

Running headline: Case-cohort with competing risks data

Stratified Proportional Subdistribution Hazards Model with Covariate-Adjusted Censoring Weight for Case-Cohort Studies

SOYOUNG KIM, YAYUN XU, MEI-JIE ZHANG,
and KWANG WOO AHN

Division of Biostatistics, Medical College of Wisconsin, U.S.A.

ABSTRACT. The case-cohort study design is widely used to reduce cost when collecting expensive covariates in large cohort studies with survival or competing risks outcomes. A case-cohort study data set consists of two parts: i) a random sample; and ii) all cases or failures from a specific cause of interest. Clinicians often assess covariate effects on competing risks outcomes. The proportional subdistribution hazards model of **Fine and Gray (1999)** directly evaluates the effect of a covariate on the cumulative incidence function. They studied the asymptotic distribution of the estimators under the non-covariate-dependent censoring assumption for the full cohort study. However, the non-covariate-dependent censoring assumption is often violated in many biomedical studies. In this paper, we propose a proportional subdistribution hazards model for case-cohort studies with stratified data with covariate-adjusted censoring weight. We further propose an efficient estimator when extra information from the other causes is available under case-cohort studies. The proposed estimators are shown to be consistent and asymptotically normal. Simulation studies show (i) the proposed estimator is unbiased when the censoring distribution depends on covariates; and (ii) the proposed efficient estimator gains estimation efficiency when using extra information from the other causes. We analyze a bone marrow transplant data set and a coronary heart disease data set using the proposed method.

Key words: Case-cohort design; Competing risks data; Efficiency; Hazard of subdistribution; Inverse probability of censoring weight

1 Introduction

In large observational studies, collecting exposure information from all subjects over a long study period may be costly. The majority of cost and effort mainly involve measuring and assembling expensive exposure information. The case-cohort design is widely used to handle such difficulties by reducing cost while accomplishing the same goal of large observational studies (Prentice, 1986).

The extensive work has been done for analyzing case-cohort data with survival outcomes. For a univariate failure time, a pseudo-likelihood approach was proposed by Prentice (1986) and Self and Prentice (1988). In order to improve efficiency, Barlow (1994) and Kulich and Lin (2004) proposed a robust estimator using a time-varying weight and a class of weighted estimating functions using all available information, respectively. When there are several diseases of interest, one often studies them using the same subcohort under multiple case-cohort studies. For such multiple case-cohort studies, Kang and Cai (2009) developed a joint model with multivariate failure time. However, they did not use extra information from the other diseases when estimating the effect of risk factors for a disease of interest. Kim et al. (2013) proposed a more efficient estimation method with a new weight to make full use of information from the other diseases.

Borgan et al. (2000), Samuelsen et al. (2007), Breslow and Wellner (2007), and Kim et al. (2018) considered a stratified case-cohort design by selecting the subcohort based on stratified sampling to increase estimation efficiency. The stratified case-cohort design assumes the common baseline hazard function for all strata.

The case-cohort design is often used to study competing risks data in which only one occurrence of failure from one cause can be observed because it hinders the occurrence of failure from the other causes. The censoring distribution often depends on covariates under such design. For example, the Atherosclerosis Risk in Communities (ARIC) study investigated the effect of high-sensitivity C-reactive protein (hs-CRP) on coronary heart disease (CHD) (Ballantyne et al., 2004). The case-cohort design was used to reduce the cost of obtaining hs-CRP from the subjects. The subcohort was selected using stratified sampling based on age, gender, and race. Death prior to CHD was a competing risk for CHD. As shown in Section 6, several covariates were associated with the censoring distribution based on the proportional hazards model.

For such competing risks data, a direct evaluation of covariates on the cumulative incidence function of a given cause is often of clinical interest (Saber et al., 2015). There is

rich literature on such modeling for the full cohort study. [Fine and Gray \(1999\)](#) proposed a proportional subdistribution hazards model that directly assesses the effect of covariates on the cumulative incidence function. The proportional subdistribution hazards assumption often does not hold for certain covariates. Ignoring nonproportional hazards structure in data analysis could lead to biased parameter estimation. To address such aspect, [Zhou et al. \(2011\)](#) extended [Fine and Gray \(1999\)](#) to a stratified proportional subdistribution hazards model by allowing different baseline hazard functions for different strata. However, [Fine and Gray \(1999\)](#) and [Zhou et al. \(2011\)](#) did not study the asymptotics of the estimators under the covariate-dependent censoring. [Scheike et al. \(2008\)](#) proposed a direct binomial modeling based on the inverse probability weighting technique. [He et al. \(2016\)](#) proposed a proportional subdistribution hazards model with covariate-adjusted censoring weight. They estimated the censoring probability given covariates based on the proportional hazards model ([Cox, 1972](#)) and the Breslow estimator ([Breslow, 1972](#)) and used it for the weight function in the estimating equation. [Mao and Lin \(2017\)](#) proposed semiparametric transformation models for the cumulative incidence of competing risks based on the non-parametric maximum likelihood estimation. On the other hand, there is limited literature on competing risks modeling under the case-cohort design. [Sorensen and Andersen \(2000\)](#) studied the cause-specific hazards model under a single case-cohort study. However, the cause-specific hazards model method does not explain a direct relationship between the estimated covariate effects and the cumulative incidence of a given cause. In addition, they did not address the case-cohort design with the nonproportional hazard structure. [Pintilie et al. \(2010\)](#) considered a pseudo-likelihood approach based on [Fine and Gray \(1999\)](#) to accommodate a single case-cohort study. However, they assumed covariate-independent censoring and did not establish the theoretical properties of the estimators. To the best of our knowledge, there is no methodology that models a direct relationship between covariates and the cumulative incidence and allows covariate-dependent censoring for case-cohort data with the nonproportional subdistribution hazard structure.

Motivated by [Zhou et al. \(2011\)](#) and [He et al. \(2016\)](#), we propose a stratified subdistribution hazards model under the case-cohort design with stratified sampling so that it can be used even when the proportional hazards assumption does not hold for some covariates. In addition, the proposed model allows covariate-dependent censoring. When multiple causes are of interest, multiple case-cohort studies may be conducted. Under multiple case-cohort studies, expensive covariate information from the other causes is also available. By incorpo-

rating such extra information into estimation, we propose a more efficient estimator. Sections 2 to 4 include the proposed method, the asymptotic properties of the proposed estimators, and the estimation of the cumulative incidence function. Simulation studies are conducted in Section 5. We apply the proposed method for a bone marrow transplant data set and the Atherosclerosis Risk in Communities study data set in Section 6. A brief conclusion is provided in Section 7.

2 Model and Estimation

2.1 Model definitions and assumptions

Suppose the full cohort consists of n subjects with K causes of failure $\epsilon \in \{1, \dots, K\}$, where ϵ denotes a cause of failure. We assume the primary cause of interest is $\epsilon = 1$. Let T , C , and $\mathbf{Z} = (Z_1, \dots, Z_p)^T$ be the failure time, the censoring time, and a $p \times 1$ vector of covariates, respectively, where \mathbf{Z} consists of time-dependent external covariates which are not affected by the causes of failure process (Kalbfleisch and Prentice, 2002). Hereafter we suppress its dependence on time for simplicity. We assume the (T, ϵ) 's are independent of the C 's given \mathbf{Z} . For right censored data, let $X = T \wedge C$ and $\Delta = I(T \leq C)$ denote the observed time and the failure indicator, respectively, where $I(\cdot)$ is an indicator function and $a \wedge b = \min(a, b)$. Assume we observe stratified data $(X_{li}, \Delta_{li}, \Delta_{li}\epsilon_{li}, \mathbf{Z}_{li})$ for subject i in stratum l , $i = 1, \dots, n_l$, $l = 1, \dots, L$, and $\sum_{l=1}^L n_l = n$. The number of strata L is finite. Non-stratified data can be handled as a special case of stratified data with $L = 1$. We assume subjects within-strata are independent and identically distributed and subjects between strata are independent. The study period is $[0, \tau]$. Our primary interest is evaluating the effect of covariates on the cumulative incidence function of cause 1, $F_{1l}(t|\mathbf{Z}_{li})$, where $F_{1l}(t|\mathbf{Z}_{li}) = P(T_{li} \leq t, \epsilon_{li} = 1|\mathbf{Z}_{li})$. A proportional subdistribution hazards model for cause 1 given \mathbf{Z}_{li} is

$$\lambda_{1l}(t|\mathbf{Z}_{li}) = \lambda_{1l0}(t) \exp(\boldsymbol{\beta}_0^T \mathbf{Z}_{li}), \quad (1)$$

where $\lambda_{1l0}(t)$ is an unspecified baseline subdistribution hazard function in stratum l and $\boldsymbol{\beta}_0$ is a p -dimensional parameter vector of interest (Zhou et al., 2011). Thus, the proposed model allows different baseline subdistribution hazard functions for different strata and assume all strata have the same covariate effect $\boldsymbol{\beta}_0$. A direct relationship between the subdistribution hazard function and the cumulative incidence function is $F_{1l}(t|\mathbf{Z}_{li}) = 1 - \exp\{-\Lambda_{1l}(t|\mathbf{Z}_{li})\}$, where $\Lambda_{1l}(t|\mathbf{Z}_{li}) = \int_0^t \lambda_{1l}(u|\mathbf{Z}_{li}) du$.

In many biomedical studies, the censoring time C may depend on the covariate vector \mathbf{Z} . In such case, we consider the proportional hazards model for the censoring distribution: $\lambda_l^C(t|\mathbf{Z}_{li}^C) = \lambda_{l0}^C(t) \exp(\boldsymbol{\gamma}_0^T \mathbf{Z}_{li}^C)$, where \mathbf{Z}_{li}^C is the covariate which is associated with the censoring distribution and can be a subset of \mathbf{Z}_{li} , $\lambda_{l0}^C(t)$ is an unspecified baseline censoring hazard function, and $\boldsymbol{\gamma}_0$ is an unknown parameter vector. Let $Y_{li}^C(t) = I(X_{li} \geq t)$ and $N_{li}^C(t) = I(X_{li} \leq t, \Delta_{li} = 0)$ denote at-risk indicator and counting process for the censoring time of subject i in stratum l , respectively.

2.2 Estimation under a single case-cohort study

Suppose we randomly select a subcohort with fixed size \tilde{n}_l from stratum l of the full cohort. Let ξ_{li} denote an indicator for the subcohort membership, i.e. $\xi_{li} = 1$ if subject i in stratum l is selected into the subcohort; otherwise 0. Let $\alpha_l = Pr(\xi_{li} = 1) = \tilde{n}_l/n_l$ denote the probability of selecting subject i in stratum l for the subcohort. Under the case-cohort design, expensive covariate information \mathbf{Z}_{li} is available for subcohort members and subjects that have failures from cause 1 outside the subcohort. Thus, we have records on $(X_{li}, \Delta_{li}, \Delta_{li}\epsilon_{li}, \xi_{li}, \mathbf{Z}_{li})$ when $\xi_{li} = 1$ or $\Delta_{li}I(\epsilon_{li} = 1) = 1$; and $(X_{li}, \Delta_{li}, \Delta_{li}\epsilon_{li}, \xi_{li})$ when $\xi_{li} = 0$ and $\Delta_{li}I(\epsilon_{li} = 1) = 0$. We call this design as a single case-cohort study.

Zhou et al. (2011) extended the subdistribution hazards model of Fine and Gray (1999) to right-censored stratified competing risks data. They assumed the C_{li} 's have a common censoring distribution $G_l(t)$ within stratum l to study the asymptotics of the estimators. Denote $\bar{G}_l(t)$ as the Kaplan-Meier estimator of $G_l(t)$ in stratum l . Let $N_{li}^1(t) = I(T_{li} \leq t, \epsilon_{li} = 1)$ and $Y_{li}^1(t) = 1 - N_{li}^1(t^-)$ denote the underlying counting process and risk process, respectively. Zhou et al. (2011) proposed the following weighted score equation to estimate $\boldsymbol{\beta}_0$ for the full cohort study:

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \mathbf{E}_l(\boldsymbol{\beta}, u)\} \hat{w}_{li}^{KM}(u) dN_{li}^1(u) = \mathbf{0}, \quad (2)$$

where $\mathbf{E}_l(\boldsymbol{\beta}, t) = \mathbf{S}_l^{(1)}(\boldsymbol{\beta}, t)/S_l^{(0)}(\boldsymbol{\beta}, t)$, $\mathbf{S}_l^{(d)}(\boldsymbol{\beta}, t) = n_l^{-1} \sum_{i=1}^{n_l} \hat{w}_{li}^{KM}(t) Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}$ for $d = 0, 1, 2$, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, $a^{\otimes 2} = aa^T$, and $\hat{w}_{li}^{KM}(t) = I(C_{li} \geq T_{li} \wedge t) \bar{G}_l(t) / \bar{G}_l(X_{li} \wedge t)$. We denote the estimator of $\boldsymbol{\beta}_0$ obtained by solving (2) as $\hat{\boldsymbol{\beta}}_F$. The weighted score equation (2) is reduced to the estimating equation of Fine and Gray (1999) when $L = 1$.

Under a single case-cohort design, covariate information is not available for subjects

without failure from cause 1 and outside the subcohort. Thus, we use an inverse probability weighting scheme to account for it. We propose the following weight function for a single case-cohort study with competing risks data:

$$\rho_{li} = \Delta_{li}I(\epsilon_{li} = 1) + \{1 - \Delta_{li}I(\epsilon_{li} = 1)\}\xi_{li}\hat{\alpha}_l^{-1},$$

where $\hat{\alpha}_l = \sum_{i=1}^{n_l} \xi_{li}\{1 - \Delta_{li}I(\epsilon_{li} = 1)\} / \sum_{i=1}^{n_l} \{1 - \Delta_{li}I(\epsilon_{li} = 1)\}$ is an estimator of the true subcohort selection probability α_l in stratum l , that is, the proportion of sampled subjects that do not have failure from cause 1 in stratum l . We have $\rho_{li} = 1$ for subjects who experienced a failure from cause 1 regardless of their subcohort membership and $\rho_{li} = \hat{\alpha}_l^{-1}$ for subjects without a failure from cause 1 in the subcohort. This type of weight function was used for survival outcomes under a single case-cohort study (Kalbfleisch and Lawless, 1988).

When the censoring distribution depends on covariates, as in He et al. (2016) we propose to use covariate-adjusted weight function $w_{li}^{Cox}(t) = I(C_{li} \geq T_{li} \wedge t)G_l(t|\mathbf{z}_{li}^C)/G_l(X_{li} \wedge t|\mathbf{z}_{li}^C)$. Following He et al. (2016), we assume the C_{li} 's in stratum l follow the proportional hazards model: $G_l(t|\mathbf{z}_{li}^C) = P(C_{li} > t|\mathbf{Z}_{li}^C = \mathbf{z}_{li}^C) = \exp\{-\Lambda_{l0}^C(t) \exp(\boldsymbol{\gamma}_0^T \mathbf{z}_{li}^C)\}$. Because a single case-cohort study consists of the subcohort and all cases with cause 1, not all censored observations have expensive covariate information. In other words, when we treat censoring as an event, expensive covariate information is available for a subset of subjects with censoring and subjects without censoring. Therefore, to estimate censoring survival probabilities given covariates, we propose to use a weighted estimating equation approach for generalized case-cohort data that allows a fraction of cases (Kim et al., 2018). Thus, we can estimate the censoring survival probability $G_l(t|\mathbf{z}_{li}^C)$ by $\hat{G}_l(t|\mathbf{z}_{li}^C) = \exp\{-\hat{\Lambda}_{l0}^C(t) \exp(\hat{\boldsymbol{\gamma}}^T \mathbf{z}_{li}^C)\}$, where $\hat{\boldsymbol{\gamma}}$ is the estimator of $\boldsymbol{\gamma}_0$ obtained by the proportional hazards model for generalized case-cohort data (Kim et al., 2018) and $\hat{\Lambda}_{l0}^C(t) = \int_0^t \sum_{i=1}^{n_l} \rho_{li} dN_{li}^C(u) / \sum_{j=1}^i \rho_{li} Y_{li}^C(u) \exp\{\hat{\boldsymbol{\gamma}}^T \mathbf{z}_{li}^C\}$ is a weighted Breslow-Aalen-type estimator for the cumulative baseline censoring hazard $\Lambda_{l0}^C(t) = \int_0^t \lambda_{l0}^C(u) du$. We denote $\hat{w}_{li}^{Cox}(t) = I(C_{li} \geq T_{li} \wedge t) \hat{G}_l(t|\mathbf{z}_{li}^C) / \hat{G}_l(X_{li} \wedge t|\mathbf{z}_{li}^C)$.

Using weight function ρ_{li} and covariate-adjusted censoring weighting $w_{li}^{Cox}(t)$, we propose the following pseudo-log-likelihood score equation for a single case-cohort study with competing risks data under model (1):

$$\hat{U}(\boldsymbol{\beta}) = \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \hat{\mathbf{E}}_l(\boldsymbol{\beta}, u)\} \hat{w}_{li}^{Cox}(u) dN_{li}^1(u) = \mathbf{0}, \quad (3)$$

where $\widehat{\mathbf{E}}_l(\boldsymbol{\beta}, t) = \widehat{\mathbf{S}}_l^{(1)}(\boldsymbol{\beta}, t)/\widehat{\mathbf{S}}_l^{(0)}(\boldsymbol{\beta}, t)$ and $\widehat{\mathbf{S}}_l^{(d)}(\boldsymbol{\beta}, t) = n_l^{-1} \sum_{i=1}^{n_l} \rho_{li} \widehat{w}_{li}^{Cox}(t) Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}$ for $d = 0, 1, 2$. Denote $\widehat{\boldsymbol{\beta}}_I$ as the solution to equation (3).

To estimate the baseline cumulative subdistribution hazard $\widehat{\Lambda}_{1l0}^I(\widehat{\boldsymbol{\beta}}_I, t)$, we propose a Breslow-Aalen-type estimator as follows:

$$\widehat{\Lambda}_{1l0}^I(\widehat{\boldsymbol{\beta}}_I, t) = \frac{1}{n_l} \sum_{i=1}^{n_l} \int_0^t \frac{\widehat{w}_{li}^{Cox}(u) dN_{li}^1(u)}{\widehat{\mathbf{S}}_l^{(0)}(\widehat{\boldsymbol{\beta}}_I, u)}. \quad (4)$$

For covariate-independent censoring, we can estimate regression coefficients and the baseline cumulative subdistribution hazard function by replacing $\widehat{w}_{li}^{Cox}(u)$ with $\widehat{w}_{li}^{KM}(u)$ in (3) and (4).

2.3 Efficient estimation under multiple case-cohort studies

When there are multiple causes of interest, several case-cohort studies can be conducted by using the same subcohort (Langholz and Thomas, 1990; Wacholder et al., 1991). Under multiple case-cohort studies, covariate information is available for the following two groups of subjects: (i) a randomly selected subcohort from the full cohort; (ii) all cases from any causes outside the subcohort. Thus, the information available under multiple-case cohort studies is $(X_{li}, \Delta_{li}, \Delta_{li}\epsilon_{li}, \xi_{li}, \mathbf{Z}_{li})$ when $\xi_{li} = 1$ or $\Delta_{li}I(\epsilon_{li} = k) = 1$ for $k = 1, \dots, K$; and $(X_{li}, \Delta_{li}, \Delta_{li}\epsilon_{li}, \xi_{li})$ when $\xi_{li} = 0$ and $\Delta_{li}I(\epsilon_{li} = k) = 0$ for $k = 1, \dots, K$.

Under multiple case-cohort studies, Kim et al. (2013) proposed an efficient estimation approach for multivariate survival outcomes by using the collected information on subjects who have other diseases outside the subcohort. By incorporating the extra information into estimation, they showed their method improved estimation efficiency compared to the method ignoring the extra information. Motivated by Kim et al. (2013), we propose the following efficient weight function π_{li} for multiple case-cohort studies with competing risks data:

$$\pi_{li} = \sum_{k=1}^K \Delta_{li}I(\epsilon_{li} = k) + \{1 - \sum_{k=1}^K \Delta_{li}I(\epsilon_{li} = k)\} \xi_{li} \widetilde{\alpha}_l^{-1},$$

$\widetilde{\alpha}_l = \sum_{i=1}^{n_l} \xi_{li} \{1 - \sum_{k=1}^K \Delta_{li}I(\epsilon_{li} = k)\} / \sum_{i=1}^{n_l} \{1 - \sum_{k=1}^K \Delta_{li}I(\epsilon_{li} = k)\}$ is the proportion of sampled subjects who do not have cases from any causes in stratum l . Thus, we have $\pi_{li} = 1$ when $\Delta_{li}I(\epsilon = k) = 1$ for some k and $\pi_{li} = \widetilde{\alpha}_l^{-1}$ when $\xi_{li} = 1$ and $\Delta_{li}I(\epsilon_i = k) = 0$ for all

$k = 1, \dots, K$ and $l = 1, \dots, L$. The weight function π_{li} takes cases from the other causes into consideration so that our proposed estimator can use the extra information on cases from the other causes $k = 2, \dots, K$. Thus, we propose the following pseudo-log-likelihood score equation for model (1):

$$\tilde{U}(\boldsymbol{\beta}) = \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, u) \} \tilde{w}_{li}^{Cox}(u) dN_{li}^1(u) = \mathbf{0}, \quad (5)$$

where $\tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) = \tilde{\mathbf{S}}_l^{(1)}(\boldsymbol{\beta}, t) / \tilde{S}_l^{(0)}(\boldsymbol{\beta}, t)$, $\tilde{\mathbf{S}}_l^{(d)}(\boldsymbol{\beta}, t) = n_l^{-1} \sum_{i=1}^{n_l} \pi_{li} \tilde{w}_{li}^{Cox}(t) Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}$ for $d = 0, 1, 2$. For $\tilde{w}_{li}^{Cox}(t) = I(C_{li} \geq T_{li} \wedge t) \tilde{G}_l(t | \mathbf{z}_{li}^C) / \tilde{G}_l(X_{li} \wedge t | \mathbf{z}_{li}^C)$, $\tilde{G}_l(t | \mathbf{z}_{li}^C)$ can be obtained using Kim et al. (2018) for multiple case-cohort studies and $\tilde{\Lambda}_{l0}^C(t) = \int_0^t \sum_{i=1}^{n_l} \pi_{li} dN_{li}^C(u) / \sum_{j=1}^i \pi_{lj} Y_{lj}^C(u) \exp\{\hat{\boldsymbol{\gamma}}^T \mathbf{z}_{li}^C\}$ similarly to Section 2.2. Let $\hat{\boldsymbol{\beta}}_{II}$ denote the solution to equation (5).

For multiple case-cohort studies, we propose the following Breslow-Aalen-type estimator for the baseline cumulative hazard function:

$$\hat{\Lambda}_{l0}^{II}(\hat{\boldsymbol{\beta}}_{II}, t) = \frac{1}{n_l} \sum_{i=1}^{n_l} \int_0^t \frac{\tilde{w}_{li}^{Cox}(u) dN_{li}^1(u)}{\tilde{S}_l^{(0)}(\hat{\boldsymbol{\beta}}_{II}, u)}. \quad (6)$$

For covariate-independent censoring, we can estimate regression coefficients and the baseline cumulative subdistribution hazard function by replacing $\tilde{w}_{li}^{Cox}(u)$ with $\hat{w}_{li}^{KM}(u)$ in (5) and (6).

3 Asymptotic properties

3.1 Asymptotic properties of proposed estimators

In this section, we study the asymptotic properties of the proposed estimators $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\beta}}_{II}$. Define $M_{li}^1(t) = N_{li}^1(t) - \int_0^t Y_{li}^1(u) d\Lambda_{l1}(u)$ and the martingale for the censoring process $M_{li}^C(t) = N_{li}^C(t) - \int_0^t Y_{li}^C(u) d\Lambda_l^C(u)$.

We make the following assumptions:

C1 For all l , $\int_0^\tau \lambda_{l10}(t) dt < \infty$ and $P\{Y_{li}^1(t) = 1\} > 0$ and $\int_0^\tau \lambda_{l10}^C(t) dt < \infty$ and $P\{Y_{li}^C(t) = 1\} > 0$ for $t \in [0, \tau]$, $i = 1, \dots, n_l$.

C2 $|Z_{lij}(0)| + \int_0^\tau |dZ_{lij}(t)| < D_z < \infty$ $l = 1, \dots, L, i = 1, \dots, n_l, j = 1, \dots, p$, almost surely and D_z is a constant.

C3 For $d = 0, 1, 2$, there exists a neighborhood \mathcal{B} of β_0 such that $\mathbf{s}_i^{(d)}(\beta, t)$ are continuous functions and $\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|\mathbf{S}_i^{(d)}(\beta, t) - \mathbf{s}_i^{(d)}(\beta, t)\| \xrightarrow{p} 0$ where $\mathbf{s}_i^{(d)}(\beta, t) = E[\mathbf{S}_i^{(d)}(\beta, t)]$ and there exists a neighborhood \mathcal{R} of γ_0 such that $\mathbf{s}_{C,l}^{(d)}(\gamma, t)$ are continuous functions and $\sup_{t \in [0, \tau], \gamma \in \mathcal{R}} \|\mathbf{S}_{C,l}^{(d)}(\gamma, t) - \mathbf{s}_{C,l}^{(d)}(\gamma, t)\| \xrightarrow{p} 0$, where $\mathbf{S}_{C,l}^{(d)}(\gamma, t) = n_l^{-1} \sum_{i=1}^{n_l} Y_{li}^C(t) \mathbf{Z}_{li}^{C \otimes d} e^{\gamma^T \mathbf{Z}_{li}^C}$ and $\mathbf{s}_{C,l}^{(d)}(\gamma, t) = E[\mathbf{S}_{C,l}^{(d)}(\gamma, t)]$.

C4 For all $\beta \in \mathcal{B}$ where \mathcal{B} is a neighborhood of β_0 , $t \in [0, \tau]$, and $l \in \{1, \dots, L\}$, we have $\mathbf{s}_i^{(1)}(\beta, t) = \partial \mathbf{s}_i^{(0)}(\beta, t) / \partial \beta$, and $\mathbf{s}_i^{(2)}(\beta, t) = \partial^2 \mathbf{s}_i^{(0)}(\beta, t) / \partial \beta \partial \beta^T$, where $\mathbf{s}_i^{(d)}(\beta, t)$, $d = 0, 1, 2$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$. $s_i^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$. For all $\gamma \in \mathcal{R}$ where \mathcal{R} is a neighborhood of γ_0 , $t \in [0, \tau]$, and $l \in \{1, \dots, L\}$, we have $\mathbf{s}_{C,l}^{(1)}(\gamma, t) = \partial \mathbf{s}_{C,l}^{(0)}(\gamma, t) / \partial \gamma$, and $\mathbf{s}_{C,l}^{(2)}(\gamma, t) = \partial^2 \mathbf{s}_{C,l}^{(0)}(\gamma, t) / \partial \gamma \partial \gamma^T$, where $\mathbf{s}_{C,l}^{(d)}(\gamma, t)$, $d = 0, 1, 2$ are continuous functions of $\gamma \in \mathcal{R}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{R} \times [0, \tau]$. $s_{C,l}^{(0)}$ is bounded away from zero on $\mathcal{R} \times [0, \tau]$.

C5 The matrix $\mathbf{A}_l(\beta_0) = \int_0^\tau \mathbf{v}_l(\beta_0, t) s_l^{(0)}(\beta_0, t) \lambda_{l0}(t) dt$ is positive definite for $l = 1, \dots, L$, where $\mathbf{v}_l(\beta, t) = \mathbf{s}_i^{(2)}(\beta, t) / s_i^{(0)}(\beta, t) - \mathbf{e}_l(\beta, t)^{\otimes 2}$ with $\mathbf{e}_l(\beta, t) = \mathbf{s}_i^{(1)}(\beta, t) / s_i^{(0)}(\beta, t)$; The matrix $\mathbf{A}_{C,l}(\gamma_0) = \int_0^\tau \mathbf{v}_{C,l}(\gamma_0, t) s_{C,l}^{(0)}(\gamma_0, t) \lambda_{l0}^C(t) dt$ is positive definite for $l = 1, \dots, L$, where $\mathbf{v}_{C,l}(\gamma, t) = \mathbf{s}_{C,l}^{(2)}(\gamma, t) / s_{C,l}^{(0)}(\gamma, t) - \mathbf{e}_{C,l}(\gamma, t)^{\otimes 2}$ with $\mathbf{e}_{C,l}(\gamma, t) = \mathbf{s}_{C,l}^{(1)}(\gamma, t) / s_{C,l}^{(0)}(\gamma, t)$.

C6 $\lim_{n \rightarrow \infty} \tilde{n}_l / n_l = \alpha_l^*$, where α_l^* is a positive constant for $l = 1, \dots, L$.

C7 $\lim_{n \rightarrow \infty} n_l / n = q_l$, where q_l is a positive constant for $l = 1, \dots, L$.

We have the following theorem on $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$:

Theorem 1. Under assumptions C1 – C7, for $w = I, II$, $\widehat{\beta}_w$ converges in probability to β_0 and $n^{1/2}(\widehat{\beta}_w - \beta_0)$ converges in distribution to a zero-mean normal distribution with

covariance matrix $\mathbf{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}^w(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)^{-1}$, where

$$\begin{aligned}
\mathbf{A}(\boldsymbol{\beta}_0) &= \sum_{l=1}^L q_l \mathbf{A}_l(\boldsymbol{\beta}_0), \quad \boldsymbol{\Sigma}^w(\boldsymbol{\beta}_0) = \sum_{l=1}^L q_l (\mathbf{V}_l^0 + \mathbf{V}_l^w), \\
\mathbf{V}_l^0 &= E\{(\boldsymbol{\eta}_{1,l1} + \boldsymbol{\eta}_{2,l1})^{\otimes 2}\}, \quad \mathbf{V}_l^w = \frac{1 - \alpha_l^*}{\alpha_l^*} E\{(\boldsymbol{\eta}_{3,l1}^w)^{\otimes 2}\}, \\
\boldsymbol{\eta}_{1,li} &= \int_0^\tau \{\mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}_0, t)\} w_{li}^{Cox}(t) dM_{li}^1(t), \quad \boldsymbol{\eta}_{2,li} = \int_0^\tau \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t), \\
\boldsymbol{\eta}_{3,li}^I &= \int_0^\tau \{1 - \Delta_{li} I(\epsilon_{li} = 1)\} \left[w_{li}^{Cox}(t) Y_{li}^1(t) \{\mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}_0, t)\} d\Lambda_{li0}(t) - \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t) \right], \\
\boldsymbol{\eta}_{3,li}^{II} &= \int_0^\tau \left\{ 1 - \sum_{k=1}^K \Delta_{li} I(\epsilon_{li} = k) \right\} \left[w_{li}^{Cox}(t) Y_{li}^1(t) \{\mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}_0, t)\} d\Lambda_{li0}(t) - \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t) \right], \\
\mathbf{h}_C(t, u, \mathbf{Z}^C) &= e^{\gamma^T \mathbf{Z}^C} \int_{v=u}^t \left\{ \mathbf{Z}^C - \frac{\mathbf{s}_{C,l}^{(1)}(\gamma_0, u)}{\mathbf{s}_{C,l}^{(0)}(\gamma_0, u)} \right\} d\Lambda_{0l}^C(v), \\
\mathbf{q}_{li}^{(1)}(u) &= - \lim_{n_l \rightarrow \infty} \left[\frac{1}{n_l} \sum_{j=1}^{n_l} \int_{t=X_{lj}}^\tau \{\mathbf{Z}_{lj} - \mathbf{e}_l(\boldsymbol{\beta}_0, t)\} w_{lj}^{Cox}(t) \frac{e^{\gamma_0^T \mathbf{Z}_{lj}^C} I(u \leq t)}{\mathbf{s}_{C,l}^{(0)}(\gamma_0, u)} dM_{lj}^1(t) \right. \\
&+ \left. \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \int_{t=X_{lj}}^\tau \{\mathbf{Z}_{lj} - \mathbf{e}_l(\boldsymbol{\beta}_0, t)\} w_{lj}^{Cox}(t) \mathbf{h}_C^T(t, X_{lj}, \mathbf{Z}_{lj}^C) \mathbf{A}_C^{-1} \{\mathbf{Z}_{li}^C - \mathbf{e}_{C,l}(\gamma_0, u)\} dM_{hj}^1(t) \right], \\
\mathbf{A}_C &= \sum_{l=1}^L q_l \mathbf{A}_{C,l}(\gamma_0).
\end{aligned}$$

The matrix $\boldsymbol{\Sigma}^w(\boldsymbol{\beta}_0)$ for $w = I, II$ consists of two parts. The first part \mathbf{V}_l^0 is the contribution from the full cohort and the second part \mathbf{V}_l^w is due to sampling the subcohort. When we select all subjects for the subcohort (i.e. $\alpha_l^* = 1$), the second variance term vanishes. The asymptotics of the estimators with $\widehat{w}_{li}^{KM}(u)$ can be shown similarly to Theorem 1 and thus its proof is omitted. The asymptotic covariance matrices of the estimators with $\widehat{w}_{li}^{KM}(u)$ for a single and multiple case-cohort studies are provided in the Appendix C of the Supplementary Materials.

Remark. We considered the same strata for event from cause 1 and censoring times for mathematical simplicity in Theorem 1. This avoids the abuse of complicated notations and subscripts. However, one can show the asymptotic results of Theorem 1 even when strata are

different between the model for cause 1 and the model for censoring using similar arguments to the proof of Theorem 1 after redefining all notations that allow different strata between the model for cause 1 and the model for censoring. When strata for event from cause 1 and censoring times are different, one can do the following to obtain parameter estimates and their standard errors: i) fit the stratified proportional hazards model for censoring using strata for censoring and estimate a censoring survival probability for each subject based on the fitted model; ii) plug the estimated censoring survival probability corresponding to subject i in stratum l to estimate $w_{li}^{Cox}(t)$, where $l = 1, \dots, L$ are strata for the model for cause 1; iii) estimate β using the estimated $w_{li}^{Cox}(t)$; iv) similarly to estimating $w_{li}^{Cox}(t)$, estimate $M_{li}^C(t)$, $\mathbf{q}_{li}^{(1)}(u)$, $\boldsymbol{\eta}_{1,li}$, $\boldsymbol{\eta}_{2,li}$, and $\boldsymbol{\eta}_{3,li}$ by plugging the estimated β and $w_{li}^{Cox}(t)$ into the expressions in Theorem 1; and v) obtain the standard error using the asymptotic formula of Theorem 1. In short, one can fit the stratified proportional hazards model using strata for censoring, estimate all censoring-related terms for each subject, and plug them into the asymptotic formula of Theorem 1. We conducted a simulation study with different strata between the model for cause 1 and the model for censoring in the Appendix F of the Supplementary Materials. Table 3 of the Appendix F of the Supplementary Materials shows little bias and empirical coverage rates close to 95%.

We have the following theorem on the cumulative hazard function:

Theorem 2. Under assumptions C1 - C7, $\widehat{\Lambda}_{110}^w(t; \mathbf{z}_0)$ is a consistent estimator for $\Lambda_{110}(t; \mathbf{z}_0)$ for $t \in [0, \tau]$ and $n_l^{1/2}\{\widehat{\Lambda}_{110}^w(t; \mathbf{z}_0) - \Lambda_{110}(t; \mathbf{z}_0)\}$ converges weakly to the Gaussian process with mean zero and the following covariance matrix between $\mathcal{H}^w(t)$ and $\mathcal{H}^w(s)$ for $w = I, II$, where

$$E\{\zeta_{1,1l}(\boldsymbol{\beta}_0, t)\zeta_{1,1l}(\boldsymbol{\beta}_0, s)\} + \frac{1 - \alpha_l^*}{\alpha_l^*} E\{\zeta_{2,1l}^w(\boldsymbol{\beta}_0, t)\zeta_{2,1l}^w(\boldsymbol{\beta}_0, s)\},$$

for $t, s \in [0, \tau]$. The explicit forms of $\zeta_{1,li}(\boldsymbol{\beta}_0, t)$ and $\zeta_{2,li}^w(\boldsymbol{\beta}_0, t)$ are provided in the Appendix. The details of the proofs for Theorem 1 and Theorem 2 are provided in the Appendix A and B of the Supplementary Materials, respectively.

3.2 Efficiency gain

We compare the asymptotic variances of $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ and evaluate asymptotic efficiency gain in this section. Theorem 1 shows the sandwich covariance matrices of $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$

depend on the first derivative of the weighted estimating function, $\mathbf{A}(\boldsymbol{\beta}_0)$, and the asymptotic variance of the weighted estimating functions, $\boldsymbol{\Sigma}^I(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}^{II}(\boldsymbol{\beta}_0)$. Since $\mathbf{A}_i(\boldsymbol{\beta}_0) = \int_0^\tau \mathbf{v}_i(\boldsymbol{\beta}_0, t) s_i^{(0)}(\boldsymbol{\beta}_0, t) \lambda_{i0}(t) dt$ is independent of the weight functions, the difference in the asymptotic variances of $\widehat{\boldsymbol{\beta}}_I$ and $\widehat{\boldsymbol{\beta}}_{II}$ comes from $\boldsymbol{\Sigma}^I(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}^{II}(\boldsymbol{\beta}_0)$. By Theorem 1, the difference between $\boldsymbol{\Sigma}^I(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}^{II}(\boldsymbol{\beta}_0)$ is due to the difference between $E\{(\boldsymbol{\eta}_{3,l1}^I)^{\otimes 2}\}$ and $E\{(\boldsymbol{\eta}_{3,l1}^{II})^{\otimes 2}\}$. Define $EG = (1 - \alpha_i^*)/\alpha_i^*[E\{(\boldsymbol{\eta}_{3,l1}^I)^{\otimes 2}\} - E\{(\boldsymbol{\eta}_{3,l1}^{II})^{\otimes 2}\}]$. As explained in the Appendix D of the Supplementary Materials, the asymptotic efficiency gain gets bigger as $(1 - \alpha_i^*)/\alpha_i^*E\{\sum_{k=2}^K \Delta_{l1} I(\epsilon_{l1} = k)\}$ becomes larger. Thus, the asymptotic efficiency gain is positively associated with smaller subcohort selection probability and larger other cause rates. More specifically, for fixed α^* , more extra information collected on subjects with the other causes leads to efficiency gain. When event rates from the other causes are fixed, a smaller subcohort selection probability induces a larger ratio of case to control. As a result, it increases efficiency gain.

4 Predicting Cumulative incidence

In this section, we estimate the cumulative incidence at time t for an individual with covariate vector $\mathbf{Z} = \mathbf{z}_0$ under the case-cohort design. We can estimate the cumulative subdistribution hazard by $\widehat{\Lambda}_{1l}^I(t; \mathbf{z}_0) = \int_0^t \exp\{\widehat{\boldsymbol{\beta}}_I^T \mathbf{z}_0\} d\widehat{\Lambda}_{1l0}^I(u)$ under a single case-cohort study and $\widehat{\Lambda}_{1l}^{II}(t; \mathbf{z}_0) = \int_0^t \exp\{\widehat{\boldsymbol{\beta}}_{II}^T \mathbf{z}_0\} d\widehat{\Lambda}_{1l0}^{II}(u)$ under multiple case-cohort studies. By Theorem 2, $n_l^{1/2}\{\widehat{\Lambda}_{1l0}^w(t; \mathbf{z}_0) - \Lambda_{1l0}(t; \mathbf{z}_0)\}$ converges weakly to a Gaussian process on $[0, \tau]$ for $w = I, II$.

The cumulative incidence function for a single case-cohort study ($w = I$) and multiple case-cohort studies ($w = II$) can be estimated as follows:

$$\widehat{F}_{1l}(t|\mathbf{z}_0) = 1 - \exp\left\{-\int_0^t \exp\{\widehat{\boldsymbol{\beta}}_w^T \mathbf{z}_0\} d\widehat{\Lambda}_{1l0}^w(\widehat{\boldsymbol{\beta}}_w, u)\right\}.$$

For multiple case-cohort studies, by the functional delta method, $n_l^{1/2}[\widehat{F}_{1l}(t|\mathbf{z}_0) - F_{1l}(t|\mathbf{z}_0)]$ converges weakly to a Gaussian process with mean zero and asymptotic variance, which can be consistently estimated by $n_l\left\{1 - \widehat{F}_{1l}(t|\mathbf{z}_0)\right\}^2 \sum_i \left\{\widehat{\mathbf{W}}_{F,li}(t|\mathbf{z}_0)\right\}^2$, where

$$\begin{aligned} \widehat{\mathbf{W}}_{F,li}(t|\mathbf{z}_0) &= \exp\{\widehat{\boldsymbol{\beta}}_{II}^T \mathbf{z}_0\} [\widehat{\Lambda}_{1l0}^{II}(t) \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}_{II})^{-1} \widetilde{\mathbf{U}}(\widehat{\boldsymbol{\beta}}_{II}) \mathbf{z}_0 + \widehat{H}(t; \mathbf{z}_0)], \\ \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}_{II}) &= \sum_{l=1}^L \frac{n_l}{n} \int_0^\tau \left[\frac{\widetilde{\mathbf{S}}_l^{(2)}(\widehat{\boldsymbol{\beta}}_{II}, t)}{\widetilde{\mathbf{S}}_l^{(0)}(\widehat{\boldsymbol{\beta}}_{II}, t)} - \frac{\widetilde{\mathbf{S}}_l^{(1)}(\widehat{\boldsymbol{\beta}}_{II}, t)}{\widetilde{\mathbf{S}}_l^{(0)}(\widehat{\boldsymbol{\beta}}_{II}, t)} \right] \widetilde{\mathbf{S}}_l^{(0)}(\widehat{\boldsymbol{\beta}}_{II}, t) d\widehat{\Lambda}_{1l0}^{II}(t), \end{aligned}$$

$$\widehat{H}(t; \mathbf{z}_0) = n_l^{-1/2} \sum_{i=1}^{n_l} \{ \zeta_{1,li}(\widehat{\beta}_{II}, t) + (1 - \xi_{li}/\alpha_l) \zeta_{2,li}^{II}(\widehat{\beta}_{II}, t) \}.$$

The asymptotics of $\widehat{F}_{l1}(t|\mathbf{z}_0)$ for a single case-cohort study can be similarly established. Its detailed variance formula can be found and in the Appendix E of the Supplementary Materials.

5 Simulation

We conducted a simulation study to evaluate the performance of the proposed estimators $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ for two case-cohort studies with stratified sampling. We considered two causes. The cumulative incidence function for cause 1 and cause 2 given Z_{li} in stratum l had the following form:

$$\begin{aligned} F_{l1}(t|Z_{li}) &= 1 - \{1 - q + qe^{-\psi_l t^{\kappa_l}}\}^{\exp(\beta Z_{li})}, \\ F_{l2}(t|Z_{li}) &= \{1 - q\}^{\exp(\beta Z_{li})} \{1 - e^{-t \exp(\theta Z_{li})}\}, \end{aligned}$$

where $F_{l1}(t|Z_{li})$ is a Weibull mixture with mass $1 - q$ at ∞ when $Z_{li} = 0$ and uses the proportional subdistribution hazards model to obtain the subdistribution for nonzero covariate values. Two parameters ψ_l and κ_l allow to generate stratified data.

We considered two strata with $(\psi_1, \psi_2)^T = (1, 1)^T$ and $(\kappa_1, \kappa_2)^T = (1, 2)^T$. One covariate Z was considered. Covariate Z_{1i} in stratum 1 and Z_{2i} in stratum 2 were generated from the Bernoulli distribution with $Pr(Z_{1i} = 1) = 0.4$ and $Pr(Z_{2i} = 1) = 0.6$, respectively. We set $(\beta_0, \theta_0)^T$ to $(1, -1)^T$. We generated censoring time from the following two scenarios: S1) the uniform distribution; and S2) the proportional hazards model with the constant baseline hazard rate, $\gamma = 2.53$, and $Z_{li}^C = Z_{li}$. We set the failure rate for cause 1 to 20%. We considered 20% and 40% of failure rates for cause 2. The sample size of the full cohort in each stratum was set to 1000. Two subcohort sizes were examined: 100 ($\alpha_1 = \alpha_2 = 0.1$) and 200 ($\alpha_1 = \alpha_2 = 0.2$) in each stratum. For each configuration, 2000 iterations were conducted. For each simulation study, we compared the performance of using covariate-unadjusted weight and covariate-adjusted weight.

Table 1 reports the average bias of the estimates $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$, the average of the estimated standard error (SE), empirical standard deviation (SD), average of standardized bias (STB)

defined as $E\{\|\widehat{\beta} - \beta_0\|/SE\}$, the empirical coverage rate (%) with the nominal 95% confidence interval for various simulation settings. The simulation results show all estimators are approximately unbiased when the censoring time does not depend on covariates (scenario S1). The average of the estimated standard errors for both proposed estimators $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ are close to their empirical standard deviations, which indicates the proposed estimated standard errors provide a consistent estimate of the true variability of $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$ regardless of censoring weights.

When the censoring time depends on covariates (scenario S2), the covariate-unadjusted estimators are significantly biased while the covariate-adjusted estimators are approximately unbiased for both $\widehat{\beta}_I$ and $\widehat{\beta}_{II}$. Moreover, the empirical coverage rates for the majority of the covariate-unadjusted estimators are not close to 95%. As the censoring rate becomes smaller and the subcohort selection probability gets larger, the average of estimated standard error is decreased. The empirical coverage rates for the covariate-adjusted estimators are between 94% and 96%. Because the censoring time depends on binary covariates, one can nonparametrically estimate the censoring survival function separately for each level of the covariate as suggested in [Fine and Gray \(1999\)](#). The results with this nonparametric covariate-level-specific weight are similar to those with the covariate-adjusted weight in [Table 1](#). These results are provided in [Table 2](#) of the [Appendix F](#) of the [Supplementary Materials](#).

In both scenarios S1 and S2, all sample relative efficiency (SRE) values, defined as the empirical standard deviation for $\widehat{\beta}_I$ divided by that for $\widehat{\beta}_{II}$, are larger than 1. This indicates $\widehat{\beta}_{II}$ using extra information on covariates collected from subjects with failure from cause 2 is more efficient than $\widehat{\beta}_I$ ignoring such extra covariate information. The efficiency gain ranges between 14% and 91%. The efficiency gain is larger as subcohort sizes get smaller and failure rates for cause 2 become larger as discussed in [Section 3.2](#).

We also conducted a simulation study for non-stratified case-cohort data. The results were similar to [Table 1](#) and are provided in [Table 1](#) of the [Appendix F](#) of the [Supplementary Materials](#). To examine the robustness of using the proportional hazards model for censoring, we conducted a small simulation study, where we generated censoring times from an additive hazards model, but used the proportional hazards model to model censoring. Detailed simulation settings can be found in the [Appendix F](#) of the [Supplementary Materials](#). As in [Table 4](#) of the [Appendix F](#) of the [Supplementary Materials](#), the proposed method showed robustness against model misspecification including approximately unbiased estimates and empirical coverage rates close to 95%. A similar robust result against model misspecification

for censoring was observed in [He et al. \(2016\)](#) for the full cohort study.

6 Data analysis

We applied the proposed methods to a real bone marrow transplant study data set and the Atherosclerosis Risk in Communities study data set ([Ghosh et al., 2016](#); [Ballantyne et al., 2004](#)). The former data set had two competing risks outcomes: relapse/progression and non-relapse mortality. Because the bone marrow transplant data set was not case-cohort data, we used it to generate non-stratified case-cohort data consisting of the subcohort and all failures from relapse and non-relapse mortality. We used these data to compare the performance of $\hat{\beta}_I$ and $\hat{\beta}_{II}$.

The Atherosclerosis Risk in Communities study conducted a single case-cohort study with stratified sampling for two competing risks: coronary heart disease (CHD) and death prior to CHD. This case-cohort data set consisted of the subcohort and all cases with CHD. The subcohort included subjects who experienced death prior to CHD. Thus, we estimated $\hat{\beta}_I$ only for this data set.

6.1 Example 1: Non-stratified data under two case-cohort studies

We applied the proposed methods to a bone marrow transplant study data set collected by the Center for International Blood and Marrow Transplant Research ([Ghosh et al., 2016](#)). After excluding missing covariates, the full cohort consisted of 902 patients aged 18 years or older with Hodgkin lymphoma and aggressive non-Hodgkin lymphoma undergoing their first reduced-intensity or nonmyeloablative conditioning allogeneic hematopoietic cell transplantation between 2008 and 2013. Non-relapse mortality was an outcome of interest. Non-relapse mortality is defined as death without evidence of lymphoma relapse or progression. Thus, relapse or progression is a competing risk for non-relapse mortality. The number of subjects who had non-relapse mortality, relapse/progression, and censoring in the full cohort were 114 (13%), 344 (38%), and 444 (49%), respectively.

To compare the performance of $\hat{\beta}_I$ and $\hat{\beta}_{II}$ for case-cohort studies, we generated a data set for two case-cohort studies for non-relapse mortality and relapse/progression. The shared subcohort was selected from the full cohort using simple random sampling with selection probability 0.4. The subcohort size was 361 including 170 failure-free patients, 44 patients

with non-relapse mortality, and 147 patients with relapse or progression. Thus, the data set consisted of the shared subcohort and all cases with relapse/progression and non-relapse mortality. To obtain $\hat{\beta}_I$, we used the shared subcohort and all patients who experienced non-relapse mortality only. For obtaining $\hat{\beta}_{II}$, we used the shared subcohort and all patients who experienced non-relapse mortality or relapse/progression.

The covariates of interest were donor type (haploidentical donors vs. haplotype-identical siblings donors), standardized patient’s age, Karnofsky performance status (KPS) at transplant ($\geq 90\%$ vs. $< 90\%$), and histology (Follicular lymphoma, Diffuse large B-cell lymphoma, Mantle cell lymphoma, Mature T- and NK-cell lymphomas, and Hodgkin lymphoma). We checked the subdistribution proportional hazards assumption by testing whether the coefficient of $\log t \times Z$ is equal to zero for each variable and all p -values were greater than 0.16. We also examined whether the censoring distribution depends on covariates using the proportional hazards model. KPS and histology were significantly associated with the censoring distribution with p -values 0.0302 and 0.018, respectively. We obtained $\hat{\beta}_I$ and $\hat{\beta}_{II}$ for non-stratified data with covariate-unadjusted and covariate-adjusted weights. We also obtained $\hat{\beta}_F$ by fitting the subdistribution hazards models of [Fine and Gray \(1999\)](#) with covariate-unadjusted weight and [He et al. \(2016\)](#) with covariate-adjusted weight for the full cohort.

Table 2 presents the regression parameter estimates, their estimated standard errors, and their p -values. In general, the parameter estimates for $\hat{\beta}_{II}$ are closer to the full-cohort-based parameter estimates $\hat{\beta}_F$ than those for $\hat{\beta}_I$. All standard errors of $\hat{\beta}_{II}$ are smaller than those of $\hat{\beta}_I$. Age and KPS with covariate-adjusted weight were statistically significant at the significant level 0.05 in both $\hat{\beta}_I$ and $\hat{\beta}_{II}$, which is consistent with the result from the full cohort data analysis based on [He et al. \(2016\)](#), that is, $\hat{\beta}_F$ with covariate-adjusted weight. Compared to younger patients and patients with KPS $\geq 90\%$, older patients and patients with KPS $< 90\%$ experienced more non-relapse mortality, respectively.

We also fitted the cause-specific hazards models with the full cohort, the single case-cohort data set with ρ_{li} , and the two case-cohort data set with π_{li} ([Kim et al., 2013](#)). Their results are provided in the Appendix G of the Supplementary Materials. The magnitude of the parameter estimates from the cause-specific hazards models is similar to that in Table 2. However, the interpretations of the estimates from these two models are different. For example, consider the parameter estimate for KPS $< 90\%$ based on $\hat{\beta}_{II}$ with covariate-adjusted weight. The parameter estimate from the cause-specific hazards model is 0.749.

Thus, at any time after bone marrow transplant, patients with KPS < 90% had a hazard of death before relapse or progression $\exp(0.749) \approx 2.11$ times higher than those with KPS $\geq 90\%$ after adjusting for the other covariates, among patients who were disease-free, that is, had not experienced death or relapse or progression, at that time. On the other hand, the hazard ratio from the subdistribution hazards model is $\exp(0.555) \approx 1.74$. Thus, the cumulative incidence of death before relapse or progression was higher in patients with KPS < 90% when compared with patients with KPS $\geq 90\%$. However, the hazard ratio 1.74 is not straightforward to interpret because it is the mortality ratio before relapse or progression among patients who are alive or have been relapsed or progressed before. For more details on the interpretation of the subdistribution hazards ratio, see [Austin and Fine \(2017\)](#).

Figure 1 shows the predicted cumulative incidence curves using $\hat{\beta}_F$, $\hat{\beta}_I$, and $\hat{\beta}_{II}$ with covariate-adjusted weight for KPS < 90% when age is 50 years old, donor type is HLA-identical siblings donor, and disease subtype is Follicular lymphoma for Histology. The predicted cumulative incidence curves with the two case-cohort estimators $\hat{\beta}_I$ and $\hat{\beta}_{II}$ are close to that with the full cohort estimator $\hat{\beta}_F$. Especially, compared to the curve using the estimate with the traditional case-cohort weight $\hat{\beta}_I$, the curve using the estimate with the efficient weight $\hat{\beta}_{II}$ is closer to that based on the full cohort estimate $\hat{\beta}_F$.

6.2 Example 2: Stratified data under case-cohort studies

The Atherosclerosis Risk in Communities study is a longitudinal and large cohort study consisting of 15,792 men and women aged from 45 to 64 years at baseline. After a baseline examination during 1987–1989, subjects in this study were prospectively followed for the development of an incident coronary heart disease (CHD) and death through 1998 ([Ballantyne et al., 2004](#)). An event due to CHD is defined as definite or probable myocardial infarction, electrocardiographic evidence of silent myocardial infarction, definite coronary heart disease death, or coronary revascularization procedure. Death prior to CHD was a competing risk for CHD.

The primary interest of this study was evaluating the effect of high-sensitivity C-reactive protein (hs-CRP) on CHD ([Ballantyne et al., 2004](#)). To reduce cost and preserve the blood sample, a single case-cohort study was implemented. The values of hs-CRP were available on all subjects with CHD and the subcohort members. The subcohort was obtained based on stratified sampling with age groups (≥ 55 or < 55 years), race (Caucasian or African American), and gender as strata. We excluded the subjects who missed their second visit

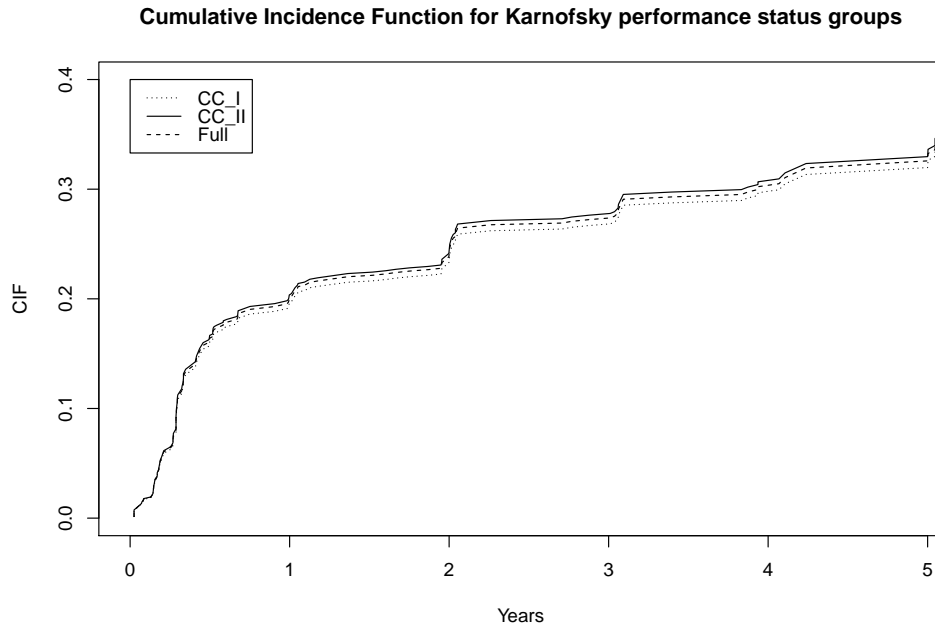


Figure 1: Predicted Cumulative incidence of non-relapse mortality for Karnofsky performance status groups

in 1990–1992, did not have information on CHD history, were under-represented minorities other than blacks, or had no valid follow-up time. There were 12,193 subjects in the full cohort consisting of 639 (5.3%) CHD cases, 965 (7.9%) deaths prior to CHD, and 10,589 (86.8%) event-free subjects. In this analysis, the total number of assayed blood sample with hs-CRP was 1409 including 818 subcohort members and 591 subjects with CHD outside the subcohort. The subcohort consisted of 48 subjects who experienced CHD, 176 subjects who died prior to CHD, 594 subjects who experienced neither CHD nor deaths. Tertiles of hs-CRP were classified into low ($< 1.0\text{mg/L}$), middle ($1.0\text{--}3.0\text{mg/L}$), and high ($> 3.0\text{mg/L}$) hs-CRP groups. The following covariates were adjusted in the analysis: smoking status, diabetes, standardized systolic blood pressure, and standardized high density lipoprotein and low density lipoprotein cholesterol. We tested the subdistribution proportional hazards assumption and whether the censoring distribution was covariate-dependent similarly to Section 6.1. All p -values from testing the subdistribution proportional assumptions were greater than 0.08. The censoring distribution depended on all covariates listed above (all p -values were less than 0.01).

Table 3 reports the regression parameter estimates and their estimated standard er-

rors from fitting the stratified subdistribution hazards model by solving (3). The results show the high hs-CRP group was significantly associated with increased risks of CHD compared with the low hs-CRP group after adjusting for smoking status, diabetes, systolic blood pressure, and high density lipoprotein and low density lipoprotein cholesterol. We also obtained covariate-unadjusted estimators and the results are very similar to those for covariate-adjusted estimators. Since the regression coefficients for the censoring distribution were from -0.0587 to 0.114 and were small in terms of magnitude, they had little impact on difference between covariate-unadjusted and covariate-adjusted estimates. Although we did not report additional simulation results in the article, we also observed the difference in covariate-unadjusted and covariate-adjusted estimates and their standard errors were small when the magnitude of the parameters for the censoring distribution was small.

The results for the cause-specific hazard models are provided in the Appendix G of the Supplementary Materials. The magnitude of the parameter estimates is similar to that of Table 3. The estimates for the two models can be interpreted similarly to Section 6.1.

Figure 2 shows the predicted cumulative incidence curves using $\widehat{\beta}_I$ with covariate-adjusted weight for the three hs-CRP groups when low and high density lipoproteins are averages, patients do not have diabetes and do not smoke in stratum with white female who are older than 55. It shows the high (low) hs-CRP group experienced the highest (lowest) CHD.

7 Discussion

We proposed a stratified subdistribution hazards model for case-cohort data with a possible covariate-dependent censoring distribution. The proposed method can be used via stratification when the proportional subdistribution hazards assumption is not valid. For multiple case-cohort studies, we proposed an efficient estimator by considering information on subjects who experienced a failure from causes other than cause of interest. Although the interpretation of the parameter estimates from the subdistribution hazards model is not straightforward, it directly estimates the effect of the covariates on the cumulative incidence function of cause of interest.

In this paper, we considered the stratified subdistribution hazards model, which allows different baseline hazards for different strata. We can modify the proposed model for the stratified case-cohort design under which the common baseline hazard is assumed for all strata. Under the stratified case-cohort design, a subdistribution hazard model is $\lambda_1(t|\mathbf{Z}_{i1}) =$

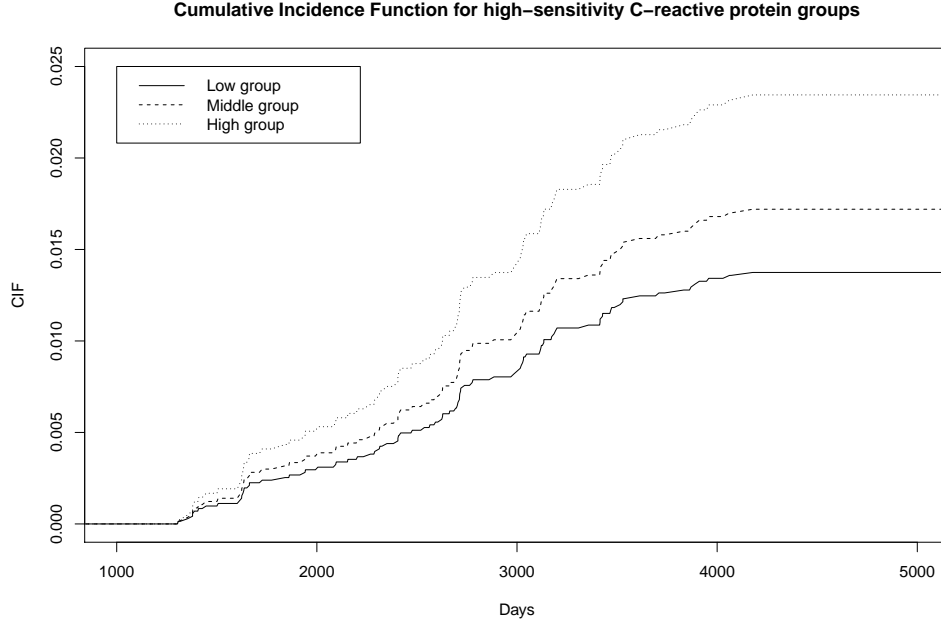


Figure 2: Predicted Cumulative incidence of Coronary heart disease for CRP groups

$\lambda_{10}(t) \exp(\beta^0 \mathbf{Z}_{li})$, and the weighted score equation with covariate-adjusted weight is

$$\mathbf{U}_*(\beta) = \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \left\{ \mathbf{Z}_{li} - \frac{\mathbf{S}_*^{(1)}(\beta, u)}{\mathbf{S}_*^{(0)}(\beta, u)} \right\} \check{w}_{li}^{Cox}(u) dN_{li}^1(u) = \mathbf{0},$$

where $\mathbf{S}_*^{(d)}(\beta, t) = n^{-1} \sum_{l=1}^L \sum_{i=1}^{n_l} \pi_{li}^* \check{w}_{li}^{Cox}(t) Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\beta^T \mathbf{Z}_{li}}$ for $d = 0, 1$. One can use $\pi_{li}^* = \rho_{li}$ and $\check{w}_{li}^{Cox}(t) = \widehat{w}_{li}^{Cox}(t)$ for a single case-cohort study, and $\pi_{li}^* = \pi_{li}$ and $\check{w}_{li}^{Cox}(t) = \widetilde{w}_{li}^{Cox}(t)$ for multiple case-cohort studies. The weight functions ρ_{li} and π_{li} under the stratified case-cohort design still remain the same as that for the stratified subdistribution hazards model. Under the stratified case-cohort design, the effect of sampling strata, e.g. a surrogate of an exposure of main interest, can be adjusted by using stratified sampling for the subcohort to improve estimation efficiency.

The proposed method uses the stratified proportional hazards model for the censoring distribution. Although the limited simulation study we conducted suggested the robustness of the proposed method against model misspecification for censoring, when the stratified proportional hazards model is inappropriate for fitting the censoring outcome, one can use different methods such as the accelerated failure time model or the additive hazards model.

In this case, the proposed method can be easily modified for different models for censoring so that it can be used when the asymptotic variance of the estimator is established. Studying the asymptotic properties of the parameter estimator when using different models for censoring is an important future research topic. Another interesting topic is examining the robustness of using the proportional hazards model for the censoring distribution when the proportional hazards model is inappropriate. Exploring this aspect requires extensive simulation studies under the accelerated failure time model or the additive model for censoring with various sample sizes.

Under our simulation settings, the performance of the model with covariate-adjusted censoring weight was better than or at least equivalent to that with covariate-unadjusted censoring weight whether the censoring distribution depended on certain covariates or not. However, when the number of covariates is large, fitting the proportional hazards model for censoring could lead to inefficient parameter estimation. In such cases, one may select significant covariates for the censoring distribution first and then use the covariate-adjusted censoring weight with the selected covariates. Investigating estimation efficiency gain from variable selection for the censoring distribution would be an interesting study.

For our bone marrow transplant data example, all patients got transplant. Thus, they are left truncated by the waiting time to transplant if we start clock at time of diagnosis of leukemia. Several methods to account for left truncation under the competing risks data have been proposed for the full cohort data (Zhang et al., 2009; Geskus, 2011; Liu et al., 2018). One normally assumes that truncation time is independent of event time to handle delayed entry (Zhang et al., 2009; Geskus, 2011). However, this assumption may be violated in practice. Developing methods for such left-truncated data is an important future research problem.

For studies with common diseases or a large number of failures from a cause of interest, sampling all cases in the original case-cohort design limits its applications (Breslow and Wellner, 2007). In this case, the generalized case-cohort design in which one can sample only a fraction of cases for exposure assessment can be used (Cai and Zeng, 2007). Developing a method to model competing risks data under such design would be a worthy future topic.

In practice, it is important for investigators to calculate the sample size before conducting the case-cohort design. Seeking simple formulae for sample size and power calculation for the case-cohort design with competing risks outcomes would be another interesting future topic.

Appendix

The following definitions are the explicit forms of $\zeta_{1,li}(\boldsymbol{\beta}_0, t)$, $\zeta_{2,li}^I(\boldsymbol{\beta}_0, t)$, $\zeta_{2,li}^{II}(\boldsymbol{\beta}_0, t)$, $\mathbf{q}_{li}^{(2)}(u, t)$, and $\mathbf{h}(t, \mathbf{z}_0)$.

$$\begin{aligned}
\zeta_{1,li}(\boldsymbol{\beta}_0, t) &= \int_0^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} w_{li}^{Cox}(u) dM_{li}^1(u) + n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^\tau \mathbf{q}_{li}^{(2)}(u, t) dM_{li}^C(u) \\
&+ \mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) \sum_{l=1}^L (\boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li}), \\
\zeta_{2,li}^I(\boldsymbol{\beta}_0, t) &= \mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) \sum_{l=1}^L \boldsymbol{\eta}_{3,li}^I \\
&+ \{1 - \Delta_{li} I(\epsilon_{li} = 1)\} \left[\int_0^t w_{li}^{Cox}(u) Y_{li}^1(u) e^{\boldsymbol{\beta}_0^T \mathbf{z}_{li}} \frac{d\Lambda_{l10}(u)}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} - \int_0^t \mathbf{q}_{li}^{(2)}(u, t) dM_{li}^C(u) \right], \\
\zeta_{2,li}^{II}(\boldsymbol{\beta}_0, t) &= \mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) \sum_{l=1}^L \boldsymbol{\eta}_{3,li}^{II} \\
&+ \left\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\right\} \left[\int_0^t w_{li}^{Cox}(u) Y_{li}^1(u) e^{\boldsymbol{\beta}_0^T \mathbf{z}_{li}} \frac{d\Lambda_{l10}(u)}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} - \int_0^t \mathbf{q}_{li}^{(2)}(u, t) dM_{li}^C(u) \right], \\
\mathbf{q}_{li}^{(2)}(u, t) &= - \lim_{n_l \rightarrow \infty} \left[\frac{1}{n_l} \sum_{j=1}^{n_l} \int_{v=X_{lj}}^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, v)} \frac{e^{\boldsymbol{\gamma}_0^T \mathbf{z}_{lj}^C} I(u \leq v)}{s_{C,l}^{(0)}(\boldsymbol{\gamma}_0, u)} w_{lj}^{Cox}(v) dM_{lj}^1(v) \right. \\
&\left. - \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \int_{v=X_{lj}}^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, v)} \mathbf{h}_C^T(v, X_{lj}, \mathbf{z}_{lj}^C) \mathbf{A}_C^{-1} \{ \mathbf{z}_{li}^C - \mathbf{e}_{C,l}(\boldsymbol{\gamma}_0, u) \} w_{hj}^{Cox}(v) dM_{hj}^1(v) \right], \\
\mathbf{h}(t, \mathbf{z}_0) &= - \int_0^t \mathbf{e}_l(\boldsymbol{\beta}_0, u) d\Lambda_{l10}(u).
\end{aligned}$$

Supplementary Material

The Supplementary Materials include the proofs for the theorems, asymptotic covariance matrix of the estimators with covariate-unadjusted weight for a single and multiple case-cohort studies, the variance formula for $\widehat{F}_{l1}(t|\mathbf{z}_0)$ for a single case-cohort study, technical details on efficiency gain, additional simulation results, and data analysis results for the cause-specific hazards model.

Acknowledgments

The authors would like to thank the Associate Editor and three anonymous reviewers for their helpful comments that significantly improved the paper. This work was supported in part by Institutional Research Grant IRG #16-183-31 from the American Cancer Society and the Medical College of Wisconsin Cancer Center, and the US National Cancer Institute (U24CA076518). This manuscript was prepared using the research materials for the Atherosclerosis Risk in Communities study obtained from the National Heart, Lung, and Blood Institute (NHLBI) Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the Atherosclerosis Risk in Communities study or the NHLBI. The authors thank Ms. Elizabeth Cho for her proof reading.

References

- Austin, P. C. and Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in Medicine* **36**, 4391–4400.
- Ballantyne, C. M., Hoogeveen, R. C., Bang, H., Coresh, J., Folsom, A. R., Heiss, G., and Sharrett, A. R. (2004). Lipoprotein-Associated phospholipase A2, High-Sensitivity C-reactive protein, and risk for incident coronary heart disease in middle-aged men and women in the Atherosclerosis Risk in Communities (ARIC) study. *Circulation* **109**, 837–842.
- Barlow, W. (1994). Robust variance estimation for the case-cohort design. *Biometrics* **50**, 1064–72.
- Borgan, O., Langholz, B., Samuelsen, S. O., G. L., and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis* **6**, 39–58.
- Breslow, N. E. (1972). Contribution to discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society: Series B* **34**, 216–217.

- Breslow, N. E. and Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scandinavian Journal of Statistics* **34**, 86–102.
- Cai, J. and Zeng, D. (2007). Power calculation for case-cohort studies with nonrare events. *Biometrics* **63**, 1288–95.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B* **34**, 187–220.
- 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.
- 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.
- Ghosh, N., Karmali, R., Rocha, V., Ahn, K., DiGilio, A., Hari, P., Bachanova, V., Bacher, U., Dahi, P., Lima, M., DSouza, A., Fenske, T., Ganguly, S., Kharfan-Dabaja, M., and Hamadani, M. (2016). Reduced-intensity transplantation for lymphomas using haploidentical related donors versus HLA-matched sibling donors: A center for international blood and marrow transplant research analysis. *Journal of Clinical Oncology* **34**, 3141–3149.
- He, P., Eriksson, F., Scheike, T., and Zhang, M. (2016). A proportional hazards regression model for the subdistribution with covariates-adjusted censoring weight for competing risks data. *Scandinavian Journal of Statistics* **43**, 103–122.
- Kalbfleisch, J. D. and Lawless, J. F. (1988). Likelihood analysis of multistate models for disease incidence and mortality. *Statistics in Medicine* **7**, 149–60.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, 2 edition.
- Kang, S. and Cai, J. (2009). Marginal hazard model for case-cohort studies with multiple disease outcomes. *Biometrika* **96**, 887–901.
- Kim, S., Cai, J., and Lu, W. (2013). More efficient estimators for case-cohort studies. *Biometrika* **100**, 695–708.

- Kim, S., Zeng, D., and Cai, J. (2018). Analysis of multiple survival events in generalized case-cohort designs. *Biometrics* **74**(4), 1250–1260.
- Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort study. *Journal of the American Statistical Association* **99**, 832–44.
- Langholz, B. and Thomas, D. (1990). Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *American Journal of Epidemiology* **131**, 169–76.
- Liu, Y., Li, J., and Zhang, X. (2018). Analysis of dependently truncated data in cox framework. *Communications in Statistics Simulation Computation* **47**(6), 1677–1695.
- Mao, L. and Lin, D. (2017). Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 573–587.
- Pintilie, M., Bai, Y., Yund, L., and Hodgson, D. (2010). The analysis of case cohort design in the presence of competing risks with application to estimate the risk of delayed cardiac toxicity among Hodgkin Lymphoma survivors. *Statistics in Medicine* **29**, 2802–2810.
- Prentice, R. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.
- Saber, W., Hamadani, M., Hashmi, S. K., and Hari, P. (2015). Innovation and intellectual property rights. In Savani, B. N. and Mohty, M., editors, *Clinical guide to transplantation in lymphoma*, chapter 2, pages 13–22. John Wiley & Sons, West Sussex, UK.
- Samuelsen, S. O., Anestad, H., and Skrondal, A. (2007). Stratified case-cohort analysis of general cohort sampling designs. *Scandinavian Journal of Statistics* **34**, 103–19.
- Scheike, T. H., Zhang, M.-J., and Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika* **95**, 205–220.
- Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Annals of Statistics* **34**, 103–19.
- Sorensen, P. and Andersen, P. K. (2000). Competing risks analysis of the case-cohort design. *Biometrika* **87**, 49–59.

Wacholder, S., Gail, M., and Pee, D. (1991). Efficient design for assessing exposure-disease relationships in an assembled cohort. *Biometrics* **47**, 63–76.

Zhang, X., Zhang, M., and Fine, J. (2009). A mass redistribution algorithm for right-censored and left-truncated time to event data. *Journal of Statistical Planning and Inference* **139**, 3329–3339.

Zhou, B., Latouche, A., Rocha, V., and Fine, J. (2011). Competing risks regression for stratified data. *Biometrics* **67(2)**, 661–670.

Soyoung Kim, Division of Biostatistics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, U.S.A.

E-mail: skim@mcw.edu

Table 1: Simulation results for stratified data

CC	(P_1, P_2, P_c)	α	Covariate-unadjusted weight				Covariate-adjusted weight							
			bias	SD	SE	STB	CR	SRE	bias	SD	SE	STB	CR	SRE
Scenario 1: Censoring distribution \sim Uniform distribution														
$\hat{\beta}_I$	(20%, 20%, 60%)	0.1	0.007	0.182	0.178	0.038	0.94	1.00	0.006	0.185	0.178	0.034	0.94	1.00
		0.2	0.011	0.139	0.141	0.081	0.95	1.00	0.012	0.140	0.141	0.083	0.95	1.00
	(20%, 40%, 40%)	0.1	0.008	0.176	0.174	0.047	0.94	1.00	0.008	0.178	0.174	0.049	0.94	1.00
		0.2	0.007	0.137	0.140	0.047	0.96	1.00	0.007	0.138	0.140	0.047	0.95	1.00
$\hat{\beta}_{II}$	(20%, 20%, 60%)	0.1	0.008	0.159	0.155	0.051	0.95	1.32	0.008	0.160	0.155	0.052	0.95	1.34
		0.2	0.013	0.127	0.129	0.099	0.95	1.19	0.013	0.127	0.128	0.103	0.95	1.21
	(20%, 40%, 40%)	0.1	0.004	0.144	0.143	0.025	0.95	1.51	0.005	0.143	0.143	0.033	0.95	1.54
		0.2	0.006	0.122	0.124	0.052	0.95	1.27	0.007	0.121	0.124	0.057	0.95	1.29
Scenario 2: Censoring distribution \sim Cox model														
$\hat{\beta}_I$	(20%, 20%, 60%)	0.1	-0.070	0.223	0.247	0.283	0.95	1.00	0.021	0.218	0.211	0.101	0.94	1.00
		0.2	-0.079	0.174	0.196	0.403	0.95	1.00	0.012	0.173	0.170	0.070	0.94	1.00
	(20%, 40%, 40%)	0.1	-0.149	0.190	0.204	0.727	0.91	1.00	0.008	0.194	0.188	0.042	0.95	1.00
		0.2	-0.154	0.147	0.163	0.941	0.86	1.00	0.007	0.154	0.149	0.044	0.94	1.00
$\hat{\beta}_{II}$	(20%, 20%, 60%)	0.1	-0.075	0.180	0.196	0.383	0.94	1.53	0.022	0.194	0.196	0.114	0.94	1.27
		0.2	-0.083	0.152	0.166	0.498	0.92	1.31	0.012	0.162	0.162	0.077	0.95	1.14
	(20%, 40%, 40%)	0.1	-0.157	0.129	0.150	1.042	0.84	2.17	0.002	0.140	0.149	0.012	0.96	1.91
		0.2	-0.157	0.116	0.134	1.178	0.80	1.60	0.003	0.125	0.128	0.023	0.95	1.51

CC, case-cohort estimators; (P_1, P_2, P_c) , probability distribution of cause 1, cause 2, and censoring; SE, the average of the estimates of standard error; SD, sample standard deviation; STB, $E\{\|\hat{\beta} - \beta_0\|/SE\}$; CR, the empirical coverage rate of the nominal 95% confidence intervals; SRE, sample relative efficiency.

Table 2: Analysis of the CIBMTR study

Variables	Full cohort			CC_I			CC_{II}		
	$\hat{\beta}_F$	SE	p -value	$\hat{\beta}_I$	SE	p -value	$\hat{\beta}_{II}$	SE	p -value
Covariate-unadjusted weight									
Age	0.485	0.141	0.001	0.437	0.158	0.006	0.454	0.155	0.003
KPS (ref = $\geq 90\%$)									
$< 90\%$	0.515	0.222	0.021	0.527	0.256	0.039	0.540	0.251	0.032
Donor type (ref = HLA)									
Haploidentical	0.310	0.254	0.222	0.347	0.292	0.236	0.328	0.277	0.237
Disease subtype (ref = FLH)									
Diffuse large B-cell lymphoma	-0.383	0.326	0.240	-0.387	0.493	0.432	-0.378	0.479	0.430
Mantle cell lymphoma	-0.002	0.313	0.996	0.077	0.494	0.876	0.032	0.472	0.946
Mature T- and NK-cell lymphomas	-0.016	0.335	0.962	-0.065	0.535	0.904	0.004	0.522	0.993
Hodgkin	0.014	0.388	0.971	0.128	0.555	0.817	-0.029	0.542	0.957
Covariate-adjusted weight									
Age	0.482	0.140	0.001	0.436	0.156	0.005	0.453	0.155	0.003
KPS (ref = $\geq 90\%$)									
$< 90\%$	0.532	0.219	0.015	0.541	0.254	0.033	0.555	0.249	0.026
Donor type (ref = HLA)									
Haploidentical	0.310	0.254	0.222	0.346	0.292	0.236	0.328	0.277	0.236
Disease subtype (ref = FLH)									
Diffuse large B-cell lymphoma	-0.355	0.326	0.276	-0.361	0.493	0.465	-0.352	0.480	0.463
Mantle cell lymphoma	-0.018	0.318	0.955	0.088	0.497	0.860	0.037	0.476	0.937
Mature T- and NK-cell lymphomas	0.012	0.343	0.973	-0.025	0.543	0.963	0.042	0.531	0.937
Hodgkin	0.028	0.388	0.943	0.152	0.551	0.782	-0.012	0.542	0.983

CC_I , a single case-cohort study; CC_{II} , two case-cohort studies; SE, standard error estimate; KPS, Karnofsky performance status at transplant; ref, reference group; The reference groups are HLA-identical siblings donor (HLA) for donor type, ≥ 90 for KPS, and Follicular lymphoma for Histology (FLH). The coefficients for the reference groups were set to 0 and therefore they were omitted from the table.

Table 3: Analysis of the ARIC study

Variables	Category	Covariate-unadjusted weight			Covariate-adjusted weight		
		$\hat{\beta}_I$	SE	p -values	$\hat{\beta}_I$	SE	p -values
hs-CRP	Middle	0.2260	0.1378	0.101	0.2256	0.1378	0.102
(ref = Low)	High	0.5397	0.1357	< 0.001	0.5395	0.1357	< 0.001
Systolic blood pressure		0.2860	0.0616	< 0.001	0.2863	0.0616	< 0.001
LDL cholesterol		0.3200	0.0499	< 0.001	0.3196	0.0499	< 0.001
HDL cholesterol		-0.3159	0.0715	< 0.001	-0.3163	0.0715	< 0.001
Diabetes (ref = No)	Yes	0.5972	0.1258	< 0.001	0.5970	0.1258	< 0.001
Smoking status (ref = No)	Yes	0.3501	0.1080	0.001	0.3496	0.1080	0.001

SE, standard error estimate; hs-CRP, high-sensitivity C-reactive protein; LDL, low density lipoprotein; HDL, high density lipoprotein; Low high-sensitivity C-reactive protein group, group without diabetes, and non-smoking group are reference groups; ref, reference group. The coefficients for the reference groups were set to 0 and therefore they were omitted from the table.

Supplementary Materials

1 Appendix A: Proofs of Theorem 1

We provide the outline of the proofs for the main theorems. The following lemma plays an important role in proving theorems.

Lemma 1 *Let $B_i(t)$, $i = 1, \dots, n$ be independent and identically distributed real-valued random process on $[0, \tau]$ and denote random process vector, $\mathbf{B}(t) = [B_1(t), \dots, B_n(t)]$ with $E\{B_i(t)\} \equiv \mu_B(t)$, $\text{var } B_i(0) < \infty$, and $\text{var } B_i(\tau) < \infty$. Let $\chi = [\chi_1, \dots, \chi_n]$ be random vector containing \tilde{n} ones and $n - \tilde{n}$ zeros with each permutation equally likely. Let χ be independent of $\mathbf{B}(t)$. Suppose that almost all paths of $B_i(t)$ have finite variation. Then $n^{-1/2} \sum_{i=1}^n \chi_i \{B_i(t) - \mu_B(t)\}$ converges weakly in $l^\infty[0, \tau]$ to a zero-mean Gaussian process, and $n^{-1} \sum_{i=1}^n \chi_i \{B_i(t) - \mu_B(t)\}$ converges in probability to zero uniformly in t .*

Lemma 1 is an extension of the proposition from Kulich and Lin (2000) and the detailed proof can be found in Lemma 2 in Kang and Cai (2010).

1.1 Preliminaries

We study the asymptotics of $\hat{\beta}_{II}$. The asymptotics of $\hat{\beta}_I$ can be shown similarly and thus its proof is omitted. We use the proportional hazards model under case-cohort studies to estimate the censoring distribution:

$$\lambda_l^C(t | \mathbf{Z}_{li}^C) = \lambda_{l0}^C(t) \exp(\boldsymbol{\gamma}_0^T \mathbf{Z}_{li}^C).$$

Define the following notations for the censoring distribution depending on \mathbf{Z}^C :

$$\begin{aligned}
G_l(t|\mathbf{Z}^C) &= \exp\{-\Lambda_{0l}^C(t) \exp(\boldsymbol{\gamma}_0^T \mathbf{Z}^C)\}, \Lambda_{0l}^C(t) = \int_0^t \lambda_{l0}^C(u) du, \\
N_{li}^C(t) &= I(X_{li} \leq t; \Delta_{li} = 0), Y_{li}^C(t) = I(X_{li} \geq t), \\
M_{li}^C(t) &= N_{li}^C(t) - \int_0^t Y_{li}^C(u) \exp\{\boldsymbol{\gamma}_0^T \mathbf{Z}^C\} d\Lambda_{0l}^C(u), \\
\mathbf{S}_{C,l}^{(d)}(\boldsymbol{\gamma}, t) &= \frac{1}{n_l} \sum_{i=1}^{n_l} \pi_{li} Y_{li}^C(t) \mathbf{Z}_{li}^{C \otimes d} \exp\{\boldsymbol{\gamma}^T \mathbf{Z}_{li}^C\} \text{ for } d = 0, 1, 2, \\
\mathbf{s}_{C,l}^{(d)}(\boldsymbol{\gamma}, t) &= E[\mathbf{S}_{C,l}^{(d)}(\boldsymbol{\gamma}, t)] \text{ for } d = 0, 1, 2, \\
e_{C,l}(\boldsymbol{\gamma}, u) &= \frac{\mathbf{s}_{C,l}^{(1)}(\boldsymbol{\gamma}, t)}{\mathbf{s}_{C,l}^{(0)}(\boldsymbol{\gamma}, t)},
\end{aligned}$$

where $M_{li}^C(t)$ is a martingale with respect to the censoring filtration. By Kim et al. (2018), we can estimate $G_l(t|\mathbf{Z}^C)$ as follows:

$$\begin{aligned}
\tilde{G}_l(t|\mathbf{Z}^C) &= \exp\{-\tilde{\Lambda}_{0l}^C(t) \exp(\hat{\boldsymbol{\gamma}}^T \mathbf{Z}^C)\}, \\
\tilde{\Lambda}_{0l}^C(t) &= \int_0^t \frac{\sum_{j=1}^{n_l} \pi_{lj} dN_{lj}^C(t)}{n_l S_{C,l}^{(0)}(\hat{\boldsymbol{\gamma}}, t)},
\end{aligned}$$

where $\hat{\boldsymbol{\gamma}}$ is an estimator for $\boldsymbol{\gamma}_0$.

By Theorem 2 of Kim et al. (2018) and the arguments of Andersen and Gill (1982), $\tilde{G}_l(t|\mathbf{Z}_{li}^C) -$

$G_l(t|\mathbf{Z}_{li}^C)$ can be written as

$$\begin{aligned}
& \tilde{G}_l(t|\mathbf{Z}_{li}^C) - G_l(t|\mathbf{Z}_{li}^C) \\
= & -\frac{G_l(t|\mathbf{Z}_{li}^C)}{n_l} \int_{u=0}^{\tau} \sum_{j=1}^{n_l} \frac{e^{\gamma_0^T \mathbf{Z}_{li}^C} I(u \leq t)}{s_{C,l}^{(0)}(\gamma_0, u)} dM_{lj}^C(u) \\
& - \frac{G_l(t|\mathbf{Z}_{li}^C)}{n_l} \int_{u=0}^{\tau} \mathbf{h}_C^T(t, 0, \mathbf{Z}_{li}^C) \mathbf{A}_C^{-1} \sum_{h=1}^L \sum_{j=1}^{n_h} \left\{ \mathbf{Z}_{hj}^C - \frac{\mathbf{s}_{C,h}^{(1)}(\gamma_0, u)}{s_{C,h}^{(0)}(\gamma_0, u)} \right\} dM_{hj}^C(u) \\
& - \frac{G_l(t|\mathbf{Z}_{li}^C)}{n_l} \sum_{j=1}^{n_l} \left(1 - \frac{\xi_{lj}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_{k=1}^K \Delta_{lj} I(\epsilon_{li} = k)\right\} \int_{u=0}^{\tau} \frac{e^{\gamma_0^T \mathbf{Z}_{li}^C} I(u \leq t)}{s_{C,l}^{(0)}(\gamma_0, u)} dM_{lj}^C(u) \\
& - \frac{G_l(t|\mathbf{Z}_{li}^C)}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \left(1 - \frac{\xi_{hj}}{\tilde{\alpha}_h}\right) \left\{1 - \sum_{k=1}^K \Delta_{hj} I(\epsilon_{hi} = k)\right\} \\
& \times \int_{u=0}^{\tau} \mathbf{h}_C^T(t, 0, \mathbf{Z}_{li}^C) \mathbf{A}_C^{-1} \left\{ \mathbf{Z}_{hj}^C - \frac{\mathbf{s}_{C,h}^{(1)}(\gamma_0, u)}{s_{C,h}^{(0)}(\gamma_0, u)} \right\} dM_{hj}^C(u) + o_p(n_l^{-1/2}), \tag{1}
\end{aligned}$$

where

$$\mathbf{h}_C(t, u, \mathbf{Z}^C) = e^{\gamma^T \mathbf{Z}^C} \int_{v=u}^t \left\{ \mathbf{Z}^C - \frac{\mathbf{s}_{C,l}^{(1)}(\gamma_0, u)}{s_{C,l}^{(0)}(\gamma_0, u)} \right\} d\Lambda_{0l}^C(v).$$

Using (1), we can show

$$\begin{aligned}
& \frac{\tilde{G}_l(t|\mathbf{Z}_{li}^C)}{\tilde{G}_l(X_{li} \wedge t|\mathbf{Z}_{li}^C)} - \frac{G_l(t|\mathbf{Z}_{li}^C)}{G_l(X_{li} \wedge t|\mathbf{Z}_{li}^C)} \\
= & I(X_{li} < t) \frac{G_l(X_{li}|\mathbf{Z}_{li}^C)\{\tilde{G}_l(t|\mathbf{Z}_{li}^C) - G_l(t|\mathbf{Z}_{li}^C)\} - G_l(t|\mathbf{Z}_{li}^C)\{\tilde{G}_l(X_{li}|\mathbf{Z}_{li}^C) - G_l(X_{li}|\mathbf{Z}_{li}^C)\}}{\tilde{G}_l(X_{li}|\mathbf{Z}_{li}^C)G_l(X_{li}|\mathbf{Z}_{li}^C)} \\
= & -I(X_{li} < t) \frac{G_l(t|\mathbf{Z}_{li}^C)}{G_l(X_{li}|\mathbf{Z}_{li}^C)} \\
\times & \left(\frac{1}{n_l} \sum_{j=1}^{n_l} \int_{u=0}^{\tau} \frac{e^{\gamma_0^T \mathbf{Z}_{li}^C} I(X_{li} < u \leq t)}{s_{C,l}^{(0)}(\gamma_0, u)} dM_{lj}^C(u) \right. \\
+ & \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_l} \int_{u=0}^{\tau} \mathbf{h}_C^T(t, X_{li}, \mathbf{Z}_{li}^C) \mathbf{A}_C^{-1} \left\{ \mathbf{Z}_{hj}^C - \frac{\mathbf{s}_{C,h}^{(1)}(\gamma_0, u)}{s_{C,h}^{(0)}(\gamma_0, u)} \right\} dM_{hj}^C(u) \\
+ & \frac{1}{n_l} \sum_{j=1}^{n_l} \left(\frac{\xi_{lj}}{\tilde{\alpha}_l} - 1 \right) \left\{ 1 - \sum_{k=1}^K \Delta_{lj} I(\epsilon_{li} = k) \right\} \int_{u=0}^{\tau} \left[\frac{e^{\gamma_0^T \mathbf{Z}_{li}^C} I(X_{li} < u \leq t)}{s_{C,l}^{(0)}(\gamma_0, u)} \right] dM_{lj}^C(u) \\
+ & \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \left(\frac{\xi_{hj}}{\tilde{\alpha}_h} - 1 \right) \left\{ 1 - \sum_{k=1}^K \Delta_{hj} I(\epsilon_{hi} = k) \right\} \\
\times & \left. \int_{u=0}^{\tau} \mathbf{h}_C^T(t, X_{li}, \mathbf{Z}_{li}^C) \mathbf{A}_C^{-1} \left\{ \mathbf{Z}_{hj}^C - \frac{\mathbf{s}_{C,h}^{(1)}(\gamma_0, u)}{s_{C,h}^{(0)}(\gamma_0, u)} \right\} dM_{hj}^C(u) \right) + o_p(n_l^{-1/2}).
\end{aligned}$$

Therefore, $\tilde{w}_{li}^{Cox}(t) - w_{li}^{Cox}(t)$ is asymptotically equivalent to

$$\begin{aligned}
& I(C_{li} \geq T_{li} \wedge t) \left\{ \frac{\tilde{G}_l(t|\mathbf{Z}_{li}^C)}{\tilde{G}_l(X_{li} \wedge t|\mathbf{Z}_{li}^C)} - \frac{G_l(t|\mathbf{Z}_{li}^C)}{G_l(X_{li} \wedge t|\mathbf{Z}_{li}^C)} \right\} \\
&= -I(C_{li} \geq T_{li} \wedge t) I(X_{li} < t) \frac{G_l(t|\mathbf{Z}_{li}^C)}{G_l(X_{li}|\mathbf{Z}_{li}^C)} \\
&\times \left(\frac{1}{n_l} \sum_{j=1}^{n_l} \int_{u=0}^{\tau} \frac{e^{\gamma_0^T \mathbf{Z}_{li}^C} I(X_{li} < u \leq t)}{s_{C,l}^{(0)}(\gamma_0, u)} dM_{lj}^C(u) \right. \\
&+ \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \int_{u=0}^{\tau} \left[\mathbf{h}_C^T(t, X_{li}, \mathbf{Z}_{li}^C) \mathbf{A}_C^{-1} \left\{ \mathbf{Z}_{lh}^C - \frac{\mathbf{s}_{C,h}^{(1)}(\gamma_0, u)}{s_{C,h}^{(0)}(\gamma_0, u)} \right\} \right] dM_{hj}^C(u) \\
&+ \frac{1}{n_l} \sum_{j=1}^{n_l} \left(\frac{\xi_{lj}}{\tilde{\alpha}_l} - 1 \right) \left\{ 1 - \sum_{k=1}^K \Delta_{lj} I(\epsilon_{li} = k) \right\} \int_{u=0}^{\tau} \frac{e^{\gamma_0^T \mathbf{Z}_{li}^C} I(X_{li} < u \leq t)}{s_{C,l}^{(0)}(\gamma_0, u)} dM_{lj}^C(u) \\
&+ \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \left(\frac{\xi_{hj}}{\tilde{\alpha}_h} - 1 \right) \left\{ 1 - \sum_{k=1}^K \Delta_{hj} I(\epsilon_{hi} = k) \right\} \\
&\times \int_{u=0}^{\tau} \mathbf{h}_C^T(t, X_{li}, \mathbf{Z}_{li}^C) \mathbf{A}_C^{-1} \left\{ \mathbf{Z}_{hj}^C - \frac{\mathbf{s}_{C,h}^{(1)}(\gamma_0, u)}{s_{C,h}^{(0)}(\gamma_0, u)} \right\} dM_{hj}^C(u) \Big) + o_p(n_l^{-1/2}) \\
&\equiv D_{li}(t|\mathbf{Z}_{li}^C) + o_p(n_l^{-1/2}). \tag{2}
\end{aligned}$$

1.2 Consistency of $\widehat{\boldsymbol{\beta}}_{II}$

We show the consistency of $\widehat{\boldsymbol{\beta}}_{II}$ with right censored data. One can write $\mathbf{S}_l^{(d)}(\boldsymbol{\beta}, t) - \widetilde{\mathbf{S}}_l^{(d)}(\boldsymbol{\beta}, t)$:

$$\begin{aligned}
& \mathbf{S}_l^{(d)}(\boldsymbol{\beta}, t) - \widetilde{\mathbf{S}}_l^{(d)}(\boldsymbol{\beta}, t) \\
&= 1/n_l \sum_{i=1}^{n_l} \widetilde{w}_{li}^{Cox}(t) \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = 1)\right\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} \\
&= 1/n_l \sum_{i=1}^{n_l} \left\{\widetilde{w}_{li}^{Cox}(t) - w_{li}^{Cox}(t)\right\} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\right\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} \\
&+ 1/n_l \sum_{i=1}^{n_l} w_{li}^{Cox}(t) \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\right\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} \\
&= 1/n_l \sum_{i=1}^{n_l} D_{li}(t | \mathbf{Z}_{li}^C) \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\right\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} \tag{3}
\end{aligned}$$

$$\begin{aligned}
&+ 1/n_l \sum_{i=1}^{n_l} w_{li}^{Cox}(t) \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\right\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}. \tag{4} \\
&+ o_p(n_l^{-1/2})
\end{aligned}$$

Based on Conditions 1 and 2, the total variation of $w_{li}^{Cox}(t) \{1 - \sum_k \Delta_{li} I(\epsilon_{li} = 1)\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}$ is finite on $[0, \tau]$. We have $1/n_l \sum_{i=1}^{n_l} w_{li}^{Cox}(t) \{1 - \sum_k \Delta_{li} I(\epsilon_{li} = 1)\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}$ converges to $E[w_{li}^{Cox}(t) \{1 - \sum_k \Delta_{l1} I(\epsilon_{l1} = k)\} Y_{l1}^1(t) \mathbf{Z}_{l1}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{l1}}]$ as $n_l \rightarrow \infty$. Furthermore, by Conditions 1–7 and (2), $D_{li}(t | \mathbf{Z}_{li}^C)$ has finite variation in $t \in [0, \tau]$ as $n_l \rightarrow \infty$. Therefore, $D_{li}(t | \mathbf{Z}_{li}^C) \{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\} Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}$ has finite variations. Thus, by Lemma 1, (3) and (4) converges to zero in probability uniformly in t . Thus, $\|\mathbf{S}_l^{(k)}(\boldsymbol{\beta}, t) - \widetilde{\mathbf{S}}_l^{(k)}(\boldsymbol{\beta}, t)\| \rightarrow 0$ in probability uniformly in t and $n_l^{1/2} \{\mathbf{S}_l^{(d)}(\boldsymbol{\beta}, t) - \widetilde{\mathbf{S}}_l^{(d)}(\boldsymbol{\beta}, t)\}$ converges weakly to a zero-mean Gaussian process. Consequently, $\widetilde{\mathbf{S}}_l^{(d)}(\boldsymbol{\beta}, t)$ and $\mathbf{S}_l^{(d)}(\boldsymbol{\beta}, t)$ converge to the same limit in probability based on Condition 3.

Since $\widetilde{\mathbf{S}}_l^{(d)}(\boldsymbol{\beta}, t)$ converges to $\mathbf{s}_l^{(d)}(\boldsymbol{\beta}, t) = E[\mathbf{S}_l^{(d)}(\boldsymbol{\beta}, t)]$ uniformly in $t \in [0, \tau]$ and $\boldsymbol{\beta} \in \mathcal{B}$, we can show $-n_l^{-1} \partial \widetilde{\mathbf{U}} / \partial \boldsymbol{\beta}$ converges in probability uniformly in $\boldsymbol{\beta} \in \mathcal{B}$ to $\mathbf{A}_l(\boldsymbol{\beta}) = \lim_{n \rightarrow \infty} -n_l^{-1} \partial \widetilde{\mathbf{U}} / \partial \boldsymbol{\beta}$ similarly to Section A.4 of He et al. (2016). From Condition C5, $\mathbf{A}_l(\boldsymbol{\beta}_0)$ is positive definite for

$l = 1, \dots, L$. As in Section 1.3, we can show $\tilde{\mathbf{U}}(\boldsymbol{\beta})$ converges in probability to zero. Therefore, by Theorem 2 of Fourtz (1977), $\hat{\boldsymbol{\beta}}_{II}$ converges in probability to $\boldsymbol{\beta}_0$.

1.3 Asymptotic normality of $\hat{\boldsymbol{\beta}}_{II}$

Next, we show the asymptotic normality of $\hat{\boldsymbol{\beta}}_{II}$. One can write $n^{-1/2}\tilde{\mathbf{U}}(\boldsymbol{\beta})$ as

$$\begin{aligned} n^{-1/2}\tilde{\mathbf{U}}(\boldsymbol{\beta}) &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, u)\} \tilde{w}_{li}^{Cox}(u) dN_{li}^1(u) \\ &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, u)\} \tilde{w}_{li}^{Cox}(u) dM_{li}^1(t) \end{aligned} \quad (5)$$

$$+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, u)\} \tilde{w}_{li}^{Cox}(u) Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t). \quad (6)$$

We can decompose (5) into two parts such that

$$\begin{aligned} &n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t)\} \tilde{w}_{li}^{Cox}(t) dM_{li}^1(t) \\ &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t)\} w_{li}^{Cox}(t) dM_{li}^1(t) \\ &+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\tilde{w}_{li}^{Cox}(t) - w_{li}^{Cox}(t)\} \{\mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, u)\} dM_{li}^1(t) \\ &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{Z}_{li} - \mathbf{E}_l(\boldsymbol{\beta}, t)\} w_{li}^{Cox}(t) dM_{li}^1(t) \end{aligned} \quad (7)$$

$$+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t)\} w_{li}^{Cox}(t) dM_{li}^1(t) \quad (8)$$

$$+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\tilde{w}_{li}^{Cox}(t) - w_{li}^{Cox}(t)\} \{\mathbf{Z}_{li} - \mathbf{E}_l(\boldsymbol{\beta}, t)\} dM_{li}^1(t) \quad (9)$$

$$+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{\tilde{w}_{li}^{Cox}(t) - w_{li}^{Cox}(t)\} \{\mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t)\} dM_{li}^1(t). \quad (10)$$

First, we show (8) and (10) converge to zero in probability as n goes to ∞ . We know $\|\mathbf{S}_l^{(k)}(\boldsymbol{\beta}, t) - \tilde{\mathbf{S}}_l^{(k)}(\boldsymbol{\beta}, t)\| \rightarrow 0$ in probability uniformly in t . Thus, due to Conditions 3 and 4, $\tilde{\mathbf{S}}^{(1)}(\boldsymbol{\beta}, t)/\tilde{\mathbf{S}}^{(0)}(\boldsymbol{\beta}, t)$ converges to $\mathbf{e}_l(\boldsymbol{\beta}, t)$ in probability uniformly in t . Using (3) and (4), $\mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t)$ can be written as

$$\begin{aligned}
& \mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \tag{11} \\
&= \frac{\mathbf{S}_l^{(1)}(\boldsymbol{\beta}, t)}{S_l^{(0)}(\boldsymbol{\beta}, t)} - \frac{\tilde{\mathbf{S}}_l^{(1)}(\boldsymbol{\beta}, t)}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}, t)} \\
&= \left[\mathbf{S}_l^{(1)}(\boldsymbol{\beta}, t) - S_l^{(0)}(\boldsymbol{\beta}, t) \frac{\tilde{\mathbf{S}}_l^{(1)}(\boldsymbol{\beta}, t)}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}, t)} \right] \frac{1}{S_l^{(0)}(\boldsymbol{\beta}, t)} \\
&= \left[\{\mathbf{S}_l^{(1)}(\boldsymbol{\beta}, t) - \tilde{\mathbf{S}}_l^{(1)}(\boldsymbol{\beta}, t)\} - \{S_l^{(0)}(\boldsymbol{\beta}, t) - \tilde{S}_l^{(0)}(\boldsymbol{\beta}, t)\} \frac{\tilde{\mathbf{S}}_l^{(1)}(\boldsymbol{\beta}, t)}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}, t)} \right] \frac{1}{S_l^{(0)}(\boldsymbol{\beta}, t)} \\
&= \left[\{\mathbf{S}_l^{(1)}(\boldsymbol{\beta}, t) - \tilde{\mathbf{S}}_l^{(1)}(\boldsymbol{\beta}, t)\} - \{S_l^{(0)}(\boldsymbol{\beta}, t) - \tilde{S}_l^{(0)}(\boldsymbol{\beta}, t)\} \mathbf{e}_l(\boldsymbol{\beta}, t) \right] \frac{1}{S_l^{(0)}(\boldsymbol{\beta}, t)} + o_p(1) \\
&= \sum_i \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_k \Delta_i I(\epsilon_{li} = k)\right\} \mathbf{Q}_{li}(\boldsymbol{\beta}, t) \frac{1}{S_l^{(0)}(\boldsymbol{\beta}, t)} + o_p(1), \tag{12}
\end{aligned}$$

where $\mathbf{Q}_{li}(\boldsymbol{\beta}, t) = w_{li}^{Cox}(t) Y_{li}^1(t) \{\mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}, t)\} e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}}$. Then, it can be shown that (8) converges to 0 in probability uniformly in t using Lemma 1 and (12).

In addition to Lemma 1, we have $\mathbf{E}_l(\boldsymbol{\beta}, t)$ and $\tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t)$ converges to the same limit $\mathbf{e}_l(\boldsymbol{\beta}, t)$, where $\mathbf{e}_l(\boldsymbol{\beta}, t) = \mathbf{s}_l^{(1)}(\boldsymbol{\beta}, t)/s_l^{(0)}(\boldsymbol{\beta}, t)$ and $\mathbf{s}_l^{(k)}(\boldsymbol{\beta}, t) = E[\mathbf{S}_l^{(k)}(\boldsymbol{\beta}, t)]$ by Conditions 2–4. Thus, similarly to (8) $\rightarrow 0$, we can show (10) converges to zero in probability uniformly in t .

The second part of $\tilde{U}(\boldsymbol{\beta})$, (6) can be written as

$$\begin{aligned}
& n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \} \tilde{w}_{li}^{Cox}(t) Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t) \\
&= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{Z}_{li} - \mathbf{E}_l(\boldsymbol{\beta}, t) + \mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \} \tilde{w}_{li}^{Cox}(t) Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t) \\
&= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{Z}_{li} - \mathbf{E}_l(\boldsymbol{\beta}, t) \} \tilde{w}_{li}^{Cox}(t) Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t) \tag{13}
\end{aligned}$$

$$+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \} \tilde{w}_{li}^{Cox}(t) Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t). \tag{14}$$

(13) is $\mathbf{0}$. Using (12) and the similar arguments to (3) $\rightarrow 0$ in probability by Lemma 1, we can write (14) as

$$\begin{aligned}
& n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \} \tilde{w}_{li}^{Cox}(t) Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t) \\
&= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \} \{ \tilde{w}_{li}^{Cox}(t) - w_{li}^{Cox}(t) \} Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t) \\
&+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \mathbf{E}_l(\boldsymbol{\beta}, t) - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \} w_{li}^{Cox}(t) Y_{li}^1(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{li}} d\Lambda_{1l0}(t) \\
&= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \left[\left(1 - \frac{\xi_{li}}{\tilde{\alpha}} \right) \left\{ 1 - \sum_k \Delta_{li} I(\epsilon_{li} = 1) \right\} \mathbf{Q}_{li}(\boldsymbol{\beta}, t) \right] d\Lambda_{1l0}(t) + o_p(1). \tag{15}
\end{aligned}$$

From (2), (9) is asymptotically equivalent to

$$\begin{aligned}
& n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \{ \tilde{w}_{li}^{Cox}(t) - w_{li}^{Cox}(t) \} \{ \mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}, t) \} dM_{li}^1(t) \\
&= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t) \\
&+ n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^\tau \left(\frac{\xi_{li}}{\tilde{\alpha}_l} - 1 \right) \left\{ 1 - \sum_{k=1}^K \Delta_{li} I(\epsilon_{li} = k) \right\} \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t) + o_p(1),
\end{aligned}$$

where

$$\begin{aligned} \mathbf{q}_{li}^{(1)}(u) &= - \lim_{n_l \rightarrow \infty} \left[\frac{1}{n_l} \sum_{j=1}^{n_l} \int_{t=X_{lj}}^{\tau} \{ \mathbf{Z}_{lj} - \mathbf{e}_l(\beta_0, t) \} w_{lj}^{Cox}(t) \frac{e^{\gamma_0^T \mathbf{Z}_{lj}^C} I(u \leq t)}{s_{C,l}^{(0)}(\gamma_0, u)} dM_{lj}^1(t) \right. \\ &\quad \left. + \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \int_{t=X_{lj}}^{\tau} \{ \mathbf{Z}_{lj} - \mathbf{e}_l(\beta_0, t) \} w_{lj}^{Cox}(t) \mathbf{h}_C^T(t, X_{lj}, \mathbf{Z}_{lj}^C) \mathbf{A}_C^{-1} \{ \mathbf{Z}_{li}^C - \mathbf{e}_{C,l}(\gamma_0, u) \} dM_{hj}^1(t) \right]. \end{aligned}$$

Combining all results for (7),(9), and (15), we have

$$\begin{aligned} n^{-1/2} \tilde{\mathbf{U}}(\boldsymbol{\beta}) &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^{\tau} \{ \mathbf{Z}_{li} - \tilde{\mathbf{E}}_l(\boldsymbol{\beta}, t) \} \tilde{w}_{li}^{Cox}(t) dN_i^1(t) \\ &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^{\tau} \{ \mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}, t) \} w_{li}^{Cox}(t) dM_{li}^1(t) \\ &\quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^{\tau} \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t) \\ &\quad + n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \int_0^{\tau} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) \left\{ 1 - \sum_{k=1}^K \Delta_{li} I(\epsilon_{li} = k) \right\} \\ &\quad \times \left\{ \mathbf{Q}_{li}(\boldsymbol{\beta}, t) d\Lambda_{li0}(t) - \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t) \right\} + o_p(1) \\ &= n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left\{ \boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li} + \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) \boldsymbol{\eta}_{3,li}^{II} \right\} + o_p(1) \\ &= \sum_{l=1}^L \left(\frac{n}{n_l} \right)^{-1/2} \sum_{i=1}^{n_l} \left\{ \boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li} + \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) \boldsymbol{\eta}_{3,li}^{II} \right\} + o_p(1), \end{aligned}$$

where $\boldsymbol{\eta}_{1,li}$, $\boldsymbol{\eta}_{2,li}$, and $\boldsymbol{\eta}_{3,li}^{II}$ are i.i.d zero-mean variables in stratum l . By central limit theorem, $n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} (\boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li})$ converges to weakly to a zero-mean normal vector with covariance $\sum_{l=1}^L q_l E(\boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li})^{\otimes 2}$. It follows from Hájek (1960)'s central limit theorem for finite sampling, Conditions 1, 6, 7 that $n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \boldsymbol{\eta}_{3,li}^{II}$ converges to weakly a zero-mean normal vector with covariance $(1 - \alpha_l^*)/\alpha_l^* E[\boldsymbol{\eta}_{3,li}^{II}]^{\otimes 2}$. Using (i) $E[n_l^{-1/2} \sum_{i=1}^{n_l} \boldsymbol{\eta}_{1,li}] = \mathbf{0}$ and $E[n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \boldsymbol{\eta}_{3,li}^{II}] = \mathbf{0}$ in stratum l ; and (ii) $(1 - \xi_{li}/\tilde{\alpha}_l)$'s are independent of history of $N_{li}^C(t)$, $Y_{li}^C(t)$, $N_{li}^1(t)$, $Y_{li}^1(t)$, and $\mathbf{Z}_{il}(t)$ for all i, l , and $t \in [0, \tau]$, we can show

$Cov(n_l^{-1/2} \sum_{i=1}^{n_l} \boldsymbol{\eta}_{1,li}, n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \boldsymbol{\eta}_{3,li}^{II}) = \mathbf{0}$ and $Cov(n_l^{-1/2} \sum_{i=1}^{n_l} \boldsymbol{\eta}_{2,li}, n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \boldsymbol{\eta}_{3,li}^{II}) = \mathbf{0}$. Since $n_l^{-1/2} \sum_{i=1}^{n_l} \boldsymbol{\eta}_{1,li}$, $n_l^{-1/2} \sum_{i=1}^{n_l} \boldsymbol{\eta}_{2,li}$, and $n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \boldsymbol{\eta}_{3,li}^{II}$ are asymptotically normal, $n_l^{-1/2} \sum_{i=1}^{n_l} \boldsymbol{\eta}_{1,li}$ and $n_l^{-1/2} \sum_{i=1}^{n_l} \boldsymbol{\eta}_{2,li}$ are independent of $n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \boldsymbol{\eta}_{3,li}^{II}$.

Combining all results, $n^{-1/2} \tilde{\mathbf{U}}(\boldsymbol{\beta})$ converges weakly to zero-mean normal vector with covariance matrix $\boldsymbol{\Sigma}^{II}(\boldsymbol{\beta}_0)$, where

$$\begin{aligned} \boldsymbol{\Sigma}^{II}(\boldsymbol{\beta}_0) &= \sum_{l=1}^L q_l \left\{ E(\boldsymbol{\eta}_{1,l1} + \boldsymbol{\eta}_{2,l1})^{\otimes 2} + \frac{1 - \alpha_l^*}{\alpha_l^*} E(\boldsymbol{\eta}_{3,l1}^{II})^{\otimes 2} \right\}, \\ \boldsymbol{\eta}_{1,li} &= \int_0^\tau \{ \mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}_0, t) \} w_{li}^{Cox}(t) dM_{li}^1(t), \\ \boldsymbol{\eta}_{2,li} &= \int_0^\tau \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t), \\ \boldsymbol{\eta}_{3,li}^{II} &= \int_0^\tau \left\{ 1 - \sum_{k=1}^K \Delta_{li} I(\epsilon_{li} = k) \right\} w_{li}^{Cox}(t) Y_{li}^1(t) \{ \mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}_0, t) \} d\Lambda_{li0}(t) \\ &\quad - \int_0^\tau \left\{ 1 - \sum_{k=1}^K \Delta_{li} I(\epsilon_{li} = k) \right\} \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t), \\ \mathbf{q}_{li}^{(1)}(u) &= - \lim_{n_l \rightarrow \infty} \left[\frac{1}{n_l} \sum_{j=1}^{n_l} \int_{t=X_{lj}}^\tau \{ \mathbf{Z}_{lj} - \mathbf{e}_l(\boldsymbol{\beta}_0, t) \} w_{lj}^{Cox}(t) \frac{e^{\boldsymbol{\gamma}_0^T \mathbf{Z}_{lj}^C} I(u \leq t)}{s_{C,l}^{(0)}(\boldsymbol{\gamma}_0, u)} dM_{lj}^1(t) \right. \\ &\quad \left. + \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \int_{t=X_{lj}}^\tau \{ \mathbf{Z}_{lj} - \mathbf{e}_l(\boldsymbol{\beta}_0, t) \} w_{lj}^{Cox}(t) \mathbf{h}_C^T(t, X_{lj}, \mathbf{Z}_{lj}^C) \mathbf{A}_C^{-1} \{ \mathbf{Z}_{li}^C - \mathbf{e}_{C,l}(\boldsymbol{\gamma}_0, u) \} dM_{hj}^1(t) \right]. \end{aligned}$$

By Condition 5, the consistency of $\hat{\boldsymbol{\beta}}_{II}$, and Taylor expansion of $\tilde{\mathbf{U}}(\hat{\boldsymbol{\beta}}_{II})$ around $\boldsymbol{\beta}_0$, $n^{-1/2}(\hat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0)$ is asymptotically normally distributed with mean zero and with covariance matrix $\mathbf{A}(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\Sigma}^{II}(\boldsymbol{\beta}_0) \mathbf{A}(\boldsymbol{\beta}_0)^{-1}$, where $\mathbf{A}(\boldsymbol{\beta}_0) = \sum_{l=1}^L q_l \mathbf{A}_l(\boldsymbol{\beta}_0)$.

Similarly, we can establish the asymptotic of $\tilde{\mathbf{U}}(\boldsymbol{\beta}_I)$ and $\hat{\boldsymbol{\beta}}_I$.

2 Appendix B: Proof of Theorem 2

In this section, we study the asymptotic distribution of $n_l^{1/2}\{\widehat{\Lambda}_{1l0}^{II}(t) - \Lambda_{1l0}(t)\}$. The asymptotics of $\widehat{\Lambda}_{1l0}^I(t)$ can be similarly shown and thus its proof is omitted.

We can decompose $n_l^{1/2}\{\widehat{\Lambda}_{1l0}^{II}(t) - \Lambda_{1l0}(t)\}$ into four parts as follows:

$$\begin{aligned} & n_l^{1/2}\{\widehat{\Lambda}_{1l0}^{II}(t) - \Lambda_{1l0}(t)\} \\ = & n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \frac{\widetilde{w}_{li}^{Cox}(u) dN_{li}^1(u)}{\widetilde{S}_l^{(0)}(\widehat{\beta}_{II}, u)} - \Lambda_{1l0}(t) \\ = & n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \left\{ \frac{1}{\widetilde{S}_l^{(0)}(\widehat{\beta}_{II}, u)} - \frac{1}{\widetilde{S}_l^{(0)}(\beta_0, u)} \right\} \widetilde{w}_{li}^{Cox}(u) dM_{li}^1(u) \end{aligned} \quad (16)$$

$$+ n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \frac{1}{\widetilde{S}_l^{(0)}(\beta_0, u)} \widetilde{w}_{li}^{Cox}(u) dM_{li}^1(u) \quad (17)$$

$$+ n_l^{-1/2} \int_0^t \left\{ \frac{1}{\widetilde{S}_l^{(0)}(\widehat{\beta}_{II}, u)} - \frac{1}{\widetilde{S}_l^{(0)}(\beta_0, u)} \right\} S_l^{(0)}(\beta_0, u) d\Lambda_{1l0}(u) \quad (18)$$

$$+ n_l^{-1/2} \int_0^t \frac{S_l^{(0)}(\beta_0, u) - \widetilde{S}_l^{(0)}(\beta_0, u)}{\widetilde{S}_l^{(0)}(\beta_0, u)} d\Lambda_{1l0}(u). \quad (19)$$

By Taylor expansion, we have

$$\frac{1}{\widetilde{S}_l^{(0)}(\widehat{\beta}_{II}, u)} - \frac{1}{\widetilde{S}_l^{(0)}(\beta_0, u)} = -\frac{\widetilde{S}_l^{(1)}(\beta^*, u)}{\widetilde{S}_l^{(0)}(\beta^*, u)^2} (\widehat{\beta}_{II} - \beta_0), \quad (20)$$

where β^* is on the line segment between $\widehat{\beta}_{II}$ and β_0 . Plugging (20) into (16), we have

$$n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \left\{ -\frac{\widetilde{S}_l^{(1)}(\beta^*, u)}{\widetilde{S}_l^{(0)}(\beta^*, u)^2} (\widehat{\beta}_{II} - \beta_0) \right\} \widetilde{w}_{li}^{Cox}(u) dM_{li}^1(u).$$

We know β^* converges to β_0 in probability and $\widehat{\beta}_{II}$ is a consistent estimator for β_0 . Since $\widetilde{S}_l^{(0)}(\beta^*, u)$ and $\widetilde{S}_l^{(1)}(\beta^*, u)$ are of bounded variation and $\widetilde{S}_l^{(0)}(\beta^*, u)$ is bounded away from 0, and $\widetilde{S}_l^{(1)}(\beta^*, u)/\widetilde{S}_l^{(0)}(\beta^*, u)^2$ is of bounded variation and can be written as sum of two monotone

functions in t . Thus, (16) converges to zero in probability uniformly in t . (17) can be written as:

$$\begin{aligned}
& n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \frac{1}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)} \{ \tilde{w}_{li}^{Cox}(u) - w_{li}^{Cox}(u) + w_{li}^{Cox}(u) \} dM_{li}^1(u) \\
&= n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \frac{1}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)} w_{li}^{Cox}(u) dM_{li}^1(u) \\
&+ n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \frac{1}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)} \{ \tilde{w}_{li}^{Cox}(u) - w_{li}^{Cox}(u) \} dM_{li}^1(u).
\end{aligned}$$

Since $\tilde{S}_l^{(0)}(\boldsymbol{\beta}, u)$ converges to $s_l^{(0)}(\boldsymbol{\beta}, u)$ by (2), Conditions 3 and 4, (17) is asymptotically equivalent to

$$\begin{aligned}
& n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} w_{li}^{Cox}(u) dM_{li}^1(u) \\
&+ n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^\tau \left[\mathbf{q}_{li}^{(2)}(u, t) + \left(\frac{\xi_{li}}{\tilde{\alpha}} - 1 \right) \left\{ 1 - \sum_{k=1}^K \Delta_{li}(\epsilon_{li} = k) \right\} \mathbf{q}_{li}^{(2)}(u, t) \right] dM_{li}^C(u),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{q}_{li}^{(2)}(u, t) &= - \lim_{n_l \rightarrow \infty} \left[\frac{1}{n_l} \sum_{j=1}^{n_l} \int_{v=X_{lj}}^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, v)} \frac{e^{\boldsymbol{\gamma}_0^T \mathbf{Z}_{lj}^C} I(u \leq v)}{s_{C,l}^{(0)}(\boldsymbol{\gamma}_0, u)} w_{lj}^{Cox}(v) dM_{lj}^1(v) \right. \\
&- \left. \frac{1}{n_l} \sum_{h=1}^L \sum_{j=1}^{n_h} \int_{v=X_{lj}}^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, v)} \mathbf{h}_C^T(v, X_{lj}, \mathbf{Z}_{lj}^C) \mathbf{A}_C^{-1} \{ \mathbf{Z}_{li}^C - \mathbf{e}_{C,l}(\boldsymbol{\gamma}_0, u) \} w_{hj}^{Cox}(v) dM_{hj}^1(v) \right].
\end{aligned}$$

Since $\boldsymbol{\beta}^*$ converges to $\boldsymbol{\beta}_0$ in probability uniformly and Conditions 3 and 4, $\tilde{S}_l^{(0)}(\boldsymbol{\beta}^*, t)$ and $S_l^{(0)}(\boldsymbol{\beta}_0, t)$ converges to $s_l^{(0)}(\boldsymbol{\beta}_0, t)$ in probability uniformly in t . By Condition 1, $d\Lambda_{l10}(u)$ is bounded. Using these results of consistency for $\hat{\boldsymbol{\beta}}_{II}$ and uniform convergence of $\tilde{S}_l^{(0)}(\boldsymbol{\beta}, t)$ and plugging $n^{-1/2}(\hat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0) = \mathbf{A}^{-1}(\boldsymbol{\beta}_0) n^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \{ \boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li} + (1 - \xi_{li}/\tilde{\alpha}_l) \boldsymbol{\eta}_{3,li} \}$ into (18),

we can show (18) becomes

$$\begin{aligned}
& n_l^{-1/2} \int_0^t \left[-\frac{\tilde{S}_l^{(1)}(\boldsymbol{\beta}_0, u)}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)^2} (\hat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0) \right] S_l^{(0)}(\boldsymbol{\beta}_0, u) d\Lambda_{l10}(u) \\
&= n_l^{-1/2} \mathbf{h}(t, \mathbf{z}_0)^T (\hat{\boldsymbol{\beta}}_{II} - \boldsymbol{\beta}_0) + o_p(1) \\
&= \mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) n_l^{-1/2} \sum_{l=1}^L \sum_{i=1}^{n_l} \left\{ \boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li} + \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \boldsymbol{\eta}_{3,li} \right\} + o_p(1),
\end{aligned}$$

where $\mathbf{h}(t, \mathbf{z}_0) = -\int_0^t \mathbf{e}_l(\boldsymbol{\beta}_0, u) d\Lambda_{l10}(u)$.

Since $\tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)$ converges to $s_l^{(0)}(\boldsymbol{\beta}_0, u)$ in probability uniformly in u and $s_l^{(0)}(\boldsymbol{\beta}_0, u)$ is bounded away from 0, then we have $\tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)^{-1}$ converges to $s_l^{(0)}(\boldsymbol{\beta}_0, u)^{-1}$. Using the similar arguments to (3) $\rightarrow 0$ in probability by Lemma 1, we can show (19) becomes

$$\begin{aligned}
& n_l^{-1/2} \int_0^t \frac{S_l^{(0)}(\boldsymbol{\beta}_0, u) - \tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)}{\tilde{S}_l^{(0)}(\boldsymbol{\beta}_0, u)} d\Lambda_{l10}(u) \\
&= n_l^{-1/2} \int_0^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} \left[\sum_{i=1}^{n_l} w_{li}^{Cox}(t) \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l}\right) \left\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\right\} Y_{li}^1(t) e^{\boldsymbol{\beta}_0^T \mathbf{z}_{li}} \right] d\Lambda_{l10}(u) + o_p(1).
\end{aligned}$$

Combining all the results, we have

$$\begin{aligned}
& n_l^{1/2} \{ \widehat{\Lambda}_{1l0}^{II}(t) - \Lambda_{1l0}(t) \} \\
= & n_l^{-1/2} \sum_{i=1}^{n_l} \left[\int_0^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} w_{li}^{Cox}(u) dM_{li}^1(u) + n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^\tau \mathbf{q}_{li}^{(2)}(u, t) dM_{li}^C(u) \right. \\
& + \left. \mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) \sum_{l=1}^L (\boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li}) \right] \\
& + n_l^{-1/2} \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) \left[\mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) \sum_{l=1}^L \boldsymbol{\eta}_{3,li}^{II} \right. \\
& + \int_0^t w_{li}^{Cox}(u) \{ 1 - \sum_k \Delta_{li} I(\epsilon_{li} = k) \} Y_{li}^1(u) e^{\boldsymbol{\beta}_0^T \mathbf{z}_{li}} \frac{d\Lambda_{1l0}(u)}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} \\
& - \left. \{ 1 - \sum_k \Delta_{li} I(\epsilon_{li} = k) \} \int_0^t \mathbf{q}_{li}^{(2)}(u, t) dM_{li}^C(u) \right] + o_p(1) \\
= & n_l^{-1/2} \sum_{i=1}^{n_l} \zeta_{1,li}(\boldsymbol{\beta}_0, t) + n_l^{-1/2} \sum_{i=1}^{n_l} \left(1 - \frac{\xi_{li}}{\tilde{\alpha}_l} \right) \zeta_{2,li}^{II}(\boldsymbol{\beta}_0, t) + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\zeta_{1,li}(\boldsymbol{\beta}_0, t) &= \int_0^t \frac{1}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} w_{li}^{Cox}(u) dM_{li}^1(u) + n_l^{-1/2} \sum_{i=1}^{n_l} \int_0^\tau \mathbf{q}_{li}^{(2)}(u, t) dM_{li}^C(u) \\
&+ \mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) \sum_{l=1}^L (\boldsymbol{\eta}_{1,li} + \boldsymbol{\eta}_{2,li}), \\
\zeta_{2,li}^{II}(\boldsymbol{\beta}_0, t) &= \mathbf{h}(t, \mathbf{z}_0)^T \mathbf{A}^{-1}(\boldsymbol{\beta}_0) \sum_{l=1}^L \boldsymbol{\eta}_{3,li}^{II} \\
&+ \left\{ 1 - \sum_k \Delta_{li} I(\epsilon_{li} = k) \right\} \left[\int_0^t w_{li}^{Cox}(u) Y_{li}^1(u) e^{\boldsymbol{\beta}_0^T \mathbf{z}_{li}} \frac{d\Lambda_{1l0}(u)}{s_l^{(0)}(\boldsymbol{\beta}_0, u)} - \int_0^t \mathbf{q}_{li}^{(2)}(u, t) dM_{li}^C(u) \right].
\end{aligned}$$

Following He et al. (2016), $H^{(1)}(t) = n_l^{-1/2} \sum_{i=1}^{n_l} \zeta_{1,li}(\boldsymbol{\beta}_0, t)$ converges weakly to the Gaussian process $\mathcal{H}^{(1)}(t)$ on $[0, \tau]$ whose mean is zero and covariance functions between $\mathcal{H}^{(1)}(t)$ and $\mathcal{H}^{(1)}(s)$ is $E\{\zeta_{1,li}(\boldsymbol{\beta}_0, t)\zeta_{1,li}(\boldsymbol{\beta}_0, s)\}$ for $t, s \in [0, \tau]$.

Next we show the weak convergence of $H^{(2)} = n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \zeta_{2,li}(\boldsymbol{\beta}_0, t)$ to a zero-mean Gaussian process. We have $s_l^{(0)}(\boldsymbol{\beta}_0, u)$ is bounded away from 0 and $\mathbf{h}(t, \mathbf{z}_0)$, $\{1 - \sum_k \Delta_{li} I(\epsilon_{li} = k)\}$, $w_{li}^{Cox}(t) Y_{li}^1(t) e^{\boldsymbol{\beta}_0^T \mathbf{Z}_{li}}$, $d\Lambda_{l10}(u)$ are bounded variations; $\mathbf{A}(\boldsymbol{\beta})$ is positive definite based on Conditions 1 to 5. Hence, it follows from Lemma 1 that the finite dimensional distribution of $H^{(2)}(t)$ is asymptotically same as that of $\mathcal{H}^{(2)}(t)$ for any finite number of time points. Combining these results, $H^{(2)}(t)$ converges weakly to the Gaussian process $\mathcal{H}^{(2)}(t)$ whose mean is zero and covariance functions between $\mathcal{H}^{(2)}(t)$ and $\mathcal{H}^{(2)}(s)$ is $(1 - \alpha^*)/\alpha^* E\{\zeta_{2,li}^{II}(\boldsymbol{\beta}_0, t) \zeta_{2,li}^{II}(\boldsymbol{\beta}_0, s)\}$ for $t, s \in [0, \tau]$.

We can show $Cov\{n_l^{-1/2} \sum_{i=1}^{n_l} \zeta_{1,li}(\boldsymbol{\beta}_0, t), n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \zeta_{2,li}^{II}(\boldsymbol{\beta}_0, t)\} = 0$ similarly to the proof of independence of $\boldsymbol{\eta}_{1,li}$ and $(1 - \xi_{li}/\alpha_{li}) \boldsymbol{\eta}_{3,li}$. Therefore, $n_l^{-1/2} \sum_{i=1}^{n_l} \zeta_{1,li}(\boldsymbol{\beta}_0, t)$ and $n_l^{-1/2} \sum_{i=1}^{n_l} (1 - \xi_{li}/\tilde{\alpha}_l) \zeta_{2,li}(\boldsymbol{\beta}_0, t)$ are independent. Thus, $H(t; \mathbf{z}_0) = H^{(1)}(t; \mathbf{z}_0) + H^{(2)}(t; \mathbf{z}_0)$ converges weakly to Gaussian process $\mathcal{H}(t)$ between $\mathcal{H}(t)$ and $\mathcal{H}(s)$ is $E\{\zeta_{1,li}(\boldsymbol{\beta}_0, t) \zeta_{1,li}(\boldsymbol{\beta}_0, s)\} + (1 - \alpha^*)/\alpha^* E\{\zeta_{2,li}(\boldsymbol{\beta}_0, t) \zeta_{2,li}(\boldsymbol{\beta}_0, s)\}$ for $t, s \in [0, \tau]$. This completes the proof of Theorem 2.

3 Appendix C: Asymptotic covariance matrix of the estimators

for case-cohort studies with $\widehat{w}_l^{KM}(t)$

Let $\bar{\beta}_I$ and $\bar{\beta}_{II}$ be the estimators with $\widehat{w}_l^{KM}(t)$ for a single and multiple case-cohort studies. Similarly to the proof of Theorem 1, we can show for $w = I, II$, $n^{-1/2}(\bar{\beta}_w - \beta_0)$ is asymptotically normally distributed with mean zero and with covariance matrix $\bar{A}(\beta_0)^{-1}\bar{\Sigma}^w(\beta_0)\bar{A}(\beta_0)^{-1}$, where $\bar{A}(\beta_0)^{-1}\bar{\Sigma}^w(\beta_0)\bar{A}(\beta_0)^{-1}$, where

$$\begin{aligned}\bar{A}(\beta_0) &= \sum_{l=1}^L q_l \bar{A}_l(\beta_0), \quad \bar{\Sigma}^w(\beta_0) = \sum_{l=1}^L q_l (\bar{V}_l^0 + \bar{V}_l^w), \\ \bar{V}_l^0 &= E\{(\bar{\eta}_{1,l1} + \bar{\eta}_{2,l1})^{\otimes 2}\}, \quad \bar{V}_l^w = \frac{1 - \alpha_l^*}{\alpha_l^*} E\{(\bar{\eta}_{3,l1}^w)^{\otimes 2}\}, \\ \bar{\eta}_{1,li} &= \int_0^\tau \{Z_{li} - \bar{e}_l(\beta_0, t)\} \widehat{w}_l^{KM}(t) dM_{li}^1(t), \quad \bar{\eta}_{2,li} = \int_0^\tau \frac{b_l(t)}{c_l(t)} dM_{li}^C(t), \\ \bar{\eta}_{3,li}^I &= \int_0^\tau \{1 - \Delta_{li} I(\epsilon_{li} = 1)\} \widehat{w}_l^{KM}(t) Y_{li}^1(t) \{Z_{li} - \bar{e}_l(\beta_0, t)\} d\Lambda_{li0}(t), \\ \bar{\eta}_{3,li}^{II} &= \int_0^\tau \left\{1 - \sum_{k=1}^K \Delta_{li} I(\epsilon_{li} = k)\right\} \widehat{w}_l^{KM}(t) Y_{li}^1(t) \{Z_{li} - \bar{e}_l(\beta_0, t)\} d\Lambda_{li0}(t), \\ \bar{A}_l(\beta_0) &= \int_0^\tau \bar{v}_l(\beta_0, t) \bar{s}_l^{(0)}(\beta_0, t) \lambda_{li}(t) dt, \\ \bar{v}_l(\beta, t) &= \bar{s}_l^{(2)}(\beta, t) / \bar{s}_l^{(0)}(\beta, t) - \bar{e}_l(\beta, t)^{\otimes 2}, \\ \bar{e}_l(\beta, t) &= \bar{s}_l^{(1)}(\beta, t) / \bar{s}_l^{(0)}(\beta, t), \\ \bar{s}_l^{(d)}(\beta, t) &= E[n_l^{-1} \sum_{i=1}^{n_l} \widehat{w}_l^{KM}(t) Y_{li}^1(t) \mathbf{Z}_{li}^{\otimes d} e^{\beta^T Z_{li}}], \quad d = 0, 1, 2, \\ b_l(u) &= \lim_{n_l \rightarrow \infty} \frac{1}{n_l} \sum_{i=1}^{n_l} \int_0^\tau \{Z_{li} - \bar{e}_l(\beta_0, t)\} \widehat{w}_l^{KM}(t) dM_{li}^1(t) I(t \geq u > X_{li}), \\ c_l(u) &= \lim_{n_l \rightarrow \infty} \frac{1}{n_l} \sum_{i=1}^{n_l} I(X_{li} \geq u).\end{aligned}$$

The detailed proof can be obtained from the authors upon request.

4 Appendix D: Technical Details of Efficiency Gain

We have

$$\begin{aligned}
EG &= \frac{1 - \alpha_l^*}{\alpha_l^*} E(\boldsymbol{\eta}_{3,l1}^I)^{\otimes 2} - \frac{1 - \alpha_l^*}{\alpha_l^*} E(\boldsymbol{\eta}_{3,l1}^{II})^{\otimes 2} \\
&= \frac{1 - \alpha_l^*}{\alpha_l^*} \left[E\{1 - \Delta_{li}I(\epsilon_{li} = k)\}^2 - E\{1 - \sum_{k=1}^K \Delta_{li}I(\epsilon_{li} = k)\}^2 \right] \\
&\times E \left[\int_0^\tau w_{li}^{Cox}(t) Y_{li}^1(t) \{ \mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}_0, t) \} d\Lambda_{l10}(t) - \int_0^\tau \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t) \right]^{\otimes 2}
\end{aligned}$$

The matrix $E[\int_0^\tau w_{li}^{Cox}(t) Y_{li}^1(t) \{ \mathbf{Z}_{li} - \mathbf{e}_l(\boldsymbol{\beta}_0, t) \} d\Lambda_{l10}(t) - \int_0^\tau \mathbf{q}_{li}^{(1)}(t) dM_{li}^C(t)]^{\otimes 2}$ is positive definite because of Conditions 1 and 2. Therefore, the asymptotic efficiency gain (EG) is associated with $(1 - \alpha_l^*)/\alpha_l^* E\{\sum_{k=2}^K \Delta_{l1}I(\epsilon_{l1} = k)\}$ which is always positive when $\alpha_l^* > 0$. Asymptotic efficiency gain is associated with smaller subcohort size and larger other causes rates.

5 Appendix E: Variance estimator for the cumulative incidence function for a single case-cohort study

For a single case-cohort studies, by the functional delta method, $n_l^{1/2}[\widehat{F}_{l1}(t|\mathbf{z}_0) - F_{l1}(t|\mathbf{z}_0)]$ converges weakly to a Gaussian process with mean zero and asymptotic variance, which can be estimated by $n_l \left\{ 1 - \widehat{F}_{l1}(t|\mathbf{z}_0) \right\}^2 \sum_i \left\{ \widehat{\mathbf{W}}_{F,li}(t|\mathbf{z}_0) \right\}^2$, where

$$\begin{aligned}
\widehat{\mathbf{W}}_{F,li}(t|\mathbf{z}_0) &= \exp\{\widehat{\boldsymbol{\beta}}_I^T \mathbf{z}_0\} [\widehat{\Lambda}_{l10}^I(t) \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}_I)^{-1} \widehat{\mathbf{U}}(\widehat{\boldsymbol{\beta}}_I) \mathbf{z}_0 + \widehat{H}(t; \mathbf{z}_0)], \\
\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}_I) &= \sum_{l=1}^L \frac{n_l}{n} \int_0^\tau \left[\frac{\widehat{\mathbf{S}}_l^{(2)}(\widehat{\boldsymbol{\beta}}_I, t)}{\widehat{\mathbf{S}}_l^{(0)}(\widehat{\boldsymbol{\beta}}_I, t)} - \frac{\widehat{\mathbf{S}}_l^{(1)}(\widehat{\boldsymbol{\beta}}_I, t)^{\otimes 2}}{\widehat{\mathbf{S}}_l^{(0)}(\widehat{\boldsymbol{\beta}}_I, t)} \right] \widehat{\mathbf{S}}_l^{(0)}(\widehat{\boldsymbol{\beta}}_I, t) d\widehat{\Lambda}_{l10}^I(t), \\
\widehat{H}(t; \mathbf{z}_0) &= n_l^{-1/2} \sum_{i=1}^{n_l} \{ \zeta_{1,li}(\widehat{\boldsymbol{\beta}}_I, t) + (1 - \xi_{li}/\alpha_l) \zeta_{2,li}^I(\widehat{\boldsymbol{\beta}}_I, t) \}.
\end{aligned}$$

6 Appendix F: Additional simulation results

We conducted simulation study for non-stratified case-cohort data. We set values for ψ_1 and γ_1 to 1, which results in a special case of a Weibull mixture distribution. One covariate Z_{1i} was generated from the Bernoulli distribution with $Pr(Z_{1i} = 1) = 0.5$. Table 1 reports the average bias of the estimates $\hat{\beta}_I$, and $\hat{\beta}_{II}$, the average of the estimated standard error (SE), empirical standard deviation (SD), average of standardized bias (STB) defined as $E\{\|\hat{\beta} - \beta_0\|/SE\}$, the empirical coverage rate (%) with the nominal 95% confidence interval for various simulation settings. The results are similar to Table 1 of the main paper: when the censoring distribution depends on covariates, the parameter estimates with the covariate-adjusted weight are approximately unbiased while those with the covariate-unadjusted weight are biased; the estimated standard errors of the covariate-adjusted estimators are close to the empirical standard deviations; The empirical coverage rates for the covariate-adjusted estimators are between 94% and 96%. All sample relative efficiencies of the covariate-adjusted estimators are greater than 1. This suggests $\hat{\beta}_{II}$ using covariate information on failures from cause 2 is more efficient than $\hat{\beta}_I$. The range of efficiency gain is from 12% to 86%.

We also conducted simulations with censoring probabilities to nonparametrically estimate the censoring survival function separately for each level of the covariate. Table 2 shows the simulation results with covariate-level-specific Kaplan-Meier-estimate-based weight for both unstratified and stratified models when the censoring time depends on covariates (scenario S2). The results are very similar to those with covariate-adjusted weight we proposed.

We conducted simulations when strata for the censoring distribution were different from those for the competing risks model. We considered stratified proportional subdistribution hazards model and two strata with $(\psi_1, \psi_2) = (1, 1)$ and $(\kappa_1, \kappa_2) = (1, 2)$. One covariate Z was generated from the Bernoulli distribution with $Pr(Z_{1i} = 1) = 0.4$ in stratum 1 and $Pr(Z_{1i} = 1) = 0.6$ in stratum 2. We set (β_0, θ_0) to $(1, -1)$. We considered strata variables for censoring distribution, which

Table 1: Simulation results for unstratified data

CC	(P_1, P_2, P_c)	α	Covariate-unadjusted weight					Covariate-adjusted weight						
			SD	SE	STB	CR	SRE	bias	SD	SE	STB	CR	SRE	
Scenario 1: Censoring distribution \sim Uniform distribution														
$\hat{\beta}_I$	(20%, 20%, 60%)	0.1	0.014	0.258	0.249	0.058	0.94	1.00	0.014	0.261	0.248	0.057	0.94	1.00
		0.2	0.011	0.195	0.198	0.053	0.95	1.00	0.010	0.196	0.197	0.053	0.95	1.00
	(20%, 40%, 40%)	0.1	0.010	0.240	0.240	0.043	0.95	1.00	0.011	0.242	0.240	0.044	0.95	1.00
		0.2	0.012	0.194	0.193	0.064	0.95	1.00	0.012	0.195	0.193	0.064	0.95	1.00
$\hat{\beta}_{II}$	(20%, 20%, 60%)	0.1	0.015	0.230	0.225	0.069	0.94	1.26	0.016	0.231	0.224	0.070	0.94	1.27
		0.2	0.012	0.183	0.185	0.065	0.96	1.14	0.012	0.184	0.184	0.066	0.95	1.14
	(20%, 40%, 40%)	0.1	0.008	0.198	0.200	0.041	0.95	1.47	0.009	0.198	0.200	0.043	0.95	1.49
		0.2	0.013	0.168	0.172	0.074	0.95	1.33	0.013	0.168	0.172	0.076	0.95	1.34
Scenario 2: Censoring distribution \sim Cox model														
$\hat{\beta}_I$	(20%, 20%, 60%)	0.1	-0.052	0.314	0.317	0.165	0.95	1.00	0.032	0.304	0.293	0.110	0.94	1.00
		0.2	-0.074	0.231	0.251	0.296	0.95	1.00	0.012	0.230	0.235	0.050	0.95	1.00
	(20%, 40%, 40%)	0.1	-0.142	0.266	0.268	0.529	0.92	1.00	0.006	0.269	0.259	0.024	0.94	1.00
		0.2	-0.146	0.206	0.206	0.709	0.89	1.00	0.005	0.212	0.203	0.027	0.94	1.00
$\hat{\beta}_{II}$	(20%, 20%, 60%)	0.1	-0.061	0.259	0.258	0.236	0.94	1.47	0.035	0.277	0.272	0.127	0.94	1.20
		0.2	-0.076	0.205	0.214	0.355	0.94	1.28	0.016	0.217	0.225	0.070	0.96	1.12
	(20%, 40%, 40%)	0.1	-0.147	0.181	0.190	0.771	0.89	2.15	0.005	0.197	0.204	0.022	0.96	1.86
		0.2	-0.153	0.158	0.163	0.941	0.85	1.69	0.000	0.170	0.174	0.001	0.95	1.55

CC, case-cohort estimators; (P_1, P_2, P_c) , probability distribution of cause 1, cause 2, and censoring; SE, the average of the estimates of standard error; SD, sample standard deviation; STB, $E\{\|\hat{\beta} - \beta_0\|/SE\}$; CR, the empirical coverage rate of the nominal 95% confidence intervals; SRE, sample relative efficiency.

Table 2: Simulation results for unstratified and stratified data with covariate-level-specific Kaplan-Meier weights

	(P_1, P_2, P_c)	α	unstratified data					stratified data						
			bias	SD	SE	STB	CR	SRE	bias	SD	SE	STB	CR	SRE
$\hat{\beta}_I$	(20%, 20%, 60%)	0.1	0.037	0.310	0.295	0.126	0.94	1.00	0.023	0.218	0.213	0.109	0.95	1.00
		0.2	0.011	0.229	0.236	0.047	0.96	1.00	0.006	0.173	0.173	0.032	0.95	1.00
	(20%, 40%, 40%)	0.1	0.012	0.266	0.262	0.047	0.94	1.00	0.015	0.196	0.196	0.076	0.95	1.00
		0.2	0.013	0.212	0.206	0.064	0.95	1.00	0.001	0.149	0.155	0.008	0.96	1.00
$\hat{\beta}_{II}$	(20%, 20%, 60%)	0.1	0.037	0.280	0.274	0.136	0.95	1.22	0.023	0.196	0.196	0.118	0.94	1.24
		0.2	0.015	0.217	0.225	0.069	0.95	1.12	0.006	0.162	0.163	0.035	0.95	1.14
	(20%, 40%, 40%)	0.1	0.011	0.197	0.208	0.055	0.96	1.82	0.011	0.141	0.153	0.074	0.96	1.92
		0.2	0.010	0.169	0.177	0.057	0.96	1.57	0.000	0.122	0.134	0.003	0.96	1.50

CC, case-cohort estimators; (P_1, P_2, P_c) , probability distribution of cause 1, cause 2, and censoring; SE, the average of the estimates of standard error; SD, sample standard deviation; STB, $E\{\|\hat{\beta} - \beta_0\|/SE\}$; CR, the empirical coverage rate of the nominal 95% confidence intervals; SRE, sample relative efficiency.

are independent of and different from those for the competing risks model. More specifically, the censoring time was generated from the stratified proportional hazards model with three different constant baseline hazard (0.11, 0.13, 0.15), $\gamma = 2.5$, and $Z_{li}^C = Z_{li}$. In other words, the competing risks model had two strata while the censoring outcome model had three strata. The sample size of the full cohort was set to 1800 and two subcohort sizes were examined: 540 ($\alpha_1 = \alpha_2 = 0.3$) and 900 ($\alpha_1 = \alpha_2 = 0.5$). For each configuration, 2000 iterations were conducted. We estimated the parameter for the censoring distribution and all censoring-related terms by fitting the stratified proportional hazards model with three strata. Then, we plugged them into the asymptotic variance formula in Theorem 1 to obtain the standard error for $\hat{\beta}$. Table 3 reports the average bias of the estimates $\hat{\beta}_I$ and $\hat{\beta}_{II}$, the average of the estimated standard error (SE), empirical standard deviation (SD), the empirical coverage rate (%) with the nominal 95% confidence interval (CR). The results show all estimates are approximately unbiased and the average of the estimated standard errors for both proposed estimators are close to their empirical standard deviations. All empirical coverage rates are close to 95%.

Table 3: Simulation results

α		bias	SD	SE	CR	SRE
0.3	$\hat{\beta}_I$	0.001	0.146	0.141	0.94	1.00
	$\hat{\beta}_{II}$	0.001	0.126	0.126	0.95	1.36
0.5	$\hat{\beta}_I$	0.000	0.129	0.126	0.95	1.00
	$\hat{\beta}_{II}$	0.000	0.119	0.119	0.95	1.16

SE, the average of the estimates of standard error; SD, sample standard deviation; CR, the empirical coverage rate of the nominal 95% confidence intervals; SRE, sample relative efficiency.

To examine the robustness of the proposed method against misspecified models for censoring, we conducted simulations when the censoring distribution follows an additive hazards model. The stratified additive hazards model for C_{li} given Z_{li}^C was

$$\lambda_{li}^A(t|Z_{li}^C) = \lambda_{0l}^A(t) + \mu_0 Z_{li}^C,$$

where $\lambda_{0l}^A(t)$ is a baseline hazard function for stratum l and μ_0 is an unknown regression parameter. We considered the stratified proportional subdistribution hazards model and two strata with $(\psi_1, \psi_2) = (1, 1)$ and $(\kappa_1, \kappa_2) = (1, 2)$. One covariate Z was generated from the Bernoulli distribution with $Pr(Z_{1i} = 1) = 0.4$ in stratum 1 and $Pr(Z_{2i} = 1) = 0.6$ in stratum 2. We set (β_0, θ_0) to $(1, -1)$. The censoring time was generated from the stratified additive hazards model with two different constant baseline hazard $(\lambda_{01}^A, \lambda_{02}^A) = (0.5, 1)$, $\mu_0 = 0.3$, and $Z_{li}^C = Z_{li}$. We set the failure rate for cause 1, cause 2, and censoring to (20%, 30%, 50%). The sample size of the full cohort was set to 1000 and two subcohort sizes were examined: 200 ($\alpha_1 = \alpha_2 = 0.2$) and 400 ($\alpha_1 = \alpha_2 = 0.4$). For each configuration, 2000 iterations were conducted. We estimated the parameter for the censoring distribution and all censoring-related terms by fitting the stratified proportional hazards model with two strata. Then, we plugged them into the asymptotic variance formula of Theorem 1 to obtain the standard error for $\hat{\beta}$. Table 4 reports the average bias of the estimates $\hat{\beta}_I$ and $\hat{\beta}_{II}$, the average of the estimated standard error (SE), empirical standard deviation (SD), the empirical coverage rate (%) with the nominal 95% confidence interval (CR). The results show all estimates are approximately unbiased and the average of the estimated standard errors for both proposed estimators are close to their empirical standard deviations. All empirical coverage rates are close to 95%. This small simulation study suggests the robustness of the proposed method against model misspecification for censoring.

Table 4: Simulation results for censoring time based on additive hazards model

α		bias	SD	SE	CR	SRE
0.2	$\hat{\beta}_I$	0.010	0.224	0.213	0.94	1.00
	$\hat{\beta}_{II}$	0.012	0.184	0.181	0.95	1.49
0.4	$\hat{\beta}_I$	0.004	0.185	0.179	0.94	1.00
	$\hat{\beta}_{II}$	0.005	0.169	0.165	0.94	1.20

SE, the average of the estimates of standard error; SD, empirical standard deviation; CR, the empirical coverage rate of the nominal 95% confidence intervals; SRE, sample relative efficiency.

7 Appendix G: Data analysis result for cause-specific hazard model

We fitted the cause-specific hazard models for the two example data: the bone marrow transplant data and the ARIC study data (Prentice et al., 1978). Table 5 reports the regression parameter estimates, their standard errors, and their p -values for the full cohort, a single case-cohort, and two case-cohort studies for the bone marrow transplant study. Table 6 presents the regression parameter estimates, their standard errors, and their p -values for the ARIC data.

Table 5: Analysis of the CIBMTR study fitting cause-specific model

Variables	Full cohort				CC_{II}				
	$\hat{\beta}$	SE	p-value	$\hat{\beta}$	SE	p-value	$\hat{\beta}$	SE	p-value
Age	0.415	0.129	0.001	0.356	0.122	0.004	0.374	0.124	0.003
KPS (ref = $\geq 90\%$)									
<90%	0.730	0.193	0.000	0.754	0.232	0.001	0.749	0.233	0.001
Donor type (ref = HLA)									
Haploidentical	0.358	0.216	0.098	0.409	0.247	0.098	0.394	0.248	0.112
Disease subtype (ref = FLH)									
Diffuse large B-cell lymphoma	-0.158	0.264	0.551	-0.173	0.318	0.587	-0.158	0.321	0.623
Mantle cell lymphoma	0.139	0.285	0.626	0.260	0.335	0.439	0.200	0.336	0.551
Mature T- and NK-cell lymphomas	-0.009	0.317	0.978	-0.002	0.356	0.995	0.045	0.360	0.902
Hodgkin	0.055	0.332	0.869	0.105	0.347	0.761	0.001	0.348	0.997

CC_I , a single case-cohort data; CC_{II} , two case-cohort data; SE, standard error estimate; KPS, Karnofsky performance status at transplant; ref, reference group; The reference groups are HLA-identical siblings donor (HLA) for donor type, ≥ 90 for KPS, and Follicular lymphoma for Histology (FLH). The coefficients for the reference groups were set to 0 and therefore they were omitted from the table.

Table 6: Analysis of the ARIC study fitting cause-specific model

Variables	Category	$\hat{\beta}$	SE	p -values
high-sensitivity C-reactive protein (ref = Low)	Middle High	0.254 0.503	0.174 0.171	0.145 0.003
Systolic blood pressure		0.267	0.074	0.001
LDL cholesterol		0.257	0.065	< .001
HDL cholesterol		-0.406	0.093	< .001
Diabetes (ref = No)		0.651	0.169	< .001
Smoking status (ref = No)		0.380	0.156	0.015

SE, standard error estimate; LDL, low density lipoprotein; HDL, high density lipoprotein; Low high-sensitivity C-reactive protein group, group without diabetes, and non-smoking group are reference groups; ref, reference group. The coefficients for the reference groups were set to 0 and therefore they were omitted from the table.

References

- Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–20.
- Fourtz, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Am. Statist. Assoc.* **72**, 147–8.
- Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 361–74.
- He, P., Eriksson, F., Scheike, T., and Zhang, M. (2016). A proportional hazards regression model for the subdistribution with covariates-adjusted censoring weight for competing risks data. *Scandinavian Journal of Statistics* **43**, 103–122.
- Kang, S. and Cai, J. (2010). Asymptotic results for fitting marginal hazards models from stratified case-cohort studies with multiple disease outcomes. *J Korean Stat Soc.* **39**, 371–385.
- Kim, S., Zeng, D., and Cai, J. (2018). Analysis of multiple survival events in generalized case-cohort designs. *Biometrics* **00**, In press.

Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika* **87**, 73–87.

Prentice, R. L., Kalbfleisch, J. D., Jr., A. V. P., 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.