the proposed optimal subsampling methods.

4.2 Real Data Example

Breast cancer is the most common malignant disease especially for women. Improvement of medical care improves the life expectancy of breast cancer patients who are then likely at risk of developing a second malignancy [34]. Thus, investigating the non-breast-cancer death among breast cancer patients is also important [35]. We choose the event of interest to be death of pneumonia and influenza among breast cancer patients since pneumonia and influenza could also cause high risk of death for cancer patients [36]. The competing risk is lung cancer [37, 38, 39]. The data come from the publicly available database, the Surveillance, Epidemiology, and End Results Program (SEER, `https://seer.cancer.gov/causespecific/`). We use an extracted cohort of breast cancer patients from 1992-2007. Deaths caused by other diseases as well as patients living at the end of the study are considered as censored cases. The numbers of patients by event are summarized as follows:

- Total patients: 177162

—**Event1:** Deaths from pneumonia and influenza: 1727(0.97%)

—**Event2:** Deaths from lung cancer: 3616 (2.04%)

—**Censored:** 177162 (97.00%)

16 features in the SEER dataset are included in our analyses. They include demographic characteristics such as age, race, gender, and morphology information such as tumor size, tumor type, cancer stage, and hormone status of breast cancers. Among the covariates, "Laterality" describes the side of a paired organ or side of the body on which the reportable tumor (breast) originated. "Sex:M" records whether a patient is male or female. "Single" and "Married" are dummy variables representing marital status at diagnosis, with "Separated" as the baseline category. "Laterality" describes the side of a paired organ or side of the body on which the reportable tumor originated, with "Bilateral" as the baseline category. "Race:White" and "Race:Black" are dummy variables representing the race of a patient, with "Asian" as the baseline category. "ER(Estrogen receptor) status" and "PR(Progesterone receptor) status" record the hormone status of breast cancers. "Sequence Number" describes the number and sequence of all reportable malignant, in situ, benign, and borderline primary tumors, which occur over the lifetime of a patient. "First malignant primary"

Table 4. Empirical Bias (EmpBias),Empirical Variance Estimates (EmpVar) , and 95% Confidence Interval (CI) for case "D", $n = 150000$, with q equal to the number of failures $\times 2$

Methods		$\beta_1^0 = 0.2$	$\beta_2^0 = -0.5$	$\beta_3^0 = 0.1$	$\beta_4^0 = -0.1$	$\beta_5^0 = 0.1$	$\beta_6^0 = -0.3$
Full	Bias(Var)	0.01(0.015)	0.0(0.015)	0.009(0.013)	-0.003(0.013)	0.005(0.014)	0.013(0.013)
	CI	[0.205,0.396]	[-0.604,-0.404]	[0.012,0.192]	[-0.298,0.112]	[0.007,0.188]	[-0.398,-0.203]
	CP	0.954	0.94	0.948	0.944	0.944	0.938
L-opt	Bias(Var)	0.009(0.016)	0.001(0.016)	0.009(0.014)	-0.002(0.014)	0.004(0.015)	0.013(0.014)
	CI	[0.194,0.403]	[-0.615,-0.391]	[0.002,0.199]	[-0.201,0.004]	[-0.002,0.198]	[-0.407,-0.193]
	CP	0.954	0.956	0.95	0.946	0.944	0.948
A-opt	Bias(Var)	0.009(0.016)	0.0(0.016)	0.01(0.014)	-0.003(0.014)	0.004(0.015)	0.012(0.013)
	CI	[0.195,0.407]	[-0.615,-0.393]	[0.001,0.204]	[-0.198,0.001]	[-0.004,0.197]	[-0.405,-0.193]
	CP	0.956	0.944	0.956	0.94	0.95	0.946
Uniform	Bias(Var)	0.1(0.016)	-0.153(0.015)	0.037(0.014)	-0.031(0.013)	0.031(0.015)	-0.08(0.013)
	CI	[0.229,0.452]	[-0.711,-0.5]	[0.003,0.214]	[-0.213,-0.002]	[-0.002,0.22]	[-0.452,-0.228]
	CP	0.888	0.768	0.946	0.946	0.946	0.912

Table 5. Empirical Bias (EmpBias),Empirical Variance Estimates (EmpVar) , and 95% Confidence Interval (CI) for case "D", $n = 150000$, with q equal to the number of failures $\times 3$

Methods		$\beta_1^0 = 0.2$	$\beta_2^0 = -0.5$	$\beta_3^0 = 0.1$	$\beta_4^0 = -0.1$	$\beta_5^0 = 0.1$	$\beta_6^0 = -0.3$
Full	Bias(Var)	-0.004(0.015)	-0.007(0.014)	-0.003(0.013)	0.01(0.014)	-0.008(0.014)	-0.0(0.013)
	CI	[0.205,0.396]	[-0.604,-0.404]	[0.012,0.192]	[-0.298,0.112]	[0.007,0.188]	[-0.398,-0.203]
	CP	0.954	0.958	0.944	0.954	0.96	0.946
L-opt	Bias(Var)	-0.004(0.015)	-0.007(0.015)	-0.003(0.014)	0.009(0.015)	-0.009(0.014)	0.0(0.013)
	CI	[0.194,0.403]	[-0.615,-0.391]	[0.002,0.199]	[-0.201,0.004]	[-0.002,0.198]	[-0.407,-0.193]
	CP	0.958	0.958	0.942	0.958	0.95	0.95
A-opt	Bias(Var)	-0.005(0.015)	-0.008(0.014)	-0.002(0.014)	0.009(0.015)	-0.008(0.015)	-0.001(0.013)
	CI	[0.195,0.407]	[-0.615,-0.393]	[0.001,0.204]	[-0.198,0.001]	[-0.004,0.197]	[-0.405,-0.193]
	CP	0.954	0.954	0.948	0.956	0.952	0.946
Uniform	Bias(Var)	0.06(0.014)	-0.118(0.014)	0.019(0.015)	-0.012(0.015)	0.013(0.015)	-0.064(0.012)
	CI	[0.229,0.452]	[-0.711,-0.5]	[0.003,0.214]	[-0.213,-0.002]	[-0.002,0.22]	[-0.452,-0.228]
	CP	0.926	0.798	0.942	0.948	0.96	0.912

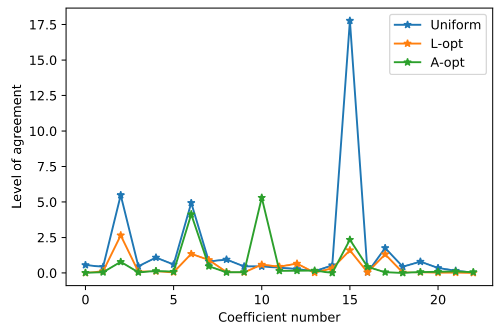

Figure 3. SEER data analysis: comparison of level of agreement for the estimated regression coefficients of each method (L-opt, A-opt, uniform) against the full-data estimator, computed as follows: the Y-axis is the ratio of difference from the full-data estimator for each method, divided by the full-data estimator, and the X-axis is the regression coefficient ordinal number

is a binary variable, taking 1 for yes and 0 for no. "EOD -size" records the largest dimension of the primary tumor (breast) in millimeters. "EOD -extent" codes the farthest documented extension of tumor away from the primary site (breast), either by contiguous extension or distant metastases. "EOD - nodes" records the highest specific lymph node chain that is involved by the tumor (breast). "Breast - Adjusted AJCC 6th Stage" records the stage of breast cancer. "Regional nodes positive" records the exact number of regional lymph nodes examined by the pathologist that were found to contain metastases.

q is set equal to the number of observed failure time. To reduce the impact of local optimization, we conduct 100 replications of the data analysis and report the average and standard deviation(SD) of each estimator. Figure 3 shows the level of agreement between the estimated regression coefficients of each method and the full-data partial-likelihood counterpart. The A-opt and L-opt method in most cases exhibit the highest level of agreement, and the uniform method the least. The estimated coefficients are displayed in Table 6.

Regarding the computing time, the full-data partial-likelihood estimator requires an average of 73.0921 seconds, and the L-opt and A-opt methods require 10.5840 and 11.6546 seconds respectively, and the uniform method requires only 3.077 seconds. Table 7 presents the running time per step. Obviously, optimal methods spend more time on computing the optimal subsampling probabilities, but still have great improvement over the full-data estimator. Uniform method requires shorter computing time but displays less agreement with the full-data estimator.

5. DISCUSSION

In this article, we developed the L-optimal and A-optimal subsampling method based on Fine-Gray model to address competing risks data. Simulations and real data analysis showed that our proposed methods produced efficient estimates and accurate results. Nonetheless, there are still some interesting problems to be investigated for future research. Firstly, to relax the assumption of Fine-Gray model, the subsampling methods for competing risks data under a model-free framework is an interesting issue. Secondly, coping with more complex and realistic situations like time-dependent covariate is another practical issue. Thirdly, as noticed, the larger the censoring rate (i.e., all competing risks are rare), the more efficient our subsampling methods. In practice, not all risks are rare. In this case, retaining all the failure observations would still cause computational inefficiency. A possible way is only retaining all the failure events of interest in order to keep the most important information and then sampling the censored observations and events of competing risks together. However, information in the competing risks may be lost. An alternative way is sampling the censored observations and events of competing risks respectively to balance the computational efficiency with loss of information. We are now investigating these interesting problems.

6. ACKNOWLEDGEMENTS

The research is supported by the National Social Science Foundation of China (Series number: 21BTU040).

The data set is provided by Surveillance Research Program, National Cancer Institute SEER*Stat software (seer.cancer.gov/seerstat) version 8.3.9.1.

We are very grateful to the two reviewers and the Editor for their detailed and important comments on our manuscript and insightful inspiration for our future work.

Received 3 January 2022

REFERENCES

[1] Keret, N. and Gorfine, M. (2023). Analyzing big EHR data–optimal Cox regression subsampling procedure with rare events. *Journal of the American Statistical Association.* **0(0)** 1–14. DOI:10.1080/01621459.2023.2209349

[2] Tian, R., Wu, W., Wang, C. et al.(2020). Clinical characteristics and survival analysis in critical and non-critical patients with COVID-19 in Wuhan, China: a single-center retrospective case control study. *Nature Resaerch: Scientifc Reports.* **10** 17524. https://doi.org/10.1038/s41598-020-74465-3

[3] Geskus, R.B. (2015). *Data Analysis with Competing Risks and Intermediate States (1st ed.).* Chapman and Hall/CRC, New York.

[4] Fine, J. P., and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association.* **94(446)** 496–509.

[5] Scrucca, L., Santucci, A., and Aversa, F. (2010). Regression modeling of competing risk using R: an in depth guide for clinicians. *Bone Marrow Transplant.* **45(9)** 1388–1395.

[6] Fu, Z., Parikh, C. R., and Zhou, B. (2017). Penalized variable selection in competing risks regression. *Lifetime Data Analysis.* **23** 353–376.

Table 6. Estimated regression coefficients(SD) of SEER data

	Full	Unif	L-opt	A-opt
Age	0.1351(0.0000)	0.0995(0.0002)	0.1355(0.0001)	0.1352(0.0002)
Sex:Male	0.4243(0.0000)	0.1866(0.0163)	0.4878(0.0143)	0.4856(0.0056)
Race:White	-0.1922(0.0000)	-0.1996(0.0065)	-0.1951(0.0044)	-0.1582(0.0039)
Race:Black	-0.0794(0.0000)	0.0129(0.0091)	-0.0889(0.0075)	-0.0278(0.0067)
Single	0.1284(0.0000)	0.1045(0.0046)	0.1282(0.0055)	0.1208(0.0081)
Married	-0.2757(0.0000)	-0.0890(0.0027)	-0.2745(0.0034)	-0.2735(0.0039)
Laterality:right	3.0558(0.3306)	0.9917(0.1033)	0.6385(0.1073)	2.3260(0.0430)
Laterality:left	2.9907(0.3306)	0.9538(0.1029)	0.5664(0.1075)	2.2825(0.0451)
Laterality:unspecified one side	-3.7117(0.3431)	-1.5131(0.1959)	-0.2927(0.1494)	-2.4295(0.0555)
Laterality:Bilateral	-3.0401(0.3264)	-0.1154(0.0963)	-0.4735(0.1925)	-2.0560(0.0482)
ER Status: Positive	0.0762(0.0000)	0.0963(0.0051)	0.0777(0.0042)	0.0868(0.0031)
PR Status: Positive	-0.0527(0.0000)	-0.0740(0.0043)	-0.0521(0.0041)	-0.0799(0.0010)
Sequence number	-0.0190(0.0000)	-0.3793(0.0048)	-0.0218(0.0090)	-0.0427(0.0041)
First malignant primary	-0.5072(0.0000)	-1.5513(0.0064)	-0.4992(0.0130)	-0.5311(0.0043)
Total number of in situ/malignant tumors for patient	-0.3909(0.0000)	-0.9574(0.0041)	-0.3804(0.0073)	-0.3844(0.0040)
Total number of benign/borderline tumors for patient	-0.3018(0.0000)	0.1440(0.0189)	-0.3053(0.0218)	-0.2871(0.0083)
EOD- extent	0.0002(0.0000)	0.0002(0.0003)	0.0005(0.0002)	0.0007(0.0002)
EOD 10 - nodes	-0.0046(0.0000)	0.0010(0.0009)	-0.0030(0.0007)	-0.0032(0.0009)
EOD 10 - size	0.0023(0.0000)	0.0028(0.0001)	0.0023(0.0000)	0.0020(0.0000)
Breast - Adjusted AJCC 6th Stage	0.0058(0.0000)	0.0040(0.0002)	0.0053(0.0001)	0.0061(0.0001)
Regional nodes positive	0.0155(0.0000)	0.0175(0.0008)	0.0158(0.0005)	0.0131(0.0002)

Table 7. Average time(SD) cost per step

Method	Time	Likelihood-step	Weight-step
Full	73.0921(1.9156)	73.0921(0)	0(0)
L-opt	10.5840(0.0985)	3.3263(0.0760)	4.1802(0.0422)
A-opt	11.6546(0.1551)	4.2051(0.1268)	4.2863(0.0755)
Unif	3.077(0.0471)	2.9951(0.0650)	0.0823(0.0432)

[7] Li, E., Mei, B., and Tian, M. (2018). Feature screening based on ultrahigh dimensional competing risks models (in Chinese). *Sci Sin Math.* **48** 1061—1086.

[8] Andersen, P. K., and Perme, M. P. (2010). Pseudo-observations in survival analysis. *Stat Methods Med Res.* **19(1)** 71–99.

[9] Binder, N., Gerds, T.A. and Andersen, P.K.(2014). Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime Data Anal.* **20** 303—315.

[10] Akhtar-Danesh, N. (2008). Some challenges in survival analysis with large datasets, *Fall North American Stata Users' Group Meetings 2008.4, Stata Users Group.*

[11] Kirby, K.A., Barnes, D.F. and Boscardin,W.J. (2013). Competing risks regression: Big Data issues from modest sample sizes. *Available at https : //www.lexjansen.com/wuss/2013/124$_P$aper.pdf.*

[12] Dhillon, P., Lu, Y., Foster, D. P., and Ungar, L. H. (2013). New subsampling algorithms for fast least squares regression. *Advances in Neural Information Processing Systems.* **26** 360–368.

[13] Ma, P., and Sun, X. (2015), Leveraging for big data regression. *WIREs Comput Stat.* **7** 70–76.

[14] Wang, H. Y., Yang M. and Stufken J.(2019) Information-Based Optimal Subdata Selection for Big Data Linear Regression, *Journal of the American Statistical Association.* **114:525** 393–405.

[15] Wang, H. Y., Zhu, R., and Ma, P. (2018). Optimal subsampling for large sample Logistic regression, *Journal of the American Statistical Association.* **113** 829–844.

[16] Wang, H. Y. (2019). More Efficient Estimation for Logistic Regression with Optimal Subsample. *Journal of Machine Learning Research.* **20(132)** 1–59

[17] Ai, M., Yu, J., Zhang, H., and Wang, H. (2021). Optimal Subsampling Algorithms for Big Data Regressions, *Statistica Sinica.* **31(2)** 749–772

[18] Ai, M., Wang, F., Yu, J. and Zhang, H. (2021). Optimal Subsampling for Large-scale Quantile Regression, *Journal of Complexity.* **62** 101–512.

[19] Wang, H. Y. and Ma, P. (2021). Optimal subsampling for quantile regression in big data, *Biometrika.* **108** 99–112.

[20] Yu, J., Wang, H., Ai, M. and Zhang, H. (2022). Optimal Distributed Subsampling for Maximum Quasi-Likelihood Estimators With Massive Data, *Journal of the American Statistical Association.* **117:537** 265–276.

[21] Fine, J.P., Jiang, H., and Chappell, R. (2001), On Semi-Competing Risks Data, *Biometrika.* **88** 907—919.

[22] Gorfine, M., Keret, N., Ben Arie, A., Zucker, D., and Hsu, L.(2021) Marginalized Frailty-Based Illness-Death Model: Application to the UK-Biobank Survival Data. *Journal of the American Statistical Association.* **116:535** 1155–1167.

[23] Zuo, L., Zhang, H., Wang, H., and Liu, L. (2021). Sampling-based estimation for massive survival data with additive hazards model. *Statistics in Medicine.* **40(2)** 441–450.

[24] Zhang, H.X., Zuo, L.L., Wang, H.Y. and Sun, L.Q. (2023) Approximating Partial Likelihood Estimators via Optimal Subsampling. *Journal of Computational and Graphical Statistic.* DOI: 10.1080/10618600.2023.2216261

[25] Gray, R.J. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk, *The Annals of Statistics.* **16** 1141–1154.

[26] Robins, J.M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *In AIDS Epidemiology Theorethodological Issues. Jewell, N. , Dietz, K. and Farewell,V.,* Birkhäuser,Boston, 24–33.

[27] Kaplan, E.L., and Meier, P. (1958). Nonparametric Estimation From Incomplete Observations. *Journal of the American Statistical Association.* **53** 457-–481.

[28] Andersen, P.K., and Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. *Ann Statist.***10(4)** 1100–1120.

[29] Ferguson, T.S. (2017). *A Course in Large Sample Theory.* Chapman & Hall, London.

[30] van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press ,Cambridge .

[31] Samuelsen, S.O. (1997). A pseudolikelihood approach to analysis of nested casecontrol studies. *Biometrika.* **84(2)** 379–394.

[32] Boos, D.D., Stefanski, L.A. (2013). *Essential Statistical Inference: Theory and Methods.* Springer Science+Business Media, New York.

[33] Pukelsheim, F. (2006). *Optimal design of experiments.* SIAM ,Philadelphia.

[34] Lorigan, P., Califano, R., Faivre-Finn, C., Howell, A., Thatcher, N.(2010) Lung cancer after treatment for breast cancer. *Lancet Oncol.* **11(12)** 1184–1192.

[35] Bush, D., Smith, B., Younger, J., Michaelson, J.S.(2011) The non-breast-cancer death rate among breast cancer patients. *Breast Cancer Res Treat.* **127(1)** 243–249.

[36] Abdel-Rahman, O.(2021) Influenza and pneumonia-attributed deaths among cancer patients in the United States; a population-based study. *Expert Rev Respir Med.* **15(3)** 393–401.

[37] Weng, C.F., Chen, L.J., Lin, C.W., Chen, H.M., Lee, H.H., Ling, T.Y., Hsiao, F.Y.(2019). Association between the risk of lung cancer and influenza: A population-based nested case-control study. *Int J Infect Dis.* **88** 8–13.

[38] Budisan, L., Zanoaga, O., Braicu, C., Pirlog, R., Covaliu, B., Esanu, V., Korban, S.S., Berindan-Neagoe, I.(2021) Links between Infections, Lung Cancer, and the Immune System. *Int J Mol Sci.* **22(17)** 9394. doi: 10.3390/ijms22179394.

[39] Cao, X., Chen P., Agyekum, E. A., Zhang, Q., Qian, X., Wu, T., Chambers, K. H., Yin, L. (2023) Lung cancer with breast metastasis: a case report and review of the literature. *J Int Med Res.* **51(7)** 3000605231188287. doi: 10.1177/03000605231188287.

Li Erqian
College of Science, North China University of Technology
China
E-mail address: li2qian@ncut.edu.cn

Tang Man-Lai
Centre of Data Innovation Research, Department of Physics, Astronomy and Mathematics, School of Physics, Engineering & Computer Science, University of Hertfordshire;
UK
E-mail address: m.l.tang@herts.ac.uk

Tian Maozai
Center for Applied Statistics, School of Statistics, Renmin University of China;
China
E-mail address: mztian@ruc.edu.cn

Yu Keming
Department of Mathematical Sciences, Brunel University London;
UK
E-mail address: Keming.Yu@brunel.ac.uk