Statistics and Its Interface Volume 12 (2019) 107–122

Fine-Gray proportional subdistribution hazards model for competing risks data under length-biased sampling

Feipeng Zhang, Heng Peng*, and Yong Zhou

In this paper, we study the Fine-Gray proportional subdistribution hazards model for the competing risks data under length-biased sampling. To exploit the special structure of length-biased sampling, we propose an unbiased estimating equation estimator, which can handle both covariate-independent censoring and the covariate-dependent censoring. The large sample properties of the proposed estimator are derived, model-checking techniques for the model adequacy are developed, and the pointwise confidence intervals and the simultaneous confidence bands for the predicted cumulative incidence functions are also constructed. Simulation studies are conducted to assess the finite sample performance of the proposed estimator. An application to the employment data illustrates the method and theory.

AMS 2000 subject classifications: Primary 62N01, 62N02; secondary 62P20.

Keywords and phrases: Competing risks data, Length-biased sampling, Fine-Gray model, Model checking techniques.

1. INTRODUCTION

In biomedical studies, medical statistics, economics, engineering, social sciences and many other areas, researchers often encounter multiple events data. In such datasets, the problem of competing risk occurs, as an individual may fail from different causes. For example, in the employment data used in Kadane and Woodworth (2004), employees may leave the company involuntarily (firing) or voluntarily for other reasons (retiring, early death, moving house, etc.). Here, voluntary and involuntary terminations are competing risks. The data include the days of termination from the start date of the study. The observed data are subject to right censoring because subjects may still be employed with the firm at the end of the study. In addition, some individuals are subject to left truncation, as they are hired after the study began. The investigators noted that employees who have longer periods of unemployment tend to work longer in the companies. That is, the sampled data are subject to length bias. In the literature, length-biased data are defined for left truncated data under the uniform distributed assumption of the truncation time. As a result, the probability of a survival time being observed is proportional to its length.

In the competing risks setting, it is of great interest to predict the cumulative incidence function, i.e., the probability of failure of a specific type in the presence of all of the competing risks. There are mainly two methods to incorporate covariate effects for the cumulative incidence functions (CIF) of competing risks data. One common method is by modeling the cause-specific hazard functions of different failure types. Prentice et al. (1978) was the first to use a proportional cause-specific hazards model when analyzing competing risks data. Benichou and Gail (1990) provided inference procedures for the cumulative incidence function, assuming that the cause-specific hazard function of interest follows a proportional hazards with an unknown constant or piecewise constant nuisance hazard function, and also assuming that the other competing risks are independent of the covariates. To relax these restrictive assumptions, Cheng et al. (1998) studied the predicted cumulative incidence function by constructing pointwise and simultaneous confidence intervals under the proportional cause-specific hazards model (Cox, 1972). As an alternative to the proportional hazards model, Shen and Cheng (1999) provided confidence bands for the cumulative incidence function under the additive risks model. Moreover, Scheike and Zhang (2003) extended this important inference to a flexible Cox-Aalen model for cause-specific hazards. For left-truncated competing risks data, the approach to analyzing cause-specific hazards functions can be generalized to the left-truncated version by adjusting the risk set (Andersen et al., 1993). Recently, for the competing risks data under length-biased sampling, Zhang et al. (2016) proposed a composite partial likelihood estimation for proportional cause-specific hazards model.

However, there is no simple one-to-one correspondence between the cause-specific hazard and the corresponding cumulative incidence function, because the cumulative incidence function depends on the rate of occurrence of all of the risks. Consequently, modeling the covariate effects on each cause-specific hazard yields a complex nonlinear relationship

*Corresponding Author.

for the cumulative incidence function. To address this issue, another popular model is the Fine-Gray proportional subdistribution hazards model (Fine and Gray, 1999), which directly links the regression coefficients with the cumulative incidence function for right-censored competing risks data. Sun et al. (2006) proposed a Cox-Aalen subdistribution hazards model for right-censored competing risks data. Scheike et al. (2008) proposed a binomial regression method for the cumulative incidence curve. For the left-truncated and right-censored competing risks data, Zhang et al. (2011) proposed two truncation-censoring probability weights for the proportional subdistribution hazards model, and Geskus (2011) proposed an alternative estimate procedure based on martingale theory. Nevertheless, to the best of our knowledge, it is unknown how to fit Fine-Gray proportional subdistribution hazards model for length-biased competing risks data.

The primary goal of the present paper is to propose an estimating equation for Fine-Gray proportional subdistribution hazards model for competing risks data under length-biased sampling. The major challenge for estimating the covariate effects on the subdistribution hazard rates under length-biased sampling is the informative censoring induced by both length-biased sampling and the presence of competing risks. There may be strong potential dependence between the failure time and the right-censoring time due to length-biased sampling. Furthermore, the model structure assumed for the target population is often different from that for the observed length-biased data. Specially, for the classic survival analysis with a single type of event, many authors have proposed semiparametric methods for the Cox model under the length-biased sampling, Wang et al. (1993), Wang (1996), Ghosh (2008), Tsai (2009), Qin and Shen (2010), Qin et al. (2011), and Huang et al. (2012), Zhang et al. (2014), and among others. However, these approaches do not provide a straightforward way to analyze competing risks data under length-biased sampling. Moreover, the developed methods for left-truncated and right-censored competing risks data, for example, Shen (2011), Zhang et al. (2011), and Geskus (2011), do not seem to have straightforward extensions to the competing risks data under length-biased sampling. Thus, some further novel method development is required to explore the length-biased data structure.

The remainder of the paper is organized as follows. In Section 2, an estimating equation estimator is derived and its large sample properties are presented. In Section 3, we also propose graphical and numerical methods for assessing the adequacy of Fine-Gray model for length-biased competing risks data, based on the cumulative residual processes. In Section 4, the prediction of the cumulative incidence function, along with its pointwise confidence intervals and simultaneous confidence bands, are presented. Substantial simulation studies and an application of the employment data are analyzed in Section 5 to evaluate the performance of the proposed method. All of the technical details are presented in the Appendix.

2. ESTIMATION PROCEDURES

2.1 Data and notations

Let W^0 be the calendar time of the initial event incidence onset, T^0 be the time from the initial event incidence to the failure event, and let $\epsilon \in \{1, ..., K\}$ be the cause of failure. The sampling time ξ is assumed to be independent of (W^0, T^0). In a prevalent population, an individual would be sampled at time ξ only if $T^0 \geq \xi - W^0 > 0$. Denote (W, T) as the random variables from the prevalent population. We drop the superscript ^{0} to emphasize that the failure time T in the prevalent population must exceed $A = \xi - W$, which is a left truncation time. Due to the end of the study or loss of follow-up, the observation of failure time $T = A + V$ in the prevalent cohort is subject to right censoring. The residual censoring time C, measured from recruitment to censoring, is usually assumed to be independent of (T, A, ϵ), conditional on a covariate vector $\mathbf{z}$. However, it is worth noting that the total censoring time $A + C$ and the survival time T are dependent, as they share the same A. Let $X = \min(T, A + C)$ be the follow-up time until failure or censoring. Let $\delta = I(T \leq A + C)$ be the indicator of censoring, where $I(\cdot)$ is the indicator function. Let $\widetilde{\epsilon} = \delta\epsilon$ be the observed cause of failure. Note that the true cause ϵ of failure can only be observed for those individuals without censoring. The observed data consist of n replicates of $(A, X, \widetilde{\epsilon})$, denoted by $(A_i, X_i, \widetilde{\epsilon}_i)$, for $i = 1, ..., n$.

To formulate the length-biased sampling, we let $f(t)$ and $S(t)$ be the density function and survival function of the latent failure time T^0, respectively. As in Huang et al. (2012), the following two assumptions are imposed throughout the paper.

Assumption 1. The variable T^0 is independent of W^0.

Assumption 2. W^0 has a constant density function, which implies that the initial event incidence occurs over calendar time at a constant rate.

Lancaster (1992) showed the joint density function of (A, T) evaluated at (a, t) is

$$(T, A)|\mathbf{Z} = \mathbf{z} \sim \frac{f(t|\mathbf{z})I(t > a > 0)}{\mu(\mathbf{z})},$$

where $\mu(z) = E\left[T^0|\mathbf{Z} = \mathbf{z}\right]$. Then, conditional on $\mathbf{Z} = \mathbf{z}$, the survival time T has a length-biased density function $f_T(t|\mathbf{z}) = tf(t|\mathbf{z})/\mu(\mathbf{z})$.

To proceed, let $F_k(t|\mathbf{z}) = P(T^0 \leq t, \epsilon = k|\mathbf{z})$ be the cumulative incidence function (CIF) of cause k given $\mathbf{z}$, and

$$\lambda_k(t|\mathbf{z}) = -\frac{d \log\{1 - F_k(t|\mathbf{z})\}}{dt}$$

be the subdistribution hazard function. Different from the cause-specific hazard function, the subdistribution hazard function could model the cumulative incidence function directly in the relationship $F_k(t|\mathbf{z}) = 1 - \exp\left\{-\int_0^t \lambda_k(s|\mathbf{z})ds\right\}$, see more details in Fine and Gray (1999). The Fine-Gray proportional subdistribution hazards model is given by

$$\lambda_k(t|\mathbf{z}) = \lambda_{0k}(t)e^{\boldsymbol{\beta}_{0k}^\top \mathbf{z}}, \tag{1}$$

where $\lambda_{0k}(\cdot)$ is an unknown nonnegative function in t, and $\boldsymbol{\beta}_{0k}$ is a $p \times 1$ regression parameter. Under model (1), the cumulative incidence function is given by

$$F_k(t|\mathbf{Z}) = 1 - \exp\left\{-e^{\boldsymbol{\beta}_{0k}^\top \mathbf{Z}} \int_0^t \lambda_{0k}(s)ds\right\}.$$

2.2 Unbiased estimating equation

To better understand the structure of length-biased data with competing risks, we start with the observation for A and T. It is easy to show that the joint subdensity of $(A, T, \epsilon = k)$ is

$$P(A = a, T = t, \epsilon = k|\mathbf{Z} = \mathbf{z}) = \frac{f_k(t)}{\mu(\mathbf{z})} I(t > a > 0),$$

where $f_k(t|\mathbf{Z}) = \frac{d}{dt}F_k(t|\mathbf{Z}) = \lambda_k(t|\mathbf{Z})\{1 - F_k(t|\mathbf{Z})\}$ is the subdensity function of cause k. Hence, given $\mathbf{Z} = \mathbf{z}$, $(A, V, \epsilon = k)$ shares a subdensity function with $(A, T, \epsilon = k)$, i.e.,

$$f_{A,V}(a, v, \epsilon = k|\mathbf{z}) = \frac{f_k(t|\mathbf{z})}{\mu(\mathbf{z})}, \quad t = a + v > 0.$$

In the presence of potential censoring, the probability of observing a pair of uncensored data for cause k is

$$\begin{aligned} & P(X = x, A = a, \widetilde{\epsilon} = k|\mathbf{Z} = \mathbf{z}) \\ = \; & P(A = a, V = x - a, C \geq x - a, \epsilon = k|\mathbf{Z} = \mathbf{z}) \\ = \; & P(A = a, V = x - a, \epsilon = k|\mathbf{Z} = \mathbf{z})P(C \geq x - a|\mathbf{z}) \\ = \; & \frac{f_k(x|\mathbf{z})}{\mu(\mathbf{z})} S_C(x - a|\mathbf{z})I(x > a), \end{aligned}$$

where $S_C(\cdot|\mathbf{z})$ is the survival distribution of residual censoring C conditional on $\mathbf{z}$, and the second equality holds by the conditional independence between C and (A, V, ϵ) given $\mathbf{z}$. Thus,

$$\begin{aligned} & P(X = x, \widetilde{\epsilon} = k|\mathbf{Z} = \mathbf{z}) \\ & = \int P(X = x, \delta = 1, \epsilon = k, A = a|\mathbf{Z} = \mathbf{z})da \\ & = \int_0^x \frac{f_k(x|\mathbf{z})}{\mu(\mathbf{z})} S_C(x - a|\mathbf{z})da = \frac{f_k(x|\mathbf{z})\omega_c(x|\mathbf{z})}{\mu(\mathbf{z})}, \end{aligned} \tag{2}$$

where $\omega_c(t|\mathbf{z}) = \int_0^t S_C(u|\mathbf{z})du$. Then, we can derive

$$\begin{aligned} & \mathrm{E}\left[\frac{I(X > x, \widetilde{\epsilon} = k)}{\omega_c(X|\mathbf{z})}\middle|\mathbf{Z} = \mathbf{z}\right] \\ & = \int_x^\infty \frac{1}{\omega_c(t|\mathbf{z})} \frac{f_k(t|\mathbf{z})\omega_c(t|\mathbf{z})}{\mu(\mathbf{z})} dt \\ & = \frac{F_k(\infty|\mathbf{z}) - F_k(x|\mathbf{z})}{\mu(\mathbf{z})}. \end{aligned} \tag{3}$$

Moreover, by some algebraic manipulations and (2), we can have

$$\mathrm{E}\left[\frac{I(\widetilde{\epsilon} > 0, \widetilde{\epsilon} \neq k)}{\omega_c(X|\mathbf{Z})}\middle|\mathbf{Z} = \mathbf{z}\right] = \frac{1 - F_k(\infty|\mathbf{z})}{\mu(\mathbf{z})}. \tag{4}$$

Combining equations (3) and (4), one can obtain

$$\begin{aligned} \frac{1 - F_k(x|\mathbf{Z})}{\mu(\mathbf{Z})} & = \frac{1 - F_k(\infty|\mathbf{Z})}{\mu(\mathbf{Z})} + \frac{F_k(\infty|\mathbf{Z}) - F_k(x|\mathbf{Z})}{\mu(\mathbf{Z})} \\ & = \mathrm{E}\left[\frac{Y_{ik}(x)}{\omega_c(X|\mathbf{Z})}\right], \end{aligned}$$

where $Y_{ik}(x) = [I(X_i \geq x, \widetilde{\epsilon}_i = k) + I(\widetilde{\epsilon}_i > 0, \widetilde{\epsilon}_i \neq k)]$.

To proceed, we introduce some counting process notations. We define

$$\begin{aligned} N_{ik}(t) & = I(X_i \leq t, \widetilde{\epsilon}_i = k), \\ M_{ik}(t) & = N_{ik}(t) - \int_0^t \pi_{ik}(s|\mathbf{Z}_i)e^{\boldsymbol{\beta}_{0k}^\top \mathbf{Z}_i} d\Lambda_{0k}(s), \end{aligned}$$

where $\pi_{ik}(t|\mathbf{Z}_i) = Y_{ik}(t)\omega_c(t|\mathbf{Z}_i)/\omega_c(X_i|\mathbf{Z}_i)$.

By equations (3) and (4), we have

$$\begin{aligned} & \mathrm{E}M_{ik}(t) \\ & = \mathrm{E}\Bigg[\mathrm{E}\Bigg\{I(X_i \leq t, \epsilon_i = k, C_i \geq X_i - A_i) \\ & \quad - \int_0^t Y_{ik}(t)\omega_c(s|\mathbf{Z}_i)/\omega_c(X_i|\mathbf{Z}_i)e^{\boldsymbol{\beta}_{0k}^\top \mathbf{Z}_i} d\Lambda_{0k}(s)\Bigg\}|\mathbf{Z}_i\Bigg] \\ & = \mathrm{E}\left[\int_0^t dx \int_0^x \frac{f_k(x|\mathbf{Z}_i)}{\mu(\mathbf{Z}_i)} S_c(x - a|\mathbf{Z}_i)da\right] \\ & \quad - \mathrm{E}\left[\int_0^t \mathrm{E}\left\{\frac{Y_{ik}(t)}{\omega_c(X_i|\mathbf{Z}_i)}|\mathbf{Z}_i\right\}\omega_c(s|\mathbf{Z}_i)\lambda_k(s|\mathbf{Z}_i)ds\right] \\ & = \mathrm{E}\left[\int_0^t \frac{f_k(x|\mathbf{Z}_i)}{\mu(\mathbf{Z}_i)}\omega_c(x|\mathbf{Z}_i)dx\right] \\ & \quad - \mathrm{E}\left[\int_0^t \frac{1 - F_k(s|\mathbf{Z}_i)}{\mu(\mathbf{Z}_i)}\omega_c(s|\mathbf{Z}_i)\lambda_k(s|\mathbf{Z}_i)ds\right] \\ & = 0. \end{aligned}$$

Motivated by the fact that $M_{ik}(t)$ is a zero-mean process, we propose the estimating equations,

$$\sum_{i=1}^n \int_0^\tau \mathbf{Z}_i \left[dN_{ik}(t) - \pi_{ik}(t|\mathbf{Z}_i)e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_i} d\Lambda_{0k}(t)\right] = 0, \tag{5}$$

$$\sum_{i=1}^{n}\left[dN_{ik}(t)-\pi_{ik}(t|\mathbf{Z}_i)e^{\boldsymbol{\beta}_k^\top\mathbf{Z}_i}d\Lambda_{0k}(t)\right]=0, \tag{6}$$

where τ is a predetermined constant. It follows from (6) that

$$d\Lambda_{0k}(t)=\sum_{i=1}^{n}dN_{ik}(t)\Big/\sum_{i=1}^{n}\pi_{ik}(t|\mathbf{Z}_i)e^{\boldsymbol{\beta}_k^\top\mathbf{Z}_i}.$$

If we plug this into (5), we can obtain an unbiased estimating equation for $\boldsymbol{\beta}_{0k}$,

$$\begin{aligned}&U_k(\boldsymbol{\beta}_k)\\&=\sum_{i=1}^{n}\int_0^{\tau}\left[\mathbf{Z}_i-\frac{\sum_{j=1}^{n}\mathbf{Z}_j\pi_{jk}(t|\mathbf{Z}_j)e^{\boldsymbol{\beta}_k^\top\mathbf{Z}_j}}{\sum_{j=1}^{n}\pi_{jk}(t|\mathbf{Z}_j)e^{\boldsymbol{\beta}_k^\top\mathbf{Z}_j}}\right]dN_{ik}(t)\\&=0.\end{aligned} \tag{7}$$

Remark. Note that this estimating equation reduces to the second estimating equation of Qin and Shen (2010) if $K=1$ and the censoring variable C is independent of the covariate $\mathbf{Z}$. Therefore, this paper can be viewed as an extension of Qin and Shen (2010) to length-biased data with competing risks. However, the proposed estimator is beyond such extension. Under length-biased sampling, the proposed estimating equation for Fine-Gray proportional subdistribution hazards model is more complex than the work of Qin and Shen (2010), due to the existence of competing risks. In addition, as we will show in Section 4, the cumulative incidence function is of great interest, which is more complicated since it is improper (i.e., $F_k(\infty|\mathbf{Z})<1$). We will provide the prediction of CIF, along with its pointwise confidence intervals and simultaneous confidence bands.

2.3 The proposed estimator

In the estimating equation (7), however, $S_c(t|\mathbf{z})$ is always unknown in practice. To estimate the weighting function $S_c(t|\mathbf{z})$, one can use the local Kaplan-Meier estimator $\widehat{S}_c(t|\mathbf{z})$,

$$\begin{aligned}&\widehat{S}_c(t|\mathbf{z})\\&=\prod_{j=1}^{n}\left[1-\frac{B_{nj}(\mathbf{z})}{\sum\limits_{l=1}^{n}I(R_l\geq R_j)B_{nl}(\mathbf{z})}\right]^{I(R_j\leq t,\delta_j=0)},\end{aligned} \tag{8}$$

where $R_i=X_i-A_i$, and $\{B_{nj}(\mathbf{z}),\ j=1,...,n\}$ is a sequence of nonnegative weights adding up to 1. When $B_{nj}(\mathbf{z})=1/n$ for all j, $\widehat{S}_c(t|\mathbf{z})$ reduces to the classic Kaplan-Meire estimator of the survival function. As suggested in Wang and Wang (2009), one may use

$$B_{nj}(\mathbf{z})=\mathcal{K}\left(\frac{\mathbf{z}-\mathbf{z}_j}{h_n}\right)\left[\sum_{l=1}^{n}\mathcal{K}\left(\frac{\mathbf{z}-\mathbf{z}_l}{h_n}\right)\right]^{-1},$$

where $\mathcal{K}\left(\frac{\mathbf{z}-\mathbf{z}_j}{h_n}\right)=\mathcal{K}\left(\frac{z_1-z_{j1}}{h_n},...,\frac{z_p-z_{jp}}{h_n}\right)$ is a multivariate kernel function, z_{jm} is the mth element of $\mathbf{z}_j$, and $h_n>0$ is the bandwidth. We adopt the commonly used product kernel function $\mathcal{K}(u_1,...,u_p)=\prod_{i=1}^{p}K(u_i)$ with $K(\cdot)$ being a univariate kernel function. As elaborated in Leng and Tong (2014), we can choose the bi-quadratic kernel $K(x)=\frac{15}{16}(1-x^2)^2I(|x|\leq 1)$ for the univariate covariate ($p=1$). However, for multiple continuous covariates with $p\geq 2$, we should use a product kernel function with a higher order kernel for each covariate. For example, if $p=2$, we use $K(x)=\frac{15}{32}(3-10x^2+7x^4)I(|x|\leq 1)$. More details for the higher order kernel can be found in Müller (1988).

Replacing $\omega_c(t|\mathbf{z})$ by $\widehat{\omega}_c(t|\mathbf{z})=\int_0^t\widehat{S}_c(s|\mathbf{z})ds$, where $\widehat{S}_c(t|\mathbf{z})$ is defined as in (8), one can obtain an estimator $\widehat{\boldsymbol{\beta}}_k$ for $\boldsymbol{\beta}_{0k}$ by solving the following estimating equation

$$\begin{aligned}&\widehat{U}_k(\boldsymbol{\beta}_k)\\&=\sum_{i=1}^{n}\int_0^{\tau}\left[\mathbf{Z}_i-\frac{\sum_{j=1}^{n}\mathbf{Z}_j\widehat{\pi}_{jk}(t|\mathbf{Z}_j)e^{\boldsymbol{\beta}_k^\top\mathbf{Z}_j}}{\sum_{j=1}^{n}\widehat{\pi}_{jk}(t|\mathbf{Z}_j)e^{\boldsymbol{\beta}_k^\top\mathbf{Z}_j}}\right]dN_{ik}(t)\\&=0.\end{aligned} \tag{9}$$

where $\widehat{\pi}_{jk}(t|\mathbf{Z}_j)=Y_{jk}(t)\widehat{\omega}_c(t|\mathbf{Z}_j)/\widehat{\omega}_c(X_j|\mathbf{Z}_j)$. Clearly, the proposed estimating equation requires the information of the distribution function for censoring variable C. As argued in Qin and Shen (2010), while $\widehat{S}_c(t)$ could be close to zero at the tail, the integral $\widehat{\omega}_c(X_j|\mathbf{Z}_j)$ of $\widehat{S}_c(t)$ will not go to zero at the tail. Thus, the proposed estimating equation (9) is stable, at least from our numeric studies.

For comparison, we propose a naive estimating equation for left-truncated data, which does not require estimating the survival function of the censoring variable. Let $h(\cdot)$ be the marginal density function of covariate $\mathbf{Z}$. By model (1), we have

$$\begin{aligned}&\mathrm{E}\left[\mathbf{Z}|X=x,A=a,\widetilde{\epsilon}=k\right]\\&=\frac{\int\mathbf{z}f_k(x|\mathbf{z})S_C(x-a|\mathbf{z})h(\mathbf{z})/\mu(\mathbf{z})d\mathbf{z}}{\int f_k(x|\mathbf{z})S_C(x-a|\mathbf{z})h(\mathbf{z})/\mu(\mathbf{z})d\mathbf{z}}\\&=\frac{\mathrm{E}\left[\mathbf{Z}e^{\boldsymbol{\beta}^\top\mathbf{Z}}\{1-F_k(x|\mathbf{Z})\}/\mu(\mathbf{Z})\right]}{\mathrm{E}\left[e^{\boldsymbol{\beta}^\top\mathbf{Z}}\{1-F_k(x|\mathbf{Z})\}/\mu(\mathbf{Z})\right]}.\end{aligned} \tag{10}$$

It is sufficient to re-express the term $\{1-F_k(x|\mathbf{Z})\}/\mu(\mathbf{Z})$. By assuming that residual censoring C is independent of covariate $\mathbf{Z}$, tedious calculation shows

$$\begin{aligned}&[1-F_k(x|\mathbf{Z})]\frac{\omega_c(x)}{\mu(\mathbf{Z})}\\&=\mathrm{E}\Big[I\left\{X\geq x,A\leq x,\widetilde{\epsilon}=k\right\}+I\big\{A\leq x,\widetilde{\epsilon}>0,\widetilde{\epsilon}\neq k\big\}|\mathbf{Z}\Big].\end{aligned} \tag{11}$$

Combing equations (10) and (11), we can obtain

$$\begin{aligned}&\mathrm{E}\left[\mathbf{Z}|X=x,\widetilde{\epsilon}=k,A=a\right]\\&=\frac{\mathrm{E}\left[\mathbf{Z}e^{\boldsymbol{\beta}_k^\top\mathbf{Z}}\mathrm{E}\{I(X\geq x\geq A,\widetilde{\epsilon}=k)+I(A\leq x,\widetilde{\epsilon}>0,\widetilde{\epsilon}\neq k)|\mathbf{Z}\}\right]}{\mathrm{E}\left[e^{\boldsymbol{\beta}_k^\top\mathbf{Z}}\mathrm{E}\{I(X\geq x\geq A,\widetilde{\epsilon}=k)+I(A\leq x,\widetilde{\epsilon}>0,\widetilde{\epsilon}\neq k)|\mathbf{Z}\}\right]}.\end{aligned}$$

By sample analogy, one can construct the naive estimating equation,

$$\begin{aligned} & U_L(\boldsymbol{\beta}_k) \\ &= \sum_{i=1}^n I(\widetilde{\epsilon}_i = k) \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j \exp(\boldsymbol{\beta}_k^\top \mathbf{Z}_j) \pi_j^L(X_i)}{\sum_{j=1}^n \exp(\boldsymbol{\beta}_k^\top \mathbf{Z}_j) \pi_j^L(X_i)} \right] \\ &= 0, \end{aligned} \tag{12}$$

where $\pi_j^L(t) = I(X_j \geq t \geq A_j, \widetilde{\epsilon}_j = k) + I(A_j \leq X_j, \widetilde{\epsilon}_j > 0, \widetilde{\epsilon}_j \neq k)$. It is emphasized that the summations in the fraction terms of naive estimating equation (12) can include both failure and censored times as long as the pair (A_j, X_j) satisfies the inequality condition.

2.4 Asymptotic theory

To derive the asymptotic properties of the proposed estimator, we introduce some notations. For $l = 0,\ 1,\ 2$, we define

$$\widehat{S}_l(\boldsymbol{\beta}_k, t) = \sum_{j=1}^n \mathbf{Z}_j^{\otimes l} e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_j} \widehat{\pi}_{jk}(t|\mathbf{Z}_j),$$

$$S_l(\boldsymbol{\beta}_k, t) = \sum_{j=1}^n \mathbf{Z}_j^{\otimes l} e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_j} \pi_{jk}(t|\mathbf{Z}_j),$$

$$s_l(\boldsymbol{\beta}_k, t) = \mathrm{E}\left[\mathbf{Z}_j^{\otimes l} e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_j} \pi_{jk}(t|\mathbf{Z}_j)\right].$$

where $\mathbf{z}^{\otimes} = \mathbf{z}\mathbf{z}^\top$ for any vector $\mathbf{z}$. Let $Q_l(t|\mathbf{Z}_j) = \int_0^t \xi(V_l^*, \delta_l, s, \mathbf{Z}_j) ds$, where $\xi(V_l^*, \delta_l, s, \mathbf{Z}_j)$ is defined in Lemma A.1 in the Appendix. Let $\Sigma_k = \mathrm{E}(\phi_i(\boldsymbol{\beta}_{0k})^{\otimes 2})$, where $\phi_{ik}(\boldsymbol{\beta}_{0k}) = \int_0^\tau \left[\mathbf{Z}_i - \frac{S_1(\boldsymbol{\beta}_{0k}, t)}{S_0(\boldsymbol{\beta}_{0k}, t)}\right] dM_{ik}(t) - \eta_{ik}$ with

$$\eta_{lk} = \frac{\mathbf{Z}_l e^{\boldsymbol{\beta}_{0k}^T \mathbf{Z}_l} Q_l(Y_1|\mathbf{Z}_l)}{\omega_c(X_1|\mathbf{Z}_l)} \sum_{i=1}^n \int_0^\tau \frac{\pi_{jk}(t|\mathbf{Z}_l)}{S_0(\boldsymbol{\beta}_{0k}, t)} dN_{ik}(t).$$

We also define

$$\Gamma_k(\boldsymbol{\beta}_k) = \mathrm{E}\left[I(\widetilde{\epsilon}_i = k)\left\{\frac{S_2(\boldsymbol{\beta}_k, X_i)}{S_0(\boldsymbol{\beta}_k, X_i)} - \left(\frac{S_1(\boldsymbol{\beta}_k, X_i)}{S_0(\boldsymbol{\beta}, X_i)}\right)^{\otimes 2}\right\}\right],$$

and $\Gamma_k \equiv \Gamma_k(\boldsymbol{\beta}_{0k})$.

When $S_c(\cdot|\mathbf{Z})$ is known, the estimating equation (7) can be asymptotically represented by the following independent and identical summation of the mean zero process, $U_k(\boldsymbol{\beta}_{0k}) = \sum_{i=1}^n \int_0^\tau \left[\mathbf{Z}_i - \frac{s_1(\boldsymbol{\beta}_{0k}, t)}{s_0(\boldsymbol{\beta}_{0k}, t)}\right] dM_{ik}(t)$. However, in practice, $S_c(\cdot|\mathbf{Z})$ is always unknown, and we can replace it with its consistent local Kaplan-Meier estimator $\widehat{S}_c(t|\mathbf{Z})$ for the censoring time. Thus, the estimating equation (9) can be rewritten as

$$\widehat{U}_k(\boldsymbol{\beta}_k) = \sum_{i=1}^n \int_0^\tau \left[\mathbf{Z}_i - \frac{\widehat{S}_1(\boldsymbol{\beta}_k, t)}{\widehat{S}_0(\boldsymbol{\beta}_k, t)}\right] dN_{ik}(t).$$

Using the modern empirical process theory, we can derive the consistency and asymptotic normality of the proposed estimator.

Theorem 1. *Under the regular conditions in the Appendix, there exists a unique solution $\widehat{\boldsymbol{\beta}}_k$ to the estimating equation $\widehat{U}_k(\boldsymbol{\beta}_k) = 0$, and $\widehat{\boldsymbol{\beta}}_k$ converges to $\boldsymbol{\beta}_{0k}$ in probability. Moreover, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{0k})$ is asymptotically zero-mean normal with the covariance matrix $\Gamma_k^{-1}\Sigma_k\Gamma_k^{-1}$.*

Remark. As pointed out by a referee, the local Kaplan-Meier estimator may converge at a slower rate if the dimension of the covariate is high. However, it is common to assume that the residual censoring C depends on part of the covariate $\mathbf{Z}$. Consequently, the convergence rate of the local Kaplan-Meier estimator could be slower when the dimension of the covariate is low. Fortunately, the dimension of the covariate is two both in our simulation studies and the real data example.

The variance-covariance of $\widehat{\boldsymbol{\beta}}_k$ can be consistently estimated by $\widehat{\Gamma}_k(\widehat{\boldsymbol{\beta}}_k)\widehat{\Sigma}_k^{-1}\widehat{\Gamma}_k(\widehat{\boldsymbol{\beta}}_k)^T$, where

$$\widehat{\Gamma}_k(\boldsymbol{\beta}_k) = n^{-1}\sum_{i=1}^n I(\widetilde{\epsilon}_i = k)\left[\frac{\widehat{S}_2(\boldsymbol{\beta}_k, X_i)}{\widehat{S}_0(\boldsymbol{\beta}_k, X_i)} - \left\{\frac{\widehat{S}_1(\boldsymbol{\beta}_k, X_i)}{\widehat{S}_0(\boldsymbol{\beta}_k, X_i)}\right\}^{\otimes 2}\right],$$

and $\widehat{\Sigma}_k = n^{-1}\sum_{i=1}^n \widehat{\phi}_{ik}(\widehat{\boldsymbol{\beta}}_k)^{\otimes 2}$, and $\widehat{\phi}_{ik}$ is the corresponding estimator of ϕ_{ik} by replacing the population quantities with the sample quantities.

Given the estimator $\widehat{\boldsymbol{\beta}}_k$ for $\boldsymbol{\beta}_{0k}$, a natural estimator for the cumulated baseline hazard function that is similar to Breslow's estimator can be proposed,

$$\widehat{\Lambda}_{0k}(t, \widehat{\boldsymbol{\beta}}_k) = \sum_{i=1}^n \int_0^t \frac{dN_{ik}(u)}{\widehat{S}_0(\widehat{\boldsymbol{\beta}}_k, u)}, \quad t \in [0, \tau].$$

In the Appendix, we show that $\sqrt{n}\big[\widehat{\Lambda}_{0k}(t, \widehat{\boldsymbol{\beta}}_k) - \Lambda_{0k}(t)\big]$ converges to a zero-mean Gaussian process with covariance function $\mathrm{E}\left[\varphi_{1k}(t)\varphi_{1k}(t)^\top\right]$, where $\varphi_{ik}(t)$ is defined in the Appendix.

Another important concern is the choice of the bandwidth parameter h_n, because the proposed estimator $\widehat{\boldsymbol{\beta}}_k$ involves the local Kaplan-Meier estimates. As in Wang and Wang (2009), we observe that the results are not very sensitive to h_n. In practical data analysis, one method for the bandwidth selection is based on L-fold cross-validation (Tian et al., 2005; Fan et al., 2006). Specifically, we first divide the dataset randomly into L parts with roughly equal size. For the lth part D_l, we use the rest $L-1$ parts of the data to fit the model. The lth prediction error is given by

$$\mathrm{PE}_l(h_n) = \sum_{i \in D_l} \int_0^\tau \left[N_{ik}(t) - \widehat{\mathrm{E}}N_{ik}(t)\right]^2 d\Big\{\sum_{j \in D_l} N_{jk}(t)\Big\},$$

where $\widehat{\mathrm{E}}\left(N_{ik}(t)\right) = \int_0^t \widehat{\pi}_{ik}(s|\mathbf{Z}_i) e^{\mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}}_k^{(-l)}} d\widehat{\Lambda}_{0k}(s, \widehat{\boldsymbol{\beta}}_k^{(-l)})$, and $\widehat{\boldsymbol{\beta}}_k^{(-l)}$ is estimated using the data from all of the subgroups other than D_l. The optimal bandwidth can be obtained by minimizing the total prediction error $\mathrm{PE}(h_n) = \sum_{l=1}^{L} \mathrm{PE}_l(h_n)$ with respect to h_n.

3. MODEL CHECKING TECHNIQUES

In this section, we propose model checking techniques for the adequacy of model (1) by using cumulative sums of the residuals. Although Li et al. (2015) proposed model checking Fine-Gray model with right censoring data, their method could not be applied to our context of Fine-Gray model under length-biased sampling. To proceed, we define the ith residual for cause k,

$$\widehat{M}_{ik}(t) = N_{ik}(t) - \int_0^t \widehat{\pi}_{ik}(s|\mathbf{Z}_j)e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i} d\widehat{\Lambda}_{0k}(s),$$

which is the difference between the observed and expected number of failures for subject i due to cause k by time t. Inspired by the basic idea of Lin et al. (1993), we develop a class of graphical and numerical methods by using the cumulative sums of the these residuals.

We first consider checking the functional forms of the covariates. For the jth component of $\mathbf{Z}$, we consider

$$W_{jk}(z) = n^{-1/2}\sum_{i=1}^n \int_0^\tau I(Z_{ij} \le z) d\widehat{M}_{ik}(u).$$

As shown in Appendix, the null distribution of $W_{jk}(z)$ can be approximated by the zero-mean Gaussian process

$$\begin{aligned}
&\widehat{W}_{jk}(z)\\
&= n^{-1/2}\sum_{i=1}^n \int_0^\tau I(Z_{ij} \le z) d\widehat{M}_{ik}(u)G_i\\
&- n^{-1/2}\sum_{i=1}^n \int_0^\tau I(Z_{ij} \le z)\frac{\widehat{\pi}_{ik}(u|\mathbf{Z}_i)}{\widehat{\omega}_c(X_i|\mathbf{Z}_i)}\\
&\quad \times \sum_{l=1}^n Q_l(X_i|\mathbf{Z}_i)e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i} d\widehat{\Lambda}_{0k}(u)G_i\\
&- n^{-1/2}\sum_{i=1}^n \int_0^\tau I(Z_{ij} \le z)e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i}\widehat{\pi}_{ik}(u|\mathbf{Z}_i)\\
&\quad \times \left[n^{-1}\widehat{\Gamma}_k^{-1}(\widehat{\boldsymbol{\beta}}_k)\sum_{l=1}^n \widehat{\phi}_{lk}(\widehat{\boldsymbol{\beta}})\right]^\top \mathbf{Z}_i d\widehat{\Lambda}_{0k}(u)G_i\\
&- n^{-1/2}\sum_{i=1}^n \int_0^\tau I(Z_{ij} \le z)e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i}\widehat{\pi}_{ik}(u|\mathbf{Z}_i)n^{-1}\\
&\quad \times \sum_{l=1}^n d\widehat{\varphi}_{lk}(u)G_i,
\end{aligned}$$

where $\{G_1, \ldots, G_n\}$ are independently generated from the standard normal distribution, which are independent of the observed data, and $\widehat{\varphi}_{ik}$ is the corresponding estimator of φ_{ik} by replacing the population quantities with the sample quantities. We can plot the observed process and a few simulated limiting processes under the model (1) versus the covariate values for visually checking the linearity of functional form. Moreover, the p-value of the test is obtained by firstly generating a large number of realizations from $\sup_z |\widehat{W}_{jk}(z)|$, and then calculate the percentage of those greater than the observed value of $\sup_z |W_{jk}(z)|$. For the case with more than one covariate ($p > 1$), a global checking of the model can be accomplished by $\sup_{\mathbf{z}} |W_k(\mathbf{z})|$, where $W_k(\mathbf{z}) = n^{-1/2}\sum_{i=1}^n \int_0^\tau I(\mathbf{Z}_i \le \mathbf{z}) d\widehat{M}_{ik}(u)$. Here, $\{\mathbf{Z}_i \le \mathbf{z}\}$ means that each component of $\mathbf{Z}_i$ is no greater than the corresponding component of $\mathbf{z}$.

To check the proportional hazards assumption for the jth covariate component, we consider the standard score type process

$$V_{jk}(t) = \left(\widehat{\Sigma}_{jj,k}^{-1}\right)^{1/2} n^{-1/2}\widehat{U}_{jk}(\widehat{\boldsymbol{\beta}}_k, t),$$

where $\widehat{\Sigma}_{jj,k}^{-1}$ is the jth diagonal element of $\widehat{\Sigma}_k^{-1}$ and

$$\widehat{U}_{jk}(\widehat{\boldsymbol{\beta}}_k, t) = \sum_{i=1}^n \int_0^t \left[Z_{ij} - \frac{\widehat{S}_{1j}(\widehat{\boldsymbol{\beta}}_k, u)}{\widehat{S}_0(\widehat{\boldsymbol{\beta}}_k, u)}\right] d\widehat{M}_{ik}(u),$$

is the jth component of estimating equation $\widehat{U}_k(\widehat{\boldsymbol{\beta}}_k, t)$. As shown in Appendix, the null distribution of $V_{jk}(z)$ can be approximated by the zero-mean Gaussian process

$$\begin{aligned}
&\widehat{V}_{jk}(z)\\
&= \left(\widehat{\Sigma}_{jj,k}^{-1}\right)^{1/2}\Bigg[n^{-1/2}\sum_{i=1}^n \int_0^t \left[Z_{ij} - \frac{\widehat{S}_{1j}(\widehat{\boldsymbol{\beta}}_k, u)}{\widehat{S}_0(\widehat{\boldsymbol{\beta}}_k, u)}\right] d\widehat{M}_{ik}(u)G_i\\
&- n^{-1/2}\sum_{i=1}^n \int_0^t \left[Z_{ij} - \frac{\widehat{S}_{1j}(\widehat{\boldsymbol{\beta}}_k, u)}{\widehat{S}_0(\widehat{\boldsymbol{\beta}}_k, u)}\right]\frac{\widehat{\pi}_{ik}(u|\mathbf{Z}_i)}{\widehat{\omega}_c(X_i|\mathbf{Z}_i)}\\
&\quad \times \sum_{l=1}^n Q_l(X_i|\mathbf{Z}_i)e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i} d\widehat{\Lambda}_{0k}(u)G_i\\
&- n^{-1/2}\sum_{i=1}^n \int_0^t \left[Z_{ij} - \frac{\widehat{S}_{1j}(\widehat{\boldsymbol{\beta}}_k, u)}{\widehat{S}_0(\widehat{\boldsymbol{\beta}}_k, u)}\right] e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i}\widehat{\pi}_{ik}(u|\mathbf{Z}_i)\\
&\quad \times \left[n^{-1}\widehat{\Gamma}_k^{-1}(\widehat{\boldsymbol{\beta}}_k)\sum_{l=1}^n \widehat{\phi}_{lk}(\widehat{\boldsymbol{\beta}})\right]^\top \mathbf{Z}_i d\widehat{\Lambda}_{0k}(u)G_i\\
&- n^{-1/2}\sum_{i=1}^n \int_0^t \left[\mathbf{Z}_{ij} - \frac{\widehat{S}_{1j}(\widehat{\boldsymbol{\beta}}_k, u)}{\widehat{S}_0(\widehat{\boldsymbol{\beta}}_k, u)}\right] e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i}\widehat{\pi}_{ik}(u|\mathbf{Z}_i)n^{-1}\\
&\quad \times \sum_{l=1}^n d\widehat{\varphi}_{lk}(u)G_i\Bigg].
\end{aligned}$$

The test statistic for checking the proportional hazards assumption of the jth covariate ($j = 1, \cdots, p$) is given by $\sup_{t\in[0,\tau]} |V_{jk}(t)|$. The p-value can be empirically estimated by the percentage of those $\sup_{t\in[0,\tau]} |\widehat{V}_{jk}(t)|$ greater than $\sup_{t\in[0,\tau]} |V_{jk}(t)|$ through generating many realizations of $\sup_{t\in[0,\tau]} |\widehat{V}_{jk}(t)|$. The overall test statistic for the joint additivity of all the covariates is given by $\sup_{t\in[0,\tau]} \sum_{j=1}^{p} |V_{jk}(t)|$.

4. PREDICTION OF THE CUMULATIVE INCIDENCE

In the competing risks setting, it is of great interest to predict the cumulative incidence function at time t for an individual with covariate $\mathbf{Z} = \mathbf{z}$. To address this issue, we first estimate the cumulative subdistribution hazard by

$$\widehat{\Lambda}_k(t|\mathbf{z}) = \int_0^t e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{z}} d\widehat{\Lambda}_{0k}(s, \widehat{\boldsymbol{\beta}}_k),$$

where $\widehat{\boldsymbol{\beta}}_k$ and $\widehat{\Lambda}_{0k}(t, \widehat{\boldsymbol{\beta}}_k)$ are the proposed estimators in Section 2.4. Thus, $F_k(t|\mathbf{z})$ can be consistently estimated by

$$\widehat{F}_k(t|\mathbf{z}) = 1 - \exp\{-\widehat{\Lambda}_k(t|\mathbf{z})\}.$$

The limiting distribution of the process $I_k(t|\mathbf{z}) = n^{1/2}\left(\widehat{F}_k(t|\mathbf{z}) - F_k(t|\mathbf{z})\right)$ is provided in the following theorem.

Theorem 2. *Under the regular conditions in the Appendix, as $n \to \infty$, the stochastic process $I_k(t|\mathbf{z})$ has an asymptotic representation $n^{-1/2}\sum_{i=1}^{n}\{1 - F_k(t|\mathbf{z})\}\gamma_{ik}(t|\mathbf{z}) + o_p(n^{-1/2})$, where $\gamma_{ik}(t|\mathbf{z})$ is defined in the Appendix. Moreover, $I_k(t|\mathbf{z})$ converges to a zero mean Gaussian process with the variance-covariance function,*

$$\begin{aligned}&\Sigma_{F_k}(t_1, t_2|\mathbf{z}) \\ &= \{1 - F_k(t_1|\mathbf{z})\}\{1 - F_k(t_2|\mathbf{z})\} E[\gamma_k(t_1|\mathbf{z})\gamma_k(t_2|\mathbf{z})].\end{aligned}$$

The variance function of the limiting distribution, $\sigma_k^2(t|\mathbf{z}) \equiv \Sigma_{F_k}(t, t|\mathbf{z})$, can be consistently estimated by $\widehat{\sigma}_k^2(t|\mathbf{z}) = n^{-1}\left\{1 - \widehat{F}_k(t|\mathbf{z})\right\}^2 \sum_{i=1}^{n} \widehat{\gamma}_{ik}(t|\mathbf{z})^2$, where $\widehat{\gamma}_{ik}(\cdot|\mathbf{z})$ is the corresponding estimator of $\gamma_{ik}(\cdot|\mathbf{z})$ with the population quantities of the sample quantities.

In order to construct pointwise confidence intervals and simultaneous confidence bands for $F_k(t|\mathbf{z})$, it is common to consider a class of transformed processes

$$J_k(t|\mathbf{z}) = n^{1/2} w(t, \mathbf{z}) \left[g\{\widehat{F}_k(t|\mathbf{z})\} - g\{F_k(t|\mathbf{z})\}\right],$$

where g is a known function whose derivative g' is continuous and nonzero, and $w(\cdot, \cdot)$ is a weight function that converges uniformly in probability to a nonnegative bounded function on $[\tau_1, \tau_2]$, $0 < \tau_1 < \tau_2 < \tau$. For example, let $g(t) = \log\{-\log(t)\}$, and let $w(t, \mathbf{z})$ be the reciprocal of an estimator for the standard deviation of $n^{1/2}g(\widehat{F}_k(t|\mathbf{z}))$, i.e., $w(t, \mathbf{z}) = \left[g'\{\widehat{F}_k(t|\mathbf{z})\}\widehat{\sigma}_k(t|\mathbf{z})\right]^{-1}$. By the functional delta-method, we can show that $J_k(t, \mathbf{z})$ is asymptotically equivalent to $w(t, \mathbf{z})g'\{\widehat{F}_k(t, \mathbf{z})\}I_k(t, \mathbf{z})$. Based on the asymptotic variance estimate of $J_k(t, \mathbf{z})$, one may calculate the pointwise $(1 - \alpha)$ confidence interval for $F_k(t|\mathbf{z})$ from

$$g^{-1}\left[g\left\{\widehat{F}_k(t|\mathbf{z})\right\} \pm n^{-1/2}g'\left\{\widehat{F}_k(t|\mathbf{z})\right\}\widehat{\sigma}_k(t|\mathbf{z})c_{\alpha/2}\right],$$

where g^{-1} is the inverse of g, and c_α is the α upper percentile of the standard normal distribution $N(0, 1)$. It is worth noting that this pointwise confidence interval does not depend on the weight function $w(t, \mathbf{z})$.

However, it is difficult to construct confidence bands for $F_k(t|\mathbf{z})$ based on its covariance function, because the limiting process $I_k(t|\mathbf{z})$ is very complicated. Fortunately, by using the results in Lin et al. (1993), the limiting distribution of $I_k(t|\mathbf{z})$ can be approximated by a zero-mean Gaussian process

$$\widehat{I}_k(t|\mathbf{z}) = n^{-1/2}\sum_{i=1}^{n}\left\{1 - \widehat{F}_k(t|\mathbf{z})\right\}\widehat{\gamma}_{ik}(t|\mathbf{z})G_i,$$

where $\{G_i, i = 1, ..., n\}$ are independent standard normal variables that are independent of the observed data $(A_i, X_i, \widetilde{\epsilon}_i, \mathbf{Z}_i)$ for $i = 1, ..., n$. Using similar arguments to those given in Lin et al. (2000), one can show that, conditional on the data, $\widehat{I}_k(t|\mathbf{z})$ converges weakly to a zero-mean Gaussian process.

To construct an $(1-\alpha)$ simultaneous confidence band for $F_k(t|\mathbf{z})$, one may use the simulated distribution of $\widehat{J}_k(t|\mathbf{z})$. One first needs to find the cutoff value $\widetilde{c}_{\alpha/2}$ satisfying

$$\mathbb{P}\left\{\max_{\tau_1\le t\le\tau_2}\left|w(t, \mathbf{z})g'\left\{\widehat{F}_k(t, \mathbf{z})\right\}\widehat{I}_k(t|\mathbf{z})\right| > \widetilde{c}_{\alpha/2}\right\} = \alpha.$$

In practice, $\widetilde{c}_{\alpha/2}$ can be approximated by the histogram of NB simulated realizations of $\widehat{I}_k(t|\mathbf{z})$ by repeatedly generating standard normal random samples $\{G_i\}$ but fixing the observed data $(A_i, X_i, \widetilde{\epsilon}_i, \mathbf{Z}_i)$. By using the weight function $w(t, \mathbf{z}) = \left[g'\{\widehat{F}_k(t|\mathbf{z})\}\widehat{\sigma}_k(t|\mathbf{z})\right]^{-1}$, one can construct the $(1-\alpha)$ confidence band for $F_k(t|\mathbf{z})$ on the interval $[\tau_1, \tau_2]$ as

$$g^{-1}\left[g\left\{\widehat{F}_k(t|\mathbf{z})\right\} \pm n^{-1/2}g'\left\{\widehat{F}_k(t|\mathbf{z})\right\}\widehat{\sigma}_k(t|\mathbf{z})\widetilde{c}_{\alpha/2}\right].$$

5. NUMERICAL STUDIES

5.1 Simulation examples

In this section, we demonstrate the good performance of the proposed estimator using two different examples.

Table 1. Simulation results based on 1000 replications.

		$n=200$				$n=400$			
		Proposed		Naive		Proposed		Naive	
$C\%$		$\widehat{\beta}_{11}$	$\widehat{\beta}_{12}$	$\widehat{\beta}_{11}$	$\widehat{\beta}_{12}$	$\widehat{\beta}_{11}$	$\widehat{\beta}_{12}$	$\widehat{\beta}_{11}$	$\widehat{\beta}_{12}$
					covariate-independent censoring				
35%	Bias	0.006	-0.003	-0.182	-0.002	0.004	0.000	-0.184	-0.013
	SD	0.229	0.158	0.257	0.175	0.156	0.108	0.177	0.122
	ESE	0.229	0.152	0.241	0.170	0.161	0.107	0.168	0.115
	CP	0.950	0.943	0.856	0.944	0.953	0.942	0.789	0.926
	MSE	**0.052**	**0.025**	0.099	0.031	**0.024**	**0.012**	0.065	0.015
15%	Bias	0.002	-0.001	-0.070	-0.092	-0.007	-0.001	-0.087	-0.095
	SD	0.199	0.145	0.223	0.154	0.145	0.096	0.165	0.108
	ESE	0.205	0.140	0.217	0.161	0.143	0.098	0.152	0.111
	CP	0.960	0.938	0.937	0.911	0.950	0.958	0.886	0.864
	MSE	**0.040**	**0.021**	0.055	0.032	**0.021**	**0.009**	0.035	0.021
					covariate-dependent censoring				
35%	Bias	0.050	0.051	0.085	-0.109	0.061	0.075	0.081	-0.138
	SD	0.199	0.199	0.213	0.213	0.150	0.150	0.162	0.166
	ESE	0.208	0.206	0.238	0.240	0.150	0.148	0.168	0.169
	CP	0.967	0.963	0.976	0.953	0.935	0.912	0.946	0.847
	MSE	**0.042**	**0.042**	0.052	0.057	**0.026**	**0.028**	0.033	0.047
15%	Bias	0.030	0.016	0.048	-0.139	0.030	0.038	0.049	-0.146
	SD	0.181	0.184	0.194	0.196	0.138	0.132	0.151	0.151
	ESE	0.187	0.186	0.211	0.220	0.133	0.132	0.147	0.154
	CP	0.959	0.950	0.973	0.929	0.932	0.934	0.929	0.826
	MSE	**0.034**	**0.034**	0.040	0.058	**0.020**	**0.019**	0.025	0.044

Proposed: the proposed estimator from estimating equation (9); Naive: the naive estimator from estimating equation (12); $\widehat{\beta}_1$ and $\widehat{\beta}_2$: the estimated coefficients; $C\%$: censoring ratio; Bias: the empirical bias; SD: the empirical standard error; ESE: the average estimated standard error; CP: 95% coverage probability; MSE: the average of estimated mean square error.

Example 1 (Covariate-independent censoring)**.** Data are generated from the following model. The sampling time ξ is set to 100, and W^0 is simulated from a uniform $[0, 100]$ distribution to mimic the incidence of a stable disease. With $K = 2$, as in Fine and Gray (1999) and Zhang et al. (2011), the subdistributions are given by

$$P(T^0 \le t, \epsilon = 1|\mathbf{Z}) = 1 - [1 - p\{1 - \exp(-\gamma_1 t)\}]^{\exp(\boldsymbol{\beta}_1^\top \mathbf{Z})},$$
$$P(T^0 \le t, \epsilon = 2|\mathbf{Z}) = (1-p)^{\exp(\boldsymbol{\beta}_1^\top \mathbf{Z})} \times \left\{1 - \exp(-\gamma_2 t \exp(\boldsymbol{\beta}_2^\top \mathbf{Z}))\right\},$$

where $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}) = (1, 1)^T$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}) = (0.5, -0.5)^T$, and the components of the covariate vector $\mathbf{Z} = (Z_1, Z_2)^T$ are independently generated from the Bernoulli distribution with $Pr(Z_1 = 1) = 0.5$ and the standard normal distribution, respectively. The covariate effects on the cumulative incidence function of cause 1 can be assessed via a proportional subdistribution hazard model. We set $\gamma_1 = 0.8$ and $\gamma_2 = 0.5$, and let $p = 0.75$ to generate a setting with a domain risk. To form a prevalent cohort, pairs of observations (W^0, T^0) are generated repeatedly until there are n pairs of observations satisfying the sampling constraint $W^0 + T^0 \ge \xi$. Residual censoring time C is independently generated from the Uniform(0, c) with different c's, corresponding to censoring rates of approximately 15% and 35%.

Example 2 (Covariate-dependent censoring)**.** We consider a second scenario in which the residual censoring variable C is dependent on the covariate $\mathbf{Z}$. The set-up is the same as in Example 1, except that (Z_1, Z_2) are generated from a bivariate normal distribution with mean vector $(0, 0)$ and covariance matrix $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and the residual censoring time C is generated from a Cox model, $\lambda_C(t|Z) = c\exp(Z_1 - Z_2)$, with different c's to control the censoring rates of approximately 15% and 35%.

For each scenario, 1000 datasets of size $n = 200, 400$ are generated. For comparison, the proposed estimator ("Proposed") from estimating equation (9) and the naive estimator ("naive") from estimating equation (12) are used. For Example 2 with covariate-independent censoring, the used kernel function is $K(x) = \frac{15}{32}(3 - 10x^2 + 7x^4)I(|x| \le 1)$ for two covariates. We only report the results using the bandwidth $h_n = 0.05$ for saving space, since the results using other bandwidths are similar and comparable. Table 1 reports the average biases ("Bias") of the estimators, the empirical standard errors ("SD"), the average estimated standard errors ("ESE"), the 95% coverage probabilities ("CP"), and the average of the estimated mean square errors ("MSE").

From Table 1, the biases of the proposed estimator in the covariate-independent censoring scenario are small. But the biases in the covariate-dependent censoring scenario are not very small. This is not strange because it requires the local Kaplan-Meier estimator. As a referee pointed out, the biases will shrink as the sample size increases. On the whole, the biases of the proposed estimator are of reasonable order and are generally negligible relative to the variance. The ESEs are close to SDs, and the empirical coverage probabilities are close to the nominal level. In addition, the proposed estimator outperforms the naive estimator in terms of Monte Carlo standard deviation and mean square error. In summary, the proposed estimator is more efficient than the naive estimator in all of the scenarios.

5.2 Real data analysis

In this section, we analyze the employment data available in StatLib. The data were firstly analyzed by Kadane and Woodworth (2004), who advocated a Bayesian analysis of the employment decisions. Such flow data were collected to examine a specified observation period. There were 416 subjects employed by the companies. The data consist of birth, hire and end of employment dates, and causes of termination. Here, the days of termination, counted from the first day of the study, are subjected to right censoring; that is, subjects may still be in the workforce at the end of the study period. It should be noted that some subjects are hired after the study start date. Thus, the employed times are left truncated by the searching-job times. In this example, we consider the subjects who were fired by the company before the first day of the study, and denote the delayed entry time A as the day until entry into the study period. Let the failure time T be the day in the study period when termination occurred. Thus, the observed failure time X is the exit date (or end date of the study period if still employed then) expressed in days after the study start. Obviously, the data are left truncated. To check the stationarity of the entry times, we use the formal test statistic proposed by Addona and Wolfson (2006). Unfortunately, the stationarity condition is rejected and the length-biased assumption fails to hold for the original data set. However, for those subjects with a delayed entry time greater than 200 days (the 0.15-th quantile of A), the p-value becomes 0.5324, and the truncation variable A is approximately uniformly distributed. There are 176 subjects included in this subset, and 104 subjects died before the end of study. Meanwhile, most of the subjects are censored because they are still employed at the end of the study period. The censoring rate is 40.9%.

The primary purpose of our study is to compare two different causes of termination on the cumulative incidence, voluntary and involuntary. The voluntary termination and involuntary termination are competing risks. We let $k = 1$ be the involuntary termination and $k = 2$ be the voluntary termination. We consider the following factors: Z_1 (age in years) and Z_2 (age greater than 40 years old or not). For each competing risk, we construct the Fine-Gray proportional subdistribution hazards model, $\lambda_k(t|\mathbf{z}) = \lambda_{0k}(t)\exp(\boldsymbol{\beta}_k^\top \mathbf{Z})$, $k = 1, 2$, where $\lambda_{0k}(\cdot)$ is an unknown non-negative function in t, $\boldsymbol{\beta}_k$ is a 2×1 regression parameter, and covariate $\mathbf{Z} = (Z_1, Z_2)^\top$.

Table 2. Results for employment data. Standard errors are in parentheses.

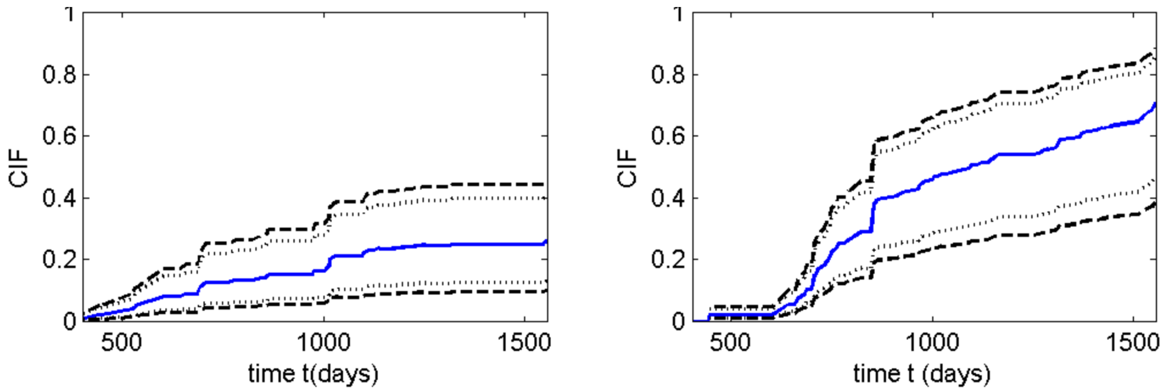
Method	Involuntary		Voluntary	
	$\widehat{\beta}_{11}$	$\widehat{\beta}_{12}$	$\widehat{\beta}_{21}$	$\widehat{\beta}_{22}$
Proposed	-0.0230	1.4393	0.0085	-0.8958
	(0.0211)	(0.6415)	(0.0198)	(0.4859)
Naive	-0.0221	1.3034	0.0089	-0.8200
	(0.0206)	(0.6485)	(0.0193)	(0.3875)

To derive the estimates of $\boldsymbol{\beta}$, we should find the solution of estimating equation (9). Note that the proposed estimator involves the local Kaplan-Meier estimate and we should choose the bandwidth parameter h_n through L-fold cross-validation, as discussed in Section 2.4. In this data set, we use $L = 4$ and select the optimal bandwidth $h_n = 0.1908$. The estimate results are summarized in Table 2, and the predicted cumulative incidence functions for $\mathbf{Z} = (25, 0)^\top$ and $\mathbf{Z} = (45, 1)^\top$ are presented in Figures 1a and 1b, respectively.

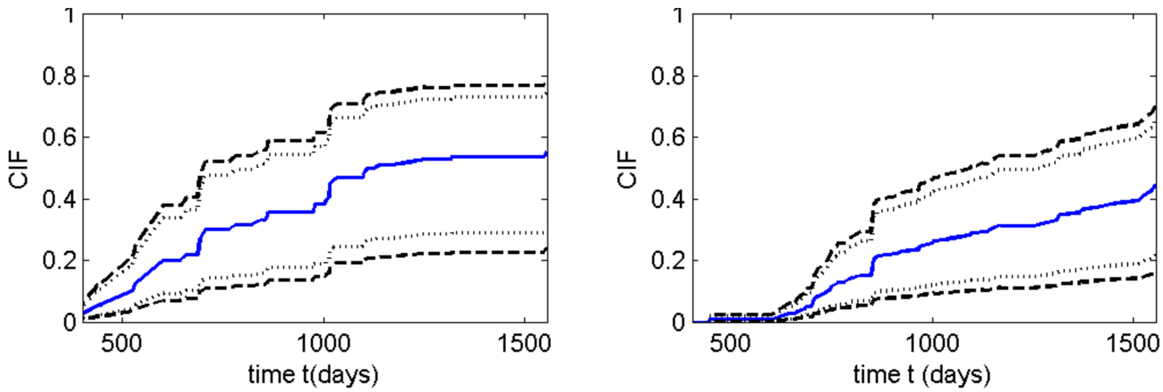
For comparison, we also use the naive estimator. All the results are tabulated in Table 2. The estimating coefficients by the naive estimator are similar to those by the proposed estimator. From these estimating results, we draw the following conclusions for this data subset:

1. For involuntary termination, young employees seem to have a higher risk of termination than those old employees, but they are not significantly different ($\widehat{\beta}_{11} = -0.0230$, p-value 0.2749). However, employees older than 40 have significantly higher risks than those under 40 ($\widehat{\beta}_{12} = 1.4393$, p-value 0.0248).
2. For voluntary termination, in contrast to involuntary termination, older employees seem to have a higher risk of termination than young employees but they are not significantly different ($\widehat{\beta}_{21} = 0.0085$, p-value 0.6661). However, those employees older than 40 have lower risks than those under 40 ($\widehat{\beta}_{22} = -0.8958$, p-value 0.0652), perhaps because young employees voluntarily leave the firm more frequently than their older colleagues.
3. Figure 1 confirms the convincing evidence for age discrimination at 40 years, which is in accordance with the conclusions of Tableman et al. (2006).

To check the adequacy of the assumed model, the proposed model-checking techniques are also applied to the data



(a) $(Z_1, Z_2) = (25, 0)$.



(b) $(Z_1, Z_2) = (45, 1)$.

Figure 1. Predicted cumulative incidence function (CIF, solid curves) for (a) $(Z_1, Z_2) = (25, 0)$ *and (b)* $(Z_1, Z_2) = (45, 1)$*, along with their* 95% *confidence intervals (dotted curves) and confidence bands (dashed curves).*

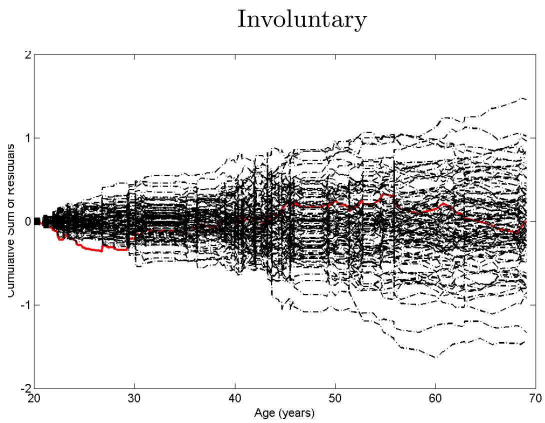
based on 1000 simulated realizations. For involuntary termination, the p-values for testing the functional forms of the covariates Z_1 and Z_2 are 0.775 and 1, respectively; the p-values for testing the additive risk assumption of Z_1 and Z_2 are 0.598 and 0.917, respectively, which suggests that the additive assumption is appropriate. Figure 2a displays the observed cumulative residuals versus Z_1, which appears to be within the normal ranges. However, we do not recommend testing the functional form for Z_2, since it dichotomous with two values and discussing the linearity is not very meaningful. Figures 2c and 2e display that the observed score processes appear to be completely covered by the first 100 simulated ones, which graphically supports that there is no evidence against the assumed model.

For voluntary termination, the p-values for testing the functional forms of the covariates Z_1 and Z_2 are 0.660 and 1, respectively. Figure 2b pertains to the functional forms of Z_1. However, Figures 2d and 2f demonstrate that the proportional hazards assumption of both Z_1 (p-values 0.037) and Z_2 (p-value 0.067) are strongly violated. More flexible regression models for the competing risks data under length-biased sampling, which allow some covariates to have time-varying effects, need to be considered in the future study.

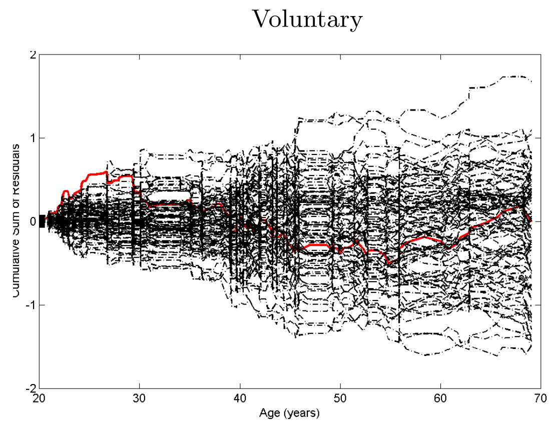
6. CONCLUDING REMARKS

In this paper, we propose an estimating procedure for Fine-Gray proportional subdistribution hazards model with length-biased competing risks data. The developed estimation and inference procedures use the competing risks structure and account for length-biased sampling. We also develop model checking techniques to assess the adequacy of Fine-Gray model based on the cumulative residual processes. As a comparison, the naive estimator from the left-truncation model has the advantages of not requiring an estimate for the censoring distribution, and it works for general left-truncated data, including length-biased data as a special case. However, the ignored component of the structure of length-biased data would cause a loss of information compared to the proposed estimator, as demonstrated in the simulation studies.

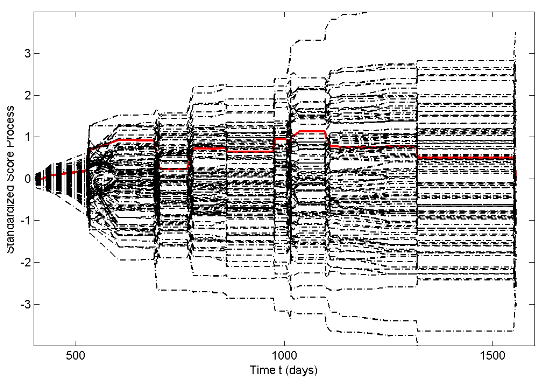
However, in the proposed estimating equation, we need to estimate the censoring distribution, which increases the burden of computation. Moreover, the integral of $S_c(t)$ may go to zero at the tail when censoring is very heavy. Further research is needed to investigate such situations, and some new methods should be developed.



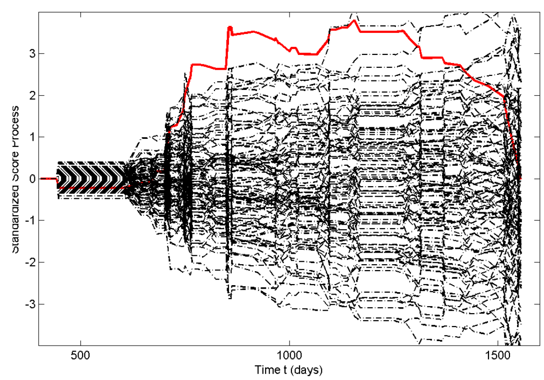
(a) Cumulative residuals versus Z_1 (p-value 0.775).



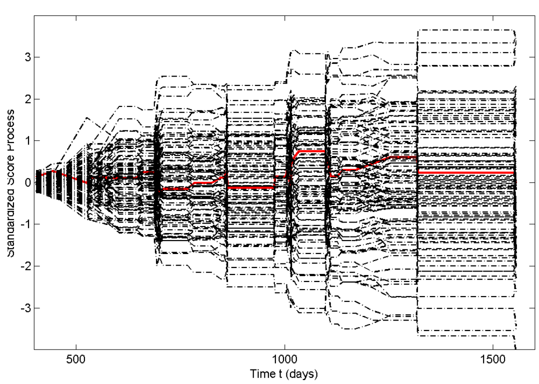
(b) Cumulative residuals versus Z_1 (p-value 0.660).



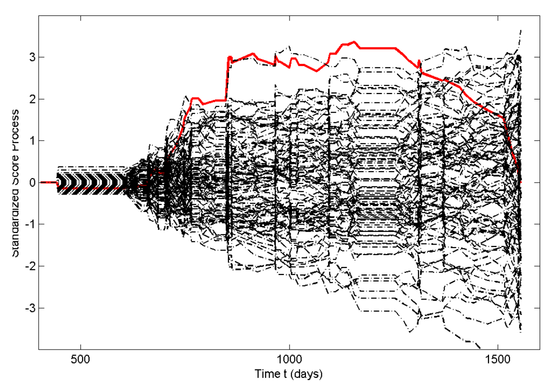
(c) Standardized score processes versus follow-up time for Z_1 (p-value 0.598).



(d) Standardized score processes versus follow-up time for Z_1 (p-value 0.037).



(e) Standardized score processes versus follow-up time for Z_2 (p-value 0.917).



(f) Standardized score processes versus follow-up time for Z_2 (p-value 0.067).

Figure 2. Plots of residual processes for the employment data. Red bold line, observed process; dashed lines, simulated processes.

APPENDIX A. TECHNIQUE DETAILS

The regularity conditions are provided as follows.

(C1) $\mathbb{P}(\min(V,C) \geq \tau) > 0$, where τ is a predetermined constant. In practice, τ is the maximum of the observations.

(C2) The parameter space $\mathcal{B}$ is a compact subset of $\mathbb{R}^p$ and the true parameter value $\boldsymbol{\beta}_{0k}$ is in the interior of $\mathcal{B}$.

(C3) $H_k(\boldsymbol{\beta}_{0k}, t) = \frac{s_2(\boldsymbol{\beta}_{0k},t)}{s_0(\boldsymbol{\beta}_{0k},t)} - \left(\frac{s_1(\boldsymbol{\beta}_{0k},t)}{s_0(\boldsymbol{\beta}_{0k},t)}\right)^{\otimes 2}$ is positive definite, and

$$\Gamma_k(\boldsymbol{\beta}_k) = \mathrm{E}\left[I(\widetilde{\epsilon}_i = k)\left\{\frac{S_2(\boldsymbol{\beta}_k, X_i)}{S_0(\boldsymbol{\beta}_k, X_i)} - \left(\frac{S_1(\boldsymbol{\beta}_k, X_i)}{S_0(\boldsymbol{\beta}_k, X_i)}\right)^{\otimes 2}\right\}\right]$$

is nonsingular.

(C4) $0 < \omega_c(\tau) < \infty$ and $\int_0^\tau \frac{\left[\int_t^\tau S_c(u)du\right]^2}{S_c^2(t)S_v(t)} dS_c(t) < \infty$, where $S_v(t) = \mathbb{P}(Y - A > t)$.

(C5) The kernel function $\mathcal{K}(\cdot)$ is a higher order kernel with order q; that is,

(i) $\mathcal{K}$ has a compact support and is a bounded kernel function,

(ii) $\mathcal{K}$ has order q, satisfying $\int \mathcal{K}(\mathbf{z})d\mathbf{z} = 1$, and

$$\int z_1^{u_1}...z_p^{u_p}\mathcal{K}(z_1,...,z_p)dz_1...dz_p \times \begin{cases} = 0, & \text{if } 0 \neq \sum_{j=1}^p u_j < q, \\ \neq 0, & \text{if } \sum_{j=1}^p u_j = q. \end{cases}$$

(C6) $\mathbf{Z}$ is bounded, and the first q partial derivatives of $f_{\mathbf{Z}}(\mathbf{z})$ with respect to $\mathbf{z}$ are uniformly bounded for $\mathbf{z}$; $f_v(t|\mathbf{z})$ and $f_c(t|\mathbf{z})$ are uniformly bounded away from infinity and have bounded first q order partial derivatives with respect to $\mathbf{z}$.

(C7) The bandwidth h_n satisfies $h_n = O(n^{-\alpha})$, where $1/2q < \alpha < 1/3p$.

Conditions C1–C4 are standard in survival analysis for length-biased data (Qin and Shen, 2010). Condition C5 is the definition of higher order kernel (Müller, 1988). Conditions C6–C7 are needed to ensure the consistency of the local Keplan-Meier estimator.

The following conclusion follows directly from Theorems 2.2 and 2.3 of Liang et al. (2012) and Corollary 1 of Leng and Tong (2014). We omit the details for saving space.

Lemma A.1. Under conditions (C1)–(C7), we have

(i) $\sup_s \sup_{\mathbf{z}} \left|\widehat{S}_c(s|\mathbf{z}) - S_c(s|\mathbf{z})\right| = O(\{\log n/(nh_n^p)\}^{3/4} + h_n^q) = o_P(n^{-1/2})$ a.s., and

(ii) $\widehat{S}_c(s|\mathbf{z}) - S_c(s|\mathbf{z}) = -\sum_{j=1}^n B_{nj}(\mathbf{z})\xi(X_j^*, \delta_j, s, \mathbf{z}) + O(\{\log n/(nh_n^p)\}^{3/4} + h_n^q)$ a.s. for $s < b$ such that $\inf_{\mathbf{z}} \{1 - F_v(b|\mathbf{z})\} S_c(b|\mathbf{z}) > 0$, where

$$\begin{aligned}\xi(V_j^*, \delta_j, s, \mathbf{z}) &= S_c(s|\mathbf{z})\Bigg[\int_0^{\min(V_j^*,s)} \frac{dS_c(u|\mathbf{z})}{\{1 - F_c(u|\mathbf{z})\}S_c^2(u|\mathbf{z})} \\ &+ \frac{I(V_j^* \leq s, \delta_j = 0)}{\{1 - F_V(V_j^*|\mathbf{z})\}S_c(V_j^*|\mathbf{z})}\Bigg],\end{aligned}$$

and $V_j^* = \min(V_j, C_j)$, $F_v(s|\mathbf{z}) = P(V \leq s|\mathbf{z})$ and $S_c(s|\mathbf{z}) = P(C \geq s|\mathbf{z})$.

A.1 Proof of Theorem 1

We divide the proof in two parts.

(1) For the consistence of $\widehat{\boldsymbol{\beta}}_k$, we need apply the Theorem 5.9 of Van der Vaart (1998). It suffices to show that

(i) $\sup_{\boldsymbol{\beta}_k \in \mathcal{B}} |n^{-1}\widehat{U}_k(\boldsymbol{\beta}_k) - u_k(\boldsymbol{\beta}_k)| \to 0$ almost surely as n goes to infinity, and

(ii) $u_k(\boldsymbol{\beta}_{0k}) = 0$ and $u_k(\boldsymbol{\beta}_k)$ is bounded away from 0.

We first show that $\sup_{\boldsymbol{\beta} \in \mathcal{B}} |n^{-1}\widehat{U}_k(\boldsymbol{\beta}_k) - n^{-1}U_k(\boldsymbol{\beta}_k)| \to 0$ almost surely, as n goes to infinity. By Lemma A.1, $\widehat{S}_c(t|\mathbf{z})$ converges almost surely to $S_c(t|\mathbf{z})$ uniformly over $t \in [0,\tau]$ and all $\mathbf{z}$. Thus, $\widehat{\omega}_c(t|\mathbf{z}) = \int_0^t \widehat{S}_c(s|\mathbf{z})ds$ converges almost surely to $\omega_c(t|\mathbf{z}) = \int_0^t S_c(s|\mathbf{z})ds$ uniformly over $t \in [0,\tau]$ and all $\mathbf{z}$. For any $a_n > 0$ with $a_n \to 0$ as $n \to \infty$, we define

$$\begin{aligned}\mathcal{F} = \{&f(t,s) = (g_1(t) - w_c(t|\mathbf{z}))\left(g_2(s) - w_c(s|\mathbf{z})^{-1}\right)\mathbf{z}^{\otimes l} \\ &\times \exp(\boldsymbol{\beta}_k^\top \mathbf{z})Y_k(s) : g_1(t) \text{ is nonnegative and} \\ &\text{nondecreasing, } g_2(s) \text{ is nonnegative and nonincreasing,} \\ &\boldsymbol{\beta}_k \in \mathcal{B}, \sup_{s,t} |(g_1(s) - w_c(s|\mathbf{z}))(g_2(s) - w_c(s|\mathbf{z})^{-1})| \leq a_n\}.\end{aligned}$$

By the definition, $\sup_{f \in \mathcal{F}} |Pf| \leq a_n \sup_{\boldsymbol{\beta}_k \in \mathcal{B}} |\mathbf{z}\exp(\boldsymbol{\beta}_k^\top \mathbf{z})|$. As $\mathcal{F}$ is a class of the product of monotone functions and indicator functions, $\mathcal{F}$ is Glivenko-Cantelli class. Hence, $\sup_{f \in \mathcal{F}} |P_n f - Pf|$ converges almost surely to zero, as $n \to \infty$. For sufficiently large n, we have

$$|n^{-1}\widehat{S}_l(\boldsymbol{\beta}, t) - S_l(\boldsymbol{\beta}, t)| \leq \sup_{f \in \mathcal{F}} |Pf| + \sup_{f \in \mathcal{F}} |P_n f - Pf|.$$

Thus, $\sup_{\boldsymbol{\beta}_k \in \mathcal{F}, t \in [0,\tau]} |n^{-1}\widehat{S}_l(\boldsymbol{\beta}_k, t) - S_l(\boldsymbol{\beta}_k, t)|$ converges almost surely to zero. It follows that $\sup_{\boldsymbol{\beta}_k \in \mathcal{B}} |n^{-1}\widehat{U}_k(\boldsymbol{\beta}_k) - U_k(\boldsymbol{\beta}_k)| \to 0$ almost surely by the permanence of product.

Next, we show that $\sup_{\boldsymbol{\beta}_k \in \mathcal{B}} |n^{-1}U_k(\boldsymbol{\beta}_k) - u_k(\boldsymbol{\beta}_k)|$ converges almost surely to zero, as $n \to \infty$, where $u_k(\boldsymbol{\beta}_k) = \mathrm{E}U_k(\boldsymbol{\beta})$. Note that $\mathcal{B}$ is compact, $m(\boldsymbol{\beta}_k) = I(\widetilde{\epsilon} = k)\left[\mathbf{z} - \frac{s_1(\boldsymbol{\beta}_k, X)}{s_0(\boldsymbol{\beta}_k, X)}\right]$ is continuous and dominated by an integrable function. Then

$\{m(\boldsymbol{\beta}_k) : \boldsymbol{\beta}_k \in \mathcal{B}\}$ is Glivenko-Cantelli class. By the uniform law of large numbers (Pollard, 1990), $\sup_{\boldsymbol{\beta}_k \in \mathcal{B}} |n^{-1}U_k(\boldsymbol{\beta}_k) - u_k(\boldsymbol{\beta}_k)|$ converges almost surely to zero, and consequently, $\sup_{\boldsymbol{\beta}_k \in \mathcal{B}} |n^{-1}\widehat{U}_k(\boldsymbol{\beta}_k) - U_k(\boldsymbol{\beta}_k)|$ converges almost surely to zero.

It is easy to show that $u(\boldsymbol{\beta}_{0k}) = 0$. Note that

$$\begin{aligned}&\frac{\partial U_k(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{0k}}\\ =&\mathrm{E}\left[I(\widetilde{\epsilon}=k)\left\{\frac{S_2(\boldsymbol{\beta}_{0k},X_i)}{S_0(\boldsymbol{\beta}_{0k},X_i)} - \left(\frac{S_1(\boldsymbol{\beta}_{0k},X_i)}{S_0(\boldsymbol{\beta}_{0k},X_i)}\right)^{\otimes 2}\right\}\right]\\ =&H(\boldsymbol{\beta}_{0k},t) > 0.\end{aligned}$$

Hence, $U_k(\boldsymbol{\beta})$ is bounded away from zero. It follows that $\widehat{\boldsymbol{\beta}}_k$ converges to $\boldsymbol{\beta}_{0k}$ in probability by the Theorem 5.9 of Van der Vaart (1998).

(2) To show the asymptotic normality of $\widehat{\boldsymbol{\beta}}_k$, we first establish the weak convergence of the process $n^{-1/2}\widehat{U}_k(\boldsymbol{\beta})$. We can write

$$\begin{aligned}&\widehat{U}_k(\boldsymbol{\beta}_{0k})\\ =&\sum_{i=1}^n \int_0^\tau \left[\mathbf{Z}_i - \frac{S_1(\boldsymbol{\beta}_{0k},t)}{S_0(\boldsymbol{\beta}_{0k},t)}\right] dM_{ik}(t)\\ &+\sum_{i=1}^n \int_0^\tau \left[\frac{S_1(\boldsymbol{\beta}_{0k},t)}{S_0(\boldsymbol{\beta}_{0k},t)} - \frac{\widehat{S}_1(\boldsymbol{\beta}_{0k},t)}{\widehat{S}_0(\boldsymbol{\beta}_{0k},t)}\right] dN_{ik}(t)\\ \equiv& I_1 + I_2.\end{aligned}$$

It requires the decomposition of I_2 as follows. By Lemma A.1 (ii), we have the almost surely i.i.d. representation:

$$\begin{aligned}\widehat{S}_c(s|\mathbf{z}) - S_c(s|\mathbf{z}) =& -\sum_{j=1}^n B_{nj}(\mathbf{z})\xi(V_j^*,\delta_j,s,\mathbf{z})\\ &+ O\left(\{\log n/(nh_n^p)\}^{3/4} + h_n^q\right) \text{ a.s.,}\end{aligned}$$

where $V_j^* = \min(V_j, C_j)$ and

$$\begin{aligned}&\xi(V_j^*,\delta_j,s,\mathbf{z})\\ =&S_c(s|\mathbf{z})\left[\int_0^{\min(V_j^*,s)} \frac{dS_c(u|\mathbf{z})}{\{1-F_v(u|\mathbf{z})\}S_c^2(u|\mathbf{z})}\right.\\ &\left.+\frac{I(V_j^* \le s, \delta_j = 0)}{\{1-F_v(V_j^*|\mathbf{z})\}S_c(V_j^*|\mathbf{z})}\right],\end{aligned}$$

are independent random variables with mean zero and finite variance for any given s and $\mathbf{z}$. Hence, by (C7),

$$\begin{aligned}&\widehat{\omega}_c(X_j|\mathbf{Z}_j) - \omega_c(X_j|\mathbf{Z}_j)\\ =& -\sum_{l=1}^n B_{nl}(\mathbf{Z}_j)\int_0^{X_j} \xi(V_l^*,\delta_l,s,\mathbf{Z}_j)ds\\ &+ O\left(\{\log n/(nh_n^p)\}^{3/4} + h_n^q\right)\\ \equiv& -\sum_{l=1}^n Q_l(X_j|\mathbf{Z}_j) + O\left(n^{-1/2}\right),\end{aligned}$$

where

$$Q_l(t|\mathbf{Z}_j) = B_{nl}(\mathbf{Z}_j)\int_0^t \xi(V_l^*,\delta_l,s,\mathbf{Z}_j)ds.$$

Using a standard change of variables and Taylor expansion arguments, we obtain

$$\begin{aligned}I_2 =& \sum_{i=1}^n \int_0^\tau \frac{1}{S_0(\boldsymbol{\beta}_{0k},t)} \sum_{j=1}^n Y_{jk}(t)\omega_c(t|\mathbf{Z}_j)e^{\boldsymbol{\beta}_{0k}^\top \mathbf{Z}_j}\mathbf{Z}_j\\ &\times\left[\frac{1}{\omega_c(X_j|\mathbf{Z}_j)} - \frac{1}{\widehat{\omega}_c(X_j|\mathbf{Z}_j)}\right] dN_{ik}(t) + o_P(n^{1/2})\\ =& \sum_{i=1}^n \int_0^\tau \frac{1}{S_0(\boldsymbol{\beta}_{0k},t)} \sum_{j=1}^n Y_{jk}(t)\omega_c(t|\mathbf{Z}_j)e^{\boldsymbol{\beta}_{0k}^\top \mathbf{Z}_j}\mathbf{Z}_j\\ &\times\left[\frac{\widehat{\omega}_c(X_j|\mathbf{Z}_j) - \omega_c(X_j|\mathbf{Z}_j)}{\omega_c^2(X_j|\mathbf{Z}_j)}\right] dN_{ik}(t) + o_P(n^{1/2})\\ =& -\sum_{l=1}^n \eta_{lk} + O_p(nh^2) + n * O\left(\{\log n/(nh_n^p)\}^{3/4} + h_n^q\right)\\ &+ o_P(n^{1/2}),\end{aligned}$$

where

$$\eta_{lk} = \sum_{j=1}^n \frac{\mathbf{Z}_j e^{\boldsymbol{\beta}_{0k}^\top \mathbf{Z}_j} Q_l(X_j|\mathbf{Z}_j}{\omega_c(X_j|\mathbf{Z}_j)} \sum_{i=1}^n \int_0^\tau \frac{\pi_{jk}(t)}{S_0(\boldsymbol{\beta}_{0k},t)} dN_{ik}(t),$$

and by Conditions (C7), the residual items are of order $o_P(n^{1/2})$. Since $\mathrm{E}\left(\xi(V_1^*,\delta_1,s,\mathbf{Z}_l)|\mathbf{Z}_l\right) = 0$, then η_{lk} are i.i.d random variables with mean zero. Thus, we have

$$\widehat{U}_k(\boldsymbol{\beta}_{0k}) = \sum_{i=1}^n \phi_{ik}(\boldsymbol{\beta}_{0k}) + o_P(n^{1/2}),$$

where $\phi_{ik}(\boldsymbol{\beta}_{0k}) = \int_0^\tau \left[\mathbf{Z}_i - \frac{S_1(\boldsymbol{\beta}_{0k},t)}{S_0(\boldsymbol{\beta}_{0k},t)}\right] dM_{ik}(t) - \eta_{ik}$.

Using the multivariate central limit theorem, $n^{-1/2}\widehat{U}_k(\boldsymbol{\beta}_{0k})$ is asymptotically normal with mean zero and covariance matrix $\Sigma_k = \mathrm{E}\left[\phi_{ik}(\boldsymbol{\beta}_{0k})^{\otimes 2}\right]$.

By the Taylor expansion of $\widehat{U}_k(\widehat{\boldsymbol{\beta}}_k)$ at $\boldsymbol{\beta}_{0k}$,

$$0 = \widehat{U}_k(\widehat{\boldsymbol{\beta}}_k) = \widehat{U}_k(\widehat{\boldsymbol{\beta}}_k) + \frac{\partial \widehat{U}_k(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{0k}),$$

where $\boldsymbol{\beta}^*$ is on the line segment between $\widehat{\boldsymbol{\beta}}_k$ and $\boldsymbol{\beta}_{0k}$. It then follows from the uniform convergence of $\widehat{\Gamma}_k(\boldsymbol{\beta}) =$

$-n^{-1}\frac{\partial \widehat{U}_k(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and the consistency of $\widehat{\boldsymbol{\beta}}_k$ that

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{0k}) &= \Gamma_k^{-1} n^{-1/2}\widehat{U}_k(\boldsymbol{\beta}_{0k}) + o_P(1) \\ &= \Gamma_k^{-1} n^{-1/2}\sum_{i=1}^n \phi_{ik}(\boldsymbol{\beta}_{0k}) + o_P(1).\end{aligned}$$

This implies that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{0k})$ is asymptotically zero-mean normal with the covariance matrix $\Gamma_k^{-1}\Sigma_k\Gamma_k^{-1}$.

A.2 Weak convergence of $\widehat{\Lambda}_{0k}$

Recall that the Breslow-type estimator

$$\widehat{\Lambda}_{0k}(t, \widehat{\boldsymbol{\beta}}_k) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\widehat{S}_0(\widehat{\boldsymbol{\beta}}_k, s)},$$

for $t \in [0, \tau]$. It is easy to see that $\widehat{\Lambda}_{0k}(t, \boldsymbol{\beta}_k)$ is continuous functional of two empirical processes, $\widehat{S}(\boldsymbol{\beta}_k, s)$ and $\sum_{i=1}^n dN_{ik}(s)$ with respect to the supremum norm. The almost sure convergence of the two processes implies that $\sup_{t\in[0,\tau],\boldsymbol{\beta}_k\in\mathcal{B}} |\widehat{\Lambda}_{0k}(t, \boldsymbol{\beta}_k) - \Lambda_{0k}(t, \boldsymbol{\beta}_k)| \to 0$, where

$$\Lambda_{0k}(t, \boldsymbol{\beta}_k) = \int_0^t \frac{dF_k^u(u)}{s_0(\boldsymbol{\beta}_k, u)},$$

and $F_k^u(t) = P(X \le u, \widetilde{\epsilon} = k)$. Then, by the functional delta method, it follows

$$\sqrt{n}\left\{\widehat{\Lambda}_{0k}(t, \widehat{\boldsymbol{\beta}}_k) - \Lambda_{0k}(t, \boldsymbol{\beta}_k)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varphi_{ik}(t) + o_P(n^{-1/2}),$$

where

$$\begin{aligned}\varphi_{ik}(t) &= \left[\frac{I(X_i \le t, \widetilde{\epsilon}_i = k)}{s_0(\boldsymbol{\beta}_{0k}, X_i)} - \int_0^t \frac{dF_k^u(u)}{s_0(\boldsymbol{\beta}_{0k}, u)}\right] \\ [4pt] &- \left[\Gamma_k^{-1}(\boldsymbol{\beta}_{0k})\int_0^t \frac{s_1(\boldsymbol{\beta}_{0k}, u)}{s_0^2(\boldsymbol{\beta}_{0k}, u)} dF_k^u(u)\right]^T \phi_{ik}(\boldsymbol{\beta}_{0k}).\end{aligned}$$

A.3 Weak convergence of W_{jk}, $W_k(\mathbf{z})$ and V_{jk}

We establish the consistency of each model-checking test statistic by using similar arguments to Lin et al. (1993). The process W_{jk}, $W_k(\mathbf{z})$ and V_{jk} are all special cases of the process

$$W(t, \mathbf{z}) = n^{-1/2}\sum_{i=1}^n \int_0^t q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) dM_{ik}(u),$$

where $q(\cdot)$ is a known bounded vector-valued function, and $I(\mathbf{Z}_i \le \mathbf{z}) = I(Z_{i1} \le z_1, \ldots, Z_{ip} \le z_p)$. By Taylor series expansion and some simple algebra, we have

$$\begin{aligned}&W(t, \mathbf{z}) \\ &= n^{-1/2}\sum_{i=1}^n \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) dM_{ik}(u) \\ &- n^{-1/2}\sum_{i=1}^n \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) \frac{\pi_{ik}(u|\mathbf{Z}_i)}{\omega_c(X_i|\mathbf{Z}_i)} \\ &\quad \times \sum_{l=1}^n Q_l(X_i|\mathbf{Z}_i) e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_i} d\Lambda_{0k}(u) \\ &- n^{-1/2}\sum_{i=1}^n \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_i} \pi_{ik}(u|\mathbf{Z}_i) \\ &\quad \times \left[\Gamma_k^{-1}(\boldsymbol{\beta}_k) n^{-1}\sum_{l=1}^n \phi_{lk}(\boldsymbol{\beta}_k)\right]^\top \mathbf{Z}_i d\Lambda_{0k}(u) \\ &- n^{-1/2}\sum_{i=1}^n \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_i} \pi_{ik}(u|\mathbf{Z}_i) \\ &\quad \times d\Big[n^{-1}\sum_{l=1}^n \varphi_{lk}(u)\Big] + o_P(1),\end{aligned}$$

By the arguments similar to those of Lin et al. (2000), one can show that $W(t, \mathbf{z})$ is tight.

By the strong law of large number, $n^{-1}\sum_{l=1}^n \phi_{lk}(\boldsymbol{\beta}_k)$ and $n^{-1}\sum_{l=1}^n \varphi_{lk}(u)$ converge almost surely to $\bar{\phi}_k(\boldsymbol{\beta}_k)$ and $\bar{\varphi}_k$, respectively. Furthermore, by the kernel theory, $\sum_{l=1}^n Q_l(x|\mathbf{z})$ converges to $\bar{Q}(x|\mathbf{z})$. Thus, we can derive $W(t, \mathbf{z}) = n^{-1/2}\sum_{i=1}^n \zeta_i(t, \mathbf{z}) + o_P(1)$, where

$$\begin{aligned}\zeta_i(t, \mathbf{z}) &= \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) dM_{ik}(u) \\ &- \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) \frac{\pi_{ik}(u|\mathbf{Z}_i)}{\omega_c(X_i|\mathbf{Z}_i)} \bar{Q}(X_i|\mathbf{Z}_i) \\ &\quad \times e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_i} d\Lambda_{0k}(u) \\ &- \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_i} \pi_{ik}(u|\mathbf{Z}_i) \\ &\quad \times \left[\Gamma_k^{-1}(\boldsymbol{\beta}_k)\bar{\phi}_k(\boldsymbol{\beta}_k)\right]^\top \mathbf{Z}_i d\Lambda_{0k}(u) \\ &- \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) e^{\boldsymbol{\beta}_k^\top \mathbf{Z}_i} \pi_{ik}(u|\mathbf{Z}_i) d\bar{\varphi}_k(u).\end{aligned}$$

By the multivariate central limit theorem and the tightness of $W(t, \mathbf{z})$, $W(t, \mathbf{z})$ converges weakly to a mean zero Gaussian process with covariance function $\mathrm{E}\left[\zeta_1(t, \mathbf{z})\zeta_1(\widetilde{t}, \widetilde{\mathbf{z}})\right]$ at $(t, \mathbf{z})$ and $(\widetilde{t}, \widetilde{\mathbf{z}})$. The covariance function can be consistently estimated by $n^{-1}\sum_{i=1}^n \widehat{\zeta}_i(t, \mathbf{z})\widehat{\zeta}_i(\widetilde{t}, \widetilde{\mathbf{z}})$, where

$$\begin{aligned}\widehat{\zeta}_i(t, \mathbf{z}) &= \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) d\widehat{M}_{ik}(u) \\ &- \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) \frac{\widehat{\pi}_{ik}(u|\mathbf{Z}_i)}{\widehat{\omega}_c(X_i|\mathbf{Z}_i)}\end{aligned}$$

$$
\begin{aligned}
&\times \sum_{l=1}^{n} Q_l(X_i|\mathbf{Z}_i) e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i} d\widehat{\Lambda}_{0k}(u) \\
&- \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i} \widehat{\pi}_{ik}(u) \\
&\times \left[\widehat{\Gamma}_k^{-1}(\widehat{\boldsymbol{\beta}}_k) n^{-1} \sum_{l=1}^{n} \widehat{\phi}_{lk}(\widehat{\boldsymbol{\beta}}_k) \right]^\top \mathbf{Z}_i d\widehat{\Lambda}_{0k}(u) \\
&- \int_0^\tau q(\mathbf{Z}_i) I(\mathbf{Z}_i \le \mathbf{z}) e^{\widehat{\boldsymbol{\beta}}_k^\top \mathbf{Z}_i} \widehat{\pi}_{ik}(u|\mathbf{Z}_i) \\
&\times d\Big[n^{-1} \sum_{l=1}^{n} \widehat{\varphi}_{lk}(u) \Big].
\end{aligned}
$$

A.4 Proof of Theorem 2

To derive the asymptotic approximation of the process $I_k(t|\mathbf{z}) = n^{1/2}[\widehat{F}_k(t|\mathbf{z}) - F_k(t|\mathbf{z})]$, we define $W_k(t|\mathbf{z}) = n^{1/2}[\widehat{\Lambda}_k(t|\mathbf{z}) - \Lambda_k(t|\mathbf{z})]$, where $\Lambda_k(t|\mathbf{z}) = \Lambda_{0k}(t)\exp(\boldsymbol{\beta}_{0k}^\top \mathbf{z})$ for $k = 1, ..., K$. By the functional delta method, we can show that $W_k(t|\mathbf{z})$ can be asymptotically equivalent to the sum of i.i.d. stochastic process $W_k(t|\mathbf{z}) = n^{-1/2} \sum_{i=1}^{n} \gamma_{ik}(t|\mathbf{z}) + o_p(n^{-1/2})$, where

$$
\gamma_{ik}(t|\mathbf{z}) = \int_0^t \exp(\boldsymbol{\beta}_0^\top \mathbf{z}) \left[d\varphi_i(u) + \mathbf{z}^\top \Gamma_k^{-1} \phi_{ik}(\boldsymbol{\beta}_{0k}) d\Lambda_{0k}(u) \right].
$$

Now, it follows from a Taylor series approximation that the process

$$
\begin{aligned}
& n^{1/2}[\widehat{F}_k(t|\mathbf{z}) - F_k(t|\mathbf{z})] \\
= \; & n^{1/2}\{1 - F_k(t|\mathbf{z})\}[\widehat{\Lambda}_k(t|\mathbf{z}) - \Lambda_k(t|\mathbf{z})] \\
= \; & \{1 - F_k(t|\mathbf{z})\} n^{-1/2} \sum_{i=1}^{n} \gamma_{ik}(t|\mathbf{z}) + o_p(n^{-1/2}).
\end{aligned}
$$

ACKNOWLEDGEMENTS

The authors are grateful to the editor and two anonymous referees for many helpful comments. Feipeng Zhang's work is partially supported by the National Natural Science Foundation of China (11771133,11401194). Heng Peng's research is supported in part by CEGR grant of the Research Grants Council of Hong Kong (No. HKBU 12302615 and HKBU 12303618), FRG grants from Hong Kong Baptist University (FRG2/16-17/042), and the Natural Science Foundation of Hunan Province, China (2017JJ3021). Yong Zhou's work is partially supported by the State Key Program of National Natural Science Foundation of China (71331006), the State Key Program in the Major Research Plan of National Natural Science Foundation of China (91546202).

Received 19 July 2017

REFERENCES

Addona, V. and Wolfson, D. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis*, 12:267–284. MR2328577

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Verlag, New York. MR1198884

Benichou, J. and Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, 46:813–826.

Cheng, S. C., Fine, J. P., and Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, 54:219–228. MR1626801

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 34:187–220. MR0341758

Fan, J., Lin, H., and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *The Annals of Statistics*, 34:290–325. MR2275243

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*, 94:496–509. MR1702320

Geskus, R. B. (2011). Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics*, 67:39–49. MR2898815

Ghosh, D. (2008). Proportional hazards regression for cancer studies. *Biometrics*, 64:141–148. MR2422828

González-Manteiga, W. and Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics*, 4:65–78. MR1366364

Huang, C. Y., Qin, J., and Follmann, D. A. (2012). A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling. *Biometrika*, 99:199–210. MR2899673

Kadane, J. B. and Woodworth, G. G. (2004). Hierarchical models for employment decisions. *Journal of Business and Economic Statistics*, 22:182–193. MR2049920

Lancaster, T. (1992). *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge. MR1167199

Leng, C. and Tong, X. (2014). Censored quantile regression via Box-Cox transformation under conditional independence. *Statistica Sinica*, 24:221–249. MR3183682

Li, J. and Scheike, T. H. and Zhang, M.-J. Checking Fine and Gray subdistribution hazards model with cumulative sums of residuals. *Lifetime Data Analysis*, 21: 197–217. MR3324227

Liang, H.-Y., de Uña-Álvarez, J., and del Carmen Iglesias-Pérez, M. (2012). Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. *Test*, 21:790–810. MR2992093

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B*, 62:711–730. MR1796287

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80:557–572. MR1248021

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Springer, New York. MR0960887

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematics Statistics, Hayward, CA. MR1089429

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34:541–554. MR0570114

Qin, J., Ning, J., Liu, H., and Shen, Y. (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association*, 106:1434–1449. MR2896847

Qin, J. and Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under cox model. *Biometrics*, 66:382–392. MR2758818

Scheike, T. H. and Zhang, M. J. (2003). Extensions and applications of the Cox-Aalen survival model. *Biometrics*, 59:1036–1045. MR2025128

Scheike, T. H., Zhang, M. J., and Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95:205–220. MR2409723

Shen, P.-S. (2011). Proportional subdistribution hazards regression for left-truncated competing risks data. *Journal of Nonparametric Statistics*, 23:885–895. MR2854244

Shen, Y. and Cheng, S. C. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55:1093–1100. MR1732719

Sun, L., Liu, J., Sun, J., and Zhang, M. (2006). Modeling the subdistribution of a competing risk. *Statistica Sinica*, 16:1367. MR2327495

Tableman, M., Stahel, W. A., Stahel, W. A., and Stahel, W. A. (2006). Nonparametric methods for termination time data with competing risks. Technical report, Seminar für Statistik.

Tian, L., Zucker, D., and Wei, L. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*, 100:172–183. MR2156827

Tsai, W. (2009). Pseudo-partial likelihood for proportional hazards models with biased-sampling data. *Biometrika*, 96:601–615. MR2538760

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. MR1652247

Wang, H. J. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, 104:1117–1128. MR2562007

Wang, M. C. (1996). Hazards regression analysis for length-biased data. *Biometrika*, 83:343–354. MR1439788

Wang, M. C., Brookmeyer, R., and Jewell, N. P. (1993). Statistical models for prevalent cohort data. *Biometrics*, 49:1–11. MR1221402

Zhang, F.P., Peng, H., and Zhou, Y. (2016). Composite partial likelihood estimation for length-biased and right-censored data with competing risks. *Journal of Multivariate Analysis*, 149:160–176. MR3507321

Zhang, F.P., Chen, X., and Zhou, Y. (2014). Proportional hazards model with varying coefficients for length-biased data. *Lifetime Data Analysis*, 20:132–157. MR3148187

Zhang, X., Zhang, M. J., and Fine, J. (2011). A proportional hazards regression model for the subdistribution with right-censored and left-truncated competing risks data. *Statistics in Medicine*, 30:1933–1951. MR2829057

Feipeng Zhang
School of Mathematics and Statistics, Hunan Normal University, Changsha, 410081, China
E-mail address: zhangfp108@163.com

Heng Peng
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
E-mail address: hpeng@math.hkbu.edu.hk

Yong Zhou
School of Statistics, East China Normal University, Shanghai, 200241, China
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
E-mail address: yzhou@amss.ac.cn